Information Theory

R.J. Marks II Class Notes
Texas Tech University (1976)

TEST #1 PLUG SHEET

- PROBABILITY: $P[A\cap B\cap C] =$  $= P(A)+P(B)+P(C)-P(AB)-P(BC)-P(AC)+P(ABC$

MARGINAL PROBABILITY: $P[A_i] = \Sigma_j P[A_i, B_j]$

CONDITIONAL PROBABILITY: $P[AB] = P[A/B]P[B]$

JOINT PROBABILITY : $P[AB]$

STATISTICAL IND : $P[ABC]=P_A P_B P_C,\ P[AB]=P_A P_B\ \ P[AC]=P_A P_C\ \ P[BC]=P_C P_B$

PDF (PROBABILITY DIST FUNC) : $= P[x \le \mathbf{X}]$

JOINT PDF : $P[\mathbf{X}\le x, \mathbf{Y}\le y]$

MARGINAL PDF : $P[\mathbf{X}\le x] = P[\mathbf{X}\le x, \mathbf{Y}\le +\infty]$

CONDITIONAL PDF : $P[\mathbf{X}\le x/\mathbf{Y}\le y] = P[\mathbf{X}\le x, \mathbf{Y}\le y]/P[\mathbf{Y}\le y]$

pdf (PROBABILITY DENSITY FUNCTION): $p(\mathbf{X}) = \frac{d}{dX}P[\mathbf{X}\le x$

JOINT pdf : $p(\mathbf{X},\mathbf{Y}) = \frac{d^2}{d\mathbf{X}d\mathbf{Y}} P[\mathbf{X}\le x, \mathbf{Y}\le y]$

MARGINAL pdf : $p(\mathbf{X}) = \int_{-\infty}^{\infty} p(\mathbf{X},\mathbf{Y})d\mathbf{Y}$

CONDITIONAL pdf : $p[\mathbf{Y}/\mathbf{X}] = p[\mathbf{X},\mathbf{Y}]/p(\mathbf{X})$

- INFORMATION MEASURE : $I[X_k] = -\lg P[X_k]$
- THE MATHEMATICS OF ENTROPY

ENTROPY DEFN: $H(S) = -\Sigma_k P[X_k]\lg P[X_k]$

ADDITIVE PROPERTY: $H(P_1,\dots P_n) = H(P_1 \dots P_{n-1}) + P_n H(\frac{q_1}{p_n},\dots,\frac{q_k}{p_n})$

- LEMMAS & COR. OF H

CONVEX FUNCTION CRITERION: $\frac{d^2 f}{dx^2} \le 0$ or $\frac{1}{2}[f(x_1)+f(x_2)] \le f[\frac{x_1+x_2}{2}]$

$\sqrt{x_1 x_2} \le \frac{1}{2}(x_1+x_2)$: INEQUALITY OF GEO. & ARITH. MEANS

EXTREMA PROPERTY: $H(x_1,\dots,x_m) \le \ln m$

$\Sigma x_i \ln \frac{1}{x_i} \le \Sigma x_i \ln \frac{1}{y_i}$

JOINT ENTROPY: $H(\mathbf{X},\mathbf{Y}) = -\Sigma_i\Sigma_j P(x_i y_j)\lg p(x_i y_j)$

$H(X,Y) \le H(X) + H(Y)$

CONDITIONAL ENTROPY: $H(X/Y) = -\Sigma_i\Sigma_j P(x_i, y_j)\ln P(x_i|y_j$

$H(x,y) = H(x) + H(y/x) = H(Y) + H(X/Y)$

$H(X/Y) \le H(X)$

- SOURCE EXTENSION

  S HAS q SYMBOLS $\Rightarrow$ S^n HAS q^n SYMBOLS

  $\sum_{S_n} = \sum_{i=1}^{q^n} = \sum_{i_1}^{q} \sum_{i_2}^{q} \cdots \sum_{i_n}^{q}$

  $p(\sigma_i) = p(s_{i_1}) p(s_{i_2}) \ldots p(s_{i_n})$

  $H^n(S) = n H(S)$

- CHANNEL ENTROPY

  JOINT PROBABILITY MATRIX: $H(X,Y)$

  CHANNEL EQUIVOCATION: $H(X/Y)$

  CONDITIONAL PROBABILITY MATRICES

● PROBABILITY THEORY REVIEW

A. TWO APPROACHES

a. FREQUENCY OF EVENTS APPROACH

LET $n(X_K)$ = # OF TIMES AN EVENT X_K OCCURS

N = TOTAL NUMBER OF EVENTS $\geq n(X_K)$

$$\Rightarrow P[X_K] = \lim_{N \to \infty} n(X_K)/N$$

b. AXIOMATIC (PROBABILITY MEASURE) APPROACH

- LET a_K BE A POINT IN A SAMPLE SPACE

$m[a_K]$ BE A REAL & SINGLE VALUED MEASURE ON a_K

$m[E] = \Sigma\, m[a_K \in E]$

- TWO EVENTS ARE DISJOINT IF $m[A \cup B] = m(A) + m(B)$

(THIS IS CALLED "ADDITIVE PROPERTY" OF THE MEASURE)

- $m(X) = 0$ IFF $X = \phi$ (NULL SET)

$m(X) = 1$ IFF $X = U$ (UNIVERSAL SET)

- ALSO
  1) $m(A) \leq m(B)$ IF $A \in B$
  2) $m(A) = m(B) - m(B - A)$ IF $A \in B$
  3) $m(\bar{A}) \triangleq m(U - A) = 1 - m(A)$
  4) $m(A \cup B) = m[\,A(A-B) \cup B\,]$
     $= m(A) + m(B) - m(AB)$
  5) $m(A \cup B \cup C) = m(A) + m(B) + m(C)$
     $- m(AB) - m(BC) - m(CA)$
     $+ m(ABC)$

B. AXIOMS & THEOREMS OF AXIOMATIC APPROACH

AXIOM 1: $P(A)$ IS A REAL NUMBER $\ni P(A) \geq 0 \; \forall$ EVENT $A \in S$

AXIOM 2: $P(S) = 1$

AXIOM 3: LET $S = \{s_1, s_2, \dots\} \ni s_i \cap s_j = \phi \; \forall \; i \neq j$ (i.e. ALL s_i ARE DISJOINT) THEN

$$P[s_1 \cup s_2 \cup s_3 + \dots] = P[s_1] + P[s_2] + \dots$$

THEOREM 1: LET S BE A SAMPLE SPACE AND P BE A PROBABILITY MEASURE ON S. THEN

$$P[\bar{A}] = P[U - A] = 1 - P[A]$$

THEOREM 2: LET S BE A SAMPLE SPACE WITH PROBABILITY MEASURE P. THEN $0 \leq P(A) \leq 1 \; \forall \; A \in S$.

THEOREM 3: LET S BE A SAMPLE SPACE WITH PROBABILITY MEASURE P AND S_0 = NULL SET. THEN $P[S_0] = 0$.

C. MARGINAL, JOINT, & CONDITIONAL PROBABILITY

LET S BE A SAMPLE SPACE WITH PROBABILITY MEASURE P. PARTITION S INTO r DISJOINT SUBSETS $\{A_1, A_2, \dots, A_r\}$. REPARTITION S INTO s DISJOINT SUBSETS $\{B_1, B_2, \dots, B_s)$.

- THE JOINT PROBABILITY OF EVENTS A_i AND B_j OCCURING IS DENOTED $P[A_i, B_j] \triangleq P[A_i \cap B_j]$

- GIVEN THE JOINT PROBABILITY, THE MARGINAL PROBABILITIES ARE $P[A_i] = \sum_{j=1}^{r} P[A_i, B_j]$ AND $P[B_j] = \sum_{j=1}^{s} P[A_i, B_j]$.

FOR THREE DISJOINT PARTITIONS OF S:

$$P[A_i, C_k] = \sum_{j=1}^{s} P[A_i, B_j, C_k]$$

AND $$P[C_k] = \sum_{j=1}^{s} \sum_{i=1}^{r} P[A_i, B_j, C_k]$$

EXTENSIONS ARE OBVIOUS

- CONDITIONAL PROBABILITY (MULT. LAW OF PROB. MEASURE)

$$P[A_i, B_j] = P[A_i / B_j]\,P[B_j] = P[B_j / A_i]\,P[A_i]$$

- RELATIVE FREQUENCY VIEW OF JOINT, MARGINAL, AND CONDITIONAL PROBABILITIES

$$S : \{A_1, A_2, \dots, A_r\} \qquad A_i \cap A_j = \phi \ \ \forall\, i \neq j$$

$$S : \{B_1, B_2, \dots, B_s\} \qquad B_i \cap B_j = \phi \ \ \forall\, i \neq j$$

$$\begin{array}{ccccc}
 & B_1 & B_2 \ \dots & B_j \ \dots & B_s \\
A_1 & n_{11} & n_{12} & n_{1j} & n_{1s} \\
A_2 & n_{21} & n_{22} & n_{2j} & n_{2s} \\
\vdots & & & & \\
A_i & n_{i1} & n_{i2} & n_{ij} & n_{is} \\
\vdots & & & & \\
A_r & n_{r1} & n_{r2} & n_{rj} & n_{rs}
\end{array}$$

LET $\sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij} = n$

$$\Rightarrow P(A_i, B_j) = P[A_i \cap B_j] = n_{ij}/n \Leftarrow \text{JOINT}$$

$$\left.\begin{aligned}
P(A_i) &= \sum_{j=1}^{s} P(A_i, B_j) = \frac{1}{n} \sum_{j=1}^{s} n_{ij} \\
P(B_j) &= \sum_{i=1}^{s} P(A_i, B_j) = \frac{1}{n} \sum_{i=1}^{s} n_{ij}
\end{aligned}\right\} \Leftarrow \text{MARGINAL}$$

$$\left.\begin{aligned}
P(A_i / B_j) &= n_{ij} \Big/ \sum_{i=1}^{r} n_{ij} \\
&= \frac{n_{ij}}{n} \div \frac{1}{n} \sum_{i=1}^{r} n_{ij} \\
&= P[A_i, B_j] / P[B_j] \\
P(B_j / A_i] &= P[A_i, B_j] / P[A_i]
\end{aligned}\right\} \Leftarrow \text{CONDITIONAL}$$

D. STATISTICAL INDEPENDENCE

- TWO EVENTS, A & B, ARE STATISTICALLY INDEPENDENT IFF $p[A,B] = P[A]P[B]$

  IT FOLLOWS THAT $P[A/B] = P[A]$

- N EVENTS $\{A_1, A_2, \ldots, A_i, \ldots A_N\}$ ARE STATISTICALLY INDEPENDENT IFF $\forall\ i \neq j \neq k \ldots$

$$P[A_i, A_j] = P(A_i)\, P(A_j)$$
$$P[A_i, A_j, A_k] = P(A_i)\, P(A_j)\, P(A_k)$$
$$\vdots$$
$$P[A_i, A_j, A_k, \ldots, A_N] = P(A_i)\, P(A_j)\, P(A_k) \ldots P(A_N)$$

E. RANDOM VARIABLE

- DEFN: A REAL VALUED FUNCTION $X(S)$ DEFINED ON A SAMPLE SPACE, S, IS A RANDOM VARIABLE IF $\forall$ REAL NUMBER a, THE SET OF POINTS FOR WHICH $X(S) \leq a$ IS ONE OF THE CLASS OF ADMISSABLE SETS FOR WHICH A PROBABILITY IS DEFINED.

F. PROBABILITY DISTRIBUTION FUNCTION

- DEFN: $P[X \leq \mathfrak{X}] =$ PDF OF R.V. X

$$P[X \leq b] - P[X \leq a] = P[a < X \leq b] \quad \forall\ b > a$$

- JOINT PDF $= P[X \leq \mathfrak{X},\ Y \leq \mathfrak{Y}]$

  $= P[$SAMPLE POINT IS IN APPROPRIATE QUADRANT$]$

- MARGINAL PDF'S

$$P[X \leq \mathfrak{X}] = P[X \leq \mathfrak{X},\ Y < \infty]$$
$$P[Y \leq \mathfrak{Y}] = P[X \leq \infty,\ Y \leq \mathfrak{Y}]$$

- CONDITIONAL PDF'S

$$P[X \leq \mathfrak{X} / Y \leq \mathfrak{Y}] = P[X \leq \mathfrak{X}, Y \leq \mathfrak{Y}] / P[Y \leq \mathfrak{Y}]$$
$$P[Y \leq \mathfrak{Y} / X \leq \mathfrak{X}] = P[X \leq \mathfrak{X}, Y \leq \mathfrak{Y}] / P[X \leq \mathfrak{X}]$$

G. CONTINUOUS R.V. / PROBABILITY DENSITY FUNCTION (pdf)

- $pdf(\bar{X}) = p(\bar{X}) \triangleq \frac{d}{dx} PDF = \frac{d}{d\bar{X}} P[x \le \bar{X}]$

$$= \lim_{\Delta x \to 0} \frac{1}{\Delta x}\left[P(x \le \bar{X}) - P(x \le \bar{X} - \Delta x)\right]$$

(PDF'S ARE RIGHT CONTINUOUS)

$$p(\bar{X})\Delta x = P[x - \Delta x < \bar{X} < x]$$

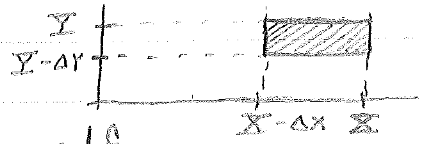
- PROPERTIES OF A pdf FOR A CONTINUOUS R.V.

(1) $p(x) \ge 0$ (2) $P[a \le x \le b] = \int_a^b p(x)dx$

(3) $\int_{-\infty}^{\infty} p(x)dx = 1$ ALSO, $p(x_0) = 0$

H. JOINT, MARGINAL, & CONDITIONAL pdf'S.

- JOINT pdf ⇒ JOINT PDF $= P[x \le \bar{X}, Y \le \bar{Y}]$, MUST BE CONTINUOUS & TWICE (MIXED) DIFFERENTIABLE THEN JOINT pdf $= \frac{\delta^2}{\delta \bar{X} \delta \bar{Y}} P[x \le \bar{X}, Y \le \bar{Y}]$



JOINT pdf

$$= p[\bar{X},\bar{Y}] = \lim_{\substack{\Delta x \to 0 \\ \Delta Y \to 0}} \frac{1}{\Delta x \Delta Y}\Big[P(x \le \bar{X}, Y \le \bar{Y}) - P(x \le \bar{X} - \Delta x, Y \le \bar{Y})$$
$$- P[x \le \bar{X}, Y \le \bar{Y} - \Delta \bar{Y}) + P(x \le \bar{X} - \Delta x, Y \le \bar{Y} - \Delta Y)\Big]$$

$$p[\bar{X},\bar{Y}]\Delta x \Delta Y = P[\bar{X} - \Delta x < x < \bar{X} ; \bar{Y} - \Delta Y \le Y \le \bar{Y}]$$

- SOME RELATIONSHIPS

$$\int_{-\infty}^{\infty}\int p(x,Y)dx dY = 1$$
$$\int_{-\infty}^{\infty} p(x,\bar{Y}) dx = p(\bar{Y}) \Leftarrow \text{MARGINAL DENSITY}$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\bar{X}} p(x,Y) dx dY = P(x \le \bar{X}) \Leftarrow \text{MARGINAL DISTRIBUTION}$$

- CONDITIONAL pdf

$$P[Y \le \bar{Y} / \bar{X} - \Delta x < x \le \bar{X}] = \frac{P[\bar{X} - \Delta x < x < \bar{X}, Y \le \bar{Y}]}{P[\bar{X} - \Delta x < x < \bar{X}]}$$
$$= \left[\int_{-\infty}^{Y}\int_{\bar{X}-\Delta x}^{\bar{X}} P(x,Y) dx dY\right] \Big/ \int_{\bar{X}-\Delta x}^{\bar{X}} p(x)dx$$
$$= P[Y \le \bar{Y}/x] = \int_{-\infty}^{\bar{Y}} p(\bar{X}, Y] dY \,/\, p(\bar{X})$$

DIFFERENTIATING W.R.T. $\bar{Y}$:

$$p(\bar{Y}/\bar{X}) = p[\bar{X},\bar{Y}]/p(\bar{X})$$

- INFORMATION MEASURE
  - CONSIDER A SAMPLE SPACE S & AN EVENT $x_K \in S$. WE WISH TO ASSOCIATE AN AMOUNT OF INFORMATION ASSOCIATED WITH THE OCCURANCE OF x_K, $I(x_K)$ HARTLEY RECOGNIZED THAT $I(x_K) = f[P(x_K)]$ INTUITIVELY, WE WOULD EXPECT

$$I[x_K, C_j] = I[x_K] + I(C_j) \ni C_j \in S, \ C_j \,\&\, x_K \text{ IND.}$$

  - SHANNON PROPOSED THE FOLLOWING INFORMATION MEASURE:

$$I[x_K] = -\ln P(x_K)$$

= APRIORI UNCERTAINTY ASSOCIATED WITH THE OCCURANCE OF x_K

= AMOUNT OF INFORMATION ASSOCIATED WITH THE OCCURANCE OF x_K

  - INFORMATION UNITS

USING LOG BASE 2, I IS IN BITS
" " " e, " " " NATS
" " " 10, " " " HARTLEYS

FOR CONVERSION, USE $\log_a x = \log_b x / \log_b a$

1 HARTLEY = 3.32 BITS, 1 NAT = 1.44 BITS

THE MATHEMATICS OF ENTROPY

A. DEFN: CONSIDER A SOURCE S WITH EVENTS $\{x_1, x_2, \ldots, x_K\}$ AND CORRESPONDING PROBABILITY (MEASURES) $\{p_1, p_2, \ldots, p_K\}$. THE SOURCE ENTROPY IS THE AVERAGE, OR EXPECTED VALUE, OF INFORMATION ASSOCIATED WITH THE OCCURANCE OF AN EVENT:

$$H(S) = \sum_{i=1}^{K} p(x_i)\, I(x_i)$$
$$= -\sum p(x_i) \ln p(x_i)$$

B. SOME PROPERTIES OF ENTROPY:

1. $H(S)$ IS CONTINUOUS WITH RESPECT TO p_i
2. $H(S)$ IS SYMMETRIC. ie, $H(p_1, p_2, \ldots, p_{n-1}, p_n) = H(p_2, p_n, \ldots, p_{n-1}, p_2)$
3. $H(S)$ IS MAXIMUM WHEN $p_i = \frac{1}{n}$; $i = 1, 2, \ldots, n$

PROOF: $$\frac{dH}{dp_K} \triangleq \sum_{i=1}^{n} \frac{\delta H}{\delta p_i} \frac{\delta p_i}{\delta p_K}$$
$$= -\frac{\delta}{\delta p_K} p_K \ln p_K \cdot \frac{dp_K}{dp_K} - \frac{d}{dp_n}(p_n \ln p_n) \frac{dp_K}{dp_n} + 0$$

ALSO, SINCE $p_n = 1 - (p_1 + p_2 + \ldots + p_{n-1})$

WE HAVE $\frac{dH}{dp_K} = \ln p_n - \ln p_K = 0 \Rightarrow p_n = p_K$

4. ADDITIVE PROPERTY: $H(p_1, p_2, \ldots, p_{n-1}, q_1, q_2, \ldots, q_m)$
$$= H(p_1, p_2, \ldots, p_n) + p_n H\left(q_1/p_n, q_2/p_n, \ldots, q_m/p_n\right)$$

WHERE $p_n = \sum_{i=1}^{m} q_i$

PROOF: $$H(p_1, p_2, \ldots, p_{n-1}, q_1, q_2, \ldots q_m) = -\sum_{i=1}^{n-1} p_i \ln p_i - \sum_{K=1}^{m} q_K \ln q_K$$
$$= -\sum_{i=1}^{n} p_i \ln p_i + p_n \ln p_n - \sum_{K=1}^{m} q_K \ln q_K$$
$$= -\sum_{i=1}^{n} p_i \ln p_i + p_n \sum_{K=1}^{m} \frac{q_K}{p_n} \ln p_n - p_n \sum_{K=1}^{m} \frac{q_K}{p_n} \ln q_K$$
$$= -\sum_{i=1}^{n} p_i \ln p_i - p_n \sum_{K=1}^{m} \frac{q_K}{p_n} \ln q_K/p_n$$
$$= H(p_1, p_2, \ldots, p_{n-1}, q_1, q_2, \ldots, q_m)$$

- LEMMAS & CORR. OF H

A. LEMMA 1: $\ln x$ IS A CONVEX FUNCTION

WE MAY PROVE THIS TWO WAYS

1. $f(x)$ IS CONVEX IFF $\frac{\delta^2 f(x)}{\delta x^2} \leq 0$

$$\frac{\delta^2}{\delta x^2} \ln x = -\frac{1}{x^2} \leq 0 \quad \forall \; x$$

2. $f(x)$ IS CONVEX IFF $\frac{1}{2}[f(x_1) + f(x_2)] \leq f\left[\frac{x_1+x_2}{2}\right]$

FOR $\ln x$, WE MUST RESTRICT x_1 & $x_2 > 0$

$$\Rightarrow \frac{1}{2}[\ln x_1 + \ln x_2] \overset{?}{\leq} \ln\left(\frac{x_1+x_2}{2}\right)$$

$$\frac{1}{2}\ln x_1 x_2 \overset{?}{\leq} \ln\left(\frac{x_1+x_2}{2}\right)$$

$$\text{GEOMETRIC MEAN} \Rightarrow \sqrt{x_1 x_2} \underset{\uparrow \text{ALWAYS}}{\leq} \frac{x_1+x_2}{2} \Leftarrow \text{ARITHMATIC MEAN}$$

GEOMETRIC MEAN $\leq$ ARITHMETIC MEAN. ∴ LEMMA IS PROVED

B. LEMMA 2: $\ln x \leq x-1$ (OR $\ln x \ln_2 e = \ln_2 x \leq (x-1)\ln_2 e$

C. EXTREMA PROPERTY OF H: $H(x_1, x_2, \ldots, x_m) \leq \ln m$

$$\text{PROOF: } H(x) - \ln m = \sum_{i=1}^{m} p_i \ln \frac{1}{p_i} + \ln \frac{1}{m}$$

$$= \sum_{i=1}^{m} p_i \ln \frac{1}{p_i} + \sum_{i=1}^{m} p_i \ln \frac{1}{m}$$

$$= \sum_{i=1}^{m} p_i \ln \frac{1}{m p_i}$$

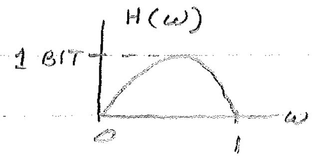
$$\leq \sum_{i=1}^{m} p_i \left(\frac{1}{m p_i} - 1\right) \quad \Leftarrow \text{FROM LEMMA 2}$$

$$\leq \sum_{i=1}^{m} \left(\frac{1}{m} - p_i\right)$$

$$\leq 1 - 1 = 0 \qquad \text{QED}$$

D. COR: THE ENTROPY FUNCTION, $H(\omega)$, IS CONVEX

$$\ni H(\omega) = -\omega \ln \omega - \bar{\omega} \ln \bar{\omega}$$



E. LEMMA 3: LET $\{X_1, X_2, \ldots, X_q\}$ AND $\{Y_1, Y_2, \ldots, Y_q)$ BE SEPERATE DISJOINT PARTITIONS OF S_0 THAT IS $\sum_{i=1}^{q} X_i = \sum_{i=1}^{q} Y_i = 1$. THEN

$$\sum_{i=1}^{q} X_i \ln \frac{1}{X_i} \le \sum_{i=1}^{q} X_i \ln \frac{1}{Y_i}$$

PROOF: $\sum_{i=1}^{q} X_i \ln \frac{Y_i}{X_i} \le \sum_{i=1}^{q} X_i \left(\frac{Y_i}{X_i} - 1 \right)$

$$\le -\sum X_i + \sum Y_i$$

$$\le 0$$

OR $\sum_{i=1}^{q} X_i \ln \frac{1}{X_i} - \sum_{i=1}^{q} X_i \ln \frac{1}{Y_i} \le 0$

$$\Rightarrow \sum X_i \ln 1/X_i \le \sum X_i \ln \frac{1}{Y_i}$$

F. LEMMA 4: RELATION OF JOINT & MARGINAL ENTROPIES

$\left\{ \text{JOINT ENTROPY: } H(X,Y) \triangleq -\sum_{i=1}^{M} \sum_{j=1}^{L} p(X_i, Y_i) \ln p(X_i, Y_j) \right\}$
$\left\{ \ni \mathbb{X} = \{X_1, \ldots, X_i, \ldots, X_M\} \ \& \ \mathbb{Y} = \{Y_1, \ldots, Y_j, \ldots, Y_L\} \right\}$

$$H(\mathbb{X}, \mathbb{Y}) \le H(\mathbb{X}) + H(\mathbb{Y})$$

PROOF: $H(\mathbb{X}) = -\sum_{i=1}^{M} p(X_i) \ln p(X_i)$

$$= -\sum_{i=1}^{M} \sum_{j=1}^{L} p(X_i, Y_j) \ln p(X_i)$$

$$H(\mathbb{Y}) = -\sum_{i=1}^{M} \sum_{j=1}^{L} p(X_i, Y_j) \ln p(Y_i)$$

$$\therefore H(\mathbb{X}) + H(\mathbb{Y}) = -\sum_i \sum_j p(X_i, Y_i) \ln P(Y_j) P(X_i)$$

$$= -\sum_i \sum_j p(X_i, Y_j) \ln q_{ij}$$

$$\ni q_{ij} \triangleq P(X_i)\, p(Y_j) \Rightarrow \sum_i \sum_j q_{ij} = 1$$

ALSO, LET $p_{ij} \triangleq p(X_i, Y_j)$

$$\Rightarrow H(\mathbb{X}, \mathbb{Y}) = -\sum_i \sum_j p_{ij} \ln p_{ij}$$

FROM LEMMA 3:

$$H(\mathbb{X}, \mathbb{Y}) = -\sum_i \sum_j p_{ij} \ln p_{ij}$$

$$\le -\sum_i \sum_j p_{ij} \ln q_{ij}$$

$$\le H(\mathbb{X}) + H(\mathbb{Y}) \qquad \text{QED}$$

(LEMMA'S & COR. OF H)

G. COR F1: $H(X_1,X_2,\dots,X_n) \le H(X_1)+H(X_2)+\dots+H(X_n)$

EQUALITY IF ALL X_i ARE STATISTICALLY INDEP.

H. COR F2: $H(X_1,X_2,\dots,X_n,Y_1,Y_2,\dots Y_m)$

$$\le H(X_1,X_2,\dots,X_n)+H(Y_1,Y_2,\dots,Y_m)$$

EQUALITY IF THE RANDOM VECTOR $\{X_1,X_2,\dots,X_n\}$ IS STATISTICALLY INDEP. OF $\{Y_1,Y_2,\dots,Y_m\}$

I. CONDITIONAL ENTROPY (DEFN)

$$H[\mathbf{Y}/\mathbf{X}=x_i] \overset{\Delta}{=} -\sum_{j=1}^{L} p(y_j/x_i)\ln p(y_i/x_i)$$

$$H[\mathbf{Y}/\mathbf{X}] \overset{\Delta}{=} E[H\{\mathbf{Y}/\mathbf{X}=x_i\}]$$

$$= -\sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_i)\ln p(y_j/x_i)$$

J. LEMMA 5:

$$H(X,Y)=H(X)+H(Y/X)=H(Y)+H(X/Y)$$

PROOF: $H(X,Y)=-\sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\ln p(x_i,y_j)$

$$=-\sum_i\sum_j p(x_i,y_j)\ln p(x_i)p(y_j/x_i)$$

$$=H(X)+H(Y/X)$$

K. LEMMA 6: $H(Y/X)\le H(Y)$

$H(X,Y)=H(X)+H(Y/X) \Leftarrow$ FROM LEMMA 5

$\le H(X)+H(Y) \Leftarrow$ FROM LEMMA 4

$\therefore H(Y/X)\le H(Y)$

● SOURCE EXTENSION

A. DEFN: UTILIZING A SOURCE OF q SYMBOLS TO GENERATE n SYMBOL WORDS DENOTES THE SOURCE AS THE n^{TH} EXTENSION.

THE n^{TH} EXTENSION OF S HAS q^n SYMBOLS

B. SOME THEOREMS

1. THE n^{TH} EXTENTION OF A SOURCE IS COMPLETE

$$S = \{s_1, s_2, \ldots, s_q\} \Rightarrow (p_1, p_2, \ldots, p_q)$$

$$S^n = \{\sigma_1, \sigma_2, \ldots, \sigma_{q^n}\} \Rightarrow (p_1', p_2', \ldots, p_{q^n}')$$

$$\sum_{S^n} p_i' = \sum_{i=1}^{q^n} p(\sigma_i) \ni p(\sigma_i) = p_{i_1} p_{i_2} \cdots p_{i_n}$$

$$\Sigma_{S^n} p(\sigma_j) = \sum_{i_1=1}^{q} \sum_{i_2=1}^{q} \cdots \sum_{i_n=1}^{q} p_{i_1} p_{i_2} \cdots p_{i_n}$$

$$= \Sigma_{i_1} p_{i_1} \sum_{i_2} p_{i_2} \cdots \Sigma_{i_n} p_{i_n} = 1$$

2. $H^n(S) = n H(S)$

PROOF: $H^n(S) = \Sigma_{S^n} p(\sigma_i) \lg \dfrac{1}{p(\sigma_i)}$

$$= \Sigma_{S_n} p(\sigma_i) \lg \frac{1}{p_{i_1} p_{i_2} \cdots p_{i_n}}$$

$$= \Sigma_{S_n} p(\sigma_i) \lg 1/p_{i_1} + \Sigma_{S_n} p(\sigma_i) \lg 1/p_{i_2}$$

$$+ \Sigma_{S_n} p(\sigma_i) \lg \frac{1}{p_{i_k}} + \cdots + \Sigma_{S_n} p(\sigma_i) \lg 1/p_{i_n}$$

CONSIDER K^{TH} TERM:

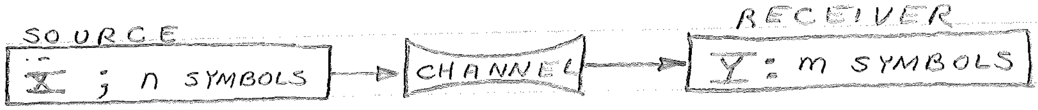
$$\Sigma_{S_n} p(\sigma_i) \lg \frac{1}{p_{i_k}} = \sum_{i_1=1}^{q} \sum_{i_2=1}^{q} \cdots \sum_{i_n=1}^{q} p_{i_1} p_{i_2} \cdots p_{i_n} \ln \frac{1}{p_{i_k}}$$

$$= \sum_{i_k=1}^{q} p_{i_k} \lg 1/p_{i_k} = H(S)$$

$$\therefore H^n(S) = n H(S)$$

• CHANNEL ENTROPY

A.



$H(X)$ = AVE INFO PER SOURCE SYMBOL

$H(Y)$ = AVE. INFO PER RECEIVER CHARACTER

$H(Y/X)$ = INFO ABOUT RECEIVED SYMBOL GIVEN TRANSMITTED SIGNAL (MEASURE OF ERROR & NOISE)

$H(X/Y)$ = CHANNEL EQUIVOCATION (MEASURE OF THE RECOVERABILITY OF THE INPUT AT THE RECEIVER)

$H(X,Y)$ = AVERAGE INFO PER CHARACTER PAIR

B. SOURCE CHARACTERIZATION

1. JOINT PROB MATRIX:

$$\begin{array}{c|cccccc|c}
 & Y_1 & Y_2, & \ldots & Y_j & \ldots & Y_m & \\
X_1 & p(X_1Y_1) & p(X_1Y_2) & & p(X_1Y_j) & & p(X_1Y_m) & p(X_1) \\
X_2 & p(X_2Y_1) & p(X_2Y_2) & & p(X_2Y_j) & & p(X_2,Y_m) & p(X_2) \\
\vdots & & & & & & & \\
X_i & p(X_i,Y_1) & p(X_i,Y_2) & & p(X_iY_j) & & p(X_iY_m) & p(X_i) \\
\vdots & & & & & & & \\
X_n & p(X_nY_1) & p(X_nY_2) & & p(X_nY_j) & & p(X_nY_m) & p(X_n) \\
\hline
 & p(Y_1) & p(Y_2) & & p(Y_j) & & p(Y_m) & 1
\end{array}$$

MARGINAL PROBABILITIES

2. CONDITIONAL PROB. MATRIX

$$p(X_i/Y_j) = p(X_i,Y_j)/p(Y_j)$$
$$p(Y_j/X_i) = p(X_i,Y_j)/p(X_i)$$

TEST # 2 PLUG SHEET

MUTUAL INFORMATION: $I(x_i, Y_j) = \lg \frac{P(x_i; Y_j)}{P(x_i)P(Y_j)}$

$I(x_i) \geq I(x_i; Y_j)$

TRANSINFORMATION: $I(x;Y) = H(x) - H(x/Y) = H(Y) - H(Y/x$

CHANNEL CAPACITY: $C = Max\, I(x;Y)$

REDUNDANCY : $R = C - I(x;Y) = 1 - \eta$

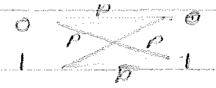
RELATIVE REDUNDANCY: $R_r = 1 - \frac{I(x;Y)}{\lg n}$

CHANNEL EFFICIENCY: $\eta = H_r(S)/\bar{L}$

TRANSMISSION RATE: $R_t = H(x) / \sum_i t_i p(x_i)$

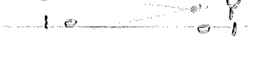
INFORMATION CHANNEL

BINARY SYMMETRIC CHANNEL:



UNIFORM CHANNELS : $P[b_j/a_i]$, THINGS ADD UP

BINARY ERASURE CHANNEL:



MUROGA'S TECHNIQUE

BLOCK CODE, NON-SINGULAR, UNIQUELY DECODABLE, INSTANTANEOUS

PREFIX PROPERTY

KRAFT'S INEQUALITY $\sum_{i=1}^{q} r^{-l_i} = \sum_{i=1}^{l} N_i r^{-i} \leq 1$

McMILLAN'S INEQUALITY

SHANNON'S FIRST THEOREM: $H_r(x) \leq \bar{L}, \lim_{n\to\infty} \frac{\bar{L}_n}{n} = H_r$

AVERAGE WORD LENGTH: $\bar{L} = \sum_i p(x_i) l_i$

CODING PROCEDURES:

SHANNON, SHANNON-FANO

HUFFMAN → $q = r + (r-1)\alpha \ni \alpha \in INT$

ALSO MAY USE $\lg_r 1/P_i \leq l_i \leq \lg_r \frac{1}{P_i} + 1$

● MUTUAL INFORMATION:

A. $I(X_i;Y_j) \triangleq \lg p(X_i/Y_i) - \lg p(X_i)$

$= \lg p(X_i/Y_i)/p(X_i) = \lg \frac{p(X_i,Y_i)}{p(X_i)\,p(Y_i)}$

-SELF INFORMATION: $I(X_i) = -\lg p(X_i)$

-A'PRIORI KNOWLEDGE X_i IS BEING XMITTED AND $\overset{\text{SOME}}{\wedge}$ Y_j IS BEING RECEIVED $= p(X_i)$

- POSTERIOR KNOWLEDGE $= P(X_i/Y_j)$

-THE DIFFERENCE OF THESE TWO IS THE GAIN IN INFORMATION $= I(X_i;Y_i) = \lg P(X_i/Y_j) - \lg p(X_i)$

B. SOME PROPERTIES OF MUTUAL INFO

(1) $I(X_i;Y_j)$ IS CONTINUOUS

(2) $I(X_i;Y_j) = I(Y_j;X_i)$ IS SYMMETRIC

(3) $I(X_i) = I(X_i;X_i) = -\lg P(X_i) \geq I(X_i;Y_j)$

ALSO, $I(Y_j) = I(Y_j;Y_j) \geq I(X_i;Y_j)$

C. AVERAGE INFORMATION GAIN

$$I(X;Y) = \overline{I(X_i;Y_j)}$$
$$= \Sigma_i \Sigma_j\, p(X_i,Y_j)\, I(X_i;Y_j)$$
$$= H(X) + H(Y) - H(X,Y)$$
$$= H(X) - H(X/Y) = H(Y) - H(Y/X)$$

∴ ON THE AVERAGE, OBSERVATION OF ANY Y GIVES US $I(X;Y)$ BITS OF INFO.

- CHANNEL PARAMETERS

A. CHANNEL CAPACITY

$$C \triangleq \text{Max } I(X;Y)$$
$$= \text{Max } [H(X) - H(X/Y)]$$

B. RATE OF INFORMATION TRANSMISSION

$$C_t = c/t \quad \text{BITS/SEC}$$

$\supset$ t = TIME THAT IT TAKES TO XMIT EACH SYMBOL

$$C_t = lg\, n/t \quad \text{FOR A NOISE FREE CHANNEL}$$

C. REDUNDANCY

- ABSOLUTE REDUNDANCY: DIFFERENCE TWIXT MAXIMUM CAPACITY AND ACTUAL INFO:

$$R = C - I(X;Y) \Longleftarrow \text{GENERAL} \quad \text{no!}$$
$$= lg\, n - H(X) \Longleftarrow \text{NOISE FREE}$$

- RELATIVE REDUNDANCY

$$R_r = \frac{1}{lg\, n}[lg\, n - H(X)] \Longleftarrow \text{NOISE FREE}$$
$$= 1 - \frac{I(X;Y)}{lg\, n} \Longleftarrow \text{GENERAL}$$

D. CHANNEL EFFICIENCY

$$\eta = I(X;Y)/lg\, n = 1 - R_r$$

E. GENERALIZED XMISSION RATE

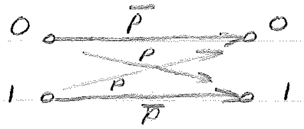
$$R_t = H(X) / \Sigma_i\, p(x_i) t_i$$
$$= -\Sigma\, p(x_i)\, lg\, p(x_i) / \Sigma_i\, p(x_i) t_i$$

THE INFORMATION CHANNEL

A. DEFN: AN INFO CHANNEL IS DESCRIBED BY GIVING AN INPUT ALPHEBET, $A; \{a_i\}, i=1,\ldots,r$; AN OUTPUT ALPHEBET $B; \{b_j\}, j=1,\ldots,s$; AND A SET OF CONDITIONAL PROBABILITIES $p(b_j/a_i) \ \forall \ i,j$. WE DENOTE $p(b_j/a_i)=p_{ij}$. CLEARLY, $\Sigma_j p_{ij}=1$

B. BINARY SYMMETRIC CHANNEL

$\bar{P}=P(0/0)=P(1/1) \qquad P=P(0/1)=P(1/0)$



C. INFO CHANNEL EXTENSION

-CONSIDER INFO CHANNEL WITH ALPHEBETS A & B AND PROBABILITY (COND) MATRIX p.

-DEFN: LET A^n & B^n BE THE n^{TH} XTENSIONS OF A & B:

$A^n: \{\alpha_i\}, i=1,\ldots,r^n \qquad B^n=\{\beta_j\}; j=1,\ldots,s^n$

THE CHANNEL MATRIX, Π, IS THEN

$$\Pi = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \Pi_{1s^n} \\ \vdots & & & \\ \Pi_{r^n 1} & \cdots & \cdots & \Pi_{r^n s^n} \end{bmatrix}$$

WHERE $\alpha_i = \{a_{i_1}, a_{i_2}, \ldots a_{i_n}\}$, $\beta_j = \{b_{j_1}, b_{j_2}, \ldots, b_{j_n}\}$)

THEN $\Pi_{ji} = P[\beta_j/\alpha_i] = P_{j_1 i_1} P_{j_2 i_2} \cdots P_{j_n i_n}$

1- FORWARD PROB $= p(b_j/a_i)$; BACKWARD $= p(a_i/b_j)$

$H(A) = -\Sigma_i \, p(a_i) \lg p(a_i)$ = APRIORI ENTROPY

$H(A/b_j) = -\Sigma_i \, p(a_i/b_j) \lg p(a_i/b_j)$ = A POSTERIORY ENT.

D. UNIFORM CHANNELS

-DEFN: AN INFO CHANNEL IS UNIFORM IF ALL ROWS & COLUMNS OF $P = \{P_{ij}\} = \{P[b_j/a_i]\}$ ARE PERTURBATIONS OF EACH OTHER.

EX

$$\begin{array}{c|cccc} & b_1 & b_2 & b_3 & b_4 \\ \hline a_1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ a_2 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{array}$$

-CHANNEL CAPACITY

$$C = \text{Max}\, I(X;Y) = \text{Max}\,[H(Y) - H(Y/X)]$$

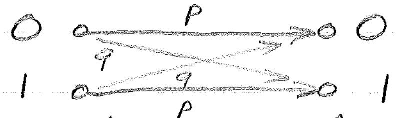
FOR A UNIFORM CHANNEL: $H(Y/X_i) = h = \text{const.}$

$$\Rightarrow H(Y/X) = \Sigma_i\, p(a_i)\, H(Y/X_i) = h$$

$H(Y)$ IS MAX FOR $p(Y_j) = \frac{1}{5} \Rightarrow H(Y) = \lg 5$

AND $C = \lg 5 - h$

E. CHANNEL CAPACITY FOR BSC



$p(0) = \alpha \quad p(1) = \bar{\alpha}$

$$H(X) = -\alpha \lg \alpha - \bar{\alpha} \lg \bar{\alpha}$$

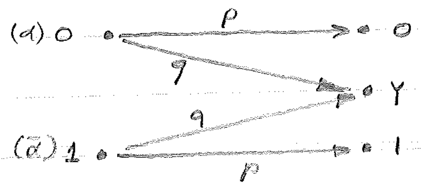
$$H(Y/X) = -p \lg p - q \lg q$$

$$I(X;Y) = H(Y) - H(Y/X)$$

$$= H(Y) + p \lg p + q \lg q$$

$$C = \text{Max}\,[I(X;Y)] = 1 + p \lg_2 p + q \lg_2 q \text{ BITS}$$

F. BEC (BINARY ERASURE CHANNEL)



$P(Y/X)$	0	Y	1
0	p	q	0
1	0	q	p

LET $P(0)=\alpha$ AND $P(1)=\bar{\alpha}$

LET'S FIND THE CHANNEL CAPACITY:

$$H(X) = -\alpha \lg \alpha - \bar{\alpha} \lg \bar{\alpha}$$

$P(X/Y)$	0	Y	1
0	$\alpha p / \alpha p = 1$	$\alpha q / q = \alpha$	0
1	0	$\bar{\alpha} q / q = \bar{\alpha}$	$\alpha p / \alpha p = 1$

$$H(X/Y) = \alpha p \lg 1 - \alpha q \lg \alpha - \bar{\alpha} q \lg \bar{\alpha} - \alpha p \lg 1$$

$$= -q(\alpha \lg \alpha + \bar{\alpha} \lg \bar{\alpha}) = q\,H(X)$$

$$\therefore I(X;Y) = H(X) - H(X/Y) = (1-q)\,H(X) = p\,H(X)$$

$$C = \text{Min}\, I(X;Y) = p\, \text{Min}\, H(X) = p \text{ BITS}$$

G. BINARY SYMETRIC CHANNEL EXTENSION:

KRONECKER MATRICES

MUROGA'S TECHNIQUE

A. CONSIDER $\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} -H(p_{11}) \\ -H(p_{22}) \end{bmatrix}$

THEN IT CAN BE SHOWN:

① $I(X;Y) = -(p_1' \lg p_1' + p_2' \lg p_2') + p_1' Q_1 + p_2' Q_2$

WHERE $p_1' = P[Y_1']$ AND $p_2' = P[Y_2']$

WE WISH TO MAXIMIZE ① VIA LAGRANGE MULTIPLIERS, μ.

LET. $U = -(p_1' \lg p_1' + p_2' \lg p_2') + p_1' Q_1 + p_2' Q_2 + \mu (p_1' + p_2)$

$\Rightarrow \frac{dU}{dp_1} = -(\lg p_1' + \lg_2 e) + Q_1 + \mu$ ②

$\frac{dU}{dp_2} = -(\lg p_2' + \lg_2 e) + Q_2 + \mu$ ③

SETTING $\frac{dU}{dp_1} = \frac{dU}{dp_2} = 0$ GIVES

$\mu = -Q_1 + (\lg_2 e + \ln p_1')$

$\mu = -Q_2 + (\lg_2 e + \lg p_2')$

SUBSTITUTING INTO ① GIVES

$C = \max I(X;Y) = Q_1 - \lg p_1' = Q_2 - \lg p_2'$

$\Rightarrow p_1' = 2^{Q_1 - C} \qquad p_2' = 2^{Q_2 - C} = 1 - p_1'$

THUS $C = \lg_2 (2^{Q_1} + 2^{Q_2})$ BITS

B. COMMENTS

(1) IN GENERAL, WE MUST SOLVE:

$$\begin{bmatrix} p_{11} & \dots & p_{1n} \\ \vdots & & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} Q_1 \\ \vdots \\ Q_n \end{bmatrix} = \begin{bmatrix} -H_1 \\ \vdots \\ -H_n \end{bmatrix}$$

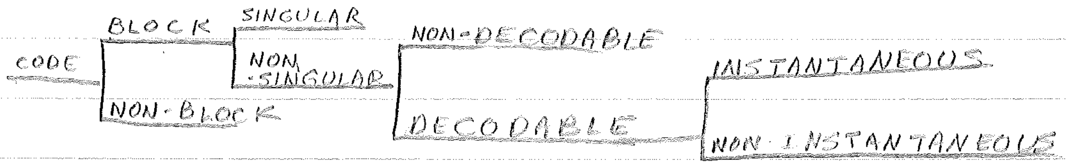
$$C = \lg_2 \sum_{i=1}^{n} 2^{Q_i}$$

(2) ONLY WORKS FOR SQUARE CHANNEL MATRICES

(3) IT'S POSSIBLE THAT THE REQUIRED INPUT PROBABILITIES FOR THE COMPUTED CAPACITY DON'T MEET $0 < p_i < 1$ OR $\sum_i p_i = 1$. i.e. WATCH OUT

(CODE CATEGORIES)

- CODE CATEGORIES AND PROPERTIES
  - CODE - LET THE SET OF SYMBOLS $S:\{s_1 s_2 \ldots s_q\}$ BE AN INPUT ALPHABET. THEN A MAPPING OF THIS TO SOME OTHER ALPHABET $X:\{x_1, \ldots x_n\}$ IS A CODE.
    - S IS THE SOURCE ALPHABET
    - X IS THE CODE ALPHABET
  - BLOCK CODE - A CODE IN WHICH S IS MAPPED DISTINCTLY (BUT NOT NECESSARILY UNIQUELY INTO X.)
  - NON-SINGULAR CODE - A BLOCK CODE WHERE EACH OUTPUT WORD IS DIFFERENT FROM ANY OTHER
  - CODE EXTENSION - THE n^{TH} EXTENSION OF A CODE MAPS THE INPUT WORD EXTENSIONS $S_{i_1} S_{i_2} \ldots S_{i_n}$ INTO THE OUTPUT WORD EXTENSION $X_{i_1} X_{i_2} \ldots X_{i_n}$.
  - DECODABLE CODE - A CODE IS DECODABLE IF ITS n^{TH} EXTENSION IS NON-SINGULAR $\forall\ n$
  - INSTANTANEOUS - A DECODABLE CODE IS INSTANT IF IT IS POSSIBLE TO DECODE EACH WORD WITHOUT REFERENCE TO THE SUCCEEDING CODE SYMBOL



● CODE PROPERTIES (CODE PROPERTIES)

- PREFIX PROPERTY - A CODE IS INSTANTANEOUS IFF (i.e. NEC. & SUF.) NO COMPLETE WORD IN A CODE BOOK IS A PREFIX OF SOME OTHER CODE WORD.
- KRAFT'S INEQUALITY

LET r BE THE # OF SYMBOLS IN THE CODE ALPHABET AND LET l_i BE THE # OF SYMBOLS IN x_i WHERE $X : \{x_1 .. x_i ... x_q\}$. THEN A NEC. & SUFF. CONDITION FOR THE EXISTANCE OF AN INSTANT CODE WITH THESE WORD LENGTHS IS

$$\sum_{i=1}^{q} r^{-l_i} \leq 1$$

PROOF: LET n_i BE THE NUMBER OF CODE WORDS OF LENGTH l_i. IF $\text{Max}[l_i] = l$, THEN

$$\sum_{i=1}^{l} n_i = q$$

$\ni q =$ # OF THE WORDS IN THE CODE BOOK.

NOW $\sum_{i=1}^{q} r^{-l_i} = \sum_{i=1}^{l} n_i r^{-i}$

WE SHOW SUFFICIENCY BY WRITING

$$\sum_{i=1}^{q} r^{-l_i} \leq 1$$

$$\Rightarrow \sum_{i=1}^{l} n_i r^{-i} \leq 1$$

$$n_1 r^{-1} + n_2 r^{-2} + \ldots + n_{l-1} r^{-(l-1)} + n_l r^{-l} \leq 1$$

OR $0 \leq n_l r^{-l} \leq 1 - n_1 r^{-1} - n_2 r^{-2} + \ldots + n_{l-1} r^{-(l-1)}$

$$0 \leq n_l \leq r^l - n_1 r^{l-1} - n_2 r^{l-2} + \ldots + n_{l-1} r$$

$$\therefore 0 \leq n_{l-1} \leq r^{l-1} - n_1 r^{l-2} - n_2 r^{l-3} - \ldots - n_{l-2} r$$

$$\vdots$$

$$0 \leq n_3 \leq r^3 - n_1 r^2 - n_2 r$$

$$0 \leq n_2 \leq r^2 - n_1 r$$

$$0 \leq n_1 \leq r$$

NOW

$n_1 \leq r$

THAT IS, WE MAY, AT MOST, BUILD r WORDS WITH LENGTH $l_1 = 1$. THUS, IF WE USE n_1 WORDS OF LENGTH $l_1 = 1$, THEN WE GOT $r - n_1$ PREFIXES LEFT TO WORK WITH. WE MAY ADD UP TO r SYMBOLS ON THESE REMAINING PREFIXES, CONSTITUTING, AT MOST, $r(r - n_1)$ WORDS OF LENGTH $n_2 = 2$. THUS

$$n_2 \leq r(r - n_1) = r^2 - n_1 r$$

SUPPOSE WE USE n_2 OF THESE. THEN, BY SIMILAR ARGUMENTS:

$$n_3 \leq [(r^2 - n_1 r) - n_2] r = r^3 - n_1 r^2 - n_2 r$$

ETC. ONE MAY PROVE THE NEC. PART OF THE THEOREM BY REVERSING THE ARGUMENTS.

- McMILLAN'S INEQUALITY

SINCE ALL INSTANT CODES ARE UNIQUELY DECODABLE (U.D), AND KRAFT'S INEQUALITY IS SUFFICIENT FOR THE EXISTANCE OF A U.D. CODE. McMILLAN'S INEQUALITY (THE SAME AS KRAFT'S) SAYS THAT KRAFT'S INEQUALITY IS ALSO NECESSARY FOR THE EXISTANCE OF A U.D. CODE.

PROOF: CONSIDER

$$(\Sigma_{i=1}^{q} r^{-l_i})^n = (r^{-l_1} + r^{-l_2} + \dots + r^{-l_q})^n$$

$$= \Sigma_{K=1}^{q^n} r^{-l_K}$$

$K = l_{i_1} + l_{i_2} + \dots + l_{i_n}$

IF $Max\, l_i = l$, THEN IT IS CLEAR THAT

$$n \le K \le nl$$

NOW, IF N_K = # OF TERMS OF r^{-l_K}, THEN

$$\Sigma_{K=1}^{q^n} r^{-l_K} = \Sigma_{K=n}^{nl} N_K r^{-K}$$

TO BE UNIQUELY DECODABLE, WE REQUIRE

$$N_K \le r^K$$

$$\Rightarrow \Sigma_{K=n}^{nl} N_K r^{-K} \le \Sigma_{K=n}^{nl} r^K r^{-K}$$

$$\le nl - n + 1 \le nl \Rightarrow (\Sigma_{i=1}^{q} r^{-l_i})^n \le nl$$

NOW, IF $x^n \le nl$, THEN $x < 1$

$$\Rightarrow \Sigma_{i=1}^{q} r^{-l_i} \le 1$$

- SHANNON'S FIRST THEOREM

A. $\bar{L}$ = AVE LENGTH = $\sum_i p(x_i) l_i$

B. THE AVE LENGTH OF A U.D. CODE IS LOWER BOUNDED BY $H(x)/\lg r$. THAT IS

$$\bar{L} \geq H(x)/\lg r$$

PROOF: $\quad -\sum_{i=1}^{n} q_i \ln q_i \leq \sum_{i=1}^{n} q_i \lg p_i$

LET $q_i = r^{-l_i} / \sum_{i=1}^{n} r^{-l_i}$

$$\Rightarrow H(x) = \sum_{i=1}^{n} p(x_i) \lg P(x_i)$$
$$\leq -\sum_{i=1}^{n} P(x_i) \lg \frac{r^{-l_i}}{\sum_{k=0}^{n} r^{-l_k}}$$
$$\leq -\sum_{i=1}^{n} P(x_i) \lg r^{-l_i} + P(x_i) \lg \sum_{k=1}^{n} r^{-l_k}$$
$$\leq \sum_{i=1}^{n} l_i P(x_i) \lg r + \sum_{i=1}^{n} p(x_i) \lg \Sigma()$$
$$\leq \bar{L} \lg r + \sum_{i=1}^{n} p(x_i) \lg \sum_{k=1}^{n} r^{-l_k}$$

FOR U.D. CODES, WE HAVE McMILLAN'S INEQ:

$$\sum_{k=1}^{n} r^{-l_k} \leq 1$$
$$\Rightarrow \sum_{i=1}^{n} p(x_i) \sum_{k=1}^{n} r^{-l_k} \leq 0$$

$\therefore \quad H(x) \leq \bar{L} \lg r$

OR $\bar{L} \geq H(x)/\lg r$

C. CRITERION FOR MEETING LOWER BOUND (BASE 2)

$$H(x) \leq \bar{L}/\lg_2 2 = \bar{L}$$
$$\Rightarrow -\lg_2 P(Y_K) \leq l_K$$
$$\frac{1}{P(Y_K)} \leq 2^{l_K}$$
$$P(Y_K) \geq 2^{-l_K}$$

$\therefore$ EQUALITY IS MET ONLY IF $P(X_i) = 2^{-l_K}$.

IN GENERAL, IF $P(X_i) = r^{-l_K}$

(SHANNON'S 1^{ST} THM)

C. $\lim_{n \to \infty} \overline{L_n}/n = H_r(S)$

$\ni$ n REFER'S TO THE 'nTH EXTENSION

AND r TO THE # OF ALPHABET SYMBOLS

PROOF: $H_r(S) \le \overline{L} \le H_r(S) + 1$

THIS IS GOOD ALWAYS. USE THE n^{TH} EXT. THEN:

$$H_r(S^n) \le \overline{L} \le H_r(S^n) + 1$$

$$n\,H_r(S^n) \le \overline{L}_n \le n\,H_r(S_n) + 1$$

$$\Rightarrow \quad H_r(S\ ) \le \overline{L}_n/n \le H_r(S\ ) + \frac{1}{n}$$

OBVIOUSLY, $\lim_{n \to \infty} \overline{L}_n/n = H_r(S)$

INTERPRET $\overline{L}_n/n$ AS THE # OF CODE SYMBOLS USED FROM THE ORIGINAL SOURCE PER SAMPLE SYMBOL (EH?)

• CODING PROCEDURES (CLASSICAL)

A. SHANNON'S BINARY CODING PROCEDURE

FOR AN OPTIMAL CODE, WE CAN ACHEIVE $H(X) \le \bar{L} \le H(X)+$

a. ARRANGE PROBABILITIES IN DECREASING ORDER

$$P_1 \le P_2 \le \ldots \le P_q$$

b. COMPUTE α_i'S :

$$\alpha_1 \triangleq 0$$

$$\alpha_2 = P(x_1)$$

$$\alpha_3 = P(x_2) + \alpha_2$$

$$\alpha_4 = P(x_3) + \alpha_3$$

$$\vdots$$

$$\alpha_q = P(x_{q-1}) + \alpha_{q-1}$$

$$\alpha_{q+1} = P(x_q) + \alpha_q = 1$$

c. FIND THE SET OF INTEGERS (SMALLEST)

WHICH SATISFIES : $2^{l_i} P(x_i) \ge 1$

d. EXPAND EACH α_i TO l_i PLACES

(IN BINARY FORM) AND NO FURTHER.

EXAPLE: $P_i = \{0.4, 0.3, 0.2, 0.1)$

$\alpha_1 \triangleq 0$ $\quad \alpha_3 = 0.7$

$\alpha_2 = 0.4$ $\quad \alpha_4 = 0.9$

$$l_1 : 2^{l_1} \frac{4}{10} \ge 1 \Rightarrow l_1 = 2$$

$$l_2 : 2^{l_2} \frac{3}{10} \ge 1 \Rightarrow l_2 = 2$$

$$l_3 : 2^{l_3} \frac{2}{10} \ge 1 \Rightarrow l_3 = 3$$

$$l_4 : 2^{l_4} \frac{1}{10} \ge 1 \Rightarrow l_4 = 4$$

$$\alpha_1 = 0 \Longrightarrow x_1 = 00$$

$$\alpha_2 = (0.4) \ge (.01)_2 \Rightarrow x_2 = 01$$

$$\alpha_3 = (0.7) \ge (.101)_2 \Rightarrow x_3 = 101$$

$$\alpha_4 = 0.9 \ge (.1110)_2 \Rightarrow x_4 = 1110$$

B. THE SHANNO-FANO CODING SCHEME

a. ARRANGE PROBABILITIES IN DECREASING ORDER

b. DIVIDE INTO r SECTIONS OF "NEARLY EQUAL" PROBABILITIES. ASIGN SAME SYMBOL TO EACH COMPONENT IN EACH SECTION.

c. SIMILARLY DIVIDE EACH SECTION AND REPEAT b

EXAMPLE: r = 2

x_1	0.25	0	0			#2
x_2	0.25	0 #1	1			
x_3	.125	1	0	0		#3
x_4	.125	1	0	1 #2		
x_5	0.0625	1	1	0	0	#4
x_6	.0625	1	1	0	1	#3
x_7	.0625	1	1	1	0	#4
x_8	.0625	1	1	1	1	

C. HUFFMAN (OR MAXIMUM EFFICIENCY) CODING

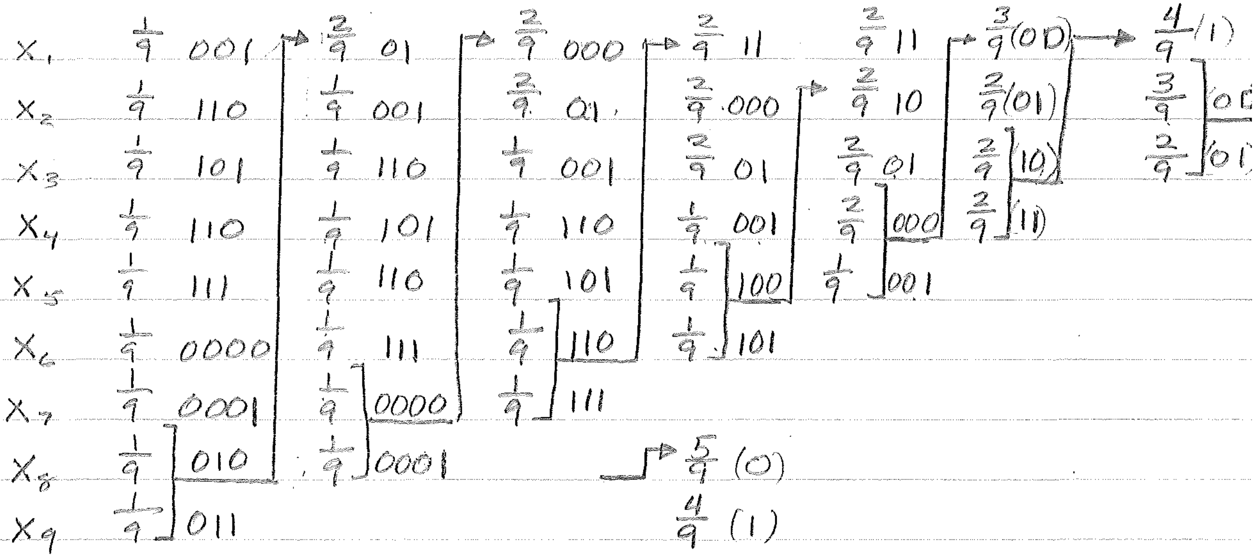
A. a. ARRANGE $P(x_i)$ IN DECENDING ORDER

b. COMBINE SMALLEST r $P(x_i)$'s

c. REPEAT

d. WORK BACKWARDS & FILL THINGS IN

B. EX. $r=2$



C. COMMENTS

- HUFFMAN CODING, THO NOT UNIQUE, WILL ALWAYS GENERATE MINIMUM $\bar{l}$
- WHEN $r > 2$, THEN THE # OF WORDS TO BE CODED <u>MUST</u> BE $q = r + (r-1)\alpha$ $\ni \alpha$ IS ANY INTEGER. USE DUMMY WORDS WITH PROB. ZERO TO COMPLETE THE BOOK

12

Marks

①

1. a. X_1 = Event on first die $\in (1, 2, 3, 4, 5, 6)$

X_2 = " " second die "

We divide our samples according to

$p = P[X_1 + X_2 = 5]$; $\bar{p} = 1 - p = P[X_1 + X_2 \neq 5]$

Clearly, X_1 & X_2 are independent.

Now, sum mutually exclusive events:

$P = P[X_1 = 2, X_2 = 3] + P[X_1 = 3, X_2 = 2]$
$+ P[X_1 = 1, X_2 = 4] + P[X_1 = 4, X_2 = 1]$

5/0

0 3

$\Rightarrow P(X_{1i} X_{2j}) = P(X_{1i}) P(X_{2j})$

and $P(X_i) = \frac{1}{6}$

$\Longrightarrow p = 4\left(\frac{1}{6}\right)\left(\frac{1}{6}\right)$
$= \frac{4}{36} = \frac{1}{9}$ ✓

Associated entropy is

$H(5) = -p \lg p - \bar{p} \lg \bar{p}$
$= +\frac{1}{9} \lg 9 + \frac{8}{9} \lg \frac{9}{8}$
$= 0.349$ NATS

$\log_2 9 = 3.17$ bits.

Self Information

-5

②

1b. $S = \left\{ \begin{matrix} \overset{\downarrow}{H}, & T \\ \frac{1}{2}; p & \frac{1}{2}; \bar{p} \end{matrix} \right\}$ # H = NUMBER OF HEADS

We have here a binomial distribution

$$p(\#H \leq h) = \sum_{k=0}^{h} \binom{10}{k} p^{k} \bar{p}^{10-k} \qquad ; h = 0, 1, 2, \ldots, 10$$

BUT $p = \bar{p} = \frac{1}{2}$

$$p(\#H \leq h) = \sum_{k=0}^{h} \binom{10}{k} 2^{-10} = 2^{-10} \sum_{k=0}^{h} \binom{10}{k}$$

THUS

$$p(\#H \leq 5) = \sum_{k=0}^{5} \binom{10}{k} 2^{-10}$$

$$= 2^{-10}\left[\binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{3} + \binom{10}{4} + \binom{10}{5}\right]$$

$$= 2^{-10}\left[1 + 10 + \frac{10 \cdot 9}{2} + \frac{\overset{5}{\cancel{10}} \cdot \overset{3}{\cancel{9}} \cdot \cancel{8}}{\cancel{3} \cdot \cancel{2}} + \frac{10 \cdot \overset{3}{\cancel{9}} \cdot \cancel{8} \cdot 7}{\cancel{4} \cdot \cancel{3} \cdot \cancel{2}} + \frac{\overset{2}{\cancel{10}} \cdot \overset{3}{\cancel{9}} \cdot \cancel{8} \cdot 7 \cdot 6}{\cancel{5} \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2}}\right]$$

$$= 2^{-10}\left[1 + 10 + 45 + 120 + 210 + 252\right]$$

$$= 638/2^{10} = 0.6230$$

That is The way to do it!

$$P[h, n] = \binom{h}{n} p^{h} \bar{p}^{n-h}$$

$p = \bar{p} = \frac{1}{2}$

$\binom{h}{n} \frac{1}{2}^{10}$

2. $X_1 =$ FIRST ~~throw~~ $= \{1, 2, \ldots, 6)$

$X_2 =$ SECOND " $= \{1, 2, \ldots, 6)$

$\because X_{1_i}$ & X_{2_j} are ind.

$\Rightarrow A$ & B are ind.

Thus (by inspection)

$\Rightarrow P(A) = \frac{1}{2}$

$\Rightarrow P(B) = \frac{1}{2}$

$\Rightarrow P(A/B) = P(A) = \frac{1}{2}$ ✓

$\Rightarrow P(B|A) = P(B) = \frac{1}{2}$

Now, consider C:

$$\begin{aligned} P[C] = {} & P[1,2] + P[1,4] + P[1,6) \\ & + P[2,1] + P[2,3] + P[2,5] \\ & + P[3,2] + P[3,4] + P[3,6] \\ & + P[4,1] + P[4,3] + P[4,5] \\ & + P[5,2] + P[5,4] + P[5,6] \\ & + P[6,1] + P[6,3] + P[6,5] \end{aligned}$$

since $P[X_{1_i}] = P[X_{2_j}] = \frac{1}{6} \quad \forall\ i$ & $j \in 1, \ldots, 6$

& $P[X_{1_i}]\,P[X_{2_j}] = P[X_{1_i}, X_{2_j}] = 1/36$,

we have

$\Rightarrow P[C] = 3 \times 6 \times \frac{1}{36} = \frac{18}{36} = \frac{1}{2}$ ✓

(CONT →)

(2 cont)

NOW

$$P[C,A] = P[1,2] + P(1,4) + P[1,6)$$
$$+ P[3,2) + P(3,4) + P(3,6)$$
$$+ P[5,2) + P(5,4) + P(5,6)$$

$$P(C,A) = 3 \times 3 \times \frac{1}{36} = \frac{9}{36} = \frac{1}{4}$$

Now

$$\Rightarrow P[C/A] = P[A,C]/P(A) \checkmark$$
$$= 1/4 \bigg/ \frac{1}{2} = \frac{1}{2}$$

$$\Rightarrow P[A/C] = P[A,C]/P(C)$$
$$= 1/4 \bigg/ \frac{1}{2} = \frac{1}{2}$$

Clearly, $P[C,B] = P[C,A] = \frac{1}{4}$

Thus

$$\Rightarrow P[B/C] = P[B,C]/P(C)$$
$$= 1/4 \bigg/ \frac{1}{2} = \frac{1}{2}$$

and

$$\Rightarrow P(C/B) = P[B,C]/P(B)$$
$$= 1/4 / \frac{1}{2} = 1/2$$

(CONT →)

(2 cont*)

TWO Events, E & C, are statistically independent iff

$$P[E, C] = P[E] P[C]$$

Here, we we associate the event

$$E = A \text{ AND } B = A, B$$

$$\Rightarrow P(E) = P(A, B)$$

We have established that A & B are independent

$$\Rightarrow P(A) P(B) = P(A, B) = \left(\tfrac{1}{2} \times \tfrac{1}{2}\right) = \tfrac{1}{4}$$

Now, obviously, if both A and B occur (ie, both X_{1i} & X_{2j} are odd), then C may never be true. That is, the sum of 2 odd #'s is even. Thus

$$P[E, C] = P[A, B, C] = 0 \neq \tfrac{1}{4}$$

Thus, the events (AB) and C are not statistically independent.

* For >2 events, stat indep. follows if all n events are pairwise, tri-wise, ... and n-wise independent

3.

$$\begin{array}{c|c|cccc|c}
 & & & & R & & \\
 & & Y_1 & Y_2 & Y_3 & Y_4 & \\
\hline
 & X_1 & 1/\cancel{4} = 10/40 & 0 & 0 & 0 & \frac{1}{4} \\
 & X_2 & \frac{1}{10} = 4/40 & 3/10 & 0 & 0 & 4/40 \\
S & X_3 & 0 & 5/100 & 1/10 & 0 & 15/100 \\
 & X_4 & 0 & 0 & 5/100 & 1/10 & 15/100 \\
 & X_5 & 0 & 0 & 5/100 & 0 & 5/100 \\
\hline
 & & \underbrace{14/40, \quad 35/100 \quad 20/100 \quad 1/10}_{P(Y_j)} & & & & \\
\end{array}
\quad \left.\right\} P(X_i$$

$$P(Y_j) = \Sigma_i \, P(X_i, Y_j) \qquad P(X_j) = \Sigma_j \, P(X_i, Y_j)$$

$$P(Y_1) = \frac{14}{40} = 7/20 \qquad P(X_1) = 1/4$$

$$P(Y_2) = \frac{35}{100} = \frac{7}{20} \qquad P(X_2) = 4/40 = 2/5$$

$$P(Y_3) = \frac{20}{100} = \frac{1}{5} \qquad P(X_3) = 3/25$$

$$P(Y_4) = \frac{1}{10} \qquad P(X_4) = 3/25$$

$$P(X_5) = 1/25$$

CONDITIONAL MATRIX

$$P(S/R) \Rightarrow p(X_i/Y_j) = p(X_i, Y_j)/p(Y_j)$$

$$\begin{array}{c|cccc}
 & Y_1 & Y_2 & Y_3 & Y_4 \\
X_1 & \frac{1}{\cancel{4}} \cdot \frac{\cancel{20}^{5}}{7} = \frac{5}{7} & 0 & 0 & 0 \\
X_2 & \frac{1}{10} \cdot \frac{20}{7} = \frac{2}{7} & \frac{3}{1\cancel{0}} \cdot \frac{2\cancel{0}}{7} = \frac{6}{7} & 0 & 0 \\
X_3 & 0 & \frac{5}{5100} \cdot \frac{20}{7} = \frac{1}{7} & \frac{1}{2} & 0 \\
X_4 & 0 & 0 & \frac{5}{100} \cdot 5 = \frac{1}{4} & 1 \\
X_5 & 0 & 0 & \frac{1}{4} & 0 \\
\hline
\Sigma \rightarrow & 1 \checkmark & 1 \checkmark & 1 \checkmark & 1 \checkmark \\
\end{array}$$

3 (CONT)

CONDITIONAL Matrix $P(R/S) \Rightarrow p(Y_j/X_i)$

$$p(Y_j/X_i) = P(X_i, Y_j)/P(X_i)$$

$$\begin{array}{c|cccc|c}
 & Y_1 & Y_2 & Y_3 & Y_4 & \Sigma \downarrow \\
X_1 & 1 \checkmark & 0 & 0 & 0 & 1 \checkmark \\
X_2 & \frac{1}{10}\cdot\frac{10}{4} = \frac{1}{4} & \frac{3}{10}\cdot\frac{10}{4} = \frac{3}{4} \checkmark & 0 & 0 & 1 \checkmark \\
X_3 & 0 \checkmark & \frac{5}{100}\cdot\frac{100}{15} = \frac{1}{3} & \frac{10}{100}\cdot\frac{100}{15} = \frac{2}{3} & 0 & 1 \checkmark \\
X_4 & 0 & 0 & \frac{5}{100}\frac{100}{15} = \frac{1}{3} \checkmark & \frac{10}{100}\cdot\frac{100}{15} = \frac{2}{3} & 1 \checkmark \\
X_5 & 0 & 0 & 1 & 0 & 1 \checkmark
\end{array}$$

Next, find $H(\overset{S}{\cancel{X}})$

$$H(\overset{S}{\cancel{X}}) = \sum_{i=1}^{5} p(x_i) \lg \frac{1}{p(x_i)}$$

$$= \frac{1}{4}\lg 4 + \frac{2}{5}\lg\frac{5}{2} + 2\cdot\frac{3}{25}\lg\frac{25}{3} + \frac{1}{25}\lg 25$$

$= 1.351$ NATS ✓

$$H(\overset{R}{\cancel{Y}}) = \sum_{j=1}^{4} p(Y_j) \lg \frac{1}{p(Y_j)}$$

$$= 2\cdot\frac{7}{20}\ln\frac{20}{7} + \frac{1}{5}\lg 5 + \frac{1}{10}\lg 10$$

$= 1.287$ NATS

$$H(\overset{RS}{\cancel{X,Y}}) = \sum_i \sum_j p(x_i, Y_j) \lg \frac{1}{p(x_i Y_j)}$$

$$= \frac{1}{4}\lg 4 + 3\times\frac{1}{10}\lg 10 + 3\,\frac{1}{20}\lg 20 + \frac{3}{10}\lg\frac{10}{3}$$

$= 1.848$ NATS

CONT $\Longrightarrow$

3 cont

b. verify $H(X,Y) \leq H(X) + H(Y)$

$1.848 \leq 1.287 + 1.351 = 2.638$

$\boxed{\text{yup}}$

continuing:

$$H(\overset{Y}{R}/\overset{X}{S}) = -\sum_i \sum_j p(X_i, Y_j) \ln p(Y_i/X_i)$$

$$= \cancel{\tfrac{1}{4} \ln 1}^{\,0} + \tfrac{1}{10} \ln 4 + \tfrac{3}{10} \ln \tfrac{4}{3} + \tfrac{5}{100} \ln 3$$
$$+ \tfrac{1}{10} \ln \tfrac{3}{2} + \tfrac{5}{100} \ln 3 + \tfrac{1}{10} \ln 3/2$$
$$+ \cancel{\tfrac{5}{100} \ln 1}^{\,0}$$

$$= \tfrac{1}{10} \ln 4 + \tfrac{1}{10} \ln 3 + \tfrac{3}{10} \ln 4/3 + \tfrac{2}{10} \ln \tfrac{3}{2}$$

$= 0.4159$ NATS

$$H(S/R) = -\sum_i \sum_j p(X_i, Y_j) \ln P(X_i/Y_j)$$

$$= \tfrac{1}{4} \ln \tfrac{7}{5} + \tfrac{1}{10} \ln \tfrac{7}{2}$$
$$+ \tfrac{3}{10} \ln \tfrac{7}{6} + \tfrac{1}{20} \ln 7 + \tfrac{1}{10} \ln 2 +$$
$$+ \tfrac{1}{20} \ln 4 + \tfrac{1}{20} \ln 4 + \cancel{\tfrac{1}{10} \ln 1}^{\,0}$$

$$= \tfrac{1}{4} \ln \tfrac{7}{5} + \tfrac{1}{10} \ln \tfrac{7}{2}$$
$$+ \tfrac{3}{10} \ln \tfrac{7}{6} + \tfrac{1}{20} \ln 7 + \tfrac{1}{10} \ln 2 + \tfrac{1}{10} \ln 4$$

$= 0.5609$ NATS ✓

$H(S/R) + H(R) = 0.5609 + 1.287$ NATS

$\cong 1.848 = H(R,S)$

✓

✓

$\bar{\lambda} = 1-\lambda$

4. $S_1 = \begin{pmatrix} s_1 & s_2 & s_3 \dots s_{q_1} \\ p_1 & p_2 & p_3 \dots p_{q_1} \end{pmatrix} \Rightarrow H(S_1) = H_1$

$S_2 = \begin{pmatrix} k_1 & k_2 & k_3 \dots k_{q_2} \\ Q_1 & Q_2 & Q_3 \dots Q_{q_2} \end{pmatrix} \Rightarrow H(S_2) = H_2$

$$S_\lambda = \begin{pmatrix} s_1 & s_2 \dots s_{q_1} & k_1 & k_2 & \dots k_{q_2} \\ \lambda P_1 & \lambda P_2 \dots \lambda P_{q_1} & \bar{\lambda} Q_1 & \bar{\lambda} Q_2 & \dots \bar{\lambda} Q_{q_2} \end{pmatrix}$$

$$H(S_\lambda) = -\sum_{i=1}^{q_1} \lambda P_i \lg \lambda P_i - \sum_{j=1}^{q_2} \bar{\lambda} Q_j \lg \bar{\lambda} Q_j$$

$$= -\lambda \Sigma_i P_i \lg \lambda - \lambda \Sigma_i P_i \lg P_i$$
$$- \bar{\lambda} \Sigma_j Q_i \lg \bar{\lambda} - \bar{\lambda} \Sigma_j Q_j \lg Q_j$$

$$\Sigma_i P_i = \Sigma_j Q_j = \underline{1}$$

$$\Rightarrow H(S_\lambda) = -\lambda \lg \lambda - \bar{\lambda} \lg \bar{\lambda}$$
$$+ \lambda H_1 + \bar{\lambda} H_2$$

The entropy, $H(\lambda)$, associated with $S = \{\lambda, \bar{\lambda}\}$ is

$$H(\lambda) = -\lambda \lg \lambda - \bar{\lambda} \lg \bar{\lambda}$$

Thus

$$H(S_\lambda) = \lambda H_1 + \bar{\lambda} H_2 + H(\lambda) \Longrightarrow$$

4 (cont)

To maximize, take

$$\frac{dH(S_Y)}{d\lambda} = 0$$

For simplicity, use natural logs & units of nats

$$\frac{dH(S_Y)}{d\lambda} = H_1 - H_2 - \frac{d}{d\lambda}\left[\lambda \lg \lambda + (1-\lambda) \lg (1-\lambda)\right)$$

$$= H_1 - H_2 - \left[(\lg \lambda + 1) + (-1) \lg (1-\lambda) + (1-\lambda)\left(\frac{-1}{1-\lambda}\right)\right)$$

$$= H_1 - H_2 - \left[(\lg \lambda \cancel{+1}) - \lg (1-\lambda) - \cancel{1}\right]$$

$$= H_1 - H_2 + \lg \frac{1-\lambda}{\lambda} = 0$$

Thus

$$H_2 - H_1 = \ln \frac{1-\lambda}{\lambda}$$

$$\frac{1-\lambda}{\lambda} = e^{H_2 - H_1}$$

~~$1 - \lambda = \lambda e^{H_2 - H_1}$~~

$$\lambda / 1-\lambda = e^{H_1 - H_2}$$

$$\lambda = (1-\lambda) e^{H_1 - H_2}$$

$$= e^{H_1 - H_2} - \lambda e^{H_1 - H_2}$$

$$\lambda (1 + e^{H_1 - H_2}) = e^{H_1 - H_2}$$

or

Ys.

$$\lambda_0 = \frac{e^{H_1 - H_2}}{1 + e^{H_1 - H_2}} = \frac{1}{e^{H_2 - H_1} + 1}$$

→

(4 cont)

$$\overline{\lambda_0} = 1 - \frac{1}{e^{H_2-H_1}+1}$$

$$= \frac{e^{H_2-H_1}+1-1}{e^{H_2-H_1}+1} = \frac{e^{H_2-H_1}}{e^{H_2-H_1}+1}$$

$$= \frac{1}{1+e^{H_1-H_2}}$$

Now

$$H(\lambda_0) = -\lambda_0 \ln \lambda_0 - \overline{\lambda_0} \ln \overline{\lambda_0}$$

$$= \frac{1}{e^{H_2-H_1}+1} \ln\left(e^{H_2-H_1}+1\right)$$

$$+ \frac{1}{1+e^{H_1-H_2}} \ln\left(e^{H_1-H_2}+1\right)$$

It is a matter of Substitution ✓

Thus

$$H(S_{\lambda_0})_e = \lambda_0 H_1 + \overline{\lambda_0} H_2 + H(\lambda)$$

$$= \frac{H_1}{e^{H_2-H_1}+1} + \frac{H_2}{e^{H_1-H_2}+1}$$

$$+ \frac{1}{e^{H_2-H_1}+1} \ln\left(e^{H_2-H_1}+1\right)$$

$$+ \frac{1}{1+e^{H_1-H_2}} \ln\left(e^{H_1-H_2}+1\right)$$

-1

$$= \frac{1}{e^{H_2-H_1}+1}\left[H_1 + \ln\left\{e^{H_2-H_1}+1\right\}\right]$$

$$+ \frac{1}{e^{H_1-H_2}+1}\left[H_2 + \ln\left(e^{H_1-H_2}+1\right)\right] \text{ NATS.}$$

where again, we have used base e.

for base r

NATS. $\lg_e x = \frac{\lg_R x}{\lg_R e}$

$$H(S_{\lambda_0})_r = H(S_{\lambda_0})_e \lg_r e \; ; \; r\text{-arys}$$

is the maximum value the entropy may achieve for this mixed source.

5. a. $H(X,Y) \le H(x) + H(Y)$ ✓

with equality ~~a~~ iff X & Y are statistically independent random vectors ✓

b. $H(Y/X) \le H(Y)$ with equality only if X & Y are statistically independent random vectors

c. In general

$$\log_b X = \log_a X / \log_a b$$

BITS is $\lg_2$

∴

$$\# BITS = \lg_2 X = \lg_{10} X / \lg_{10} 2$$

or

$$\# Hartleys = \lg_{10} X = \lg_2 X \lg_{10} 2$$
$$= (\# BITS)(0.3010)$$

$$\therefore \quad 7.2 \overset{BITS}{=} 7.2 \times 0.3013$$
$$= 2.17 \text{ Hartleys}$$ ✓

Similarly, we may show

1.44 BIT = 1. nats

Where nats are from $\lg e$

thus

$$7.2 \text{ BITS} = 7.2 \text{ BITS} \times \frac{1 \text{ NAT}}{1.44 \text{ BITS}}$$
$$= 5 \text{ NATS}$$

✓

→

(5 cont)

d. Additive property on Entropy

$$S = \begin{Bmatrix} s_1 & s_2, & \dots, & s_{n-1} & s_n \\ p_1 & p_2 & , & p_{n-1} & p_n \end{Bmatrix}$$

Divide event s_n into m disjoint ✓ events:

$$s_n = \begin{Bmatrix} r_1 & r_2, & \dots, & r_m \\ q_1 & q_2 & \dots & q_m \end{Bmatrix} \quad ①$$

And define

$$S' = \begin{Bmatrix} s_1, & s_2 \dots & s_{n-1} & r_1 & r_2 & \dots, & r_m \\ p_1 & p_2 & p_{n-1} & q_1/p_n & q_2/p_n & \dots & q_m/p_n \end{Bmatrix}$$

Then the Entropies of S & S' are related by

$$H'\left(p_1, p_2, \dots p_{n-1}, \frac{q_1}{\cancel{p_n}}, \frac{q_2}{\cancel{p_n}}, \dots q_m/\cancel{p_n}\right)$$

See class notes!

$$= H(p_1, p_2, p_3, \dots, p_{n\cancel{-1}})$$

$$+ p_n H_{s_n}\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \dots, q_m/p_n\right)$$

-3

Where H_{s_n} is the entropy associated with the ~~source~~ split. s_n in ①.

You are applying it
You are not saying it
right!

(5 cont)

c. ~~For the source shown~~

Here, we have

$$S = \begin{Bmatrix} E_1 & E_2 & E_3 & E_4 & E_5 \\ 1/10 & 2/10 & 3/10 & 1/10 & 3/10 \end{Bmatrix}$$

where $P[E_5] = P[F_2] + P[F_1]$

The corresponding entropy of S is

$$H(S) = 2 \times \frac{1}{10} \ln 10 + 2\,\frac{3}{10} \ln \frac{10}{3} + \frac{2}{10} \ln \frac{10}{2}$$

$$= 1.505 \text{ NATS}$$

Now consider the augmented source

$$S' = \begin{Bmatrix} E_1 & E_2 & E_3 & E_4 & F_1 & F_2 \\ 1/10 & 2/10 & 3/10 & 1/10 & 1/10 & 2/10 \end{Bmatrix}$$

The corresponding entropy is

$$H(S') = 3\,\frac{1}{10} \ln 10 + 2\,\frac{2}{10} \ln 5 + \frac{3}{10} \ln \frac{10}{3}$$

$$= 1.66 \text{ NATS}$$

NOW $H\left(\frac{q_1}{p_n} \ldots \frac{q_m}{p_n}\right) =$

$$\frac{q_1}{p_n} = \frac{1/10}{3/10} = \frac{1}{3}$$

$$q_2/p_n = \frac{2/10}{3/10} = \frac{2}{3}$$

$$H_{S_n} = \frac{1}{3} \ln 3 + \frac{2}{3} \ln \frac{3}{2}$$

$$= 0.6365$$

$$1.505 + \frac{3}{10}(0.6365) = 1.67 \text{ BITS}$$

6. $S = \begin{Bmatrix} x_1 & x_2 \\ p_1 & p_2 \end{Bmatrix}$

$S' = \begin{Bmatrix} x_1' & x_2' \\ p_1' & p_2' \end{Bmatrix}$

$p_1' = p_1 - \Delta P \qquad p_2' = P_2 + \Delta P$

Assume ~~that~~ $P_1 - \Delta P \geq P_2 + \Delta P \leftarrow$ ①

(which also says $P_1 \geq P_2$ FOR $P_1 > \Delta P > 0$)

ie, the probabilities in S' are numerically closer.

Now:

$H(S) = H = -p_1 \lg p_1 - p_2 \lg p_2$ ✓

$H(S') = H' = -p_1' \lg p_1' - p_2' \lg p_2'$

$= -(p_1 - \Delta P) \lg (P_1 - \Delta P)$ ✓

$- (P_2 + \Delta P) \lg (P_2 + \Delta P)$

Consider the difference

$$H - H' = +P_1 \lg (P_1 - \Delta P) - \Delta P \lg (p_1 - \Delta P) + P_2 \lg (P_2 + \Delta P) + \Delta P \lg (P_2 + \Delta P) - p_1 \lg p_1 - P_2 \lg p_2$$

✓

$$= p_1 \lg \frac{P_1 - \Delta P}{P_1} + \Delta P \lg \frac{P_2 + \Delta P}{P_1 - \Delta P} + p_2 \lg \frac{P_2 + \Delta P}{P_2}$$ ✓ ✓

Use the inequality

$\lg x \leq \cancel{1 - x}$ $(x - 1)$ See p16 Text

$$\left(\cancel{1} - \frac{\Delta P}{P_1} \cancel{-1}\right)P_1 = -\Delta P \qquad +\Delta P$$

$$\left(\cancel{1} + \frac{\Delta P}{P_2} \cancel{-1}\right)P_2$$

(6 CONT)

It follows

$$H - H' \cancel{\leq}\ \geq P_1\left[1 - \left(1 - \frac{\Delta P}{P_1}\right)\right] + P_2\left[1 - \left(1 + \frac{\Delta P}{P_2}\right)\right] + \Delta P \lg \frac{P_2 + \Delta P}{P_1 - \Delta P}$$

$$\cancel{\leq}\ \geq (\Delta P) + (-\Delta P) + \Delta P \lg \frac{P_2 + \Delta P}{P_1 - \Delta P}$$

$$\cancel{\leq}\ \geq \Delta P \lg \frac{P_2 + \Delta P}{P_1 - \Delta P}$$

You are lucky you got the right result by using the wrong inequality

Now, from ①, $\frac{P_2 + \Delta P}{P_1 - \Delta P} \leq 1.$

$$\Rightarrow \lg \frac{P_2 + \Delta P}{P_1 - \Delta P} \leq 0$$

(Since $\lg(\cdot)$ is a monotonic increasing function)

Thus, since $\Delta P > 0$

$$H - H' \leq \Delta P \lg \frac{P_2 + \Delta P}{P_1 - \Delta P} \leq 0 \checkmark$$

$$H_1 \leq H'$$

or, the entropy of the perturbed source is greater, or, equivalently, more "uncertain"

QED

$\boxed{-3}$

BOB MARKS

7. FROM TEXT: PROB. 2-5 p. 41

$\frac{86}{150}$

a. $S = \{s_1, s_2, \ldots, s_i, \ldots\}$

$P(s_i) = a\alpha^i$ ①

WE REQUIRE THAT

$$1 = \sum_{i=1}^{\infty} P(s_i) = \Sigma_i\, a\alpha^i \quad ②$$

$$= a\,\Sigma_i\,\alpha^i$$

$$= a\,\frac{\alpha}{1-\alpha} \checkmark$$

$$\Rightarrow a = \frac{1-\alpha}{\alpha} \checkmark \quad ③$$

$$\text{ie}\quad P(s_i) = \frac{1-\alpha}{\alpha}\,\alpha^i \quad ④$$

b. $H(S) \overset{\Delta}{=} \Sigma_i\, P(s_i)\,\lg 1/P(s_i)$

$$= -\Sigma_i\,\frac{1-\alpha}{\alpha}\alpha^i \lg \frac{1-\alpha}{\alpha}\alpha^i$$

$$= -\Sigma_i\,\frac{1-\alpha}{\alpha}\alpha^i\left[\lg\frac{1-\alpha}{\alpha} + \lg\alpha^i\right]$$

$$= -\Sigma_i\,\frac{1-\alpha}{\alpha}\alpha^i\left[\lg\frac{1-\alpha}{\alpha} + i\lg\alpha\right]$$

$$= -\left[\Sigma_i\,\frac{1-\alpha}{\alpha}\alpha^i\right]\lg\frac{1-\alpha}{\alpha}$$

$$+ \frac{\alpha-1}{\alpha}\lg\alpha\,\Sigma_i\, i\,\alpha^i$$

BUT $\quad \Sigma_i\,\frac{1-\alpha}{\alpha}\alpha^i = \Sigma_i\,P(s_i) = 1$

AND $\quad \Sigma_i\, i\,\alpha^i = \frac{\alpha}{(1-\alpha)^2}$

$$\Rightarrow H(S) = -\lg\frac{1-\alpha}{\alpha} - \frac{1-\alpha}{\alpha}\,\frac{\alpha}{(1-\alpha)^2}\lg\alpha$$

$$= \lg\frac{\alpha}{1-\alpha} - \frac{1}{1-\alpha}\lg\alpha$$

$$= \lg\alpha - \lg 1-\alpha - \frac{1}{1-\alpha}\lg\alpha$$

$$= \left[1 - \frac{1}{1-\alpha}\right]\lg\alpha - \lg 1-\alpha$$

$$= \frac{-\alpha}{1-\alpha}\lg\alpha - \lg(1-\alpha) \checkmark$$

$$= \frac{\alpha}{\alpha-1}\lg\alpha - \lg(1-\alpha) \checkmark \quad ⑤$$

Ace!

HENCEFORTH, USE LOG BASE e. ie $\lg \overset{\Delta}{=} \lg_e$

(NOTE THAT FOR $\alpha = \frac{1}{2}$, WE GET $H(S) = 2$ BITS)

1. $\lim\limits_{\alpha\to 0^+} H(S) = \lim\limits_{\alpha\to 0^+} -\alpha \lg\alpha$

$= \lim\limits_{\alpha\to 0^+} -\lg\alpha / 1/\alpha$

USING LA HOPITAL:

$\lim\limits_{\alpha\to 0^+} H(S) = \lim\limits_{\alpha\to 0^+} 1/x \big/ 1/x^2$

$= \lim\limits_{\alpha\to 0^+} X = 0^+$

2. $\lim\limits_{\alpha\to 1^-} H(S) = \lim\limits_{\alpha\to 1^-} \frac{\lg\alpha}{\alpha-1} - \lg(1-\alpha)$

NOW $\lim\limits_{\alpha\to 1^-} \frac{\lg\alpha}{\alpha-1} = \lim\limits_{\alpha\to 1^-} \frac{1/\alpha}{1} = \lim\limits_{\alpha\to 1^-} \frac{1}{\alpha} = \infty$

$\therefore \lim\limits_{\alpha\to 1^-} H(S) = \infty - (-\infty) = \infty$

3. CHECKING FOR EXTREMA:

$$\frac{d}{d\alpha} H(S) = \frac{d}{d\alpha}\frac{\alpha}{\alpha-1}\lg\alpha - \frac{d}{d\alpha}\lg(1-\alpha)$$

$$= \left[\frac{\alpha}{\alpha-1}\cdot\frac{1}{\alpha} + \frac{(\alpha-1)-1\alpha}{(\alpha-1)^2}\lg\alpha\right] - \frac{(-1)}{1-\alpha}$$

$$= -\frac{1}{1-\alpha} - \frac{1}{(1-\alpha)^2}\lg\alpha + \frac{1}{1-\alpha}$$

$$= -\frac{1}{(1-\alpha)^2}\lg\alpha$$

$\therefore$ NO EXREMA FOR FINITE α, SINCE $H(S)\big|_{\lambda=1} = \infty$

4. ON α

- IN ORDER FOR ALL $P(S_i)$ IN ① TO BE POSITIVE, WE REQUIRE $\alpha > 0$
- IN ORDER FOR THE (GEOMETRIC) SERIES IN ② TO CONVERGE, $|\alpha| < 1$

THUS

$$0 < \alpha < 1 \qquad ⑥$$

TO PLOT H(S) IN ⑤, USE HP-25 PROGRAM:

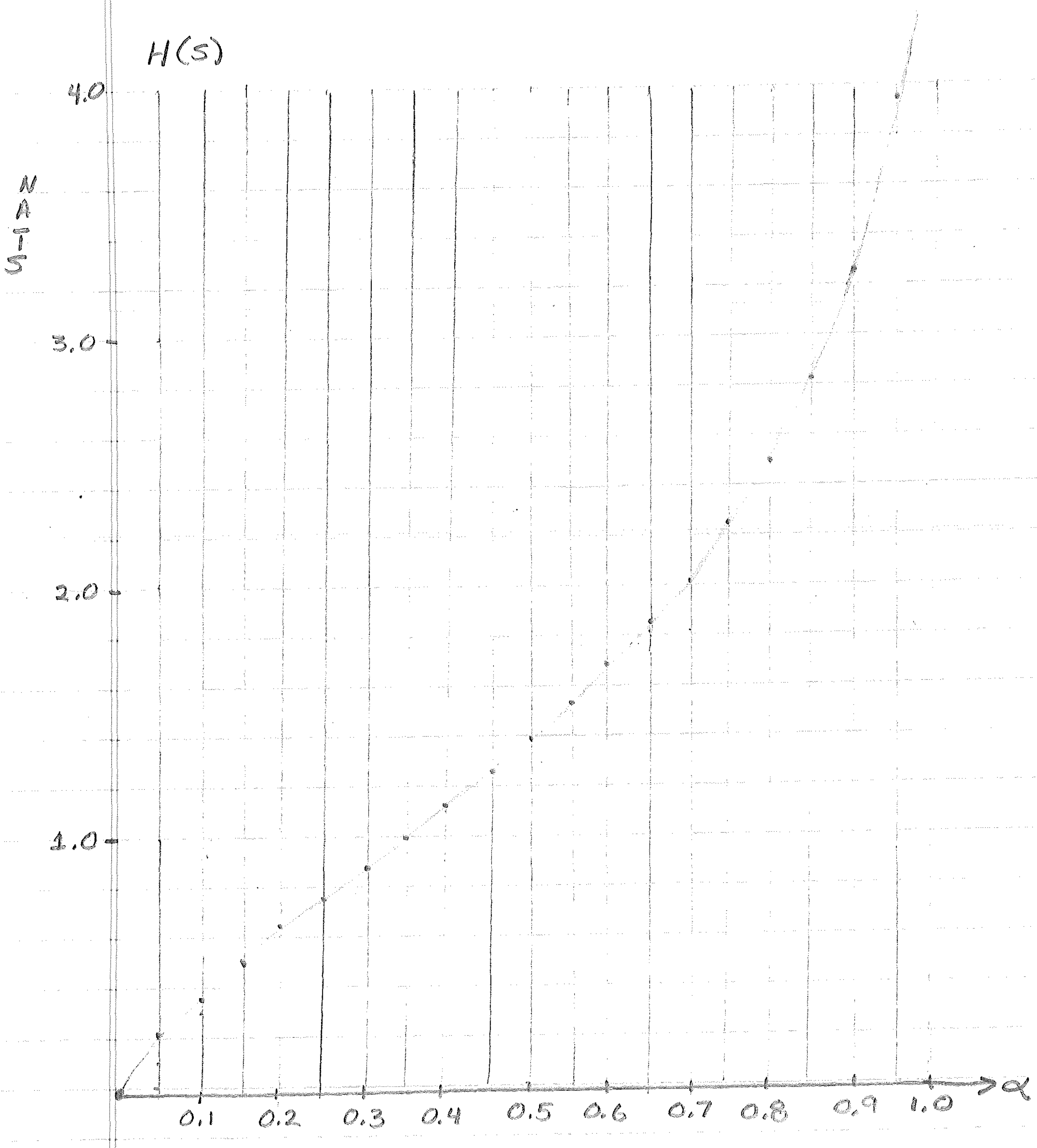
```
(α)

STO 0      -        1       ln

ln         ÷        RCL 0   +

RCL 0      RCL 0    -       GTO 00

1          x        1/x
```


H(s)
NATS
4.0
3.0
2.0
1.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
α

THE STRIKING ASPECT OF THIS PROBLEM IS THE INFINITE ENTROPY FOR α NEAR 1. THAT IS, WE CAN "TUNE" OUR SOURCE TO AS HIGH AN ENTROPY AS DESIRED BY LETTING α GO CORRESPONDINGLY CLOSE TO 1.

INTUITIVELY, WHAT IS HAPPENING IS AS FOLLOWS. FOR $\alpha \approx 0$, THE PROBABILITIES ARE ROUGHLY:

$$(P_1, P_2, P_3, \ldots, P_i, \ldots) \approx (1, 0, 0, \ldots, 0, \ldots)$$

WHERE WE HAVE INTERPRETED $H(S)\big|_{i=0} = 1 \Leftarrow 0^0$

FOR α NEAR 1, WE <u>ESSENTIALLY</u> HAVE

$$(P_1, P_2, \ldots, P_i, \ldots) \approx (\epsilon, \epsilon, \epsilon, \ldots, \epsilon, \ldots)$$

WHERE $\epsilon \ll 1$. THAT IS, WE APPROACH A CONDITION OF HAVING AN INFINITE NUMBER OF INFINTESIMALLY EQUALLY PROBABLE EVENTS. THIS, THEN, CONSTITUTES A CORRESPONDING APPROACH TO INFINITE ENTROPY.

Quiz 2, 8-6-76 EE 5325 Summer 2, 1976 J-C. Prabhakar

check [illegible]

1. During class work it was shown that the efficiency of a coding scheme could be enhanced by source extension. The purpos of this exercise is to investigate the effect of no. of symbols (r) in the encoding alphabet on the efficiency of compact codes.

Consider an ensemble of 10 messages arranged, for your convenience, in the non-increasing order of probabilities.

$$X : \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$$

$$P : \left\{\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{64}\right\}$$

Let $r = 2, 3, 4, 5$ and 7.

~~These~~ These points on the horizontal axis should enable you to figure out the dependence of efficiency on r.

Determine the compact codes by Shannon-Fano method or Huffman's procedure and determine the efficiency for each r. Make a plot and list your conclusion!

Note: When probs. are given in a certain "form", the codes resulting from Shannon-Fano method are compact.

2. A Source alphabet $S : \{S_1, S_2\}$ has prob. distribution

$$P : \left(\tfrac{3}{4}, \tfrac{1}{4}\right)$$

Derive compact codes for S, its 2nd, and 4th. extensions. Calculate the three efficiencies and verify, approximately, the result

$$\frac{\overline{L_n}/n}{H(S)} \longrightarrow 1 \text{ for "large" } n$$

$\left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2$

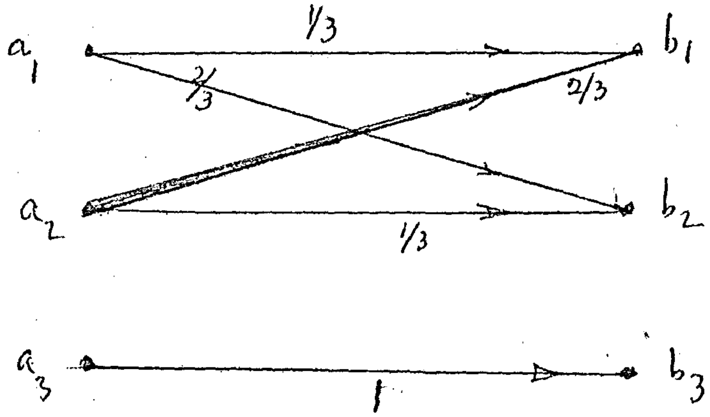
$\frac{1}{9} - \frac{4}{9}$

$-\frac{3}{9} = -\frac{1}{3}$

Note: This prob. distribution requires Huffman Coding for Compactness

3. A channel is described by the following Source-Receiver relationship. The numbers are conditional probabilities.

$\frac{1}{9} - \frac{4}{9} = (\frac{3}{9})$



Use Muroga's technique to determine the maximum of transinformation that can be associated with the arrival of the received message i.e the channel Capacity

the units to your answer.

Bailon

4. ✓ Derive the following codes for the ensemble

$$A : \{a_1, a_2, a_3, a_4, a_5, a_6, a_7\}$$

$$P : \left\{\frac{4}{10}, \frac{2}{10}, \frac{12}{100}, \frac{8}{100}, \frac{8}{100}, \frac{8}{100}, \frac{4}{100}\right\}$$

(a) Shannon Coding
(b) Shannon-Fano Coding
(c) Huffman's Optimal coding.

Use (0,1) as the alphabet and verify the conclusion arrived at in class that (c) will yield compact codes. (a) & (b) may <u>not</u> be so.

Note: Since compactness is related to the average length of a scheme, you need not evaluate $H(A)$.

5. ✓ (i) Define the following
(a) Non-singular codes
(b) Uniquely Decodable Codes
(c) Instantaneous Code
(d) Prefix property
(e) Redundancy of a Coding Scheme
(f) The average length of a Coding Scheme
(g) An Independent Channel.

(ii) An ensemble has 8 words with lengths:
No. of words of Length 1 = 0

Barton

No. of words of length 2 = 3

- - - ——— 3 = 1

- - - - - - - 4 = 4

Find the no. of symbols in the encoding alphabet required to generate an Instantaneous Scheme. Determine the resulting words

(iii) Which of the sets of word lengths shown below are acceptable for a uniquely decodable codes when

(a) The alphabet is (0, 1)

(b) - ——— (0, 1, 2)

No. of words of length li in each code	word lengths li 1	2	3	4	5
Code A	2	1	2	4	1
Code B	2	2	2	3	1
Code C	1	4	6	0	0
Code D.	2	2	2	2	3

Is the test you are applying both necessary and sufficient?

6.* Prob. 4-8 p 92 Text

7* Prob 5-17 p 146 Text.

* Prob 6, 7 are Take home due Tuesday 8-10-76.

Notes: 1. All HW* is due 8-13-76. Delay may carry a mild penalty.

2. If the average score on this quiz is less than 60, there will be a final quiz on Aug 17 or 18.

3. Reading of Shannon's paper is mandatory and will count as a HW* problem.

MARKS 1-1

-5 pwh (2)

-4 pwh 5

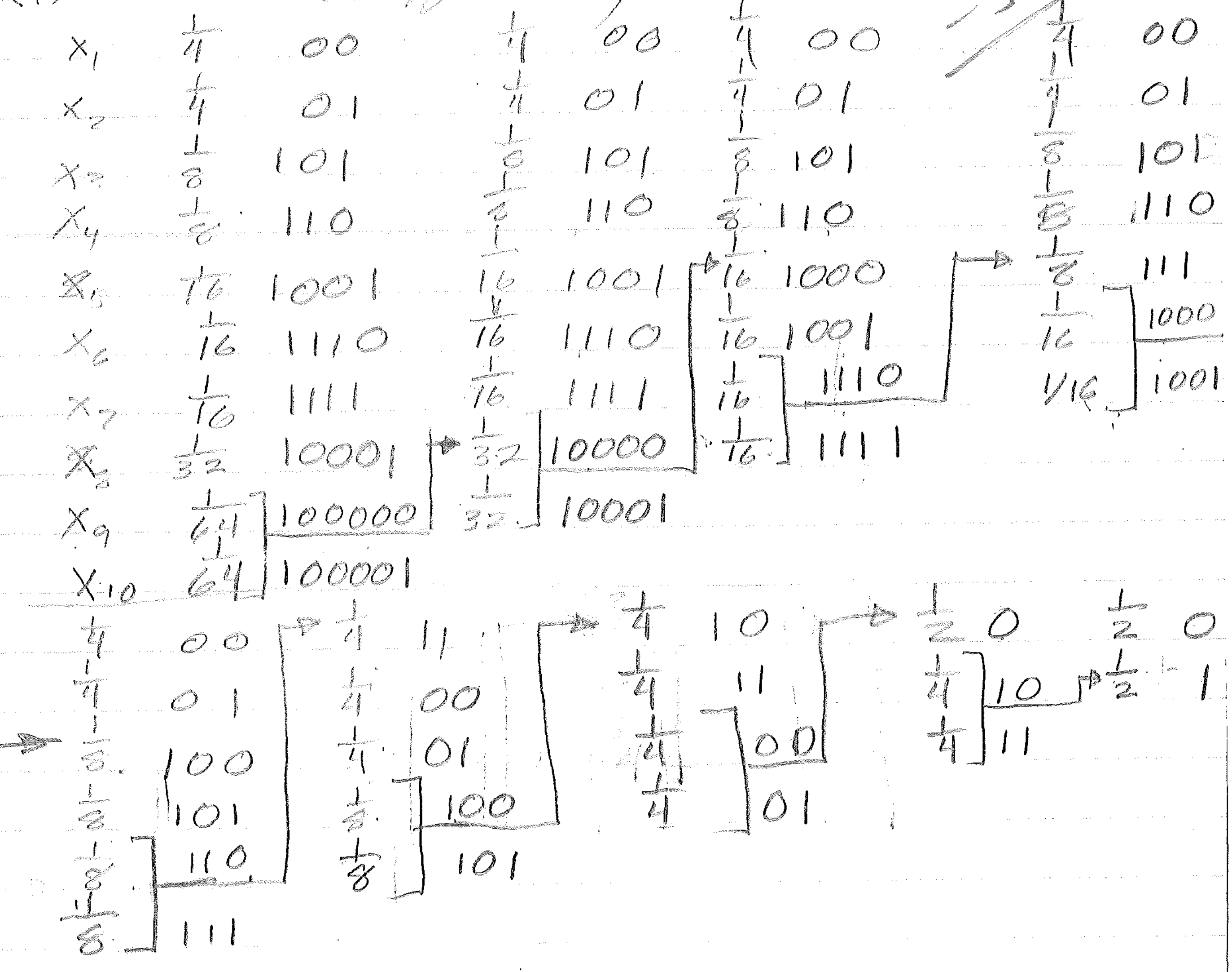
-3

1-19. In general

η = EFFICIENCY

$= \bar{L} / H_r(S)$

(a) r = 2 (Huffman)



$$\bar{L} = \frac{2}{4} + \frac{2}{4} + \frac{3}{8} + \frac{3}{8} + \frac{4}{16} + \frac{4}{16} + \frac{4}{16} + \frac{5}{32} + \frac{6}{64} + \frac{6}{64}$$

$$= 1 + \frac{6}{8} + \frac{12}{16} + \frac{5}{32} + \frac{12}{64}$$

$$= \frac{1}{32}\left[24 + 24 + 5 + 6\right]$$

$$= 59/32$$

(b) $r=3 \Rightarrow q = r+(r-1)\alpha = 3+2\alpha \Rightarrow \alpha = 4 \Rightarrow q = 11$

x_1	$\frac{1}{4}$	2
x_2	$\frac{1}{4}$	00
x_3	$\frac{1}{8}$	02
x_4	$\frac{1}{8}$	10
x_5	$\frac{1}{16}$	11
x_6	$\frac{1}{16}$	12
x_7	$\frac{1}{16}$	010
x_8	$\frac{1}{32}$	011
x_9	$\frac{1}{64}$	0120
x_{10}	$\frac{1}{64}$	0121
x_{11}	0	0122

$\frac{1}{4}$	2
$\frac{1}{4}$	00
$\frac{1}{8}$	02
$\frac{1}{8}$	10
$\frac{1}{16}$	11
$\frac{1}{16}$	12
$\frac{1}{16}$	010
$\frac{1}{32}$	011
$\frac{1}{32}$	012

$\frac{1}{4}$	2
$\frac{1}{4}$	00
$\frac{1}{8}$	01
$\frac{1}{8}$	02
$\frac{1}{8}$	10
$\frac{1}{16}$	11
$\frac{1}{16}$	12

$\frac{1}{4}$	1
$\frac{1}{4}$	2
$\frac{1}{4}$	00
$\frac{1}{8}$	01
$\frac{1}{8}$	02

$\frac{1}{2}$	0
$\frac{1}{4}$	1
$\frac{1}{4}$	2

$$\bar{L}_3 = \frac{1}{4} + \frac{2}{4} + \frac{2}{8} + \frac{2}{8} + \frac{2}{16} + \frac{2}{16} + \frac{3}{16} + \frac{3}{32} + \frac{4}{64} + \frac{4}{64}$$
$$= \frac{3}{4} + \frac{4}{8} + \frac{7}{16} + \frac{3}{32} + \frac{8}{64}$$
$$= \frac{1}{32}\left[24+16+14+3+4\right]$$
$$= 61/32$$

(c) $r=4 \Rightarrow q = r+(r-1)\alpha = 4+3\alpha \Rightarrow \alpha = 2 \Rightarrow q=10$

x_1	$\frac{1}{4}$	1	$\frac{1}{4}$	1	$\frac{3}{8}$	0
x_2	$\frac{1}{4}$	2	$\frac{1}{4}$	2	$\frac{1}{4}$	1
x_3	$\frac{1}{8}$	01	$\frac{1}{8}$	3	$\frac{1}{4}$	2
x_4	$\frac{1}{8}$	02	$\frac{1}{8}$	01	$\frac{1}{8}$	3
x_5	$\frac{1}{16}$	03	$\frac{1}{8}$	02		
x_6	$\frac{1}{16}$	04	$\frac{1}{16}$	03		
x_7	$\frac{1}{16}$	30	$\frac{1}{16}$	04		
x_8	$\frac{1}{32}$	31				
x_9	$\frac{1}{64}$	32				
x_{10}	$\frac{1}{64}$	33				

$$\bar{L} = \frac{3}{4} + \frac{4}{8} + \frac{6}{16} + \frac{2}{32} + \frac{4}{64}$$
$$= \frac{3}{4} + \frac{4}{8} + \frac{6}{16} + \frac{2}{16}$$
$$= \frac{3}{4} + \frac{1}{2} + \frac{1}{2}$$
$$= 1 + \frac{3}{4}$$
$$= 7/4$$

(d) $r = 5 \Rightarrow q = r + (r-1)\alpha = 5 + 4\alpha \Rightarrow \alpha = 2 \therefore q =$

X_1	$\frac{1}{4}$	1	$\frac{1}{4}$	1	$\frac{1}{4}$	0
X_2	$\frac{1}{4}$	2	$\frac{1}{4}$	2	$\frac{1}{4}$	1
X_3	$\frac{1}{8}$	3	$\frac{1}{8}$	3	$\frac{1}{4}$	2
X_4	$\frac{1}{8}$	4	$\frac{1}{8}$	4	$\frac{1}{8}$	3
X_5	$\frac{1}{16}$	00	$\frac{1}{16}$	00	$\frac{1}{8}$	4
X_6	$\frac{1}{16}$	01	$\frac{1}{16}$	01		
X_7	$\frac{1}{16}$	02	$\frac{1}{16}$	02		
X_8	$\frac{1}{32}$	04	$\frac{1}{32}$	03		
X_9	$\frac{1}{64}$	030	$\frac{1}{32}$	04		
X_{10}	$\frac{1}{64}$	031				
X_{11}	0	032				
X_{12}	0	033				
X_{13}	0	034				

✓ $\bar{L} = \frac{2}{4} + \frac{2}{8} + \frac{6}{16} + \frac{1}{16} + \frac{6}{64}$

$= \frac{3}{4} + \frac{7}{16} + \frac{3}{32}$

$= \frac{1}{32}\left[24 + 14 + 3\right]$

$= 41/32$

$$\begin{array}{r} 24 \\ 14 \\ 3 \\ \hline 41 \end{array}$$

e. $r=7 \Rightarrow q=7+6\alpha \Rightarrow \alpha=1 \Rightarrow q=13$

x_1	$\frac{1}{4}$	0	$\frac{1}{4}$	0
x_2	$\frac{1}{4}$	1	$\frac{1}{4}$	1
x_3	$\frac{1}{8}$	3	$\frac{1}{8}$	2
x_4	$\frac{1}{8}$	4	$\frac{1}{8}$	3
x_5	$\frac{1}{16}$	5	$\frac{1}{8}$	4
x_6	$\frac{1}{16}$	6	$\frac{1}{16}$	5
x_7	$\frac{1}{16}$	20	$1/16$	6
x_8	$\frac{1}{32}$	21		
x_9	$\frac{1}{64}$	22		
x_{10}	$\frac{1}{64}$	23		
x_{11}	0	24		
x_{12}	0	25		
x_{13}	0	26		

$$\overline{L} = \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \frac{2}{32} + \frac{4}{64}$$

$$= \frac{2}{4} + \frac{1}{4} + \frac{2}{16}$$

$$= \frac{2}{4} + \frac{1}{4} + \frac{1}{8}$$

$$= 7/8$$

Summer 2 1976 J C Prabhakar

Inf. Theory EE5325

1(a) A pair of 6-face dice are thrown and the sum of their faces = 5. What is the information content of this message.

(b) Determine the probability that at most 5 heads will occur in 10 independent tosses of a coin. Assume the elemental probabilities

$$P(\text{Head}) = P(\text{Tail}) = 1/2.$$

2. Two dice are thrown resulting in

Event A — "odd faces on first dice"

Event B — "Odd faces on second dice"

Event C — "Sum of faces odd"

Find $P(A)$, $P(B)$, $P(C)$, $P(A/B)$ $P(B/A)$, $P(C/A)$,

$P(A/c)$, $P(B/c)$, $P(C/B)$

Consider the two events (A and B) and C. Are they statistically independent? Your answer should includes the use of definition of statistical Independence

3. A communication system has a source

$$S : [x_1, x_2, x_3, x_4, x_5]$$

and a Receiver

$$R : (y_1, y_2, y_3, y_4)$$

connected by a channel with Joint Prob. Matrix

$$P(S,R) : \begin{array}{c|cccc} & y_1 & y_2 & y_3 & y_4 \\ \hline x_1 & \frac{1}{4}\checkmark & 0 & 0 & 0 \\ x_2 & \frac{1}{10}\checkmark & \frac{3}{10}\checkmark & 0 & 0 \\ x_3 & 0 & \frac{5}{100}\checkmark & 0.10\checkmark & 0 \\ x_4 & 0 & 0 & \frac{5}{100}\checkmark & 0.10\checkmark \\ x_5 & 0 & 0 & \frac{5}{100}\checkmark & 0 \end{array}$$

Determine $P(x_i)\ \forall i$, $P(y_j)\ \forall j$. Build up the Conditional matrices $P(S/R)$ and $P(R/S)$ and thence determine $H(S)$, $H(R)$, $H(R/S)$, $H(S/R$

and $H(R, S)$.

Verify that (a) $H(R,S) = H(R) + H(S/R)$

4.

Consider 2 zero memory sources

$$S_1 : (s_1, s_2, \text{——} s_{q_1}),$$

$$S_2 : (k_1, k_2, \text{——} k_{q_2})$$

with S_1 having prob. structure $P_1, P_2, \text{——} P_{q_1}$

and S_2 - - - - ——— - $Q_1, Q_2, \text{——} Q_{q_2}$.

Let $H(S_1) = H_1$, $H(S_2) = H_2$.

Form a new zero memory source $S(\lambda)$ with $q_1 + q_2$ symbols such that the first q_1 symbols of $S(\lambda)$ have probs. λP_i, $i = 1, 2, \text{——} q_1$ and the last q_2 symbols have probs $(1-\lambda) Q_i$, $i = 1, 2, \text{——} q_2$.

Find the value of λ which maximizes $H[S(\lambda)]$ in terms of H_1 and H_2. What is that max. value?

5. (i) Complete the following statements.

a). $H(X, Y) \leq H(X) + \text{- - - -}$ with equality iff - - - -

b). $H(\cdots / \cdots) \leq H(Y)$ with equality iff - - - -

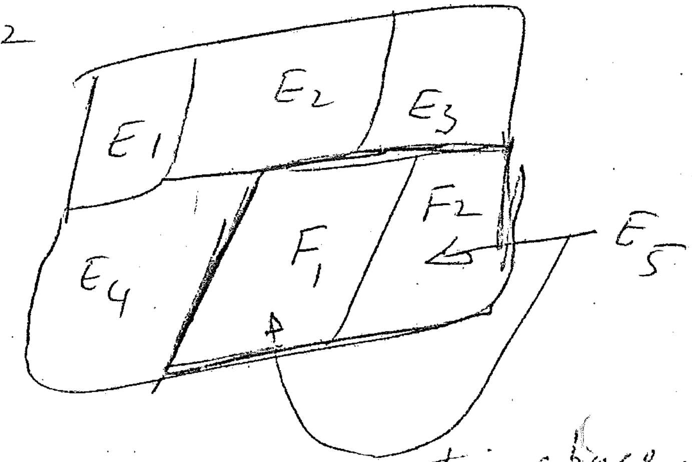
(c) 7.2 bits = ---- Hartleys = --- nats

(d) The additivity requirement on the entropy function is, – – – –

– – – – – – – – – –

– – – – – – –

(ii) Consider the prob space, as under, with E_5 further dissected into F_1 and F_2



Let $P(E_1) = P(E_4) = 0.1$

$P(E_2) = 0.2, P(E_3) = 0.3$

$P(F_1) = \frac{1}{10} \quad P(F_2) = \frac{2}{10}$.

Verify the additive ~~law~~ property of the entire space.

6. Let a source $S : (x_1, x_2)$ have Prob. Structure

$$P : (p_1, p_2).$$

Let this structure be disturbed such that

$$P' = (p_1 - \Delta p, \ p_2 + \Delta p). \ , \Delta p > 0$$

and $p_1 - \Delta p \geq p_2 + \Delta p$

Show that the disturbed source is more "uncertain".

7.* Problem 2-5 text p41 due 7-26-76

$\textcircled{12}$

Marks

①

1. a. X_1 = Event on first die $\in (1, 2, 3, 4, 5, 6)$

X_2 = " " second die "

We divide our sample according to

$p = P[X_1 + X_2 = 5]$; $\bar{p} = 1 - p = P[X_1 + X_2 \neq 5]$

Clearly, X_1 & X_2 are independent.

/5/0

Now, sum mutually exclusive events:

$$P = P[X_1 = 2, X_2 = 3] + P[X_1 = 3, X_2 = 2] + P[X_1 = 1, X_2 = 4] + P[X_1 = 4, X_2 = 1]$$

① 3

$P(X_1, X_2) = P(X_1)\,P(X_2)$

and $P(X_i) = \frac{1}{6}$

$$\Rightarrow p = 4\left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{4}{36} = \frac{1}{9} \checkmark$$

Associated entropy is

$$H(S) = -p \lg p - \bar{p} \lg \bar{p} = \frac{1}{9} \lg 9 + \frac{8}{9} \lg \frac{9}{8} = 0.349 \text{ NATS}$$

Self Information $\Longrightarrow \log_2 9 = 3.17$ bits

-5

②

1b. $S = \left\{ \overset{\downarrow}{H}, T \right\}$ # H = NUMBER OF HEADS

$\left\{ \frac{1}{2} p, \frac{1}{2} \bar{p} \right\}$

We have here a binomial distribution

$$p(\#H \leq h) = \sum_{k=0}^{h} \binom{10}{k} p^k \bar{p}^{10-k} \quad ; h = 0, 1, 2, \ldots, 10$$

BUT $p = \bar{p} = \frac{1}{2}$

$$p(\#H \leq h) = \sum_{k=0}^{h} \binom{10}{k} 2^{-10} = 2^{-10} \sum_{k=0}^{h} \binom{10}{k}$$

THUS

$$p(\#H \leq 5) = \sum_{k=0}^{5} \binom{10}{k} 2^{-10}$$

$$= 2^{-10} \left[\binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{3} + \binom{10}{4} + \binom{10}{5} \right]$$

$$= 2^{-10} \left[1 + 10 + \frac{10 \cdot 9}{2} + \frac{\overset{5}{\cancel{10}} \cdot \overset{3}{\cancel{9}} \cdot 8}{\cancel{3} \cdot \cancel{2}} + \frac{10 \cdot \overset{3}{\cancel{9}} \cdot \cancel{8} \cdot 7}{\cancel{4} \cdot \underset{3}{\cancel{3}} \cdot \underset{2}{\cancel{2}}} + \frac{\cancel{10} \cdot \cancel{9} \cdot \cancel{8} \cdot 7 \cdot 6}{\cancel{5} \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2}} \right]$$

$$= 2^{-10} \left[1 + 10 + 45 + 120 + 210 + 252 \right]$$

$$= 638/2^{10} = 0.6230$$

That is The way to

don't do it!

$$P[h, n] = \binom{h}{n} p^h \bar{p}^{n-h}$$

$$p = \bar{p} = \frac{1}{2}$$

$$\binom{h}{n} \frac{1}{\;}^{10}$$

2. X_1 = FIRST throw = $\{1, 2, \ldots, 6)$

X_2 = SECOND " = $\{1, 2, \ldots, 6)$

∵ X_{1_i} & X_{2_j} are ind.

$\Rightarrow A$ & B are ind.

Thus (by inspection)

$\Rightarrow P(A) = \frac{1}{2}$

$\Rightarrow P(B) = \frac{1}{2}$

$\Rightarrow P(A/B) = P(A) = \frac{1}{2}$ ✓

$\Rightarrow P(B|A) = P(B) = \frac{1}{2}$

Now, consider C:

$$P[C] = P[1,2] + P[1,4] + P[1,6) + P[2,1] + P[2,3] + P[2,5] + P[3,2] + P[3,4] + P[3,6] + P[4,1] + P[4,3] + P[4,5] + P[5,2] + P[5,4] + P[5,6] + P[6,1] + P[6,3] + P[6,5]$$

since $P[X_{1_i}] = P[X_{2_j}] = \frac{1}{6} \quad \forall\ i$ & $j \in 1, \ldots, 6$

& $P[X_{1_i}]P[X_{2_j}] = P[X_{1_i}, X_{2_j}] = 1/36$,

we have

$\Rightarrow P[C] = 3 \times 6 \times \frac{1}{36} = \frac{18}{36} = \frac{1}{2}$ ✓

(CONT →)

(2 cont)

NOW

$$P[C, A] = P[1, 2] + P(1, 4) + P[1, 6) + P[3, 2) + P(3, 4) + P(3, 6) + P[5, 2) + P(5, 4) + P(5, 6)$$

$$P(C, A) = 3 \times 3 \times \frac{1}{36} = \frac{9}{36} = \frac{1}{4}$$

Now

$$\Rightarrow P[C/A] = P[A, C]/P(A) \checkmark$$
$$= 1/4 \Big/ \frac{1}{2} = \frac{1}{2}$$

$$\Rightarrow P[A/C] = P[A, C]/P(C)$$
$$= 1/4 \Big/ \frac{1}{2} = \frac{1}{2}$$

Clearly, $P[C, B] = P[C, A] = \frac{1}{4}$

Thus

$$\Rightarrow P[B/C] = P[B, C]/P(C)$$
$$= 1/4 \Big/ \frac{1}{2} = \frac{1}{2}$$

and

$$\Rightarrow P(C/B) = P[B, C]/P(B)$$
$$= 1/4 \Big/ \frac{1}{2} = 1/2$$

(CONT →)

(2 cont)

TWO* Events, E & C, are statistically independent iff

$$P[E, C] = P[E]P[C]$$ ✓

Here, we we associate the event

$$E = A \text{ AND } B = A, B$$

$$\Rightarrow P(E] = P(A, B)$$

We have established that A & B are independent

$$\Rightarrow P(A)P(B) = P(A, B) = \left(\tfrac{1}{2} \times \tfrac{1}{2}\right) = \tfrac{1}{4}$$

Now, obviously, if both A and B occur (ie, both X_{1i} & X_{2j} are odd), then C may never be true. That is, the sum of 2 odd #'s is even. Thus

$$P[E, C] = P[A, B, C] = 0 \neq \tfrac{1}{4}$$

Thus, the events (AB) and C are not statistically independent.

* For >2 events, stat indep. follows if all n events are pairwise, tri-wise, ... and n-wise independent

3.

$$
\begin{array}{c|cccc|c}
 & \multicolumn{4}{c|}{R} & \\
 & Y_1 & Y_2 & Y_3 & Y_4 & \\
\hline
X_1 & 1/4 = 10/40 & 0 & 0 & 0 & \frac{1}{4} \\
X_2 & \frac{1}{10} = 4/40 & 3/10 & 0 & 0 & 4/10 \\
S \quad X_3 & 0 & 5/100 & 1/10 & 0 & 15/100 \\
X_4 & 0 & 0 & 5/100 & 1/10 & 15/100 \\
X_5 & 0 & 0 & 5/100 & 0 & 5/100 \\
\hline
 & 14/40 & 35/100 & 20/100 & 1/10 &
\end{array}
\Bigg\} P(X_i)
$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{P(Y_j)}$$

$p(Y_j) = \Sigma_i\, P(X_i, Y_j)$ $\qquad$ $P(X_j) = \Sigma_j\, P(X_i, Y_j)$

$P(Y_1) = \frac{14}{40} = 7/20$ $\qquad$ $P(X_1) = 1/4$

$P(Y_2) = \frac{35}{100} = \frac{7}{20}$ $\qquad$ $P(X_2) = 4/10 = 2/5$

$P(Y_3) = \frac{20}{100} = \frac{1}{5}$ $\qquad$ $P(X_3) = 3/25$

$P(Y_4) = \frac{1}{10}$ $\qquad$ $P(X_4) = 3/25$

$P(X_5) = 1/25$

CONDITIONAL MATRIX

$$P(S/R) \Rightarrow p(X_i/Y_j) = p(X_i, Y_j) / p(Y_j)$$

$$
\begin{array}{ccccc}
 & Y_1 & Y_2 & Y_3 & Y_4 \\
X_1 & \frac{1}{\cancel{4}} \cdot \frac{\cancel{20}^{5}}{7} = \frac{5}{7} & 0 & 0 & 0 \\
X_2 & \frac{1}{10} \cdot \frac{20}{7} = \frac{2}{7} & \frac{3}{1\cancel{0}} \cdot \frac{2\cancel{0}}{7} = \frac{6}{7} & 0 & 0 \\
X_3 & 0 & \frac{5}{5\,100} \cdot \frac{20}{7} = \frac{1}{7} & \frac{1}{2} & 0 \\
X_4 & 0 & 0 & \frac{5}{100} \cdot 5 = \frac{1}{4} & 1 \\
X_5 & 0 & 0 & \frac{1}{4} & 0 \\
\hline
\Sigma \rightarrow & 1 \checkmark & 1 \checkmark & 1 \checkmark & 1 \checkmark
\end{array}
$$

3(CONT)

CONDITIONAL Matrix $P(R/S) \Rightarrow p(Y_j/X_i)$

$p(Y_j/X_i) = P(X_i, Y_j) / P(X_i)$

$$\begin{array}{c|cccc|c}
 & Y_1 & Y_2 & Y_3 & Y_4 & \Sigma \downarrow \\
X_1 & 1 \checkmark & 0 & 0 & 0 & 1 \checkmark \\
X_2 & \frac{1}{10}\cdot\frac{10}{4}=\frac{1}{4} & \frac{3}{10}\cdot\frac{10}{4}=\frac{3}{4}\checkmark & 0 & 0 & 1 \checkmark \\
X_3 & 0 \checkmark & \frac{5}{100}\cdot\frac{100}{15}=\frac{1}{3} & \frac{10}{100}\cdot\frac{100}{15}=\frac{2}{3} & 0 & 1 \checkmark \\
X_4 & 0 & 0 & \frac{5}{100}\frac{100}{15}=\frac{1}{3}\checkmark & \frac{10}{100}\cdot\frac{100}{15}=\frac{2}{3} & 1 \checkmark \\
X_5 & 0 & 0 & 1 & 0 & 1 \checkmark
\end{array}$$

Next, find $H(\overset{S}{\cancel{X}})$

$$H(\overset{S}{\cancel{X}}) = \sum_{i=1}^{5} p(X_i) \lg \frac{1}{p(X_i)}$$

$$= \frac{1}{4}\lg 4 + \frac{2}{5}\lg\frac{5}{2} + 2\cdot\frac{3}{25}\lg\frac{25}{3} + \frac{1}{25}\lg 25$$

$= 1.351$ NATS ✓

$$H(\overset{R}{\cancel{Y}}) = \sum_{j=1}^{4} p(Y_j) \lg \frac{1}{p(Y_j)}$$

$$= 2\cdot\frac{7}{20}\ln\frac{20}{7} + \frac{1}{5}\lg 5 + \frac{1}{10}\lg 10$$

$= 1.287$ NATS

$$H(\overset{RS}{\cancel{X,Y}}) = \sum_i \sum_j p(X_i, Y_j) \lg \frac{1}{p(X_i\,Y_j)}$$

$$= \frac{1}{4}\lg 4 + 3\times\frac{1}{10}\lg 10 + 3\,\frac{1}{20}\lg 20 + \frac{3}{10}\lg \frac{10}{3}$$

$= 1.848$ NATS

CONT $\Rightarrow$

3 cont

a. verify $H(X,Y) \le H(X) + H(Y)$

$1.848 \le 1.287 + 1.351 = 2.638$

yup

continuing:

$$H(\overset{Y}{R}/\overset{X}{S}) = -\sum_i \sum_j p(x_i, y_j) \ln p(y_i/x_i)$$

$$= \cancel{\tfrac{1}{4} \ln 1}^{\,0} + \tfrac{1}{10} \ln 4 + \tfrac{3}{10} \ln \tfrac{4}{3} + \tfrac{5}{100} \ln 3 + \tfrac{1}{10} \ln \tfrac{3}{2} + \tfrac{5}{100} \ln 3 + \tfrac{1}{10} \ln 3/2 + \cancel{\tfrac{5}{100} \ln 1}^{\,0}$$

$$= \tfrac{1}{10} \ln 4 + \tfrac{1}{10} \ln 3 + \tfrac{3}{10} \ln 4/3 + \tfrac{2}{10} \ln \tfrac{3}{2}$$

$= 0.4159$ NATS

$$H(S/R) = -\sum_i \sum_j p(x_i, y_j) \ln p(x_i/y_j)$$

$$= \tfrac{1}{4} \ln \tfrac{7}{5} + \tfrac{1}{10} \ln \tfrac{7}{2} + \tfrac{3}{10} \ln \tfrac{7}{6} + \tfrac{1}{20} \ln 7 + \tfrac{1}{10} \ln 2 + \tfrac{1}{20} \ln 4 + \tfrac{1}{20} \ln 4 + \cancel{\tfrac{1}{10} \ln 1}^{\,0}$$

$$= \tfrac{1}{4} \ln \tfrac{7}{5} + \tfrac{1}{10} \ln \tfrac{7}{2} + \tfrac{3}{10} \ln \tfrac{7}{6} + \tfrac{1}{20} \ln 7 + \tfrac{1}{10} \ln 2 + \tfrac{1}{10} \ln 4$$

$= 0.5609$ NATS ✓

$H(S/R) + H(R) = 0.5609 + 1.287$ NATS

$\simeq 1.848 = H(R,S)$ ✓

✓

$\bar{\lambda} = 1 - \lambda$

4. $S_1 = \begin{pmatrix} s_1 & s_2 & s_3 \ldots s_{q_1} \\ p_1 & p_2 & p_3 \ldots p_{q_1} \end{pmatrix} \Rightarrow H(S_1) = H_1$

$S_2 = \begin{pmatrix} k_1 & k_2 & k_3 \ldots k_{q_2} \\ Q_1 & Q_2 & Q_3 \ldots Q_{q_2} \end{pmatrix} \Rightarrow H(S_2) = H_2$

$$S_\lambda = \begin{pmatrix} s_1 & s_2 \ldots & s_{q_1} & k_1 & k_2 & \ldots k_{q_2} \\ \lambda P_1 & \lambda P_2 \ldots & \lambda P_{q_1} & \bar{\lambda} Q_1 & \bar{\lambda} Q_2 & \ldots \bar{\lambda} Q_{q_2} \end{pmatrix}$$

$$H(S_\lambda) = -\sum_{i=1}^{q_1} \lambda P_i \lg \lambda P_i - \sum_{j=1}^{q_2} \bar{\lambda} Q_j \lg \bar{\lambda} Q_j$$

$$= -\lambda \Sigma_i P_i \lg \lambda - \lambda \Sigma_i P_i \lg P_i - \bar{\lambda} \Sigma_j Q_i \lg \bar{\lambda} - \bar{\lambda} \Sigma_j Q_j \lg Q_j$$

$$\Sigma_i P_i = \Sigma_j Q_j = 1$$

$$\Rightarrow H(S_\lambda) = -\lambda \lg \lambda - \bar{\lambda} \lg \bar{\lambda} + \lambda H_1 + \bar{\lambda} H_2$$

The entropy, $H(\lambda)$, associated with $S = \{\lambda, \bar{\lambda}\}$ is

$$H(\lambda) = -\lambda \lg \lambda - \bar{\lambda} \lg \bar{\lambda}$$

Thus

$$H(S_\lambda) = \lambda H_1 + \bar{\lambda} H_2 + H(\lambda) \Longrightarrow$$

✓ ✓ ✓

4 (cont)

TO maximize, take

$$\frac{dH(S_Y)}{d\lambda} = 0$$

For simplicity, use natural logs & units of nats

$$\frac{dH(S_Y)}{d\lambda} = H_1 - H_2 - \frac{d}{d\lambda}\left[\lambda \lg \lambda + (1-\lambda)\lg(1-\lambda)\right)$$

$$= H_1 - H_2 - \left[(\lg\lambda + 1) + (-1)\lg(1-\lambda) + (1-\lambda)\left(\frac{-1}{1-\lambda}\right)\right.$$

$$= H_1 - H_2 - \left[(\lg\lambda + \cancel{1}) - \lg(1-\lambda) - \cancel{1}\right]$$

$$= H_1 - H_2 + \lg\frac{1-\lambda}{\lambda} = 0$$

Thus

$$H_2 - H_1 = \ln\frac{1-\lambda}{\lambda}$$

$$\frac{1-\lambda}{\lambda} = e^{H_2-H_1}$$

~~$1-\lambda = \lambda e^{H_2-H_1}$~~

$$\lambda/1-\lambda = e^{H_1-H_2}$$

$$\lambda = (1-\lambda)e^{H_1-H_2}$$

$$= e^{H_1-H_2} - \lambda e^{H_1-H_2}$$

$$\lambda(1+e^{H_1-H_2}) = e^{H_1-H_2}$$

or

$$\lambda_0 = \frac{e^{H_1-H_2}}{1+e^{H_1-H_2}} = \frac{1}{e^{H_2-H_1}+1}$$

ys.

$\longrightarrow$

(4 cont)

$$\lambda_0 = 1 - \frac{1}{e^{H_2-H_1}+1}$$

$$= \frac{e^{H_2-H_1}+1-1}{e^{H_2-H_1}+1} = \frac{e^{H_2-H_1}}{e^{H_2-H_1}+1}$$

$$= \frac{1}{1+e^{H_1-H_2}}$$

Now

$$H(\lambda_0) = -\lambda_0 \ln \lambda_0 - \overline{\lambda_0} \ln \overline{\lambda_0}$$

$$= \frac{1}{e^{H_2-H_1}+1} \ln\left(e^{H_2-H_1}+1\right)$$

$$+ \frac{1}{1+e^{H_1-H_2}} \ln\left(e^{H_1-H_2}+1\right)$$

It is a matter of substitution ✓

Thus

$$H(S_{\lambda_0})_e = \lambda_0 H_1 + \overline{\lambda_0} H_2 + H(\lambda)$$

$$= \frac{H_1}{e^{H_2-H_1}+1} + \frac{H_2}{e^{H_1-H_2}+1}$$

$$+ \frac{1}{e^{H_2-H_1}+1} \ln\left(e^{H_2-H_1}+1\right)$$

$$+ \frac{1}{1+e^{H_1-H_2}} \ln\left(e^{H_1-H_2}+1\right)$$

-1

$$= \frac{1}{e^{H_2-H_1}+1} \left[H_1 + \ln\left\{e^{H_2-H_1}+1\right\}\right]$$

$$+ \frac{1}{e^{H_1-H_2}+1} \left[H_2 + \ln\left(e^{H_1-H_2}+1\right)\right] \text{ NATS.}$$

where again, we have used base e. for base r

NATS $= \lg_e x = \frac{\lg_R x}{\lg_R e}$

$$H(S_{\lambda_0})_r = H(S_{\lambda_0})_e \lg_r e \; ; \; r\text{-arys}$$

is the maximum value the entropy may achieve for this mixed source.

✓

5. a. $H(X,Y) \leq H(x) + H(Y)$
with equality ~~a~~ iff X & Y are statistically independent random vectors ✓

b. $H(Y/X) \leq H(Y)$ with equality only if X & Y are statistically independent random vectors

c. In general $\log_b X = \log_a X / \log_a b$
BITS is lg_2
$\therefore$
$\# BITS = lg_2 X = lg_{10} X / lg_{10} 2$
or
$\# Hartleys = lg_{10} X = lg_2 X \, lg_{10} 2$
$= (\# BITS)(0.3010)$
$\therefore \quad 7.2^{BITS} = 7.2 \times 0.3013$
$= 2.17$ Hartleys ✓

Similarly, we may show
1.44 BIT $\doteq 1.$ nats
Where nats are from lg_e
thus
$7.2 \text{ BITS} = 7.2 \text{ BITS} \times \frac{1 \text{ NAT}}{1.44 \text{ BITS}}$
$= 5$ NATS

$\longrightarrow$

✓

(5 cont)

d. Additive property on Entropy

$$S = \left\{ \begin{array}{cccccc} s_1 & s_2 & , & \dots, & s_{n-1} & s_n \\ p_1 & p_2 & , & & , p_{n-1} & p_n \end{array} \right)$$

Divide event s_n into m disjoint ✓ events:

$$s_n = \left\{ \begin{array}{cccc} r_1 & r_2, & \dots, & r_m \\ q_1 & q_2 & \dots & q_m \end{array} \right) \qquad ①$$

And define

$$S' = \left\{ \begin{array}{cccccccc} s_1, & s_2 & \dots & s_{n-1} & r_1 & r_2 & \dots, & r_m \\ p_1 & p_2 & & p_{n-1} & q_1/p_n & q_2/p_n & \dots & q_m/p_n \end{array} \right\}$$

Then the Entropies of S & S' are related by

$$H'(p_1, p_2, \dots p_{n-1}, \frac{q_1}{\cancel{p_n}}, \frac{q_2}{\cancel{p_n}}, \dots q_m/\cancel{p_n})$$

See class notes!

$$= H(p_1, p_2, p_3, \dots, p_{n\cancel{1}})$$

$$+ p_n H_{s_n}(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \dots, q_m/p_n)$$

(-3)

Where H_{s_n} is the entropy associated with the source s_n in ①.

You are applying it right.
You are not saying it right!

(5 cont)

c. ~~For the source shown~~ Here, we have

$$S = \begin{Bmatrix} E_1 & E_2 & E_3 & E_4 & E_5 \\ 1/10 & 2/10 & 3/10 & 1/10 & 3/10 \end{Bmatrix}$$

where $P\{E_5\} = P[F_2] + P[F_1]$

The corresponding entropy of S is

$$H(S) = 2 \times \frac{1}{10} \ln 10 + 2\frac{3}{10} \ln \frac{10}{3} + \frac{2}{10} \ln \frac{10}{2}$$

$$= 1.505 \text{ NATS}$$

Now consider the augmented source

$$S' = \begin{Bmatrix} E_1 & E_2 & E_3 & E_4 & F_1 & F_2 \\ 1/10 & 2/10 & 3/10 & 1/10 & 1/10 & 2/10 \end{Bmatrix}$$

The corresponding entropy is

$$H(S') = 3\frac{1}{10} \ln 10 + 2\frac{2}{10} \ln 5 + \frac{3}{10} \ln \frac{10}{3}$$

$$= 1.66 \text{ NATS}$$

NOW $H\left(\frac{q_1}{p_n} \dots \frac{q_m}{p_n}\right) =$

$$\frac{q_1}{p_n} = \frac{1/10}{3/10} = \frac{1}{3}$$

$$q_2/p_n = \frac{2/10}{3/10} = \frac{2}{3}$$

$$H_{S_n} = \frac{1}{3} \ln 3 + \frac{2}{3} \ln \frac{3}{2}$$

$$= 0.6365$$

$$1.505 + \frac{3}{10}(0.6365) = 1.67 \text{ BITS}$$

6. $S = \begin{Bmatrix} x_1 & x_2 \\ p_1 & p_2 \end{Bmatrix}$

$S' = \begin{Bmatrix} x_1' & x_2' \\ p_1' & p_2' \end{Bmatrix}$

$p_1' = p_1 - \Delta P \qquad p_2' = P_2 + \Delta P$

Assume that $P_1 - \Delta P \geq P_2 + \Delta P \leftarrow$ ①

(which also says $P_1 \geq P_2$ FOR $P_1 > \Delta P > 0$)

ie, the probabilities in S' are numerically closer.

Now:

$H(S) = H = -p_1 \lg p_1 - p_2 \lg p_2$ ✓

$H(S') = H' = -p_1' \lg p_1' - p_2' \lg p_2'$

$= -(p_1 - \Delta P) \lg (P_1 - \Delta P) - (P_2 + \Delta P) \lg (P_2 + \Delta P)$ ✓

Consider the difference

$$H' - H' = +P_1 \lg (P_1 - \Delta P) - \Delta P \lg (p_1 - \Delta P) + P_2 \lg (P_2 + \Delta P) + \Delta P \lg (P_2 + \Delta P) - p_1 \lg p_1 - p_2 \lg p_2$$

$$= p_1 \lg \frac{P_1 - \Delta P}{P_1} + \Delta P \lg \frac{P_2 + \Delta P}{P_1 - \Delta P} + p_2 \lg \frac{P_2 + \Delta P}{P_2}$$ ✓ ✓ ✓

Use the inequality

$\lg x \leq \cancel{1-x}$ $(x-1)$ See p16 Text

$\left(\cancel{1} - \frac{\Delta P}{P_1} \cancel{-1}\right) P_1 = -\Delta P$ $\qquad +\Delta P$

$\left(\cancel{1} + \frac{\Delta P}{P_2} \cancel{-1}\right) P_2$

(6 CONT)

It follows

$$H - H' \overset{\geq}{\cancel{\leq}} P_1 \left[\cancel{1} - \left(1 - \frac{\Delta P}{P_1}\right)\right] + P_2 \left[\cancel{1} - \left(1 + \frac{\Delta P}{P_2}\right)\right] + \Delta P \lg \frac{P_2 + \Delta P}{P_1 - \Delta P}$$

$$\overset{\geq}{\cancel{\leq}} (\Delta P) + (-\Delta P) + \Delta P \lg \frac{P_2 + \Delta P}{P_1 - \Delta P}$$

$$\overset{\geq}{\cancel{\leq}} \Delta P \lg \frac{P_2 + \Delta P}{P_1 - \Delta P}$$

You are lucky you got the right result by using the wrong inequality

Now, from ①, $\frac{P_2 + \Delta P}{P_1 - \Delta P} \leq 1.$

$$\Rightarrow \lg \frac{P_2 + \Delta P}{P_1 - \Delta P} \leq 0$$

(Since $\lg(\cdot)$ is a monotonic increasing function)

Thus, since $\Delta P > 0$

$$H - H' \leq \Delta P \lg \frac{P_2 + \Delta P}{P_1 - \Delta P} \leq 0 \checkmark$$

$$H_1 \leq H'$$

or, the entropy of the perturbed source is greater, or, equivilantly, more "uncertain"

QED

$\boxed{-3}$

BOB MARKS

85/10

7. FROM TEXT: PROB. 2-5 p. 41

a. $S = \{s_1, s_2, \ldots, s_i, \ldots\}$

$$P(s_i) = a\alpha^i \quad ①$$

WE REQUIRE THAT

$$1 = \sum_{i=1}^{\infty} P(s_i) = \Sigma_i \, a\alpha^i \quad ②$$

$$= a \, \Sigma_i \, \alpha^i$$

$$= a \frac{\alpha}{1-\alpha} \checkmark$$

$$\Rightarrow a = \frac{1-\alpha}{\alpha} \checkmark \quad ③$$

$$\therefore \; P(s_i) = \frac{1-\alpha}{\alpha} \alpha^i \quad ④$$

b. $H(S) \triangleq \Sigma_i \, P(s_i) \, \lg {}^1/P(s_i)$

$$= -\Sigma_i \frac{1-\alpha}{\alpha} \alpha^i \lg \frac{1-\alpha}{\alpha} \alpha^i$$

$$= -\Sigma_i \frac{1-\alpha}{\alpha} \alpha^i \left[\lg \frac{1-\alpha}{\alpha} + \lg \alpha^i \right]$$

$$= -\Sigma_i \frac{1-\alpha}{\alpha} \alpha^i \left[\lg \frac{1-\alpha}{\alpha} + i \lg \alpha \right]$$

$$= -\left[\Sigma_i \frac{1-\alpha}{\alpha} \alpha^i \right] \lg \frac{1-\alpha}{\alpha}$$

$$+ \frac{\alpha-1}{\alpha} \lg \alpha \; \Sigma_i \, i \, \alpha^i$$

BUT $\quad \Sigma_i \frac{1-\alpha}{\alpha} \alpha^i = \Sigma_i P(S_i) = 1$

AND $\quad \Sigma_i \, i \, \alpha^i = \frac{\alpha}{(1-\alpha)^2}$

$$\Rightarrow H(S) = -\lg \frac{1-\alpha}{\alpha} - \frac{1-\alpha}{\alpha} \frac{\alpha}{(1-\alpha)^2} \lg \alpha$$

$$= \lg \frac{\alpha}{1-\alpha} - \frac{1}{1-\alpha} \lg \alpha$$

$$= \lg \alpha - \lg 1-\alpha - \frac{1}{1-\alpha} \lg \alpha$$

$$= \left[1 - \frac{1}{1-\alpha} \right] \lg \alpha - \lg 1-\alpha$$

$$= \frac{-\alpha}{1-\alpha} \lg \alpha - \lg (1-\alpha) \checkmark$$

$$= \frac{\alpha}{\alpha-1} \lg \alpha - \lg (1-\alpha) \checkmark \quad ⑤$$

HENCEFORTH, USE LOG BASE e. ie $\lg \triangleq \lg_e$

(NOTE THAT FOR $\alpha = \frac{1}{2}$, WE GET $H(S) = 2$ BITS)

1. $\lim_{\alpha \to 0^+} H(s) = \lim_{\alpha \to 0^+} -\alpha \lg \alpha$

$= \lim_{\alpha \to 0^+} - \lg \alpha / 1/\alpha$

USING LA HOPITAL:

$\lim_{\alpha \to 0^+} H(s) = \lim_{\alpha \to 0^+} 1/x \big/ 1/x^2$

$= \lim_{\alpha \to 0^+} X = 0^+$

2. $\lim_{\alpha \to 1^-} H(s) = \lim_{\alpha \to 1^-} \frac{\lg \alpha}{\alpha - 1} - \lg(1-\alpha)$

NOW $\lim_{\alpha \to 1^-} \frac{\lg \alpha}{\alpha - 1} = \lim_{\alpha \to 1^-} \frac{1/\alpha}{1} = \lim_{\alpha \to 1^-} \frac{1}{\alpha} = \infty$

$\therefore \lim_{\alpha \to 1^-} H(s) = \infty - (-\infty) = \infty$

3. CHECKING FOR EXTREMA:

$$\frac{d}{d\alpha} H(s) = \frac{d}{d\alpha} \frac{\alpha}{\alpha-1} \lg \alpha - \frac{d}{d\alpha} \lg(1-\alpha)$$

$$= \left[\frac{\alpha}{\alpha-1} \cdot \frac{1}{\alpha} + \frac{(\alpha-1)-1\alpha}{(\alpha-1)^2} \lg \alpha\right] - \frac{(-1)}{1-\alpha}$$

$$= -\frac{1}{1-\alpha} - \frac{1}{(1-\alpha)^2} \lg \alpha + \frac{1}{1-\alpha}$$

$$= \frac{-1}{(1-\alpha)^2} \lg \alpha$$

$\therefore$ NO EXREMA FOR FINITE α, SINCE $H(s)\big|_{\lambda=1} = \infty$

4. ON α

- IN ORDER FOR ALL $P(S_i)$ IN ① TO BE POSITIVE, WE REQUIRE $\alpha > 0$
- IN ORDER FOR THE (GEOMETRIC) SERIES IN ② TO CONVERGE, $|\alpha| < 1$

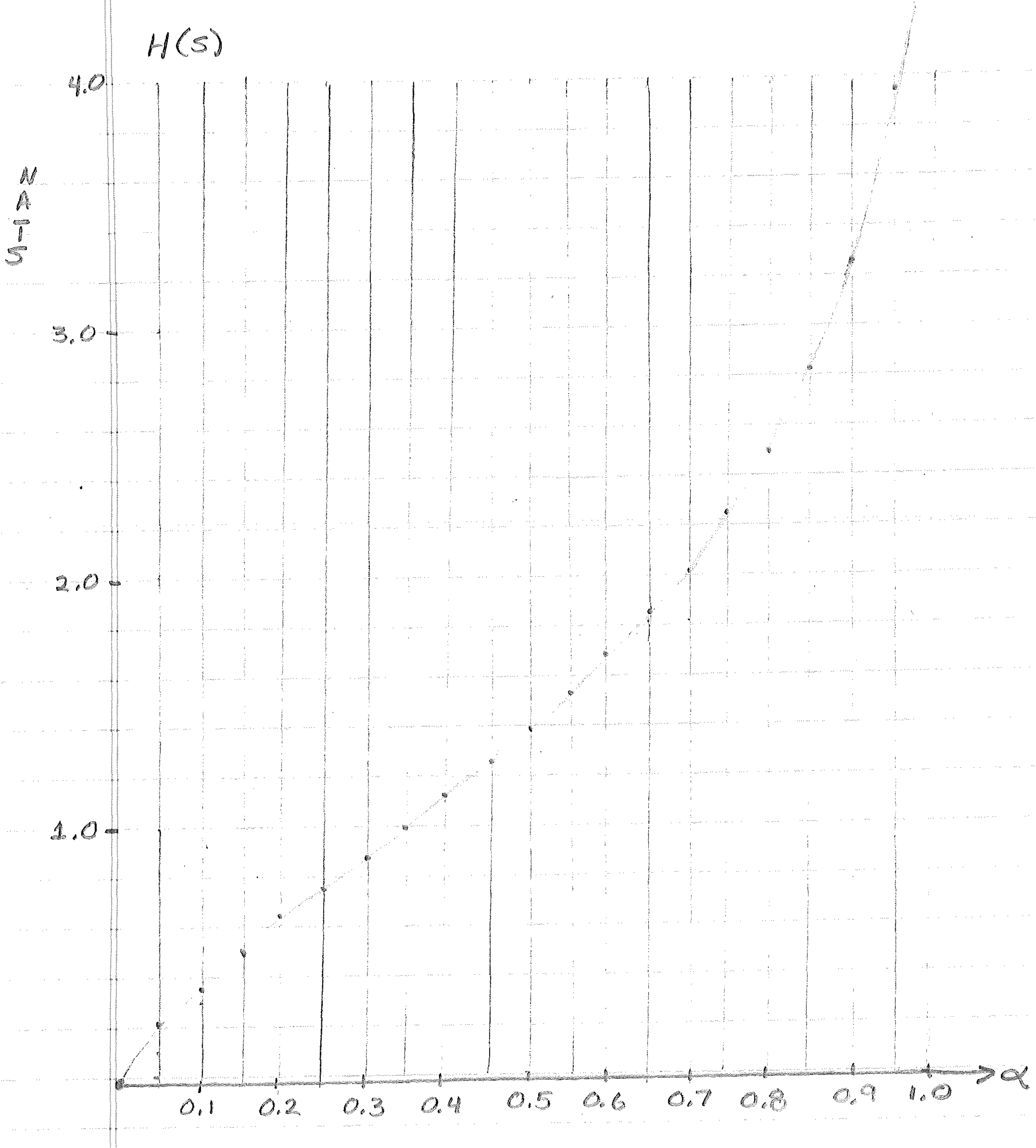
THUS

$$0 < \alpha < 1 \qquad ⑥$$

TO PLOT H(S) IN ⑤, USE HP-25 PROGRAM:

(α)

```
STO 0     -       1      ln
ln        ÷       RCL 0  +
RCL 0     RCL 0   -      GTO 00
1         x       1/x
```


H(s)
NATS
4.0
3.0
2.0
1.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
α

THE STRIKING ASPECT OF THIS PROBLEM IS THE INFINITE ENTROPY FOR α NEAR 1. THAT IS, WE CAN "TUNE" OUR SOURCE TO AS HIGH AN ENTROPY AS DESIRED BY LETTING α GO CORRESPONDINGLY CLOSE TO 1.

INTUITIVELY, WHAT IS HAPPENING IS AS FOLLOWS. FOR $\alpha \simeq 0$, THE PROBABILITIES ARE ROUGHLY:

$$(P_1, P_2, P_3, \ldots, P_i, \ldots) \approx (1, 0, 0, \ldots, 0, \ldots)$$

WHERE WE HAVE INTERPRETED $H(S)\big|_{i=0} = 1 \Leftarrow 0^{0}$

FOR α NEAR 1, WE <u>ESSENTIALLY</u> HAVE

$$(P_1, P_2, \ldots, P_i, \ldots) \approx (\epsilon, \epsilon, \epsilon, \ldots, \epsilon, \ldots)$$

WHERE $\epsilon \ll 1$. THAT IS, WE APPROACH A CONDITION OF HAVING AN INFINITE NUMBER OF INFINTESIMALLY EQUALLY PROBABLE EVENTS. THIS, THEN, CONSTITUTES A CORRESPONDING APPROACH TO INFINITE ENTROPY.

Quiz 2, 8-6-76 EE 5325 Summer 2, 1976 J-C. Prabhakar

1\. During class work it was shown that the efficiency of a coding scheme could be enhanced by source extension. The purpo of this exercise is to investigate the effect of no. of symbols (r) in the encoding alphabet on the efficien of compact codes.

Check H_i's

Consider an ensemble of 10 messages arranged, for you convenience, in the non-increasing order of probabilities.

$$X : \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$$
$$P : \left\{\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{64}\right\}$$

Let $r = 2, 3, 4, 5$ and 7.

~~These~~ These points on the horizontal axis should enable you to figure out the dependence of efficiency on r.

Determine the compact codes by Shannon-Fano method or Huffman's procedure and determine the efficiency for each r. Make a plot and list your conclusion!

Note: When probs. are given in a certain "form", the codes resulting from Shannon-Fano method are compact.

2. A Source alphabet $S : \{S_1, S_2\}$ has prob. distribution

$$P : \left(\tfrac{3}{4}, \tfrac{1}{4}\right)$$

Derive compact codes for S, its 2nd, and 4th extensions. Calculate the three efficiencies and verify, approximately, the result

$$\frac{\overline{L}_n / n}{H(S)} \longrightarrow 1 \text{ for "large" } n$$

$\left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2$

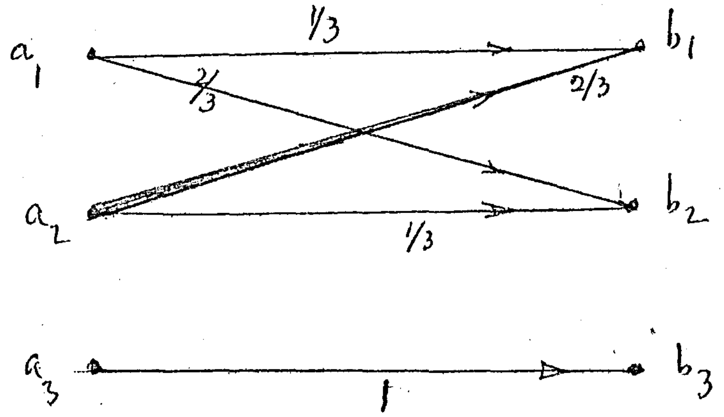
$\frac{1}{9} - \frac{4}{9}$

$-\frac{3}{9} = -\frac{1}{3}$

Note: This prob. distribution requires Huffman Coding for Compactness

3. A channel is described by the following Source-Receiver relationship. The numbers are conditional probabilities.

$\frac{1}{9} - \frac{4}{9} = \left(-\frac{3}{9}\right)$



Use Murogas' technique to determine the maximum of the transinformation that can be associated with the arrival of the received message i.e the channel Capacity.
units to your answer.

Barlow

4. ✓ Derive the following codes for the ensemble

$$A : \{a_1, a_2, a_3, a_4, a_5, a_6, a_7\}$$

$$P : \left\{\frac{4}{10}, \frac{2}{10}, \frac{12}{100}, \frac{8}{100}, \frac{8}{100}, \frac{8}{100}, \frac{4}{100}\right\}$$

(a) Shannon Coding
(b) Shannon-Fano Coding
(c) Huffman's Optimal coding.

Use (0,1) as the alphabet and verify the conclusion arrived at in class that (c) will yield compact Codes. (a) & (b) may <u>not</u> be so.

Note: Since compactness is related to the average length of a scheme, you need not evaluate $H(A)$.

5. ✓ (i) Define The following
(a) Non-singular Codes
(b) Uniquely Decodable Codes
(c) Instantaneous Code
(d) Prefix property
(e) Redundancy of a Coding Scheme
(f) The average Length of a Coding Scheme
(g) An Independent Channel.

(ii) An ensemble has 8 words with lengths:

No. of words of Length 1 = 0

No. of words of length 2 = 3

– – – ——— 3 = 1

– – – – – – – 4 = 4

Find the no. of symbols in the encoding alphabet required to generate an Instantaneous Scheme. Determine the resulting words

(iii) Which of the sets of word lengths shown below are acceptable for a uniquely decodable codes when

(a) The alphabet is (0, 1)

(b) – ——— (0, 1, 2)

No. of words of length li in each code	word lengths li 1	2	3	4	5
Code A	2	1	2	4	1
Code B	2	2	2	3	1
Code C	1	4	6	0	0
Code D.	2	2	2	2	3

Is the test you are applying both necessary and sufficient?

6.* Prob. 4-8 p 92 Text

7.* Prob. 5-17 p 146 Text.

* Prob. 6, 7 are Take home due Tuesday 8-10-76.

Notes: 1. All HW* is due 8-13-76. Delay may carry a mild penalty.

2. If the average score on this quiz is less than 60, there will be a final quiz on Aug 17 or 18.

3. Reading of Shannon's paper is mandatory and will count as a HW* problem.

MARKS 1-1

~~7~~
~~9~~
-5 puls (2)
puls 5
-4
-3

-19. In general

η = EFFICIENCY

$= \bar{L} / H_r(S)$

(a) r = 2 (Huffman)

x_1	1/4	00	1/4	00	1/4	00	1/4	00
x_2	1/4	01	1/4	01	1/4	01	1/4	01
x_3	1/8	101	1/8	101	1/8	101	1/8	101
x_4	1/8	110	1/8	110	1/8	110	1/8	110
x_5	1/16	1001	1/16	1001	1/16	1000	1/8	111
x_6	1/16	1110	1/16	1110	1/16	1001	1/16	1000
x_7	1/16	1111	1/16	1111	1/16	1110	1/16	1001
x_8	1/32	10001	1/32	10000	1/16	1111		
x_9	1/64	100000	1/32	10001				
x_{10}	1/64	100001						

1/4	00	1/4	11	1/4	10	1/2	0	1/2	0
1/4	01	1/4	00	1/4	11	1/4	10	1/2	1
1/8	100	1/4	01	1/4	00	1/4	11		
1/8	101	1/8	100	1/4	01				
1/8	110	1/8	101						
1/8	111								

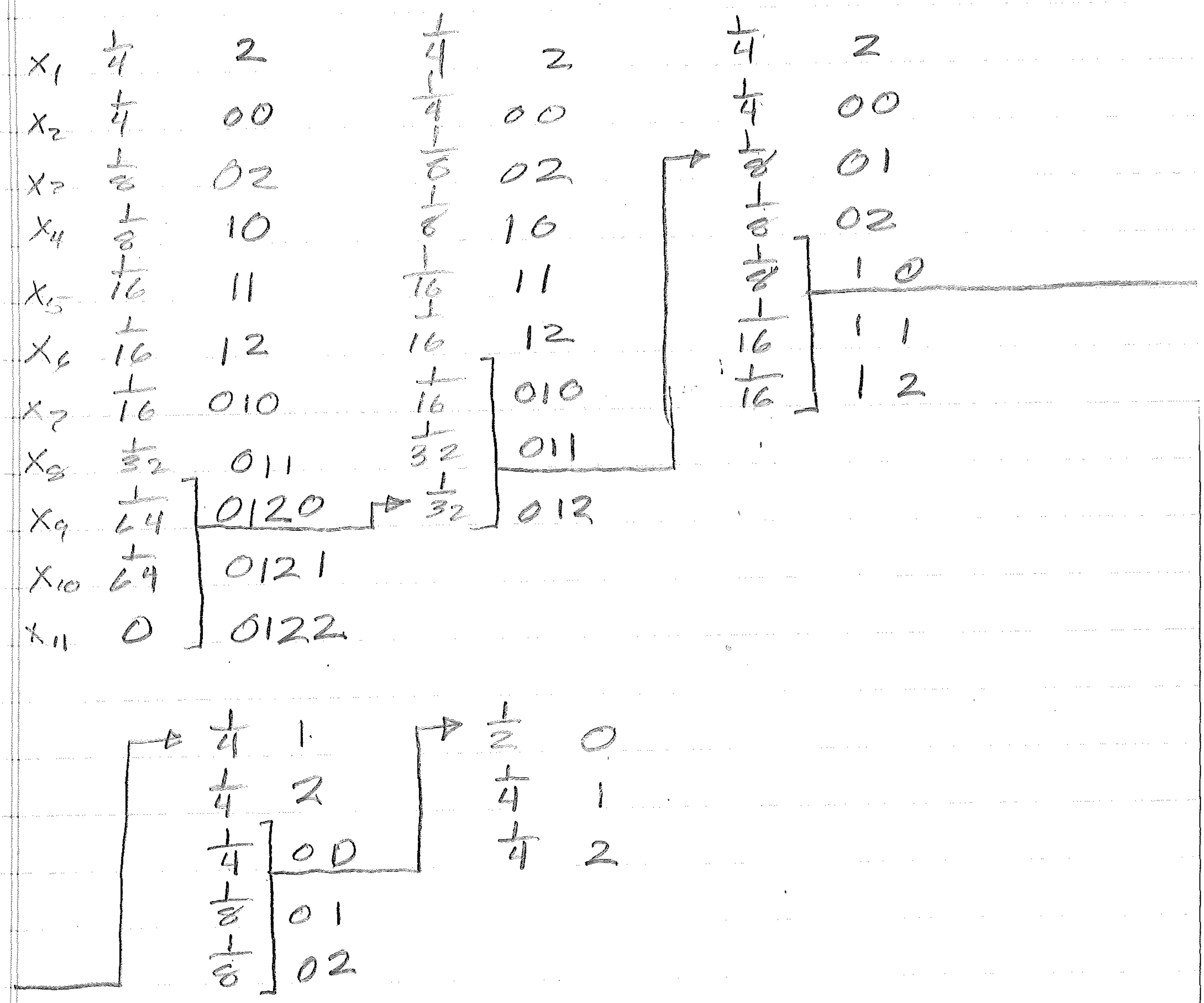
$$\bar{L} = \frac{2}{4} + \frac{2}{4} + \frac{3}{8} + \frac{3}{8} + \frac{4}{16} + \frac{4}{16} + \frac{4}{16} + \frac{5}{32} + \frac{6}{64} + \frac{6}{64}$$

$$= 1 + \frac{6}{8} + \frac{12}{16} + \frac{5}{32} + \frac{12}{64}$$

$$= \frac{1}{32}\left[24 + 24 + 5 + 6\right]$$

$$= 59/32$$

(b) $r=3 \Rightarrow q = r+(r-1)\alpha = 3+2\alpha \Rightarrow \alpha = 4 \Rightarrow q = 11$



x_1	$\frac{1}{4}$	2
x_2	$\frac{1}{4}$	00
x_3	$\frac{1}{8}$	02
x_4	$\frac{1}{8}$	10
x_5	$\frac{1}{16}$	11
x_6	$\frac{1}{16}$	12
x_7	$\frac{1}{16}$	010
x_8	$\frac{1}{32}$	011
x_9	$\frac{1}{64}$	0120
x_{10}	$\frac{1}{64}$	0121
x_{11}	0	0122

$\frac{1}{4}$	2
$\frac{1}{4}$	00
$\frac{1}{8}$	02
$\frac{1}{8}$	10
$\frac{1}{16}$	11
$\frac{1}{16}$	12
$\frac{1}{16}$	010
$\frac{1}{32}$	011
$\frac{1}{32}$	012

$\frac{1}{4}$	2
$\frac{1}{4}$	00
$\frac{1}{8}$	01
$\frac{1}{8}$	02
$\frac{1}{8}$	10
$\frac{1}{16}$	11
$\frac{1}{16}$	12

$\frac{1}{4}$	1
$\frac{1}{4}$	2
$\frac{1}{4}$	00
$\frac{1}{8}$	01
$\frac{1}{8}$	02

$\frac{1}{2}$	0
$\frac{1}{4}$	1
$\frac{1}{4}$	2

$$\bar{L}_3 = \frac{1}{4} + \frac{2}{4} + \frac{2}{8} + \frac{2}{8} + \frac{2}{16} + \frac{2}{16} + \frac{3}{16} + \frac{3}{32} + \frac{4}{64} + \frac{4}{64}$$

$$= \frac{3}{4} + \frac{4}{8} + \frac{7}{16} + \frac{3}{32} + \frac{8}{64}$$

$$= \frac{1}{32}\left[24+16+14+3+4\right]$$

$$= 61/32$$

(c) $r=4 \Rightarrow q = r+(r-1)\alpha = 4+3\alpha \Rightarrow \alpha = 2 \Rightarrow q=10$

x_1	$\frac{1}{4}$	1	$\frac{1}{4}$	1	$\frac{3}{8}$	0
x_2	$\frac{1}{4}$	2	$\frac{1}{4}$	2	$\frac{1}{4}$	1
x_3	$\frac{1}{8}$	01	$\frac{1}{8}$	3	$\frac{1}{4}$	2
x_4	$\frac{1}{8}$	02	$\frac{1}{8}$	01	$\frac{1}{8}$	3
x_5	$\frac{1}{16}$	03	$\frac{1}{8}$	02		
x_6	$\frac{1}{16}$	04	$\frac{1}{16}$	03		
x_7	$\frac{1}{16}$	30	$\frac{1}{16}$	04		
x_8	$\frac{1}{32}$	31				
x_9	$\frac{1}{64}$	32				
x_{10}	$\frac{1}{64}$	33				

$$\bar{L} = \frac{3}{4} + \frac{4}{8} + \frac{6}{16} + \frac{3}{32} + \frac{4}{64}$$
$$= \frac{3}{4} + \frac{4}{8} + \frac{6}{16} + \frac{2}{16}$$
$$= \frac{3}{4} + \frac{1}{2} + \frac{1}{2}$$
$$= 1 + \frac{3}{4}$$
$$= 7/4$$

(d) $r=5 \Rightarrow q = r + (r-1)\alpha = 5 + 4\alpha \Rightarrow \alpha = 2 \therefore q = 13$

x_1	$\frac{1}{4}$	1	$\frac{1}{4}$	1	$\frac{1}{4}$	0
x_2	$\frac{1}{4}$	2	$\frac{1}{4}$	2	$\frac{1}{4}$	1
x_3	$\frac{1}{8}$	3	$\frac{1}{8}$	3	$\frac{1}{4}$	2
x_4	$\frac{1}{8}$	4	$\frac{1}{8}$	4	$\frac{1}{8}$	3
x_5	$\frac{1}{16}$	00	$\frac{1}{16}$	00	$\frac{1}{8}$	4
x_6	$\frac{1}{16}$	01	$\frac{1}{16}$	01		
x_7	$\frac{1}{16}$	02	$\frac{1}{16}$	02		
x_8	$\frac{1}{32}$	04	$\frac{1}{32}$	03		
x_9	$\frac{1}{64}$	030	$\frac{1}{32}$	04		
x_{10}	$\frac{1}{64}$	031				
x_{11}	0	032				
x_{12}	0	033				
x_{13}	0	034				

✓ $\bar{L} = \frac{2}{4} + \frac{2}{8} + \frac{6}{16} + \frac{1}{16} + \frac{6}{64}$

$= \frac{3}{4} + \frac{7}{16} + \frac{3}{32}$

$= \frac{1}{32}\left[24 + 14 + 3\right]$

$= 41/32$

24
14
3
41

e. $r = 7 \Rightarrow q = 7 + 6\alpha \Rightarrow \alpha = 1 \Rightarrow q = 13$

x_1	$\frac{1}{4}$	0	$\frac{1}{4}$	0
x_2	$\frac{1}{4}$	1	$\frac{1}{4}$	1
x_3	$\frac{1}{8}$	3	$\frac{1}{8}$	2
x_4	$\frac{1}{8}$	4	$\frac{1}{8}$	3
x_5	$\frac{1}{16}$	5	$\frac{1}{8}$	4
x_6	$\frac{1}{16}$	6	$\frac{1}{16}$	5
x_7	$\frac{1}{16}$	20	1/16	6
x_8	$\frac{1}{32}$	21		
x_9	$\frac{1}{64}$	22		
x_{10}	$\frac{1}{64}$	23		
x_{11}	0	24		
x_{12}	0	25		
x_{13}	0	26		

$$\bar{L} = \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \frac{2}{32} + \frac{4}{64}$$
$$= \frac{2}{4} + \frac{1}{4} + \frac{2}{16}$$
$$= \frac{2}{4} + \frac{1}{4} + \frac{1}{8}$$
$$= 7/8$$

(b) OUR CHANNEL NOW IS

$$P(Y_j/X_i) \Rightarrow \begin{bmatrix} \bar{p}-\epsilon & p-\epsilon & 2\epsilon & 0 \\ p-\epsilon & \bar{p}-\epsilon & 0 & 2\epsilon \end{bmatrix} \quad (14)$$

AGAIN, LET

$$q = P(x_1) \qquad \bar{q} = P(x_2)$$

THUS

$$P(X_i, Y_j) = \begin{bmatrix} q(\bar{p}-\epsilon) & q(p-\epsilon) & 2q\epsilon & 0 \\ \bar{q}(p-\epsilon) & \bar{q}(p-\epsilon) & 0 & 2\bar{q}\epsilon \end{bmatrix} \quad (15)$$

COMPARING (14) AND (15) WITH ① AND ③ RESPECTIVELY, WE SEE THAT THE CONDITIONAL ENTROPY IS HERE THE SAME AS IS IN PART (a). FROM ⑨:

$$H(Y/X) = -(\bar{p}-\epsilon)\lg(\bar{p}-\epsilon) - (p-\epsilon)\lg(p-\epsilon) - 2\epsilon\lg 2\epsilon \quad (16)$$

FROM (15)

$$P(Y_1) = (q+p-2pq)-\epsilon \qquad (\text{SAME AS } ④)$$

$$P(Y_2) = 1-(p+q-2pq)-\epsilon \qquad (\text{SAME AS } ⑤)$$

$$P(Y_3) = 2q\epsilon$$

$$P(Y_4) = 2\bar{q}\epsilon = 2\epsilon - 2q\epsilon$$

THUS

$$\begin{aligned} -H(Y) = {} & [(q+p-2pq)-\epsilon]\lg[1\ (p+q-2pq)-\epsilon] \quad (17) \\ & + [1-(p+q-2pq)-\epsilon]\lg[1-(p+q-2pq)-\epsilon] \\ & + 2q\epsilon\lg 2q\epsilon + (2\epsilon-2q\epsilon)\lg(2\epsilon-2q\epsilon) \end{aligned}$$

NOW, USING ⑯ & ⑰:

$$I(x;Y) = H(Y/x) - H(Y) \qquad (18)$$

AND, AS BEFORE

$$\frac{d}{dq} I(x;Y) = -\frac{d}{dq} H(Y)$$

FROM ⑰:

$$\begin{aligned}\frac{d}{dq} I(x;Y) = {} & (1-2p)\,\lg\left[(p+q-2pq)-\epsilon\right] + (1\cancel{-2}p) \\ & -(1-2p)\,\lg\left[1-(p+q-2pq)-\epsilon\right] - (\cancel{1-2}p) \\ & +2\epsilon \lg 2q\epsilon + \cancel{2\epsilon} \\ & -2\epsilon \lg (2\epsilon - 2q\epsilon) - \cancel{2\epsilon}\end{aligned}$$

$$= (1-2p)\,\lg \frac{(p+q-2pq)-\epsilon}{1-(p+q-2pq)-\epsilon} + 2\epsilon \lg \frac{2q\epsilon}{2\epsilon(1-q)}$$

SETTING THIS TO ZERO YIELDS

$$(1-2p)\,\lg \frac{(p+q-2pq)-\epsilon}{1-(p+q-2pq)-\epsilon} = 2\epsilon \lg \frac{2\epsilon(1-q)}{2q\epsilon}$$

BOTH SIDES ARE ZERO FOR $q=\frac{1}{2}$.

$$\therefore q = \bar{q} = \frac{1}{2}$$

SUBSTITUTING INTO ⑱:

$$
\begin{aligned}
C &= I(X;Y)\big|_{q=\frac{1}{2}} \\
&= -(\bar{p}-\epsilon)\lg(\bar{p}-\epsilon) - (p-\epsilon)\lg(p-\epsilon) - 2\epsilon\lg 2\epsilon \\
&\quad + \left(\tfrac{1}{2}-\epsilon\right)\lg\left(\tfrac{1}{2}-\epsilon\right) + \left(\tfrac{1}{2}-\epsilon\right)\lg\left(\tfrac{1}{2}-\epsilon\right) \\
&\quad + \epsilon\lg\epsilon + \epsilon\lg\epsilon \\
&= -(\bar{p}-\epsilon)\lg(\bar{p}-\epsilon) - (p-\epsilon)\lg(p-\epsilon) \\
&\quad + 2\left(\tfrac{1}{2}-\epsilon\right)\lg\left(\tfrac{1}{2}-\epsilon\right) \\
&\quad + \cancel{2\epsilon\lg\epsilon} - 2\epsilon\lg 2 - \cancel{2\epsilon\lg\epsilon} \\
&= -(\bar{p}-\epsilon)\lg(\bar{p}-\epsilon) - (p-\epsilon)\lg(p-\epsilon) \qquad ⑲ \\
&\quad + 2\left(\tfrac{1}{2}-\epsilon\right)\lg\left(\tfrac{1}{2}-\epsilon\right) - 2\epsilon\lg 2
\end{aligned}
$$

(c) DENOTE THE CAPACITY IN PART (a) (⑬) BY C_a:

$$C_a = -(\bar{p}-\epsilon)\lg(\bar{p}-\epsilon) - (p-\epsilon)\lg(p-\epsilon) + 2\left(\tfrac{1}{2}-\epsilon\right)\lg\left(\tfrac{1}{2}-\epsilon\right)$$

AND THAT IN PART (b) (⑲) BY C_b:

$$C_b = -(\bar{p}-\epsilon)\lg(\bar{p}-\epsilon) - (p-\epsilon)\lg(p-\epsilon) + 2\left(\tfrac{1}{2}-\epsilon\right)\lg\left(\tfrac{1}{2}-\epsilon\right) - 2\epsilon\lg 2$$

ITS CLEAR THAT

$$C_a - C_b = 2\epsilon \lg 2 > 0$$

THUS, CHANNEL a IS ALWAYS BETTER THAN b. (EVEN FOR ϵ NEAR 0)

(NOTE THAT, IF C_a AND C_b WERE MEASURED IN BITS, THEN $C_a - C_b = 2\epsilon$ BITS)

BOB MARKS

TEST #3 (DUE 8/16/76 (MON))

PROBLEMS

1. CODE THE DATE 8-15-1947 USING HAMMING'S SINGLE ERROR CORRECTING CODE. USE THE SAME NUMBER OF BINITS FOR EACH OF THE THREE NUMBERS. IN THE RESULTING CODE SEQUENCE, INFLICT AN ERROR IN THE 11TH LEAST SIGNIFICANT BINIT. CORRECT IT IN THE MANNER THE RECEIVER WOULD.
2. HAND OUT #9 (UTILIZATION OF FANO BOUND)
3. PROB 6-1, p 81 OF TEXT
4. FOR $m=5$, FIND CORRESPONDING k. CHOOSE TWO OF THE POSSIBLE 2^5 NUMBERS, AND PERFORM A HAMMING CODE (SINGLE ERROR CORRECTION). COMPUTE

   P_r[ERROR WITH HAMMING CODE]

   P_r[" WITHOUT HAMMING CODE]

   ALSO, EVALUATE THE CORRESPONDING FIGURE OF MERIT.

1. WE WISH TO USE HAMMING'S SINGLE ERROR CORRECTING CODE TO CODE THE DATE 8-15-1947

NOW

$$(1947)_{10} = (1111\ 0011\ 0\ 11)_2$$

$$(8)_{10} = (1000)_2$$

$$(15)_2 = (1111)_2$$

(CONT →)

WE WILL USE THREE SEPARATE CODES FOR EACH NUMBER. EACH CODE WILL HAVE EQUAL LENGTH.

NOW $2^K \geq m+k+1$

$m=11 \Rightarrow K=4$

CODING $(8)_{10} = (1000)_2$

USING EVEN 1's PARITY

$$\begin{array}{cccccccc|ccccccc}
012 & 12 & 02 & 2 & 01 & 1 & 0 & & 012 & 12 & 02 & 2 & 01 & 1 & 0 \\
15 & 14 & 13 & 12 & 11 & 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\
 & & & & & & & \uparrow & & & & \uparrow & & \uparrow & \uparrow \\
 & & & & & & & P_3 & & & & P_2 & & P_1 & P_0
\end{array}$$

$P_0 \oplus 0 \oplus 0 \oplus 1 \oplus 0 \oplus 0 \oplus 0 \oplus 0 = 0 \Rightarrow P_0 = 1$

$P_1 \oplus 0 \oplus 0 \oplus 1 \oplus 0 \oplus 0 \oplus 0 \oplus 0 = 0 \Rightarrow P_1 = 1$

$P_2 \oplus 0 \oplus 0 \oplus 1 \oplus 0 \oplus 0 \oplus 0 \oplus 0 = 0 \Rightarrow P_2 = 1$

$P_3 \oplus 0 \oplus \ldots \oplus 0 = 0 \Rightarrow P_3 = 0$

CODING $(15)_2 = (1111)_2$

$$\begin{array}{ccccccccccccccc}
\multicolumn{4}{c}{\overbrace{\hphantom{15\quad 14\quad 13\quad 12}}} & & & & & \multicolumn{4}{c}{\overbrace{\hphantom{7\quad 6\quad 5\quad 4}}} & & & \\
\multicolumn{2}{c}{\overline{15\quad 14}} & 13 & 12 & \multicolumn{2}{c}{\overline{11\quad 10}} & 9 & 8 & \multicolumn{2}{c}{\overline{7\quad 6}} & 5 & 4 & \multicolumn{2}{c}{\overline{3\quad 2}} & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 & & & & & & & \uparrow & & & & \uparrow & & \uparrow & \uparrow \\
 & & & & & & & P_3 & & & & P_2 & & P_1 & P_0
\end{array}$$

$P_0 \oplus 1 \oplus 1 \oplus 1 \oplus 0 \ldots \oplus 0 = 0 \Rightarrow P_0 = 1$

$P_1 \oplus 1 \oplus 1 \oplus 1 \oplus 0 \ldots \oplus 0 = 0 \Rightarrow P_1 = 1$

$P_2 \oplus 1 \oplus 1 \oplus 1 \oplus 0 \ldots \oplus 0 = 0 \Rightarrow P_2 = 1$

$P_3 \oplus 0 \ldots \oplus 0 = 0 \Rightarrow P_3 = 0$

CODING 1947:

$$\begin{array}{ccccccccccccccc} 15 & 14 & 13 & 12 & 11 & 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{array}$$

$p_0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 0 \oplus 1 \oplus 1 = 0 \Rightarrow p_0 = 0$

$p_1 \oplus 1 \oplus 0 \oplus 1 \oplus 0 \oplus 0 \oplus 1 \oplus 1 = 0 \Rightarrow p_1 = 0$

$p_2 \oplus 1 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 1 = 0 \Rightarrow p_2 = 0$

$p_3 \oplus 1 \oplus 0 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 = 0 \Rightarrow p_3 = 1$

OUR FINAL CODE FOR 8-15-1947 IS

0000000010010 11·000000001111111111100111010100

WE WISH TO IMPOSE AN ERROR OF THE 11TH LEAST SIGNIFICANT BINIT OF THIS CODE, AND CORRECT IT. THIS AMOUNTS TO IMPOSING AN INCORRECT BINIT IN THE 11TH LEAST SIGNIFICANT BINIT OF 1947. AS SUCH, (FOR TRACTIBILITY), LETS DEAL ONLY WITH THE CODE FOR 1947 ⇒

15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
1	1	1	1	1	0	1	1	1	0	1	0	1	0	0
				↑			↑ P_3				↑ P_2		↑ P_1	↑ P_0

ERROR TO BE SPOTTED

$0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 1 = 0 \Rightarrow P_0 = 1$

$0 \oplus 1 \oplus 0 \oplus 1 \oplus 0 \oplus 1 \oplus 1 \oplus 1 = 0 \Rightarrow P_1 = 1$

$0 \oplus 1 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 1 = 0 \Rightarrow P_2 = 0$

$1 \oplus 1 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 1 = 0 \Rightarrow P_3 = 1$

ERROR IS @ $P_3 P_2 P_1 P_0 = (1011)_2 = 11^{TH}$ PLACE.

AS EXPECTED, THIS IS CORRECT.

TO FIX THE CODE, WE MERELY CHANGE THE 1 IN "11" TO A 0.

2. THIS PROBLEM IS SOLVED IN ASH'S BOOK (THEM 3.7.3 p 82). WE HERE PARAPHRASE IT. WE USE ASH'S NOTATION.

a. IN GENERAL, WE MAY WRITE*

$$I(\mathbf{X}|\mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}) \quad ①$$

WHERE

$$\mathbf{X} = (X_1, \ldots X_n) \qquad \mathbf{Y} = (Y_1 \ldots Y_n) \quad ②$$

(ie, AN n^{TH} EXTENSION)

IF WE RANDOMLY CHOOSE OUR INPUT SIGNALS (EQUALLY LIKELY) SO THAT

$$p(X_i) = 1/s \quad ③$$

WHERE THERE ARE s INPUT WORDS THE i^{TH} OF WHICH IS X_i.

IT FOLLOWS THAT

$$H(\mathbf{X}) = \Sigma_{i=1}^{s} \frac{1}{s} \lg s = \lg s \quad ④$$

THUS, ① BECOMES

$$I(\mathbf{X}|\mathbf{Y}) = \lg s - H(\mathbf{X}|\mathbf{Y}) \quad ⑤$$

WE ARE GIVEN THAT

$$I(\mathbf{X}|\mathbf{Y}) \leq \Sigma_{i=1}^{n} I(X_i|Y_j) \quad ⑥$$

THUS, WE MAY WRITE ⑤ AS

$$\lg s - H(\mathbf{X}|\mathbf{Y}) \leq \Sigma_{i=1}^{n} I(X_i|Y_j) \quad ⑦$$

---

* ASH'S $I(\mathbf{X}|\mathbf{Y})$ IS OBVIOUSLY OUR $I(\mathbf{X};\mathbf{Y})$

NOW, THE CHANNEL CAPACITY IS

$$C = Max\, I(X_i/Y_j)$$

THUS

$$\Sigma_{i=1}^{n} I(X_i/Y_j) \le \Sigma_{i=1}^{n} C = nC$$

AND ⑦ BECOMES

$$\lg s - H(\mathbf{X}|\mathbf{Y}) \le nC \quad ⑧$$

AT THIS POINT, WE UTILIZE FANO'S BOUND* (THEM 3.7.1 IN ASH):

$$H(\mathbf{X}|\mathbf{Y}) \le H_E + \overline{P(E)}\lg s-1 \quad ⑨$$

WHERE THE ERROR ENTROPY IS

$$H_E = -P_E \lg P_E - (1-P_E)\lg(1-P_E) \quad ⑩$$

NOTE THAT

$$Max\, H_E = \lg 2 = 1 \text{ BIT} \quad ⑪$$

REARRANGING ⑧:

$$H(\mathbf{X}|\mathbf{Y}) \ge \lg s - nC$$

THUS, USING ⑨:

$$H_E + \overline{P(E)}\lg s-1 \ge \lg s - nC$$

OR, USING ⑪

$$\lg 2 + \overline{P(E)}\lg s-1 \ge \lg s - nC$$

OR, SINCE

$$\lg s > \lg s-1$$

WE HAVE

$$\lg 2 + \overline{P(E)}\lg s \ge \lg s - nC$$

---

* WE MUST USE $\overline{P[E]}$ (EXPECTED VALUE) AS OPPOSED TO $P[E]$, SINCE THE INPUT CODING, & THUS THE $P_r[\text{ERROR}]$, ARE RANDOM PROCESSES.

SOLVING FOR $\lg S$:

$$\lg S\ [\overline{P(E)} - 1] \geq \lg 2 - nc$$

$$\lg S \geq \frac{\lg 2 - nc}{P(E) - 1}$$

OR

$$\lg S \geq \frac{nc - \lg 2}{1 - P(E)}$$

<u>QED</u>

(WE HERE ABANDON ASM)

(b) SHOW IF $S \geq 2^{n(c+\delta)}$ $\delta > 0$, THEN

$$\overline{P(E)} \geq 1 - \frac{C + 1/n}{C+\delta}$$

AND, AS $n \to \infty$, $\overline{P(E)} \longrightarrow 0$

FROM PART (a):

$$\lg S \leq \frac{nC + \lg 2}{1 - P(E)}$$

SOLVE FOR $\overline{P(E)}$

$$1 - \overline{P(E)} \leq \frac{nc + \lg 2}{\lg S}$$

$$-\overline{P(E)} \leq \frac{nc + \lg 2}{\lg S} - 1$$

$$\overline{P(E)} \geq 1 - \frac{nc + \lg 2}{\lg S}$$

NOW, IF $S \geq 2^{n(c+\delta)} = (\text{LOG BASE})^{n(c+\delta)\lg 2}$

$$\lg S \geq n(c+\delta)\lg 2 \quad (\lg \text{ IS MONOTONIC} \uparrow)$$

$$\frac{1}{\lg S} \leq \frac{1}{n(c+\delta)\lg 2}$$

$$\frac{-1}{\lg S} \geq \frac{-1}{n(c+\delta)\lg 2}$$

AND

$$\overline{P(E)} \geq 1 - \frac{nc + \lg 2}{n(c+\delta)\lg 2}$$

$$\geq 1 - \frac{c\lg 2 + 1/n}{c+\delta}$$

LETS USE $\lg(\cdot) = \log_2(\cdot)$. THUS

$$\overline{P(E)} \geq 1 - \frac{c + 1/n}{c+\delta} \quad \forall\, n$$

AS $n \to \infty$, $\frac{1}{n} \to 0$, AND

$$\overline{P(E)} \geq 1 - c/c+\delta \underbrace{> 0}_{\text{THIS FOLLOWS FROM } \delta > 0}$$

(c) SOLVING AGAIN FOR $\overline{P(E)}$ IN PART (a):

$$\overline{P(E)} \geq 1 - \frac{nC+1}{lg\, S}$$

WHERE, AGAIN, $lg(\cdot) = log_2(\cdot)$

CONSIDER, THEN,

$$S = 2^{nR} \Rightarrow lg\, S = nR$$

THUS,

$$\overline{P(E)} \geq 1 - \frac{nC+1}{nR}$$
$$\geq 1 - \left(\frac{C+1/n}{R}\right)$$

LET $R = C + \Delta \ni \Delta$ IS A FIXED POSITIVE NUMBER. THEN:

$$\overline{P(E)} \geq 1 - \left(\frac{C+1/n}{C+\Delta}\right)$$

CLEARLY, AS $n \to \infty$, WE HAVE

$$\overline{P(E)} \geq 1 - \frac{C}{C+\Delta}$$

WHICH, FOR A FIXED Δ, REMOVES THE POSSIBILITY FOR EQUALITY.

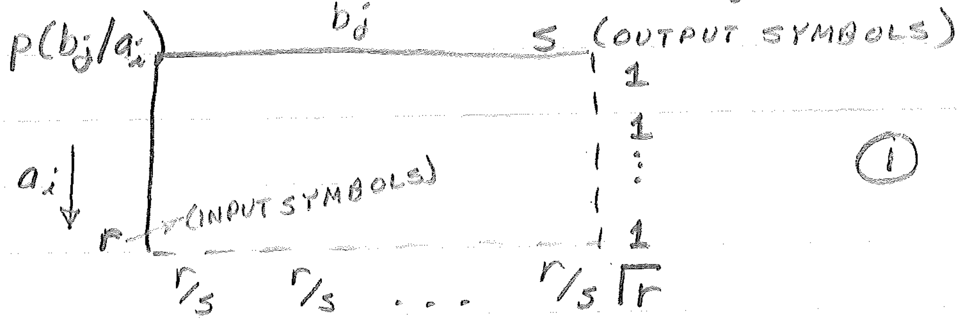
THUS:

$$\overline{P(E)} > 1 - \frac{1}{1+\Delta/C}$$

CLEARLY, THE LARGER Δ, THE WORST ASYMPTOTICALLY $\overline{P(E)}$. AND, OBLIOUSLY, $\lim_{n \to \infty} \overline{P(E)} \neq 0$

NOTE: ON Pg. 83, ASH SHOWS THAT IS WORSE THAN THIS, THAT, IN FACT, FOR $R > C$, THAT $\overline{P_n(E)} \Longrightarrow 1$.

3. WE BEGIN BY WRITING DOWN THE CONDITIONAL MATRIX $P[b_j/a_i]$



①

SINCE THE CHANNEL IN UNIFORM, EACH OF ITS COLUMNS MUST ADD TO r/s.

NOW, WE ARE GIVEN THAT

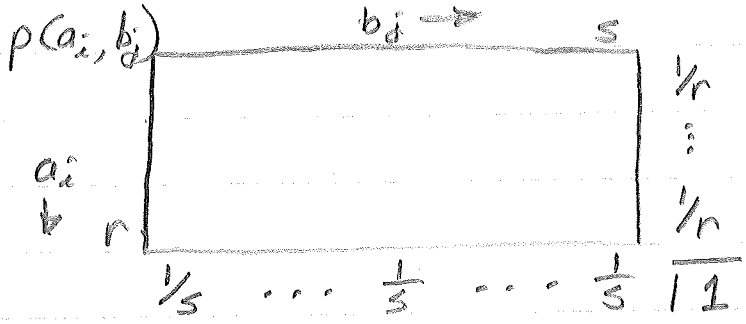
$$p(a_i) = 1/r \qquad ②$$

THUS, THE EQUIVOCATION MAY BE WRITTEN:

$$H(A/B) = -\sum_{j=1}^{s}\sum_{i=1}^{r} p(a_i, b_j) \lg p(a_i/b_j) \qquad ③$$

$$= -\sum_j \sum_i P(b_j) P(a_i/b_j) \lg p(a_i/b_j)$$

LETS NEXT GENERALIZE THE JOINT MATRIX BY MULTIPLYING ALL ELEMENTS IN ① BY ②:



④

THUS

$$p(b_j) = 1/5 \qquad ⑤$$

AND ③ BECOMES

$$H(A/B) = -\frac{1}{5}\sum_j \sum_i P(a_i/b_j) \lg p(a_i/b_j)$$

NOW, SINCE ALL COLUMNS HAVE EQUIVALENT ELEMENTS,* WE MAY WRITE

$$H(A/B) = -\frac{1}{5}\left[\sum_i 5\, P(a_i/b_1) \lg p(a_i/b_1)\right]$$
$$= -\sum_i P(a_i/b_1) \lg p(a_i/b_1) \qquad ⑥$$

WHERE WE HAVE ARBITRARILY CHOSEN THE $j=1$ COLUMN FOR REPRESENTATION.

NOW, LET'S CHOOSE THE MAX LIKELIHOOD (IDEAL OBSERVER) DECISION RULE:

$$p(a^*/b_j) \geq p(a_i/b_j) \quad \forall\, i$$

THE VALUE, $p(a^*/b_j)$ WILL OF COURSE BE THE SAME $\forall\, j$.

THE PROB. OF ERROR IS

$$P_E = 1 - P(a^*/b_j) = 1 - P(a^*/b_1)$$

---

* IN DIFFERING ORDER

LET THE ENTROPY ASSOCIATED WITH P_E BE

$$H_E \triangleq -P_E \lg P_E - (1-P_E) \lg (1-P_E)$$

BY THE ADDITIVE RULE OF ENTROPIES, IT IS OBVIOUS THAT

$$H_E \leq -\Sigma_{i=1}^{r} p(a_i/b_1) \lg p(a_i/b_j)$$

THUS, SUBSTITUTING INTO ⑥, WE HAVE OUR DESIRED BOUND

$$H(A/B) \geq H_E$$

NOTE THAT THIS BOUND IS MET FOR THE $P(a_i/b_j)$ MATRIX

$$\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

THUS, BY ALSO UTILIZING THE FANO BOUND, WE CAN WRITE (FOR THE GIVEN UNIFORM CHANNEL):

$$H_E \leq H(A/B) \leq H_E + P_E \lg r-1$$

4. FOR $m=5$ $(\Rightarrow k=4)^{*}$, WE HAVE 32 PERMISSABLE CODES. WE'LL CHOOSE TWO, AND HAMMING CODE THEM. AN EASY ONE TO CODE IS

0 0 0 0 0

THE CODE FOR WHICH IS CLEARLY**

$$\begin{array}{ccccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & \uparrow & & & & \uparrow & & \uparrow & \uparrow \\ & P_3 & & & & P_2 & & P_1 & P_0 \end{array}$$

FOR OUR SECOND CODE, CHOOSE

$(31)_2 = 1\,1\,1\,1\,1$

CODING

$$\begin{array}{ccccccccc} 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ & \uparrow & & & & \uparrow & & \uparrow & \uparrow \\ & P_3 & & & & P_2 & & P_1 & P_0 \end{array}$$

$$P_0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 = 0 \Rightarrow P_0 = 0$$
$$P_1 \oplus 1 \oplus 1 \oplus 1 = 0 \Rightarrow P_1 = 1$$
$$P_2 \oplus 1 \oplus 1 \oplus 1 = 0 \Rightarrow P_2 = 1$$
$$P_3 \oplus 1 = 0 \Rightarrow P_3 = 1$$

WE COULD, OF COURSE, CONTINUE AND ENCODE ALL POSSIBLE 32 INPUT WORDS.

* FROM CLASS

** EVEN 1 PARITY

TO COMPUTE THE ERROR PROBABILITIES, WE WILL ASSUME ALL 32 WORDS ARE IN THE INPUT ALPHABET (THIS PRECLUDES, FOR EXAMPLE, A 00001 BEING INTERPR. AS A 00000 WHEN ONLY, SAY, 00000 AND 11111 ARE IN THE INPUT ALPHABET)

FOR THE UNCODED WORD, THE PROBABILITY OF MAKING ONE OR MORE ERROR IS

$$P_E = \binom{n}{1} p\bar{p}^{n-1} + \binom{n}{2} p^2 \bar{p}^{n-2} + \dots + \binom{n}{5} p^n$$
$$= 1 - (1-p)^5 \quad , (\text{ie } n=5)$$

WHERE p IS THE PROB. A 0 IS RECEIVED AS A 1 AND VISA VERSA. FOR $p = 1/100$:

$$P_E = 1 - \left(\frac{99}{100}\right)^5 = 4.901 \times 10^{-2} = .04901$$

FOR THE CODED SCHEME

$$P_E^{(c)} = \binom{n}{2} p^2 \bar{p}^{n-2} + \binom{n}{3} p^3 \bar{p}^{n-3} + \dots + \binom{n}{9} p^n$$
$$= \sum_{i=0}^{n} \binom{n}{i} p^i \bar{p}^{n-i} - \binom{n}{0} p^0 \bar{p}^n - \binom{n}{1} p^1 \bar{p}^{n-1}$$
$$= (p + \bar{p})^n - \bar{p}^n - np\bar{p}^{n-1}$$
$$= 1 - (1-p)^n - np(1-p)^{n-1}$$

FOR $n = 9$, $p = \frac{1}{100}$:

$$P_E^{(c)} = 1 - \left(\frac{99}{100}\right)^9 - \frac{9}{100}\left(\frac{99}{100}\right)^8$$
$$= 0.003436 > 0.04901 \quad (\text{ie, HAMMING'S BETTER})$$

$$\text{FIGURE OF MERIT} = \frac{0.04901}{0.003436} = 14.265$$

7/14/76
(1)

HOMEWORK

PROB: WRITE OUT $P(A_i, B_j, C_k)$ IN TERMS OF VARIOUS CONDITIONAL & MARGINAL PROBABILITIES

FOR TWO EVENTS:

$$P(A,B) = P(A/B)P(B)$$

IT FOLLOWS THAT

$$\begin{aligned} P(A,B,C) &= P(A,B/C)\,P(C) \\ &= P(A/B,C)\;P(B,C) \\ &= P(A/B,C)\;P(B/C)P(C) \\ &= P(A/B,C)\;P(C/B)P(B) \end{aligned}$$

CLEARLY, WE MAY INTERCHANGE A, B, & C IN ANY DESIRED FASION

7/14/76
(2)

H.W.

PROB: YOU HAVE 6 RED BALLS AND 4 BLACK BALLS. YOU MAKE TWO DRAWS, X_1 AND X_2. FIND JOINT, MARGINAL, AND CONDITIONAL PROBABILITIES.

FIRST OFF, $P[X_1, X_2] = \frac{\text{TOTAL RELEVENT EVENTS}}{\text{TOTAL \# OF EVENTS}}$

DEFINING: ${}^{n}C_{m} = \frac{n!}{(n-m)!}$

TOTAL # RELEVENT EVENTS = ${}^{10}C_2 = 10 \cdot 9 = 90$

TOTAL WAYS TO GET R,R = ${}^{6}C_2 = 6 \cdot 5 = 30$
TOTAY WAYS TO GET R,B = $6 \times 4 = 24$
TOTAL WAYS TO GET B,R = $6 \times 4 = 24$
TOTAL WAYS TO GET B,B = ${}^{4}C_2 = 12$

$X_1 \backslash X_2$	R	B
R	$1/3$	$\frac{24}{90} = \frac{4}{15}$
B	$4/15$	$12/90 = 2/15$

$$P[X_2 = R] = P[X_2 = R, X_1 = B] + P[X_2 = R, X_1 = R]$$
$$= \frac{1}{3} + \frac{4}{15} = 9/15 = 3/5$$

$$P[X_2 = B] = 1 - P[X_2 = R] = 2/5$$

$X_1 \backslash X_2$	R	B	
R	$5/15$	$4/15$	$9/15 \Leftarrow P[X_1 = R]$
B	$4/15$	$2/15$	$6/15 \Leftarrow P[X_1 = B]$
	$9/15$	$6/15$	

$\Uparrow$

$P[X_2 = R]$ $P[X_1 = R]$

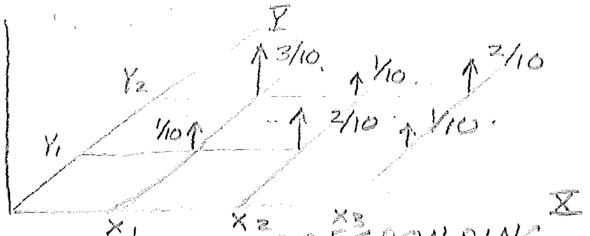
CONDITIONAL MATRIX: $P[X_2/X_1] = P(X_1, X_2)/P(X_1)$

$X_1 \backslash X_2$	R	B
R	$5/9$	$4/9$
B	$4/6$	$2/6$

HERE, $P[X_2/X_1] = P[X_1/X_2]$

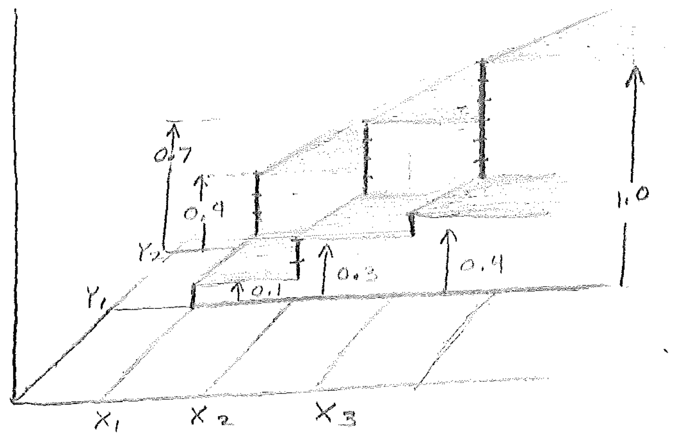
7/15/76
(1)

PROB: CONSIDER THE JOINT pdf:



SKETCH THE CORRESPONDING PDF

$$F(\mathbb{X},\mathbb{Y}) = P(X \le \mathbb{X}, Y \le \mathbb{Y})$$



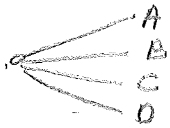
9-20-76

HOMEWORK.

DEMONSTRATE THE FACT THAT THE AVERAGE UNCERTAINTY OF A SYSTEM IS NOT AFFECTED BY THE ARRANGEMENT OF EVENTS

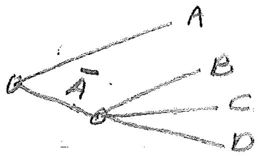
$$P(A) = \frac{1}{2} \qquad P(B) = \frac{1}{4} \qquad P(C) = P(D) = \frac{1}{8}$$

CASE 1:



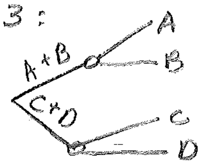
$$H(S) = -P(A)\ln P(A) - P(B)\ln P(B) - P(C)\ln P(C) - P(D)\ln P(D)$$
$$= \frac{1}{4}\ln 4 + \frac{1}{2}\ln 2 + \frac{1}{8}\ln 8 + \frac{1}{8}\ln 8$$
$$= \frac{1}{4}\,2 + \frac{1}{2}\,1 + \frac{1}{8}\,3 + \frac{1}{8}\,3$$
$$= \frac{1}{2} + \frac{1}{2} + \frac{3}{4} = 1\,{}^{3}/_{4} \text{ BITS}$$

CASE 2:



$$H(S) = H(P_A, P_{\bar{A}}) + P(\bar{A})\, H\left(\frac{P(B)}{P(\bar{A})}, \frac{P(C)}{P(\bar{A})}, \frac{P(D)}{P(\bar{A})}\right)$$
$$= \frac{1}{2}\ln 2 + \frac{1}{2}\ln 2$$
$$+ \frac{1}{2}\left[\frac{1/4}{1/2}\ln\frac{1/2}{1/4} + 2\,\frac{1/8}{1/2}\ln\frac{1/2}{1/8}\right]$$
$$= 1 + \frac{1}{2}\left[\frac{1}{2}\ln 2 + \frac{1}{2}\ln 4\right]$$
$$= 1 + \frac{1}{2}\left[\frac{1}{2} + 1\right] = 1 + \frac{3}{4} = 1\frac{3}{4} \text{ BITS}$$

CASE 3:



$$H(S) = H[P(A+B), P(C+D)]$$
$$+ P(A+B)\, H\left[\frac{P(B)}{P(A+B)}, \frac{P(A)}{P(A+B)}\right] + P(C+D)\, H\left[\frac{P(C)}{P(C+D)}, \frac{P(D)}{P(C+D)}\right]$$

$$P[A+B] = 3/4 \qquad P[C+D] = 1/4$$

$$\Rightarrow H(S) = \frac{3}{4}\ln\frac{4}{3} + \frac{1}{4}\ln 4 + \frac{3}{4}\left[\frac{1/4}{3/4}\ln 3 + \frac{2/4}{3/4}\ln\frac{3}{2}\right]$$
$$+ \frac{2}{4}\left[\frac{1/8}{2/8}\ln 2\right] +$$
$$= \frac{3}{4}\cdot 2 - \frac{3}{4}\ln 3 + \frac{1}{2} + \frac{3}{4}\left[\frac{1}{3}\ln 3 + \frac{2}{3}\ln 3 - \frac{2}{3}\right] + \frac{1}{4}$$
$$= \frac{3}{2} + \frac{1}{2} - \frac{1}{2} + \frac{1}{4}$$
$$= \frac{7}{4} = 1\frac{3}{4} \text{ BITS}$$

7/21/76

FROM TEXT
(2-3) p. 41

$S_1 \Rightarrow (P_1, P_2, P_3, \ldots, P_{q_1})$; H_1
$S_2 \Rightarrow (Q_1, Q_2, Q_3, \ldots, Q_{q_2})$; H_2
$S(\lambda) = (r_1, r_2, \ldots, r_{q_1} ; s_1, s_2, \ldots s_{q_2})$
$\Rightarrow (\lambda P_1, \lambda P_2, \ldots \lambda P_{q_1} ; \bar\lambda Q_1, \bar\lambda Q_2, \bar\lambda Q_3, \ldots \bar\lambda Q_{q_2})$

a) SHOW THAT

$$H[S(\lambda)] = \lambda H_1 + \bar\lambda H_2 + H(\lambda)$$

$$H(S(\lambda)) = -\sum_{i=1}^{q_1} \lambda P_i \ln \lambda P_i - \sum_{j=1}^{q_2} \bar\lambda Q_j \ln \bar\lambda Q_j$$
$$= -\lambda \sum_{i=1}^{q_1} P_i \ln P_i - \lambda \sum_{i=1}^{q_1} P_i \ln \lambda$$
$$- \bar\lambda \Sigma_j Q_j \ln Q_j - \bar\lambda \Sigma_j Q_j \ln \bar\lambda$$
$$= \lambda H_1 - \lambda \ln \lambda + \bar\lambda H_2 - \bar\lambda \ln \bar\lambda$$
$$= \lambda H_1 + \bar\lambda H_2 - \lambda \ln \lambda - \bar\lambda \ln \lambda$$

NOW $H(\lambda) = -\lambda \ln \lambda - \bar\lambda \ln \lambda$ (ie, $\lambda + \bar\lambda = 1$)

$$\Rightarrow H[S(\lambda)] = \lambda H_1 + \bar\lambda H_2 + H(\lambda)$$

INTERPRETATION: BY TAKING TWO SOURCES, S_1 & S_2 TOGETHER, AND LET ONLY ONE GENERATE A SYMBOL. THE EXPECTED PROPORTION OF TIME S_1 GENERATES IS λ, AND S_2 IS $\bar\lambda = 1 - \lambda$. THIS PROBLEM GIVES RESULTING ENTROPY.

b. FIND λ_0 THAT MAXIMIZES $H[S(\lambda)]$

$$\frac{d}{d\lambda} H[S(\lambda)] = H_1 - H_2 + \frac{d}{d\lambda} H(\lambda)$$
$$\frac{d}{d\lambda} H(\lambda) = -\frac{d}{d\lambda} \lambda \ln \lambda - \frac{d}{d\lambda} (1-\lambda) \ln (1-\lambda)$$
$$= -\ln \lambda - 1 - \frac{d}{d\lambda} \ln (1-\lambda) + \frac{d}{d\lambda} \lambda \ln (1-\lambda)$$
$$= -\ln \lambda - 1 + \frac{1}{1-\lambda} + \ln (1-\lambda) - \frac{\lambda}{1-\lambda}$$
$$= -\ln \lambda + \ln (1-\lambda) + \frac{1-\lambda}{1-\lambda} - 1$$
$$= \ln \frac{1-\lambda}{\lambda}$$
$$\Rightarrow \frac{d}{d\lambda} H[S(\lambda)] = 0 = H_1 - H_2 + \ln \frac{1-\lambda}{\lambda}$$
$$\Rightarrow \ln \frac{\lambda}{1-\lambda} = H_1 - H_2 \Rightarrow \frac{\lambda}{1-\lambda} = e^{H_1 - H_2}$$
$$\lambda_0 = (1-\lambda_0) e^{H_1 - H_2} \Rightarrow \lambda_0 (1 + e^{H_1 - H_2}) = e^{H_1 - H_2}$$

OR $\lambda_0 = \dfrac{e^{H_1 - H_2}}{1 - e^{H_1 - H_2}} = \dfrac{1}{e^{H_2 - H_1} - 1}$

$$H[S(\lambda_0)] = \frac{H_1}{e^{H_2 - H_1}} + \left[1 - \frac{1}{e^{H_2 - H_1} - 1}\right] H_2$$
$$- \lambda_0 \ln \lambda_0 - \bar\lambda_0 \ln \bar\lambda_0$$

ETC.

-39-
TEXT

(2-14) $S = \{s_1, s_2, \dots, s_q\} \Rightarrow \{P_1, P_2, \dots, P_q)$

$S' = \{r_1, r_2, \dots, r_q, r_{q+1}, \dots\ r_{2q}\} \Rightarrow \{P_1', P_2', \dots, P_{2q}')$

$$P_i' = \begin{cases} (1-\epsilon) P_i & ; i = 1, 2, \dots, q \\ \epsilon P_{i-q} & ; i = q+1, q+2, \dots, 2q \end{cases}$$

$$H[S] = -\sum_{i=1}^{q} p_i \lg p_i$$

$$H[S'] = -\sum_{i=1}^{q} (1-\epsilon) p_i \ln (1-\epsilon) p_i - \sum_{i=1}^{q} \epsilon p_i \ln \epsilon p_i$$

$$= -(1-\epsilon) \Sigma_i p_i \ln (1-\epsilon) - (1-\epsilon) \Sigma_i p_i \ln p_i - \epsilon \Sigma_i p_i \ln \epsilon - \epsilon \Sigma_i p_i \ln p_i$$

$$= -(1-\epsilon) \ln (1-\epsilon) - \epsilon \ln \epsilon + (1-\epsilon) H(S) + \epsilon H(S)$$

$$= H(\epsilon) + H(S)$$

LET $X = \{x_1, x_2\}$ $P: (\frac{1}{4}, \frac{3}{4})$

$Y = \{Y_1, Y_2, Y_3\}$

$P(Y_1/X_1) = \frac{1}{4}$ $P(Y_1/X_2) = 0.10$

$P(Y_2/X_1) = 0.35$ $P(Y_2/X_2) = 0.7$

$P(Y_3/X_1) = 0.40$ $P(Y_3/X_2) = 0.2$

1. FIND $H(x)$

$H(X) = \frac{1}{4} \ln 4 + \frac{3}{4} \ln \frac{4}{3} = 0.532$ NATS $= 0.811$ BITS

2. FIND $H(X,Y)$

$P(Y_1, X_1) = P(Y_1/X_1) P(X_1) = (0.25)(0.25) = 0.0625$

$P(Y_1, X_2) = P(Y_1/X_2) P(X_2) = (0.10)(0.75) = 0.075$

$P(Y_2, X_1) = (0.35)(0.25) = 0.0875$

$P(Y_2, X_2) = (0.7)(0.75) = 0.525$

$P(Y_3, X_1) = (0.4)(0.25) = 0.1$

$P(Y_3, X_2) = (0.2)(0.75) = 0.15$

$H(X,Y) = 1.43$ NATS $= 2.07$ BITS

3. FIND $H(Y)$

$P(Y_1) = 0.0625 + 0.075 = 0.1375$

$P(Y_2) = 0.0875 + 0.525 = 0.6125$

$P(Y_3) = 0.25$

$H(Y) = 0.920$ NATS $= 1.33$ BITTS

4. $H(Y/X) = H(X,Y) - H(X)$

$= 2.07 - 0.811 = 1.60$ BITS

5. $H(X/Y) = H(X,Y) - H(Y)$

$= 2.07 - 1.33 = 0.74$ BITS

9/21/

– 39 –
(ASH)
1-3.

$S = \begin{pmatrix} p & \bar{p} \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix}$ $p \Rightarrow$ PASS

$S = \begin{pmatrix} c & \bar{c} \\ & \end{pmatrix}$ $c \Rightarrow$ OWN CAR

$p(c/p) = 0.1$

$S = \begin{pmatrix} f & \bar{f} \\ & \end{pmatrix}$ $f \Rightarrow$ FRAT MEMBER

$p(c/\bar{p}) = 0.5$

$p(c/f) = 1.0$

$p(f/\bar{c}, p) = 1.0$

$p(f/c, \bar{p}) = 0.4$

a. FIND $H(S_c) = p_c \lg \frac{1}{p_c} + p_{\bar{c}} \lg \frac{1}{p_{\bar{c}}}$

$p(c) = p(c,p) + p(c,\bar{p})$

$= p(c/p)\,p(p) + p(c/\bar{p})\,p(\bar{p})$

$= (0.1)(0.75) + (0.5)(\frac{1}{4}) = 0.2$

$H(S_c) = 0.2 \lg 5 + 0.8 \lg 1.25$

$= 0.5$ NATS $= 0.722$ BITTS

b. FIND $H(S_f) = p_f \lg \frac{1}{p_f} + p_{\bar{f}} \lg \frac{1}{p_{\bar{f}}}$

$p(f) = p(f,c,p) + p(f,\bar{c},p) + p(f,c,\bar{p}) + p(f,\bar{c},\bar{p})$

ETC

7/21/76

BLISH THAT

$Y, Z/X) \le H(Y/Z) + H(Z/X)$

$Y, Z/X) = H(XYZ) - H(X)$

$H(Y/Z) = H(YZ) - H(Z)$

$H(Z/X) = H(XZ) - H(X)$

$H(XYZ) - H(X) \overset{?}{\le} [H(XZ) + H(YZ)] - [H(X) + H(Z)]$

$H(XYZ) \overset{?}{\le} H(XZ) + H(YZ) - H(Z)$

$\overset{?}{\le} H(XZ) + H(YZ/Z)$

$\overset{?}{\le} H(XZ) + H(Y)$

SINCE, BY LEMMA, $H(a,b) \le H(a) + H(b)$,

OUR PROOF IS COMPLETE

c. $H(Z/X,Y) \le H(Z/X)$

$H(Z/X,Y) = H(XYZ) - H(X,Y)$

$H(Z/X) = H(XZ) - H(X)$

$H(XYZ) - H(X,Y) \overset{?}{\le} H(XZ) - H(X)$

$H(XYZ) \overset{?}{\le} H(XY) + H(XZ) - H(X)$

$\overset{?}{\le} H(XY) + H(ZX/X)$

$\overset{?}{\le} H(XY) + H(Z)$

BY SAME ARGUMENT, WE'RE DONE

$H(Y,Z/X) = H(Y/X) + H(Z/XY)$

$H(Y,Z/X) = H(XYZ) - H(X)$

$H(Y/X) = H(Y) - H(X)$

$H(Z/XY) = H(XYZ) - H(XY)$

$H(XYZ) \overset{?}{=} H(Y) + H(XYZ) - H(XY)$

$0 \overset{?}{=} H(Y) + H(XYZ/XY)$

$\overset{?}{=} H(Y) + H(Z)$

NO!

PROBLEM SHOULD HAVE READ

$H(Y,Z/X) \ge H(Y/X) + H(Z/XY)$

7/22/76

H.W. ESTABLISH S^2 WHERE

$$S = \begin{Bmatrix} s_1 & s_2 & s_3 \\ 1/2 & 1/4 & 1/4 \end{Bmatrix}$$

S^2 WILL HAVE $3^2 = 9$ SYMBOLS

$$S^2 = \begin{Bmatrix} s_1s_1 & s_1s_2 & s_1s_3 & s_2s_1 & s_2s_2 & s_2s_3 & s_3s_1 & s_3s_2 & s_3s_3 \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{16} & \frac{1}{16} & \frac{1}{8} & \frac{1}{16} & \frac{1}{16} \end{Bmatrix}$$

$$H(s) = \frac{1}{2} \lg 2 + \frac{1}{4} \lg 4 + \frac{1}{4} \lg 4$$

$$= \frac{1}{2}(1) + 2 \, \frac{1}{4} \cdot 2$$

$$= \frac{1}{2} + 1 = 1\frac{1}{2}$$

$$H^2(s) = \frac{1}{4} \lg 4 + 4 \, \frac{1}{8} \lg 8 + 4 \, \frac{1}{16} \lg 16$$

$$= \frac{1}{4} \cdot 2 + 4 \times \frac{1}{8} \times 3 + \frac{4}{16} \, 4$$

$$= \cancel{\frac{4}{8} + \frac{15}{8} + \frac{7}{16} = \frac{8 + 30 + 7}{16} = \frac{45}{16}}$$

$$= \frac{1}{2} + \frac{3}{2} + 1 = 3$$

THAT IS $H^2(s) = n \, H(s)$ AS EXPECTED

7/22/76

P: WORK OUT CONDITIONAL MATRICES IN TERMS OF JOINT MATRICE

THE JOINT MATRIX IS

$$\begin{array}{c|ccccccc|c} & Y_1 & Y_2 & \cdots & Y_j & \cdots & Y_m & \\ x_1 & p(x_1,Y_1) & p(x_1,Y_2) & & p(x,Y_j) & & p(x_1Y_m) & p(x_1) \\ x_2 & p(x_2,Y_1) & p(x_2,Y_2) & & p(x_2Y_j) & & p(x_2Y_m) & p(x_2) \\ \vdots & & & & & & & \vdots \\ x_i & p(x_i,Y_1) & p(x_i,Y_2) & & p(x_i,Y_j) & & p(x_i,Y_m) & P(x_i) \\ \vdots & & & & & & & \\ x_n & P(x_n,Y_1) & P(x_n,Y_2) & & P(x_nY_j) & & P(x_nY_m) & P(x_n) \\ \hline & P(Y_1) & P(Y_2) & \cdots & P(Y_j) & \cdots & P(Y_n) & \end{array}$$

FOR CONDITIONAL MATRIX $H(X/Y)$, REPLACE $(i,j)^{TH}$ ELEMENT BY $p(x_i/Y_j) = \frac{p(x_i Y_j)}{p(Y_j)}$

& FOR $H(Y/X)$ BY $p(Y_j/x_i) = p(x_i \cdot Y_j)/p(x_i)$

7/22/76

① DISCRETE NOISE FREE CHANNEL: X & Y INDEPENDENT

GIVEN $P(X_i Y_j)$ MATRIX

$$
\begin{aligned}
H(X,Y) &= -\Sigma_i \Sigma_j\, p(x_i Y_j) \ln p(x_i, Y_j) \\
&= -\Sigma_i \Sigma_j\, p(x_i) p(Y_j) \ln p(x_i) p(Y_j) \\
&= -\Sigma_i \Sigma_j\, p(x_i) p(Y_i) \ln p(x_i) \\
&\qquad - \Sigma_i \Sigma_j\, p(x_i) p(Y_j) \ln p(Y_j) \\
&= H(X) + H(Y)
\end{aligned}
$$

② DISCRETE NOISY CHANNEL WITH NOISE

$$
\begin{array}{ccccccc}
 & Y_1 & Y_2 & \ldots & Y_j & \ldots & Y_m & \underline{P(X_i)} \\
x_1 & p_1 & p_1 & & p_1 & & p_1 & n p_1 \\
x_2 & p_2 & p_2 & & p_2 & & p_2 & n p_2 \\
\vdots & & & & & & & \\
x_i & p_i & p_i & & p_i & & p_i & n p_i \\
\vdots & & & & & & & \\
x_n & p_n & p_n & & p_n & & p_n & n p_j \\
 & p_Y & p_Y & & p_Y & & p_Y &
\end{array}
$$

$$\Rightarrow p_Y = \sum_{j=1}^{m} p_i = \frac{1}{m}$$

$$\Rightarrow H(Y) = + \lg m = -\Sigma_i\, p_Y \lg p_Y$$

$$
\begin{aligned}
H(X) &= -n \sum_{i=1}^{n} p_i \lg n p_i \\
&= -n\, \Sigma_i\, p_i \lg n - n\, \Sigma_i\, p_i \lg p_i \\
&= -n \left(\frac{1}{n}\right) \lg n - n\, \Sigma_i\, p_i \lg p_i \\
&= -\lg n \ldots
\end{aligned}
$$

TEXT

MARKS 7/26/76

PROB 5-4; p. 142:

PROVE THAT $I(A^n;B^n) = n\,I(A;B)$

$$\begin{cases} A = \{a_1, a_2, \ldots, a_r) \\ B = \{b_1, b_2, \ldots, b_s) \end{cases}$$

$$\begin{cases} A^n = \{\alpha_1, \alpha_2, \ldots, \alpha_{r^n}\} \\ B^n = \{\beta_1, \beta_2, \ldots, \beta_{s^n}\} \end{cases}$$

NOTATION

$$\begin{cases} P[a_i] = p_i & p[b_j] = p_j \\ p[a_i, b_j] = p_{ij} & p[a_i/b_j] = p_{i/j} \quad \text{ETC.} \end{cases}$$

$$\begin{cases} p(\alpha_i) = p_i^n & p(\beta_j) = p_j^n \\ p(\alpha_i, \beta_j) = p_{ij}^n & p(\alpha_i/\beta_j) = p_{i/j}^n \quad \text{ETC.} \end{cases}$$

NOW

$$I(A;B) = H(A) - H(A/B) \qquad ①$$

IT FOLLOWS THAT

$$I(A^n;B^n) = H(A^n) - H(A^n/B^n) \qquad ②$$

WE HAVE SHOWN (EQ. 2-18 ON pg 21) THAT

$$H(A^n) = n\,H(A) \qquad ③$$

THUS, IT REMAINS TO SHOW THAT

$$H(A^n/B^n) = n\,H(A/B)$$

NOW

$$H(A/B) = -\sum_{i=1}^{r}\sum_{j=1}^{s} p_{ij} \lg p_{i/j} \qquad ④$$

THUS

$$H(A^n/B^n) = -\sum_{i=1}^{r^n}\sum_{j=1}^{s^n} p_{ij}^n \lg p_{i/j}^n$$

NOW

$$\begin{aligned} p_{ij}^n &= p[a_{i_1} a_{i_2} \ldots a_{i_n};\ b_{j_1} b_{j_2} \ldots b_{j_n}] \\ &= p[a_{i_1} b_{j_1}]\, p[a_{i_2} b_{j_2}] \ldots p[a_{i_n} b_{j_n}] \\ &= p_{i_1 j_1}\, p_{i_2 j_2} \ldots p_{i_n j_n} \end{aligned}$$

7/22/76

ALSO

$$p^n_{i/j} = P[\alpha_i/\beta_j]$$
$$= P[a_{i_1} a_{i_2} \dots a_{i_n} / b_{j_1} b_{j_2} \dots b_{j_n}]$$
$$= P[a_{i_1}/b_{j_1}]\,P[a_{i_2}/b_{j_2}] \dots P[a_{i_n}/b_{j_n}]$$
$$= P_{i_1/j_1} P_{i_2/j_2} \dots P_{i_n/j_n}$$

THUS

$$H(A^n/B^n) = -\sum_{i=1}^{r^n} \sum_{j=1}^{s^n} [P_{i_1 j_1} P_{i_2 j_2} \dots P_{i_n j_n}]$$
$$\times \lg [P_{i_1/j_1} P_{i_2/j_2} \dots P_{i_n/j_n}]$$

BUT

$$\sum_{i=1}^{r^n} \sum_{j=1}^{s^n} = \sum_{i_1=1}^{r} \sum_{i_2=1}^{r} \dots \sum_{i_n=1}^{r} \sum_{j_1=1}^{s} \sum_{j_2=1}^{s} \dots \sum_{j_n=1}^{s}$$

$$\therefore H(A^n/B^n) = -\sum_{i_1=1}^{r} \sum_{j_1=1}^{s} P_{i_1 j_1} \sum_{i_2=1}^{r} \sum_{j_2=1}^{s} P_{i_2 j_2} \dots \sum_{i_n=1}^{r} \sum_{j_n=1}^{s} P_{i_n j_n}$$

$$\times [\lg P_{i_1/j_1} + \lg P_{i_2/j_2} + \dots + \lg P_{i_n/j_n}]$$

$$= -\sum_{i_1=1}^{r} \sum_{j_1=1}^{s} P_{i_1 j_1} \lg P_{i_1/j_1} \sum_{i_2=1}^{r} \sum_{j_2=1}^{s} P_{i_2 j_2} \dots \sum_{i_n=1}^{r} \sum_{j_n=1}^{s} P_{i_n j_n}$$

$$- \sum_{i_1=1}^{r} \sum_{j_1=1}^{s} P_{i_1 j_1} \sum_{i_2=1}^{r} \sum_{j_2=1}^{s} P_{i_2 j_2} \lg P_{i_2/j_2} \dots \sum_{i_n=1}^{r} \sum_{j_n=1}^{s} P_{i_n j_n}$$

$$- \dots - \sum_{i_1=1}^{r} \sum_{j_1=1}^{s} P_{i_1 j_1} \sum_{i_2=1}^{r} \sum_{j_2=1}^{s} P_{i_2 j_2} \dots \sum_{i_n=1}^{r} \sum_{j_n=1}^{s} P_{i_n j_n} \lg P_{i_n/j_n}$$

BUT $\displaystyle\sum_{i_K=1}^{r} \sum_{j_K=1}^{s} P_{i_K j_K} = 1$

$$\Rightarrow H(A^n/B^n) = -\sum_{i_1=1}^{r} \sum_{j_1=1}^{s} P_{i_1 j_1} \lg P_{i_1/j_1} - \sum_{i_2=1}^{r} \sum_{j_2=1}^{s} P_{i_2 j_2} \lg P_{i_2/j_2}$$

$$- \dots - \sum_{i_n=1}^{r} \sum_{j_n=1}^{s} P_{i_n j_n} \lg P_{i_n/j_n}$$

7/26/76

FROM ④:

$$H(A/B) = -\sum_{i_k=1}^{r} \sum_{j_k=1}^{s} P_{i_k j_k} \lg P_{i_k j_k}$$

THUS

$$H(A^n/B^n) = n\,H(A/B)$$

SUBSTITUTING THIS AND ③ INTO ② COMPLETES THE PROOF:

$$I(A^n;B^n) = n\,H(A) - n\,H(A/B)$$
$$= n\,I(A;B)$$

7/26/76

PROB: USING THE DEF. OF MUTUAL INFORMATION,
SHOW THAT: $I(X;Y) = H(X) + H(Y) - H(X,Y)$ (a)

$= H(X) - H(X/Y)$ (b)

$= H(Y) - H(Y/X)$ (c)

$$I(X;Y) = \overline{I(x_i, y_j)}$$

$$= \Sigma_i \Sigma_j\, p(x_i, y_j)\, I(x_i, y_j) \quad ①$$

$$= \Sigma_i \Sigma_j\, p(x_i, y_j) \lg p(x_i/y_j)/p(x_i)$$

b. $I(X;Y) = \Sigma_i \Sigma_j\, p(x_i, y_j) \lg p(x_i/y_j)/p(x_i)$

$$= \Sigma_i \Sigma_j\, p(x_i, y_j) \lg p(x_i/y_j) - \Sigma_i \Sigma_j\, p(x_i, y_j) \lg p(x_i)$$

$$= \Sigma_i \Sigma_j\, p(x_i, y_j) \lg p(x_i/y_j) - \Sigma_i\, p(x_i) \lg p(x_i)$$

$$= -H(X/Y) + H(X)$$

$$= H(X) - H(X/Y)$$

a) $I(X;Y) = \Sigma_i \Sigma_j\, p(x_i, y_j) \lg \frac{p(x_i/y_j)}{p(x_i)}$

$$= \Sigma_i \Sigma_j\, p(x_i, y_j) \lg \frac{p(x_i y_j)}{p(x_i)\, p(y_j)} \quad ②$$

$$= \Sigma_i \Sigma_j\, p(x_i, y_j) \lg p(x_i y_j) - \Sigma_i \Sigma_j\, p(x_i, y_j) \lg p(x_i) - \Sigma_i \Sigma_j\, p(x_i, y_j) \lg p(y_j)$$

$$= -H(X,Y) - \Sigma_i\, p(x_i) \lg p(x_i) - \Sigma_j\, p(y_j) \lg p(y_j)$$

$$= -H(X,Y) + H(X) + H(Y)$$

$$= H(X) + H(Y) - H(X,Y)$$

b. STARTING FROM ②:

$I(X;Y) = \Sigma_i \Sigma_j\, p(x_i, y_j) \lg \frac{p(y_j/x_i)}{p(y_j)}$

$$= \Sigma_i \Sigma_j\, p(x_i, y_j) \lg p(y_j/x_i) - \Sigma_i \Sigma_j\, p(x_i, y_j) \lg p(y_j)$$

$$= -H(Y/X) - \Sigma_j\, p(y_j) \lg p(y_j)$$

$$= -H(Y/X) + H(Y)$$

$$= H(Y) - H(Y/X)$$

7-28-76

USING MUROGA'S TECHNIQUE, FIND THE CHANNEL CAPACITY FOR

$$p = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix} \Rightarrow \begin{matrix} p_1' = 3/4 \\ p_2' = 5/4 \end{matrix}$$

WE WANNA SOLVE

$$[P]\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} -H(\frac{1}{2}) \\ -H(\frac{1}{4}) \end{bmatrix}$$

$-H(\frac{1}{2}) = 1$ BIT

$$+H(\tfrac{1}{4}) = \tfrac{1}{4} \lg 4 + \tfrac{3}{4} \lg \tfrac{4}{3}$$

$$= \lg 4 - \tfrac{3}{4} \lg 3$$

$$= 0.811278 \qquad (0)$$

$$p^{-1} = C^T / \det p$$

$$\det p = \tfrac{3}{8} - \tfrac{1}{8} = \tfrac{1}{4}$$

$$C = \begin{bmatrix} 3/4 & -1/4 \\ -1/2 & 1/2 \end{bmatrix} \qquad C^T = \begin{bmatrix} 3/4 & -\frac{1}{2} \\ -\frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

$$\Rightarrow p^{-1} = \begin{bmatrix} 3 & -2 \\ -1 & 2 \end{bmatrix}$$

$$\therefore \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} 3 & -2 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ -0.811278 \end{bmatrix}$$

$$= \begin{bmatrix} -3 + 2(0.811278) \\ 1 - 2(0.811278) \end{bmatrix}$$

$$= \begin{bmatrix} -1.3774438 \\ -0.622556 \end{bmatrix} \begin{matrix} (2) \\ (3) \end{matrix}$$

THEN

$$C = lg_2\, 2^{Q_1} + 2^{Q_2}$$
$$= lg_2\, 1.0344192$$
$$= 0.0488210 \text{ BITS}$$

USING THE GRAPH

$P_{11} = \frac{1}{2} \qquad P_{22} = 3/4 \Rightarrow Q_1 \simeq -1.35$

$Q_2 \simeq -0.6$

$C \simeq 0.048$ BITS

✓

7/28/76

GIVEN

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = -\begin{bmatrix} H(p_{11}) \\ H(p_{22}) \end{bmatrix} = \begin{bmatrix} p_{11} \lg p_{11} + p_{12} \lg p_{12} \\ p_{21} \lg p_{21} + p_{22} \lg p_{22} \end{bmatrix}$$

$p_1' = p_{11} + p_{12}$ $\qquad$ $p_2' = p_{22} + p_{21}$

SHOW THAT

$$I(X;Y) = -p_1 \lg p_1' - p_2' \lg p_2' + p_1' Q_1 + p_2' Q_2$$

NOW $\quad H(Y) = -p_1 \lg p_1' - p_2' \lg p_2'$

$$\begin{aligned} \Rightarrow I(X;Y) &= H(Y) + p_1' Q_1 + p_2' Q_2 \\ &= H(Y) + (p_{11} Q_1 + p_{12} Q_2) + (p_{21} Q_1 + p_{22} Q_2) \\ &= H(Y) + [-p_{11} \lg p_{11} - p_{12} \lg p_{12} \\ &\qquad - p_{21} \lg p_{21} - p_{22} \lg p_{22}] \\ &= H(Y) - \Sigma_i \Sigma_j\ p_{ij} \lg p_{ij} \\ &= H(Y) - H(Y/X) \end{aligned}$$

7-29-7

START WITH {10} AND BUILD AN INSTANT CODE

FOR $S = \{s_1, s_2 \ldots s_5\}$

s_1	10
s_2	11
s_3	00
s_4	010
s_5	011

FOR $S : \{s_1, \ldots s_8\}$

s_1	011
s_2	111
s_3	001
s_4	000
s_5	101
s_6	100
s_7	1100
s_8	1101

7/30/76

TEXT: PROB 3-2, p. 63

(a, b) INSTANTANEOUS CODES ARE

$\mathcal{A}, \mathcal{C}$ & $\mathcal{E}$

ALL INSTANT CODES ARE DECODABLE. THUS $\mathcal{A}, \mathcal{C}$, & $\mathcal{E}$ ARE U.D. $\mathcal{B}$ IS ALSO U.D. (BUT DOES NOT OBEY THE PREFIX PROPERTY). $\mathcal{F}$ IS NOT U.D SINCE $s_1 s_1 s_2 = s_6 s_1 s_1 = 00100$

D IS NOT U.D. SINCE $s_2 s_3 = s_1 s_5 = 10110$

TO FIND AVERAGE WORD LENGTHS, USE HP

$\mathcal{A}: \bar{L} = 3$ $\qquad$ $\bar{L} = \Sigma p_i l_i$

$\mathcal{B}: \bar{L} = 2.1250$

$\mathcal{C}: \bar{L} = 2.1250$

$\mathcal{D}: \bar{L} = 1.9375$

$\mathcal{E}: \bar{L} = 1.9375$

$\mathcal{F}: \bar{L} = 2$

[illegible] WHAT

7/30/76

FROM TEXT: # 3-3 pg 63

a. MUST CHECK VIA KRAFT

$$\Sigma_{i=1}^{q} r^{-l_i} = \Sigma_{i=1}^{l} N_i r^{-i} \le 1$$

HERE $r = 3$

```
( )      3      R/S      +
CHS      X≷Y    X        STO 0
ENT      Y^X    RCL 0    STO 00
```

a: 0.91

B: 1.004

C: 0.9999

D: 0.99999

b.

	a	C	D
s_1	0	0	0
s_2	1	10	1
s_4	20	11	20
s_4	210	12	21
s_5	211	20	220
s_6	22100	220	221
s_7	2201	221	2220
s_8	2210	222	2221
s_9	2211	210	22220
s_{10}	222111	211	22221
			22222

8/2/76

DEVISE TWO INSTANT CODES FOR

$n_i = \{0, 3, 0, 5\}$ AND $r = \{0, 1, 2)$

	CODE #1	CODE #2
s_1	01	00
s_2	02	01
s_3	10	10
s_4	2000	1100
s_5	1100	1101
s_6	1101	1110
s_7	1110	1111
s_8	1111	2100

8/2/76

VERIFY THE FOLLOWING CODE LENGTHS WILL WORK FOR $r=3$. WE MUST SATISFY KRAFT'S INEQUALITY:

$$\Sigma_{i=1}^{q} r^{-l_i} = \Sigma_{i=1}^{l} N_i r^{-i} \leq 1$$

1. $i = l_i = \{1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7)$

   $N_i = \{1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 1 \quad 0)$

   $\Sigma_{i=1}^{l} N_i r^{-i} = 0.48$

2. $l_i = 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7$

   $N_i = 1 \quad 2 \quad 0 \quad 3 \quad 2 \quad 1 \quad 0$

   $\Sigma_{i=1}^{l} N_i r^{-i} = 0.60$

3. $N_i = 1 \quad 0 \quad 2 \quad 0 \quad 1 \quad 2 \quad 3$

   $l_i = 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7$

   $= 0.42$

8-3-76

COMPUTE $\bar{L}$ AND $H(x)$ FOR

x_1	0.4	00
x_2	0.3	01
x_3	0.2	101
x_4	0.1	1110

$\bar{L} = (0.4)(2) + (0.3)(2) + (0.2)3 + 0.1(4)$

$= 2.40$

$H(x) = 1.85$ BITS

AND, AS EXPECTED $\bar{L} \geq H(x)$

TEXT: p.91 # 4-7: REQUIRE $q=3+2\alpha \Rightarrow$ LET $\alpha=3 \Rightarrow q=2$

a.

S_1	0.4	0	0.4	0	0.4	0	0.4	0
S_2	0.2	2	0.2	2	0.2	2	0.4	1
S_3	0.1	11	0.1	11	0.2	10	0.2	2
S_4	0.1	12	0.1	12	0.1	11		
S_5	0.05	101	0.1	100	0.1	12		
S_6	0.05	102	0.05	101				
S_7	0.05	1000	0.05	102				
S_8	0.05	1001						
S_9	0.0	1002						

ANOTHER CODE WOULD BE

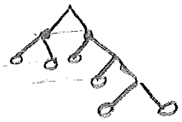
S_1	.4	1	.4	1	.4	1	.4	0
S_2	.2	00	.2	00	.2	2	.4	1
S_3	.1	02	.1	01	.2	00	.2	2
S_4	.1	20	.1	02	.1	01		
S_5	.05	21	.1	20	.1	02		
S_6	.05	22	.05	21				
S_7	.05	010	.05	22				
S_8	.05	011						
S_9	.0	012						

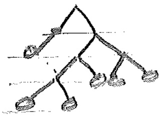
ETC.

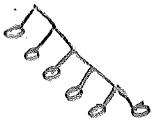
8/4/76

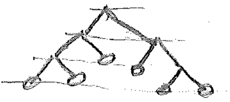
TEXT p.92 (4-12)

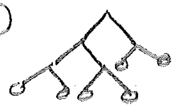
(a) FOR $q=6$, WE SHOULD GET FIVE TREES

①


②


③


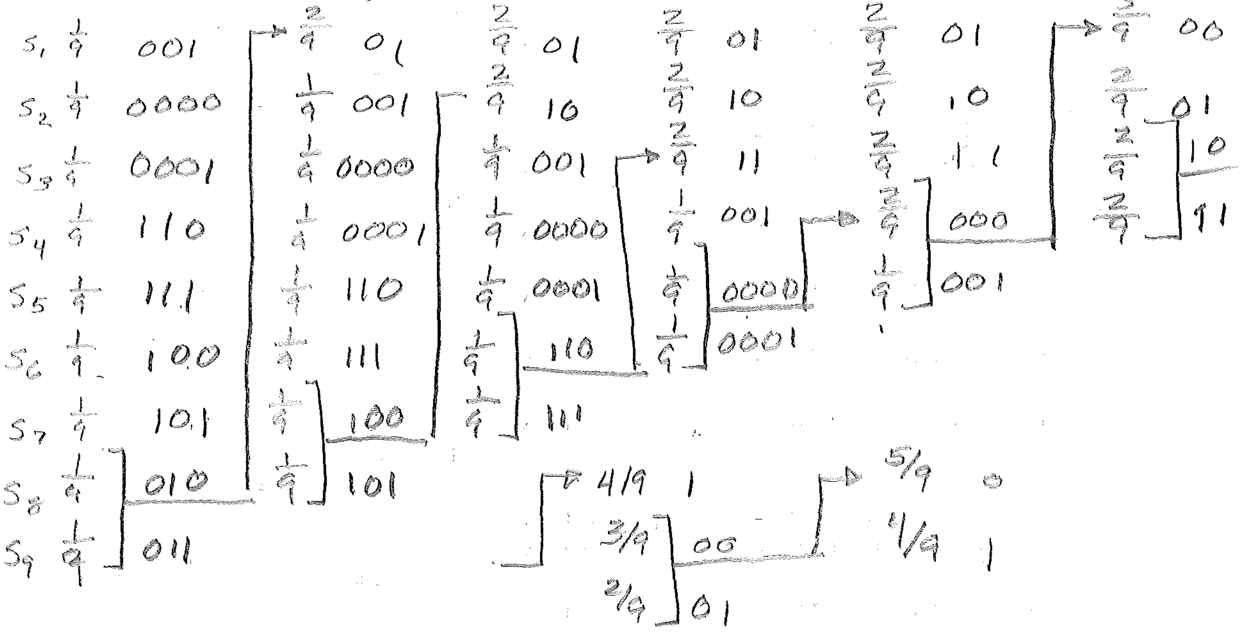
④


⑤


ETC

8/5/76

CODE (HUFFMANN)



$H(S) = 9 \times \frac{1}{9} \lg 9 = 3.17$ BITS

$\bar{L} = 7 \times 3 \frac{1}{9} + 2 \times 4 \frac{1}{9} = \frac{21+8}{9} = \frac{29}{9} = 3.222$

$\Rightarrow \eta = \bar{L} \,/ H_2(S) = 0.984$

STARRED PROBLEMS

✓ H.O # 1 (7-15-76)
✓ 2-4 FROM TEXT, (7-21-76)
✓ H·O #2 PROB 1-6 (7-21)
✓ H-O # 2 PROB 1-12 (7-21)
✓ MAROGA (pg 64 OF NOTES) (7-28-76)
✓ H·O # 3 (MATRICES ON p. 62 OF NOTES) (7-28-76)
✓ H·O # 7 (BINARY MULTIPLICITIVE CHANNEL)
✓ 4-6 FROM TEXT (8-4-76)
Pg. 88 OF NOTES. PROVE THAT, FOR A HUFFMANN CODE, $H(x) \le \bar{L} \le \bar{H}(x) + 1 - 2p_{min}$ (8-4-76)
Pg 97 OF NOTES. (8-5-76) DATE CODING
✓ H·O #8 → COMPUTER PROGRAM
✓ READING SHANNON'S PAPER

TAKE HOME TEST PROBLEMS

7/16/76
#1 ✓

$HW^{\forall}$ (1).

Experiment A has M mutually exclusive possible outcomes A_m, Experiment B has N mutually exclusive possible outcomes B_n. Show that

$$P\left(B_n / A_m\right) = \frac{P\left(A_m / B_n\right) P(B_n)}{\sum_{i=1}^{N} P\left(A_m / B_i\right) P(B_i)}$$

#2

HW **1.6** A change that tends in the following sense to equalize a set of probabilities $p_1, \dots, p_M$ always results in an increase in uncertainty:

Suppose $p_1 > p_2$. Define

$$p_1' = p_1 - \Delta p$$
$$p_2' = p_2 + \Delta p$$

where $\Delta p > 0$ and $p_1 - \Delta p \geq p_2 + \Delta p$. Show that $H(p_1', \dots, p_M') > H(p_1, \dots, p_M)$.

1.7 (Feinstein 1958). Let $A = [a_{ij}]$ be a doubly stochastic matrix, that is, $a_{ij} \geq 0$ for all i, j; $\sum_{j=1}^{M} a_{ij} = 1$, $i = 1, \dots, M$; $\sum_{i=1}^{M} a_{ij} = 1, j = 1, \dots, M$. Given a set of probabilities $p_1, \dots, p_M$, define a new set of probabilities $p_1', \dots, p_M'$ by

$$p_i' = \sum_{j=1}^{M} a_{ij} p_j, \qquad i = 1, 2, \dots, M.$$

Show that $H(p_1', \dots, p_M') \geq H(p_1, \dots, p_M)$ with equality if and only if $(p_1', \dots, p_M')$ is a rearrangement of $(p_1, \dots, p_M)$. Show also that Problem 1.6 is a special case of this result.

1.8 Given a discrete random variable X with values $x_1, \dots, x_M$, define a random variable Y by $Y = g(X)$ where g is an arbitrary function. Show that $H(Y) \leq H(X)$. Under what conditions on the function g will there be equality?

1.9 Let X and Y be random variables with numerical values $x_1, \dots, x_M$; $y_1, \dots, y_L$ respectively. Let $Z = X + Y$.

a. Show that $H(Z \mid X) = H(Y \mid X)$: hence if X and Y are independent, $H(Z \mid X) = H(Y)$ so that $H(Y) \leq H(Z)$, and similarly $H(X) \leq H(Z)$.

b. Give an example in which $H(X) > H(Z)$, $H(Y) > H(Z)$.

1.10 Prove the *generalized grouping axiom.*

$$\begin{aligned} H(p_1, \dots, p_{r_1}; p_{r_1+1}, \dots, p_{r_2}; \dots; p_{r_{k-1}+1}, \dots, p_{r_k}) \\ = H(p_1 + \cdots + p_{r_1}, p_{r_1+1} + \cdots + p_{r_2}, \dots, p_{r_{k-1}+1} + \cdots + p_{r_k}) \\ + \sum_{i=1}^{k} (p_{r_{i-1}+1} + \cdots + p_{r_i}) H\left(\frac{p_{r_{i-1}+1}}{\sum_{j=r_{i-1}+1}^{r_i} p_j}, \dots, \frac{p_{r_i}}{\sum_{j=r_{i-1}+1}^{r_i} p_j} \right) \end{aligned}$$

1.11 Show that if $h(p)$, $0 < p \leq 1$, is a continuous function such that $\sum_{i=1}^{M} p_i h(p_i) = -C \sum_{i=1}^{M} p_i \log p_i$ for all M and all $p_1, \dots, p_M$ such that $p_i > 0$, $\sum_{i=1}^{M} p_i = 1$, then $h(p) = -C \log p$.

HW* ✓**1.12** Given a function $h(p)$, $0 < p \leq 1$, satisfying

a. $h(p_1 p_2) = h(p_1) + h(p_2), \quad 0 < p_1 \leq 1, \quad 0 < p_2 \leq 1.$

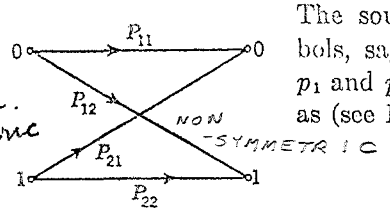
b. $h(p)$ is a monotonically decreasing and continuous function of p, $0 < p \leq 1$.

Show that the only function satisfying the given conditions is $h(p) = -C \log_b p$ where $C > 0$, $b > 1$.

Our first application of t
will be to the problem of
"noiseless" channel, that is
input to output. Thus we d
our only concern is to maxi
over the channel in a given
to be transmitted are gener
$x_1, \dots, x_M$. A noiseless ch
that accepts an input from
and reproduces the input
error. (The formal definiti
until Chapter 3; it will not
communicated properly, ea
symbols chosen from the se
code characters to each x_i
Since the problem of error
cation would involve transm
time. If the rate at which the
fixed, the requirement of effi
as short as possible. In cal
cation, the *average* length of
which we choose to minimiz
To summarize, the ingredi

1. A *random variable* X,
abilities $p_1, \dots, p_M$ respecti
and over again, thus generat
the set $\{x_1, \dots, x_M\}$; such
2. A set $\{a_1, \dots, a_D\}$ cal
alphabet; each symbol x_i i
characters called the *code* w
correspond to $a_1 a_2$, and x_2 t
is called a *code.* The code w

~~Channels~~. The problem undertaken in this section is the evaluation of the maximum rate of transmission of information of binary channels.



metric

FIG. 3-17. BC.

The source transmits independently two symbols, say 0 and 1, with respective probabilities p_1 and p_2. The channel characteristic is known as (see Fig. 3-17)

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

Cond. Matrix

In order to evaluate the capacity of such a channel, when the entropy curve is available a simple geometric procedure can be devised (see Fig. 3-18).

The points A_1 and A_2 on the segment OM are selected so that

$$MA_1 = p_{11} \qquad OA_2 = p_{22}$$

The ordinates of the entropy curve at A_1 and A_2 are

$$\overline{B_1A_1} = H(p_{11}) \qquad \overline{B_2A_2} = H(p_{22})$$

Now, for any given channel output probabilities such as $OA = p$ and

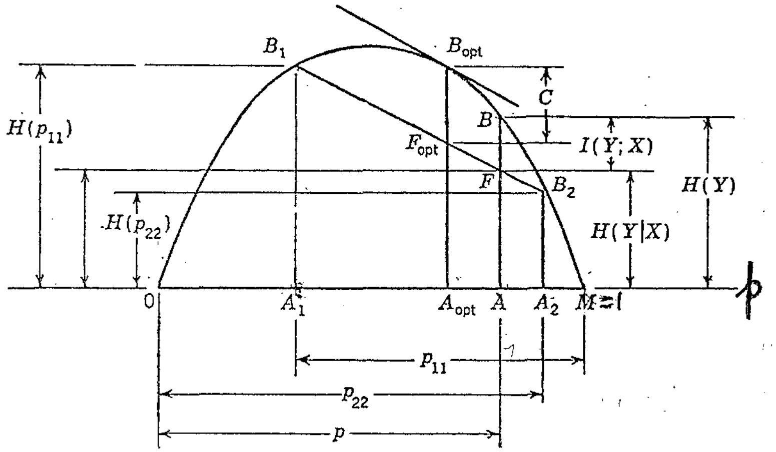


FIG. 3-18. A geometric determination of different entropies, transinformation, and channel capacity of a BC.

$MA = 1 - p$, where p is the probability of receiving 1, the transinformation can be geometrically identified. In fact,

$$I(X;Y) = H(Y) - H(Y|X)$$
$$I(X;Y) = H(p) - p_1H(p_{11}) - p_2H(p_{22})$$
$$I(X;Y) = \overline{BA} - \overline{FA}$$

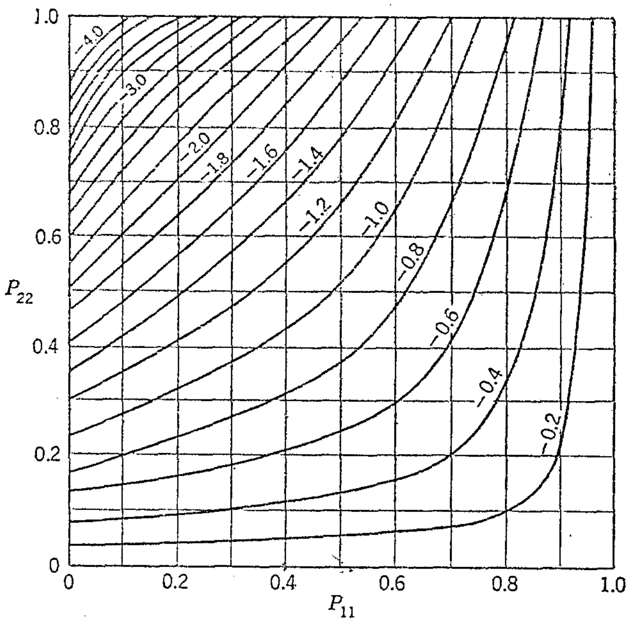
Of course, the point A corresponding to the desired mode of operation is not known. A glance at Fig. 3-18 suggests that the largest value of

transinformation is obtained when the probabilities at the receiving end are represented by point A_{opt} corresponding to point B_{opt}. The tangent of the entropy curve at point B_{opt} is parallel to B_1B_2. At B_{opt} the vertical segment representing the transinformation assumes its largest value. The corresponding source probabilities can be derived in a direct manner.

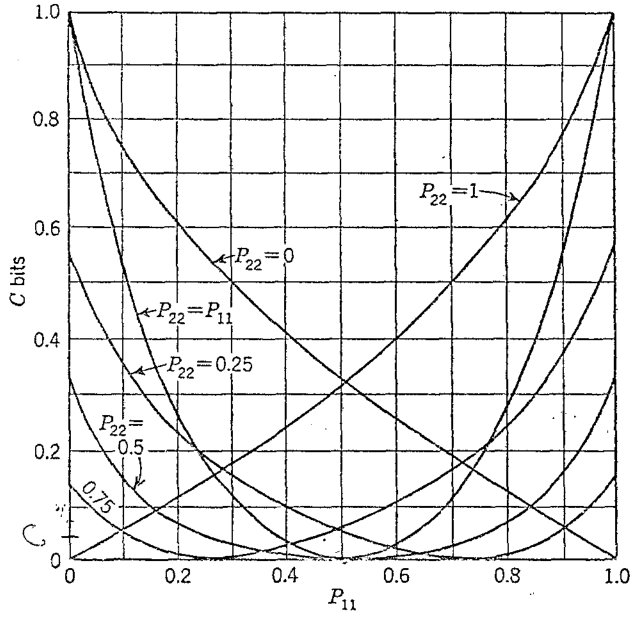
Ref HW + assigned

1-28-1976

#3



A chart for determining values of Q_1 in terms of P_{11} and P_{22} for binary channels. The corresponding value of Q_2 is obtained by an interchange of P_{11} and P_{22}.



Capacity of a binary channel in terms of P_{11} and P_{22}.

T
B

T

La
res
cha
(II
No
wh

7-28-76

Ex
direc
three

(a,
(b)
(c)

Sol

(a)

Direct

This cl

(b)

The no

Any inp
the outp
and 3-2(

HAND-OUT #4

HW* 7-29-76.

Barton, Bell, Blackketter, Golden, Froehlich

The following is a program to determine the ch. cap for a 4x4 channel using Muroga's Technique. Use it to determine C for

$$\begin{vmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{vmatrix}$$

#5

```
C
C     THIS PROGRAM COMPUTES THE CHANNEL CAPACITY OF AN NXN CHANNEL WHEN
C     THE CHANNEL (CONDITIONAL PROBABILITY) MATRIX P(B/A) IS KNOWN.
C
C
C     THE CHANNEL CAPACITY IS DETERMINED USING MUROGA'S TECHNIQUE.
C
C     THE SYSTEM OF LINEAR EQUATIONS THAT IS GENERATED BY THIS TECHNIQUE
C     IS SOLVED BY MEANS OF A GAUSS-JORDAN REDUCTION.
C
C
C
C     ****************************  WARNING  ***************************
C
C     ONE MUST ARRANGE THE ROWS OF THE CHANNEL (CONDITIONAL PROBABILITY)
C     MATRIX P(B/A) SO THAT THE MAIN DIAGONAL ELEMENTS ARE NON-ZERO.
C     REARRANGING THE ROWS DOES NOT AFFECT THE VALUE OF THE CHANNEL
C     CAPACITY AND HELPS TO INSURE THAT THE GAUSS-JORDAN REDUCTION
C     WILL YIELD THE PROPER SOLUTION TO THE SYSTEM OF LINEAR EQUATIONS.
C
C     ******************************************************************
C
C
C     THE CHANNEL (CONDITIONAL PROBABILITY) MATRIX P(B/A) IS READ IN BY
C     ROWS,I.E.,A(1,1),A(1,2),....,A(1,N),A(2,1),A(2,2),...,A(N,N).
C
C
C                    DEFINITION OF VARIABLES
C
C
C     N              DIMENSION(N) OF THE NXN CHANNEL MATRIX P(B/A)
C
C     EPS            PARAMETER FOR CHECKING THE SINGULARITY OF THE
C                    COEFFICIENT MATRIX
C
```

```
C      A(I,J)          ELEMENT IN THE ITH ROW AND JTH COLUMN OF THE CHANNEL
C                      MATRIX
C
C      SUM             ELEMENT IN THE (J+1)TH COLUMN OF THE AUGMENTED CHANNEL
C                      MATRIX,(SUM = (1.0/ALOG(2.))*(A(I,1)*ALOG(A(I,1)) +
C                      A(I,2)*ALOG(A(I,2)) + ... + A(I,N)*ALOG(A(I,N)))
C
C      MAGIC           FACTOR FOR CONVERTING LOGARITHMS TO THE BASE 2 TO
C                      LOGARITHMS TO THE BASE E.(MAGIC = ALOG(2.))
C
C      C               CHANNEL CAPACITY (C = MAGIC*ALOG(2**X(1) + 2**X(2) + ...
C                      + 2**X(N)))WHERE X(I) IS THE ELEMENT IN THE ITH ROW
C                      &(J+1)TH COLUMN OF THE SOLUTION MATRIX
C      DIMENSION A(50,51)
       REAL MAGIC
       MAGIC = (1./0.59315)
C  THE MATRIX WHICH IS READ IS THE CHANNEL MATRIX AND THE ELEMENTS ARE
C  READ IN BY ROWS
       WRITE(6,398)
   398 FORMAT('1')
       READ(5,100)N,EPS
   100 FORMAT(I4,E10.5)
       NP1 = N + 1
       WRITE(6,509)
   509 FORMAT(/)
       WRITE(6,405)
   405 FORMAT(15X,'***',34X,'***',/15X,'*',38X,'*',/15X,'*',38X,'*')
       DO 2 I=1,N
       SUM = 0.0
       READ(5,101)(A(I,J),J=1,N)
       DO 3 J=1,N
   101 FORMAT(4E20.8)
       IF(A(I,J))4,3,4
     4 SUM = A(I,J)*MAGIC*ALOG(A(I,J)) + SUM
     3 CONTINUE
       A(I,N+1) = SUM
       WRITE(6,201)(A(I,J),J = 1,NP1)
   201 FORMAT(17X,5F7.3/)
     2 CONTINUE
       WRITE(6,417)
   417 FORMAT(15X,'*',38X,'*'/15X,'*',38X,'*'/15X,'***',34X,'***')
       WRITE(6,509)
       WRITE(6,411)
   411 FORMAT(23X,'AUGMENTED CHANNEL MATRIX')
       WRITE(6,209)
   209 FORMAT(///)
C  BEGIN GAUSS-JORDAN REDUCTION
       DETER = 1.0
       DO 9 K=1,N
       DETER = DETER*A(K,K)
       IF(DABS(A(K,K)).GT.EPS) GO TO 5
       WRITE(6,202)
       GO TO 111
   202 FORMAT( 8X,'PIVOT ELEMENT SMALL - MATRIX MAY BE SINGULAR')
     5 KP1 = K+1
       DO 6 J =KP1,NP1
     6 A(K,J) = A(K,J)/A(K,K)
       A(K,K) = 1.
       DO 9 I = 1,N
       IF(I.EQ.K.OR.A(I,K).EQ.0.) GO TO 9
       DO 8 J = KP1,NP1
     8 A(I,J)= A(I,J) - A(I,K)*A(K,J)
       A(I,K)=0.
    9 CONTINUE
      WRITE(6,405)
      DO 10 I = 1,N
   10 WRITE(6,201)(A(I,J),J=1,NP1)
      WRITE(6,417)
      WRITE(6,509)
      WRITE(6,412)
  412 FORMAT(25X,'SOLUTION MATRIX FOR X')
      WRITE(6,209)
```

```
  203 FORMAT(E20.8,10X,I4,10X,I4)
C     CALCULATION OF CHANNEL CAPACITY
      SUMX=0.0
      DO 12 I=1,N
      SUMX = 2.**A(I,NP1)+ SUMX
   12 CONTINUE
      IF(SUMX)14,15,14
   15 C=0.0
      GO TO 16
   14 C = MAGIC*ALOG(SUMX)
   16 CONTINUE
      WRITE(6,413) C
  413 FORMAT( 8X,'C = LOG2 (  2**X(1) + 2**X(2) + ... + 2**X(N)   )  =
     1',F6.3)
  111 STOP
      END
C
C                    INPUT DATA FOR EXAMPLE PROBLEM
C
   41.00000-10
C
        .25000000+00        .50000000+00        .25000000+00
                           1.0
                                                1.0
                            .33333333+00                                  .66666667
C
```

```
               ***                                      ***
               *                                          *
               *                                          *
                    .250   .500   .250   .000 -1.500

                    .000  1.000   .000   .000   .000

                    .000   .000  1.000   .000   .000

                    .000   .333   .000   .667  -.918

               *                                          *
               *                                          *
               ***                                      ***



                       AUGMENTED CHANNEL MATRIX



               ***                                      ***
               *                                          *
               *                                          *
                   1.000   .000   .000   .000 -6.000

                    .000  1.000   .000   .000   .000

                    .000   .000  1.000   .000   .000

                    .000   .000   .000  1.000 -1.377

               *                                          *
               *                                          *
               ***                                      ***



                         SOLUTION MATRIX FOR X



        C = LOG2 (  2**X(1) + 2**X(2) + ... + 2**X(N)   )  =   1.263
```

#5

#6

EE 5325 2nd Summer, 1976

Sometime after 2nd quiz, problems in Quizes 1 and 2 will be put on the board by students as per the following:

Quiz 1 (7/23/76)

Problem 1. Kirbie
2. Golden
3. ~~Huang~~ Mote.
4. Froehlich
5. Redus
6. Barton
7. Hill

The schedule for 2nd Quiz will be circulated after August 4. Please bring your answer sheets with you so you may know where exactly things went wrong!

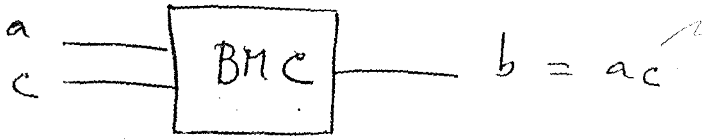
J. Prabhakar
J. C. Prabhakar
7/27/76

HW#

5325

Summer 2 1976

The following is a sketch for Binary Multiplica Channel (BMC). This channel is modelled as a Zero-memory channel with 4 possible inputs making up the alphabet $A : \{00, 01, 10, 11\}$



(a) Develop the output alphabet and the channel (Conditional) matrix

(b) Let the primary symbols be 0, 1 and the corresponding probs. be p & q respectively. Derive an expression for $I(A,B)$, the Trans-information.

(c) Hence determine the Ch: Capacity, if you can!

Due 8/13/76

HO #8
8/5/76

The following is a decoding algorithm for a Huffman Compact Code! Program This to determine the sequence of letters of English Language transmitted if the received sequence is the one at the end of p 131. The Coding <u>is</u> noiseless.

0.1859

```
C   THIS PROGRAM DECODES A HUFFMAN CODE USING A GIVEN TABLE
C   AND THE TREE SEARCH METHOD
C   ****************************************************
C   PROB. *LETTER*HUFF. CODE    *  PROB. *LETTER*HUFF. CODE
C   ****************************************************
C   0.1569* SPACE* 000           *  0.0574*   N  * 1001
C   0.0642*   A  * 0100          *  0.0632*   O  * 0110
C   0.0127*   B  * 011111        *  0.0152*   P  * 011110
C   0.0218*   C  * 11111         *  0.0008*   Q  * 0111001101
C   0.0317*   D  * 01011         *  0.0484*   R  * 1101
C   0.1031*   E  * 101           *  0.0514*   S  * 1100
C   0.0208*   F  * 001100        *  0.0796*   T  * 0010
C   0.0152*   G  * 011101        *  0.0228*   U  * 11110
C   0.0467*   H  * 1110          *  0.0083*   V  * 0111000
C   0.0575*   I  * 1000          *  0.0175*   W  * 001110
C   0.0008*   J  * 0111001110    *  0.0013*   X  * 0111001100
C   0.0049*   K  * 01110010      *  0.0164*   Y  * 001111
C   0.0321*   L  * 01010         *  0.0005*   Z  * 0111001111
C   0.0198*   M  * 001101        *  AVERAGE LENGTH = 4.1195
C   ****************************************************
C
C   THE ABOVE TABLE WAS TAKEN FROM THE FOLLOWING REFERENCE
C   REFERENCE:  REZA, 'AN INTRODUCTION TO INFORMATION THEORY'
C
C   M(I) IS THE BINARY BIT (0 OR 1) BEING OBSERVED    I=1,2,--L
C   Y(J) IS THE ALPHABET LETTER (OR SPACE) BEING FORMED FROM
C   THE CODE.
C
      DIMENSION M(320),Y(100)
      READ(5,99980)L
99980 FORMAT(I3)
99981 READ(5,99982) (M(I),I=1,L)
99982 FORMAT(60I1)
      WRITE(6,99983)
99983 FORMAT('1','THE MESSAGE TO BE DECODED IS:'/)
      WRITE(6,99984) (M(I),I=1,L)
99984 FORMAT(' ',60I1)
      WRITE(6,99985)
99985 FORMAT('0',23X,'I',5X,'J',5X,'Y(J)')
      N=L+1
      I=0
      J=0
      GO TO 99003
99001 WRITE(6,99002)I,J,Y(J)
99002 FORMAT(' ',21X,I3,5X,I3,5X,A1)
      IF(I.EQ.L)GO TO 99991
      IF(I.GT.L)GO TO 99989
99003 I=I+1
      J=J+1
1     IF(M(I).EQ.1)GO TO 12
11    I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 22
21    I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 32
```

Declare Statement

#8

```
      GO TO 99001
2     I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 24
23    I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 36
35    I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 410
409   Y(J)='I'
      GO TO 99001
36    Y(J)='E'
      GO TO 99001
410   Y(J)='N'
      GO TO 99001
24    I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 38
37    I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 414
413   Y(J)='S'
      GO TO 99001
414   Y(J)='R'
      GO TO 99001
38    I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 416
415   Y(J)='H'
      GO TO 99001
416   I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 532
531   Y(J)='U'
      GO TO 99001
532   Y(J)='C'
      GO TO 99001
22    I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 34
33    I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 406
405   Y(J)='A'
      GO TO 99001
406   I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 512
511   Y(J)='L'
      GO TO 99001
512   Y(J)='D'
      GO TO 99001
34    I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 408
407   Y(J)='O'
      GO TO 99001
408   I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 516
515   I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 630
629   I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 758
757   Y(J)='V'
      GO TO 99001
630   Y(J)='G'
      GO TO 99001
758   I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 8116
8115  Y(J)='K'
      GO TO 99001
8116  I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 9232
9231  I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 10462
10461 Y(J)='X'
      GO TO 99001
10462 Y(J)='Q'
      GO TO 99001
9232  I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 10464
10463 Y(J)='J'
      GO TO 99001
10464 Y(J)='Z'
      GO TO 99001
516   I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 632
631   Y(J)='P'
      GO TO 99001
632   Y(J)='B'
      GO TO 99001
32    I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 404
403   Y(J)='T'
      GO TO 99001
404   I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 508
507   I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 614
613   Y(J)='F'
      GO TO 99001
614   Y(J)='M'
      GO TO 99001
508   I=I+1
      IF(I.EQ.N)GO TO 99989
      IF(M(I).EQ.1)GO TO 616
615   Y(J)='W'
      GO TO 99001
616   Y(J)='Y'
      GO TO 99001
99989 WRITE(6,99990)
99990 FORMAT('0','THIS IS NOT A VALID ENCODING OF A MESSAGE')
      GO TO 99999
99991 WRITE(6,99992)
99992 FORMAT('0'///)
      WRITE(6,99993)
99993 FORMAT('0','THE DECODED MESSAGE IS:'/)
      K=J
      WRITE(6,99994) (Y(J),J=1,K)
99994 FORMAT(' ',60A1)
99999 STOP
      END
```

```
306
010001111111111000010111010011000000111011110100000001110011
100111001001010000001101100101100000111100111001101110100011
000010111100000111000001110011100110000000111101110011110000
000111001111001111000011100110000111001110000001111000101100
000110101110011010111100000110100100110100001010011100100111
001110
```

```
0001                DIMENSION M(320)
0002                INTEGER Y(100),ALPHA(27)
0003                DATA ALPHA/' ','A','B','C','D','E','F','G','H','I','J',
                   A'K','L','M','N','O','P','Q','R','S','T','U','V',
                   B'W','X','Y','Z'/
0004                READ (5,99980)L
0005         99980 FORMAT(I3)
0006         99981 READ(5,99982) (M(I),I=1,L)
0007         99982 FORMAT(60I1)
0008                WRITE(6,99983)
0009         99983 FORMAT('1','THE MESSAGE TO BE DECODED IS:'/)
0010                WRITE(6,99984)(M(I),I=1,L)
0011         99984 FORMAT(' ',60I1)
0012                WRITE(6,99985)
0013         99985 FORMAT('0',23X,'I',5X,'J',5X,'Y(J)')
0014                N=L+1
0015                I=0
0016                J=0
0017                GO TO 99003
0018         99001 WRITE(6,99902)I,J,Y(J)
0019         99902 FORMAT(' ',21X,I3,5X,I3,5X,A1)
0020                IF(I .EQ.L)GO TO 99991
0021                IF(I .GT.L)GO TO 99989
0022         99003 I=I+1
0023                J=J+1
0024         1      IF(M(I).EQ.1)GO TO 12
0025         11     I=I+1
0026                IF(I .EQ.N)GO TO 99989
0027                IF(M(I).EQ.1)GO TO 22
0028         21     I=I+1
0029                IF(I .EQ.N)GO TO 99989
0030                IF(M(I).EQ.1)GO TO 32
0031         31     Y(J)=ALPHA(1)
0032                GO TO 99001
0033         12     I=I+1
0034                IF(I .EQ.N)GO TO 99989
0035                IF(M(I).EQ.1)GO TO 24
0036         23     I=I+1
0037                IF(I .EQ.N)GO TO 99989
0038                IF(M(I).EQ.1)GO TO 36
0039         35     I=I+1
0040                IF(I .EQ.N)GO TO 99989
0041                IF(M(I).EQ.1)GO TO 410
0042         409    Y(J)=ALPHA(10)
0043                GO TO 99001
0044         36     Y(J)=ALPHA(6)
0045                GO TO 99001
0046         410    Y(J)=ALPHA(15)
0047                GO TO 99001
0048         24     I=I+1
0049                IF(I .EQ.N)GO TO 99989
0050                IF(M(I).EQ.1)GO TO 38
0051                IF(I .EQ.N)GO TO 99989
0052         37     I=I+1
0053                IF(I .EQ.N)GO TO 99989
0054                IF(M(I).EQ.1)GO TO 414
0055         413    Y(J)=ALPHA(20)
0056                GO TO 99001
```

← NOTE CHANGES

COMPILER ON VS 370
CANT HANDLE Y(J) = 'A' FORM

Wesley D. Redus

Robert J. Markell

*HW

```
THE MESSAGE TO BE DECODED IS:

010001111111111000010111010011000000111011110100000001110011
100111001001010000001101100101100000111100111001101110100011
000010111100000111000001110011100110000000111101110011110000
000111001111001111000011100110000111001110000001111000101100
000110101110011010111100000110100100110100001010011100100111
001110

                         I       J      Y(J)
                         4       1       A
                        10       2       B
                        15       3       C
                        18       4
                        23       5       D
                        26       6       E
                        32       7       F
                        35       8
                        41       9       G
                        45      10       H
                        49      11       I
                        52      12
                        62      13       J
                        70      14       K
                        75      15       L
                        78      16
                        84      17       M
                        88      18       N
                        92      19       O
                        95      20
                       101      21       P
                       111      22       Q
                       115      23       R
                       118      24
                       122      25       S
                       126      26       T
                       131      27       U
                       134      28
                       141      29       V
                       147      30       W
                       157      31       X
                       160      32
                       166      33       Y
                       176      34       Z
                       179      35
                       182      36
                       192      37       Z
                       198      38       Y
                       201      39
                       211      40       X
                       217      41       W
                       224      42       V
                       227      43
                       232      44       U
                       236      45       T
                       240      46       S
                       243      47
                       247      48       R
                       257      49       Q
                       263      50       P
                       266      51
                       270      52       O
                       274      53       N
```

```
                                280         54        M
                                283         55
                                288         56        L
                                296         57        K
                                306         58        J




THE DECODED MESSAGE IS:

ABC DEF GHI JKL MNC PGR STU VWX YZ  ZY XWV UTS RQP ONM LKJ
```

BOB MARKS

Final Grade A

HANDOUT #1
-18-

7/15/76*
(2)

*HW1: EXPERIMENT A HAS M MUTUALLY EXCLUSIVE POSSIBLE OUTCOMES, A_m. EXPERIMENT B HAS N POSSIBLE MUTUALLY EXCLUSIVE OUTCOMES B_n. SHOW THAT

$$P[B_n/A_m] = \frac{P[A_m/B_n]P[B_n]}{\sum_{i=1}^{N} P[A_m/B_i]P[B_i]} \quad ①$$

WE SHALL MAKE USE OF

$$P[A_i/B_j] = \frac{P[A_i, B_j]}{P[B_j]} \Leftarrow \text{MULTIPLICATIVE LAW OF PROBABILITY MEASURE} \quad ②$$

$$P[A_i] = \sum_{\text{all } j} P[A_i, B_j] \Leftarrow \text{MARGINAL DENSITY DEFN.} \quad ③$$

OUR EVENTS ARE:

$A : A_1, A_2, \ldots, A_m \ldots A_M$
$B : B_1, B_2, \ldots B_n, \ldots, B_N$

START BY REWRITING ② AS

$$P[B_n/A_m] = P[A_m, B_n]/P[A_m] \quad ④$$

② MAY ALSO BE WRITTEN AS

$$P[A_m, B_n] = P[A_m/B_n]P[B_n] \quad (= P[B_n, A_m]) \quad ⑤$$

FROM ③ WE WRITE

$$P[A_m] = \sum_{i=1}^{N} P[A_m, B_i] .$$

WHICH, FROM ② MAY BE WRITTEN

$$P[A_m] = \sum_{i=1}^{N} P[A_m/B_i]P[B_i]$$

SUBSTITUTING THIS AND ⑤ INTO ④ GIVES THE DESIRED RESULT:

$$P[B_n/A_m] = \frac{P[A_m/B_n]P[B_n]}{\sum_{i=1}^{N} P[A_m/B_i]P[B_i]}$$

7/21/76

FROM TEXT

*(2-4) GENERALIZE PART a OF (2-3) TO n SOURCES. WE NOW HAVE n SOURCES THE K'TH OF WHICH IS CAPABLE OF GENERATING q_K SYMBOLS:

$$S_1 = \{s_{11}, s_{12}, s_{13}, \ldots, s_{1q_1}\} \Leftarrow q_1 \text{ SYMBOLS}$$
$$S_2 = \{s_{21}, s_{22}, s_{23}, \ldots, s_{2q_2}\} \Leftarrow q_2 \text{ SYMBOLS}$$
$$\vdots$$
$$S_K = \{s_{K1}, s_{K2}, s_{K3}, \ldots, s_{Kq_K}\} \Leftarrow q_K \text{ SYMBOLS}$$
$$\vdots$$
$$S_n = \{s_{n1}, s_{n2}, s_{n3}, \ldots, s_{nq_n}\} \Leftarrow q_n \text{ SYMBOLS}$$

ASSOCIATE A PROBABILITY $p_{Kj} = P[s_{Kj}]$ WITH THE j^{TH} ELEMENT OF THE K^{TH} SOURCE SUCH THAT

$$\sum_{j=1}^{q_K} p_{Kj} = 1 \quad \forall \quad K \qquad ①$$

THE ENTROPY OF THE K^{TH} SOURCE IS

$$H_K = -\sum_{j_K=1}^{q_K} p_{Kj_K} \log p_{Kj_K} \qquad ②$$

NOW CONSIDER THE PROBABILITY SET $\{\lambda_K\} \ni$

$$\sum_{K=1}^{n} \lambda_K = 1$$

WE NOW FORM A MIXED SOURCE, $S(\lambda)$, FROM THE SOURCES S_K ; $K = 1, 2, \ldots, n$. THE PROBABILITIES ASSOCIATED WITH THIS NEW SOURCE ARE

$$\{\lambda_1 p_{11}, \lambda_1 p_{12}, \lambda_1 p_{13}, \ldots, \lambda_1 p_{1q_1}; \quad \Leftarrow \text{FROM } S_1$$
$$\lambda_2 p_{21}, \lambda_2 p_{22}, \ldots, \lambda_2 p_{2q_2}; \ldots \quad \Leftarrow \text{FROM } S_2$$
$$\vdots$$
$$\lambda_K p_{K1}, \lambda_K p_{K2}, \ldots, \lambda_K p_{Kq_K}, \ldots, \quad \Leftarrow \text{FROM } S_K$$
$$\vdots$$
$$\lambda_n p_{n1}, \lambda_n p_{n2}, \ldots, \lambda_n p_{nq_n}\} \quad \Leftarrow \text{FROM } S_n$$

(CONT →)

THE ENTROPY, $H[S(\lambda)]$, ASSOCIATED WITH THE MIXED SOURCE FOLLOWS AS

$$H[S(\lambda)] = -\sum_{j_1=1}^{q_1} \lambda_1 P_{1j_1} \log \lambda_1 P_{1j_1} - \sum_{j_2=1}^{q_2} \lambda_2 P_{2j_2} \log \lambda_2 P_{2j_2} -$$

$$\ldots - \sum_{j_K=1}^{q_K} \lambda_K P_{Kj_K} \log \lambda_K P_{Kj_K} -$$

$$\ldots - \sum_{j_n=1}^{q_n} \lambda_n P_{nj_n} \log \lambda_n P_{nj_n}$$

$$= -\sum_{K=1}^{n} \sum_{j_K=1}^{q_K} \lambda_K P_{Kj_K} \log \lambda_K P_{Kj_K}$$

$$= -\sum_{K=1}^{n} \lambda_K \sum_{j_K=1}^{q_K} P_{Kj_K} [\log \lambda_K + \log P_{Kj_K}]$$

$$= -\sum_{K=1}^{n} \lambda_K \left\{ \sum_{j_K=1}^{q_K} P_{Kj_K} \right\} \log \lambda_K$$

$$- \sum_{K=1}^{n} \lambda_K \sum_{j_K=1}^{q_K} P_{Kj_K} \log P_{Kj_K}$$

FROM ① AND ②, THIS BECOMES:

$$H[S(\lambda)] = -\sum_{K=1}^{n} \lambda_K \log \lambda_K + \sum_{K=1}^{n} \lambda_K H_K \qquad ③$$

THE FIRST TERM IS RECOGNIZED AS THE ENTROPY OF $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_n)$. THAT IS

$$H(\lambda) = -\sum_{K=1}^{n} \lambda_K \log \lambda_K$$

THUS ③ BECOMES

$$H[S(\lambda)] = H(\lambda) + \sum_{K=1}^{n} \lambda_K H_K \qquad ④$$

THIS IS THE ENTROPY OF THE MIXED SOURCE. THE MIXED SOURCE IS FORMED BY COMBINING THE INDIVIDUAL SOURCES, THE K^{TH} OF WHICH GENERATES A SYMBOL $100\lambda_K$ % OF THE TIME. THAT IS, THE PROBABILITY THAT A SYMBOL GENERATED FROM THE MIXED SOURCE ORIGINATED FROM THE K^{TH} COMPONENT SOURCE IS λ_K.

7/21/76

(ASH)

*(1-6) THE INITIAL ENTROPY IS

$$H = H(p_1, p_2, \dots p_M) = -\sum_{K=1}^{M} p_K \log p_K$$

$$= -p_1 \log p_1 - p_2 \log p_2 - \sum_{K=3}^{M} p_K \log p_K \qquad ①$$

WE PERTURB TWO PROBABILITIES. WITHOUT LOSS OF GENERALITY, LET THESE BE p_1 AND p_2 WHERE $p_1 > p_2$. LET THE PERTURBED PROBABILITIES BY

$$p_1' = p_1 - \Delta P$$

$$p_2' = p_2 + \Delta P \qquad ②$$

SUCH THAT

$$p_1 - \Delta p \geq p_2 + \Delta P \qquad ③$$

WHERE $\Delta P > 0$ MUST SATISFY $\Delta P < P_1$. THE ENTROPY OF THE PERTURBED SOURCE IS

$$H' = H(p_1', p_2', p_3, p_4, \dots, p_M)$$

$$= -P_1' \log P_1' - P_2' \log P_2' - \sum_{K=3}^{M} p_K \log p_K \qquad ④$$

WE WISH TO SHOW THAT, BY MAKING THE DIFFERENCE $P_1 - P_2$ SMALLER, WE INCREASE THE OVERALL ENTROPY.

CONSIDER, THEN, THE ENTROPY DIFFERENCE BETWEEN ① AND ④:

$$H - H' = P_1' \log P_1' + P_2' \log P_2'$$

$$- P_1 \log P_1 - P_2 \log P_2$$

OR, FROM ②:

$$H - H' = (P_1 - \Delta P) \log (P_1 - \Delta P) + (P_2 + \Delta P) \log (P_2 + \Delta P)$$

$$- P_1 \log P_1 - P_2 \log P_2$$

$$= P_1 \log (P_1 - \Delta P) - \Delta P \log (P_1 - \Delta P)$$

$$+ P_2 \log (P_2 + \Delta P) + \Delta P \log (P_1 + \Delta P)$$

$$- P_1 \log P_1 - P_2 \log P_2$$

CONT →

COMBINING THE LOGS:

$$H - H' = P_1 \log \frac{P_1 - \Delta P}{P_1} + P_2 \log \frac{P_2 + \Delta P}{P_2} + \Delta P \log \frac{P_2 + \Delta P}{P_1 - \Delta P} \qquad \textcircled{5}$$

NOW, FROM THE LEMMA (LEMMA 1):

$$\log X \leq 1 - X$$

THUS, WE MAY WRITE ⑤ AS

$$H - H' \leq P_1 \left[1 - \left(1 - \frac{\Delta P}{P_1} \right) \right] + P_2 \left[1 - \left(1 + \frac{\Delta P}{P_2} \right) \right] + \Delta P \log \frac{P_2 + \Delta P}{P_1 - \Delta P}$$

$$\leq (+\Delta P) + (-\Delta P) + \Delta P \log \frac{P_2 + \Delta P}{P_1 - \Delta P} \qquad \textcircled{6}$$

$$\leq \Delta P \log \frac{P_2 + \Delta P}{P_1 - \Delta P}$$

NOW, FROM ③

$$\frac{P_2 + \Delta P}{P_1 - \Delta P} \leq 1$$

SINCE $\log(\cdot)$ IS A MONOTONICALLY INCREASING FUNCTION, IT FOLLOWS THAT

$$\log \frac{P_2 + \Delta P}{P_1 - \Delta P} \leq 0$$

FURTHERMORE, SINCE $\Delta P > 0$, WE MAY REWRITE ⑥ AS

$$H - H' \leq 0$$

OR

$$H' \geq H \qquad \textcircled{7}$$

WHICH WAS TO BE PROVED. FROM ⑤, EQUALITY IS ACHEIVED FOR $\Delta P = 0$. IT FOLLOWS THAT, SINCE BOTH H' AND H ARE CONVEX, THAT H' IS STRICTLY GREATER THAN H FOR Δp STRICTLY GREATER THAN ZERO. THAT IS

$$H' > H \quad \text{FOR} \quad \Delta p > 0$$

7/21/76 *

-39
(ASH)
HO#2

(1.12) GIVEN FUNCTION $h(p)$; $0<p\le 1$ ①

$\ni$ $h(p_1p_2) = h(p_1) + h(p_2)$

$h(p)$ IS MONOTONICALLY DECREASING

SHOW THE ONLY FUNCTION SATISFYING

THIS CONDITION IS $h(p) = -C\lg_b p$ ②

③

$\ni$ $C>0$ ④

& $b>1$

PROOF: ⑤

DEFINE $H(p) = b^{h(p)}$ ⑥

WHERE, $b>0$

ALSO, DEFINE ⑦

$$G(p) = \frac{d}{dp}H(p)$$

PLUG ⑤ INTO ①

$$\begin{aligned} H(p_1p_2) &= b^{h(p_1p_2)} \\ &= b^{h(p_1)}\, b^{h(p_2)} \\ &= H(p_1)H(p_2) \end{aligned}$$

FROM ⑦:

$$\begin{aligned} \frac{d}{dp_1}H(p_1p_2) &= G(p_1p_2)\frac{d}{dp_1}(p_1p_2) \\ &= p_2\, G(p_1p_2) \\ &= H(p_2)\frac{d}{dp_1}H(p_1) \end{aligned}$$

OR ⑧

$$G(p_1p_2) = \frac{1}{p_2}H(p_2)\frac{d}{dp_1}H(p_1)$$

SIMILARLY, WE CAN SHOW ⑨

$$G(p_1p_2) = \frac{1}{p_1}H(p_1)\frac{d}{dp_2}H(p_2)$$

COMBINING ⑧ & ⑨:

$$\frac{1}{p_1} H(p_1) \frac{d}{dp_2} H(p_2) = \frac{1}{p_2} H(p_2) \frac{d}{dp_1} H(p_1)$$

OR, EQUIVALENTLY

$$\frac{p_1}{H(p_1)} \frac{d}{dp_1} H(p_1) = \frac{p_2}{H(p_2)} \frac{d}{dp_2} H(p_2)$$ ⑩

NOW, THE NUMBER TO WHICH THESE RELATIONS ARE EQUAL MUST BE INDEPENDENT OF BOTH p_1 AND p_2. AS SUCH, LET

$$\frac{p_1}{H(p_1)} \frac{d}{dp_1} H(p_1) = C' = \text{CONSTANT}$$

OR

$$\frac{d}{dp} H(p) = \frac{c'}{p} H(p)$$ ⑪

THE SOLN' TO THIS DIFFERENTIAL EQUATION IS UNIQUE:

$$H(p) = d\,p^{c'}$$

WHERE d = REAL CONSTANT. FROM ⑤

$$d\,p^{c'} = b^{h(p)}$$

$$\Rightarrow h(p) = \lg_b d\,p^{c'}$$
$$= \lg_b d + c' \lg_b p$$ ⑫

THIS RELATIONSHIP MUST SATISFY ①

$$\Rightarrow d = 1$$

AND $$h(p) = c' \lg_b p$$ ⑬

THIS RELATION IS THE <u>ONLY</u> CONTINUOUS DIFFERENTIABLE FUNCTION SATISFYING ①. THIS FOLLOWS FROM THE UNIQUE SOLN' FOR THE DIFF. EQ IN ⑪

IT REMAINS TO SATISFY THE MONOTONE DECREASING NATURE OF $h(p)$ WHICH MAY BE STATED AS

$$\frac{dh(p)}{dp} < 0$$

FROM ⑬

$$\frac{d}{dp} h(p) = c' \frac{d}{dp} \lg_b p = \frac{c'}{\ln b} \frac{d}{dp} \ln p$$
$$= c'/\ln b \times \frac{1}{p} < 0$$

SINCE $\frac{1}{p} > 0$, WE HAVE TWO ALTERNATIVES

① REQUIRE $b > 1$ AND $C' = -C < 0$, IN WHICH CASE OUR THEOREM STATEMENT IS TRUE

② REQUIRE $0 < b < 1$ AND $C' > 0$. CAN SEE NO REASON WHY THIS ALSO CAN'T SATISFY BOTH MONOTONICITY AND ①.

$$\lg_{\frac{1}{b}} x = \lg_b x / \lg_b \frac{1}{b} = -\lg_b x$$

NOTE: IN APPENDIX 2 OF SHANNON'S PAPER, ANOTHER PROOF OF THIS IS GIVEN, BUT, WITH DIFFERENT INITIAL ASSUMPTIONS.

7-28-76

WE WISH TO FIND THE CHANNEL CAPACITY FOR THE CHANNEL DESCRIBED BY

$$P = \begin{bmatrix} 3/4 & 1/8 & 1/8 & 0 \\ 1/8 & 3/4 & 0 & 1/8 \\ 1/8 & 1/8 & 3/4 & 0 \\ 0 & 0 & 1/4 & 3/4 \end{bmatrix}$$

BEGIN BY FINDING

$$P^{-1} = C^T / |P|$$

WHERE C IS THE MATRIX OF COFACTORS.

LET'S FIND THESE FIRST:

$$C_{11} = \begin{vmatrix} 3/4 & 0 & 1/8 \\ 1/8 & 3/4 & 0 \\ 0 & 1/4 & 3/4 \end{vmatrix}$$

$$= \left[\left(\frac{3}{4}\right)^3 + \frac{1}{4}\left(\frac{1}{8}\right)^2\right] - [0]$$

$$= \frac{27}{64} + \frac{1}{256} = \frac{108+1}{256} = \frac{109}{256}$$

$$C_{12} = -\begin{vmatrix} \frac{1}{8} & 0 & 1/8 \\ \frac{1}{8} & 3/4 & 0 \\ 0 & 1/4 & 3/4 \end{vmatrix}$$

$$= -\left[\left(\frac{3}{4}\right)^2 \frac{1}{8} + \frac{1}{4}\left(\frac{1}{8}\right)^2\right] - 0$$

$$= -\left[\frac{9}{128} + \frac{1}{256}\right] = -\frac{19}{256}$$

$$C_{13} = \begin{vmatrix} \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & 0 \\ 0 & 0 & 3/4 \end{vmatrix}$$

$$= \frac{3}{4}\left(\frac{1}{8}\right)^2 - \frac{1}{8}\left(\frac{3}{4}\right)^2$$

$$= \frac{3}{256} - \frac{9}{128}$$

$$= \frac{-15}{256}$$

$$C_{14} = -\begin{vmatrix} \frac{1}{8} & \frac{3}{4} & 0 \\ \frac{1}{8} & \frac{1}{8} & 3/4 \\ 0 & 0 & 1/4 \end{vmatrix}$$

$$= -\left\{ \left[\left(\frac{1}{8}\right)^2 \frac{1}{4} \right] - \frac{1}{8} \cdot \frac{3}{4} \cdot \frac{1}{4} \right\}$$

$$= -\left[\frac{1}{256} - \frac{6}{256} \right]$$

$$= \frac{5}{256}$$

$$C_{21} = -\begin{vmatrix} \frac{1}{8} & \frac{1}{8} & 0 \\ \frac{1}{8} & 3/4 & 0 \\ 0 & 1/4 & 3/4 \end{vmatrix}$$

$$= -\left[\left(\frac{3}{4}\right)^2 \frac{1}{8} - \left(\frac{1}{8}\right)^2 \frac{3}{4} \right]$$

$$= -\left[\frac{9}{128} - \frac{3}{256} \right]$$

$$= \frac{-15}{256}$$

$$C_{22} = \begin{vmatrix} 3/4 & 1/8 & 0 \\ 1/8 & 3/4 & 0 \\ 0 & 1/4 & 3/4 \end{vmatrix}$$

$$= \left[\left(\frac{3}{4}\right)^3 - \frac{3}{4} \left(\frac{1}{8}\right)^2 \right]$$

$$= \frac{27}{64} - \frac{3}{256}$$

$$= \frac{105}{256}$$

$$C_{23} = -\begin{vmatrix} 3/4 & 1/8 & 0 \\ 1/8 & 1/8 & 0 \\ 0 & 0 & 3/4 \end{vmatrix}$$

$$= -\left[\frac{1}{8} \left(\frac{3}{4}\right)^2 - \frac{3}{4} \left(\frac{1}{8}\right)^2 \right]$$

$$= -\left[\frac{9}{128} - \frac{3}{256} \right] = -\frac{15}{256}$$

$$C_{24} = \begin{vmatrix} 3/4 & \frac{1}{8} & \frac{1}{8} \\ 1/8 & 1/8 & 3/4 \\ 0 & 0 & 1/4 \end{vmatrix}$$

$$= \left[\frac{1}{4} \frac{3}{4} \frac{1}{8} - \left(\frac{1}{8}\right)^2 \frac{1}{4} \right]$$

$$= \frac{6}{256} - \frac{1}{256}$$

$$= \frac{5}{256}$$

$$C_{31} = \begin{vmatrix} \frac{1}{8} & \frac{1}{8} & 0 \\ \frac{3}{4} & 0 & \frac{1}{8} \\ 0 & \frac{1}{4} & 3/4 \end{vmatrix}$$

$$= -\left(\frac{1}{8}\right)^2 \frac{1}{4} - \frac{1}{8}\left(\frac{3}{4}\right)^2$$

$$= \frac{-1}{256} - \frac{9}{128}$$

$$= \frac{-19}{256}$$

$$C_{32} = -\begin{vmatrix} \frac{3}{4} & \frac{1}{8} & 0 \\ \frac{1}{8} & 0 & \frac{1}{8} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{vmatrix}$$

$$= -\left[\frac{-1}{4}\,\frac{3}{4}\,\frac{1}{8} - \left(\frac{1}{8}\right)^2\left(\frac{3}{4}\right)\right]$$

$$= \frac{3}{128} + \frac{3}{256}$$

$$= \frac{9}{256}$$

$$C_{33} = \begin{vmatrix} \frac{3}{4} & \frac{1}{8} & 0 \\ \frac{1}{8} & 3/4 & 1/8 \\ 0 & 0 & 3/4 \end{vmatrix}$$

$$= \left[\left(\frac{3}{4}\right)^3 - \left(\frac{1}{8}\right)^2\left(\frac{3}{4}\right)\right]$$

$$= \frac{27}{64} - \frac{3}{256}$$

$$= \frac{105}{256}$$

$$C_{34} = -\begin{vmatrix} \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{vmatrix}$$

$$= -\left[\left(\frac{3}{4}\right)^2 \frac{1}{4} - \left(\frac{1}{8}\right)^2 \frac{1}{4}\right]$$

$$= -\left[\frac{9}{64} - \frac{1}{256}\right]$$

$$= -\left[\frac{35}{256}\right]$$

$$C_{41} = -\begin{vmatrix} \frac{1}{8} & \frac{1}{8} & 0 \\ \frac{3}{4} & 0 & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{4} & 0 \end{vmatrix}$$

$$= -\left[\left(\frac{1}{8}\right)^3 - \left(\frac{1}{8}\right)^2 \frac{3}{4} \right]$$

$$= -\left[\frac{1}{512} - \frac{6}{512} \right]$$

$$= 5/512$$

$$C_{42} = \begin{vmatrix} \frac{3}{4} & \frac{1}{8} & 0 \\ \frac{1}{8} & 0 & 1/8 \\ \frac{1}{8} & \frac{3}{4} & 0 \end{vmatrix}$$

$$= \left(\frac{1}{8}\right)^3 - \left(\frac{3}{4}\right)^2 \frac{1}{8}$$

$$= \frac{1}{512} - \frac{9}{128}$$

$$= \frac{1}{512} - \frac{36}{512}$$

$$= -35/512$$

$$C_{43} = -\begin{vmatrix} \frac{3}{4} & \frac{1}{8} & 0 \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & 0 \end{vmatrix}$$

$$= -\left[\left(\frac{1}{8}\right)^3 - \left(\frac{1}{8}\right)^2 \frac{3}{4} \right]$$

$$= -\left[\frac{1}{512} - \frac{6}{512} \right]$$

$$= \frac{5}{512}$$

$$C_{44} = \begin{vmatrix} \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{4} & 0 \\ \frac{1}{8} & \frac{1}{8} & \frac{3}{4} \end{vmatrix}$$

$$= \left[\left(\frac{3}{4}\right)^3 + \left(\frac{1}{8}\right)^3 - \left(\frac{1}{8}\right)^2 \left(\frac{3}{4}\right) - \left(\frac{1}{8}\right)^2 \left(\frac{3}{4}\right) \right]$$

$$= \frac{27}{64} + \frac{1}{516} - \frac{6}{256}$$

$$= \frac{216 + 1 - 12}{512}$$

$$= 205/512$$

THUS

$$C = \begin{bmatrix} \frac{109}{256} & \frac{-19}{256} & \frac{-15}{256} & \frac{5}{256} \\ \frac{-15}{256} & \frac{105}{256} & \frac{-15}{256} & \frac{5}{256} \\ \frac{-19}{256} & \frac{9}{256} & \frac{105}{256} & \frac{-35}{256} \\ \frac{5}{512} & \frac{-35}{512} & \frac{5}{512} & \frac{205}{512} \end{bmatrix}$$

THUS

$$C^T = \begin{bmatrix} \frac{109}{256} & \frac{-15}{256} & \frac{-19}{256} & \frac{5}{512} \\ \frac{-19}{256} & \frac{105}{256} & \frac{9}{256} & \frac{-35}{512} \\ \frac{-15}{256} & \frac{-15}{256} & \frac{105}{256} & \frac{5}{512} \\ \frac{5}{256} & \frac{5}{256} & \frac{-35}{256} & \frac{205}{512} \end{bmatrix}$$

$$= \frac{1}{512} \begin{bmatrix} 218 & -30 & -38 & 5 \\ -38 & 210 & 18 & -35 \\ -30 & -30 & 210 & 5 \\ 10 & 10 & -70 & 205 \end{bmatrix}$$

IT REMAINS TO FIND THE DETERMINANT OF p:

$$|p| = \frac{1}{8} C_{24} + \frac{3}{4} C_{44}$$

$$= \frac{1}{8}\left(\frac{5}{256}\right) + \frac{3}{4}\left(\frac{205}{512}\right)$$

$$= \frac{5 + 615}{2048} = \frac{620}{2048} = \frac{155}{512}$$

$$\therefore p^{-1} = \frac{1}{155} \begin{bmatrix} \overset{\checkmark}{218} & -30 & \overset{\checkmark}{-38} & 5 \\ -38 & 210 & 18 & -35 \\ -30 & -30 & 210 & 5 \\ 10 & 10 & -70 & 205 \end{bmatrix}$$

CHECK

$$pp^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

NOW TO FIND THE ENTROPIES OF EACH ROW OF p:

$H_1 = H\left(\frac{3}{4}, \frac{1}{8}, \frac{1}{8}\right) = 1.061278$ BITS (0)

$H_2 = H\left(\frac{1}{8}, \frac{3}{4}, \frac{1}{8}\right) = 1.061278$ BITS

$H_3 = H\left(\frac{1}{8}, \frac{1}{8}, \frac{3}{4}\right) = 1.061278$ BITS

$H_4 = H\left(\frac{1}{4}, \frac{3}{4}\right) = 0.8112781$ BITS (1)

NOW

$$[p] \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{bmatrix} = \begin{bmatrix} -H_1 \\ -H_2 \\ -H_3 \\ -H_4 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{bmatrix} = [p^{-1}] \begin{bmatrix} -H_1 \\ -H_2 \\ -H_3 \\ -H_4 \end{bmatrix}$$

$$= \frac{-1}{155} \begin{bmatrix} 218 & -30 & -38 & 5 \\ -38 & 210 & 18 & -35 \\ -30 & -30 & 210 & 5 \\ 10 & 10 & -70 & 205 \end{bmatrix} \begin{bmatrix} +H_1 \\ +H_2 \\ +H_3 \\ +H_4 \end{bmatrix}$$

$$= \begin{bmatrix} -1.0532 \\ -1.1178 \\ -1.0532 \\ -0.7306 \end{bmatrix} \begin{matrix} (7) \\ (6) \\ (4) \\ (5) \end{matrix}$$

THUS

$$C = lg_2\ 2^{Q_1} + 2^{Q_2} + 2^{Q_3} + 2^{Q_4}$$

$$= lg_2\ 2 \times 2^{Q_1} + 2^{Q_2} + 2^{Q_4}$$

$$= lg_2\ 2.02724$$

$$= 1.01952 \text{ BITS}$$

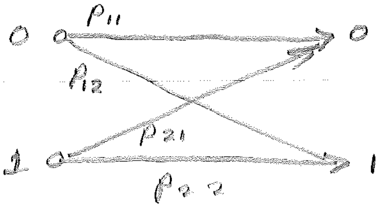
H.O # 3

7-28-76 *

WE WISH TO FIND THE MAXIMUM TRANSINFORMATION WITH THE FOLLOWING CONDITIONAL MATRICES:

$$(a)\ \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} \qquad (b)\ \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

THESE MATRICIES CORRESPOND TO

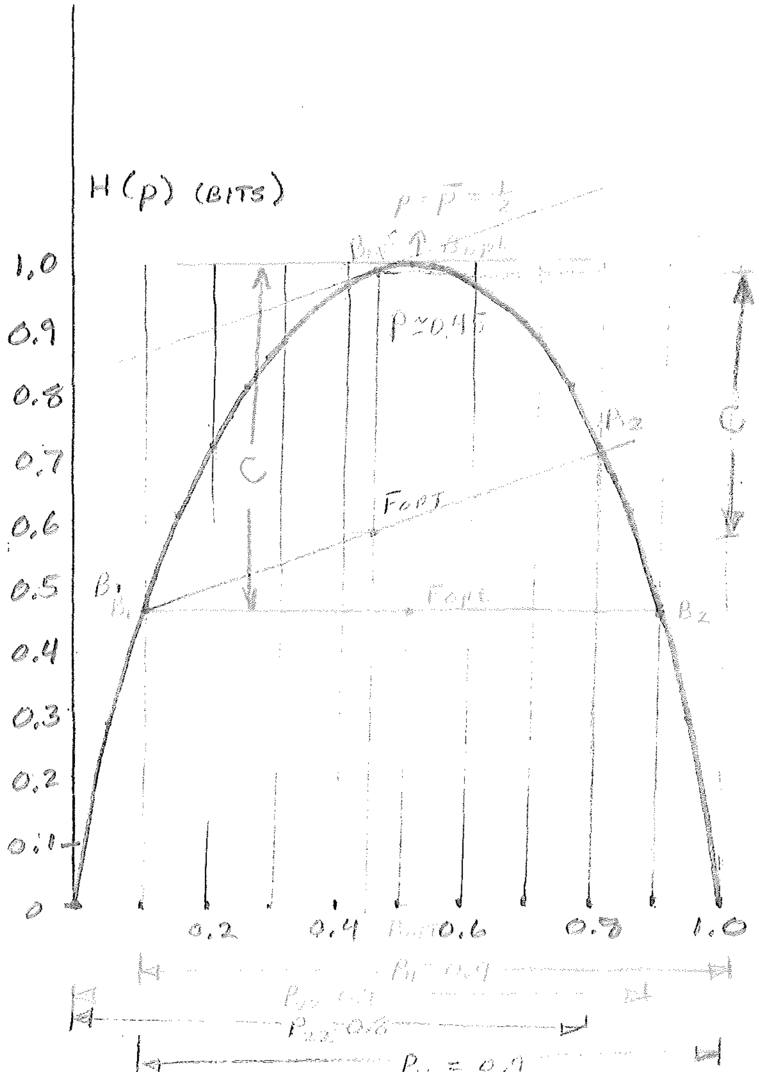


$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

WE SHALL FIRST CONSTRUCT A GRAPH OF THE ENTROPY FUNCTION

$$H(p) = p \lg_2 1/p + \bar{p} \lg_2 1/\bar{p}$$

WHERE $\bar{p} = 1 - p$



HP-25 PROGRAMS

```
(p)

STO 0
1
-
CHS
STO 1 (p̄)
1/X
ln
2
ln
÷
RCL 1
X

RCL 0
1/X
ln
2
ln
÷
RCL 0
X
+
GTO 00
```

a. TO FIND THE CHANNEL CAPACITY, IT REMAINS TO FOLLOW THE INSTRUCTIONS ON HAND-OUT #3. WE WILL WORK MATRIX (a) IN RED. AS CAN BE SEEN ON THE ENTROPY CURVE, THE CHANNEL CAPACITY FOR MATRIX (a) IS ABOUT 0.54 BITS. ALSO, WE SEE $P(Y=0) = P(Y=1) = \frac{1}{2}$ ARE THE CORRESPONDING RECEIVER PROBABILITIES.

b. WE WORK MATRIX (b) IN GREEN. FROM THE ENTROPY CURVE:

$$P[Y=0] \simeq 0.45 \qquad P[Y=1] \simeq 0.55$$

AND $C \simeq 0.40$ BITS

(CONT) ⟹

FINDING THE SOURCE PROBS:

LET $q = P(X=0)$. THEN

$$\begin{cases} P(Y=0) = P(Y=0/X=0)P(X=0) + P(Y=0/X=1)P(X=1) \\ P(Y=1) = P(Y=1/X=0)P(X=0) + P(Y=1/X=1)P(X=1) \end{cases}$$

$$p = p_{11}q + p_{21}\bar{q}$$
$$\bar{p} = p_{12}q + p_{22}\bar{q}$$

BUT $p_{11} + p_{12} = 1 \Rightarrow p_{12} = 1 - p_{11}$

AND $p_{21} + p_{22} = 1 \Rightarrow p_{21} = 1 - p_{22}$

$$\Rightarrow \begin{cases} p = p_{11}q + (1-p_{22})\bar{q} & ① \\ \bar{p} = (1-p_{11})q + p_{22}\bar{q} & ② \end{cases}$$

LETS SEE IF THESE EQUATIONS ARE CONSISTANT. ① BECOMES

$$\begin{aligned} \bar{p} &= 1 - p_{11}q - (1-p_{22})\bar{q} \\ &= 1 - p_{11}q - (1-p_{22})(1-q) \\ &= 1 - p_{11}q - (1 - p_{22} - q + qp_{22}) \\ &= \cancel{1} - p_{11}q - \cancel{1} + p_{22} + \underline{q} - qp_{22} \\ &= (1-p_{11})q + p_{22}(1-q) \\ &= (1-p_{11})q + p_{22}\bar{q} \end{aligned}$$

WHICH IS THE SAME AS ②. THUS, ① AND ② ARE CONSISTANT (ie EQUIVALENT)

LET'S SOLVE FOR q IN ①

$$\begin{aligned} p &= p_{11}q + (1-p_{22})(1-q) \\ &= p_{11}q + 1 - p_{22} - q + qp_{22} \\ &= q(p_{11} - 1 + p_{22}) + 1 - p_{22} \end{aligned}$$

$$\Rightarrow \quad q(p_{11} - 1 + p_{22}) = p - 1 + p_{22}$$

$$\therefore q = \frac{p - 1 + p_{22}}{p_{11} - 1 + p_{22}}$$

FOR MATRIX (a), WE HAVE

$p = \frac{1}{2}$ $\quad p_{11} = p_{22} = 0.9$

$\therefore \; q = (0.5 - 1 + 0.9)/(0.9 - 1 + 0.9)$

$= 0.500$

$\bar{q} = 0.500$

FOR MATRIX (b)

$p = 0.45 \quad p_{11} = 0.9 \quad p_{22} = 0.8$

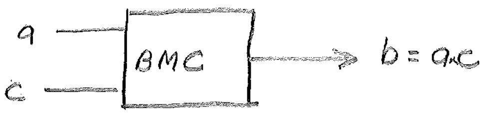
$\therefore \; q = (0.45 - 1 + 0.8)/(0.9 - 1 + 0.8)$

$= 0.357$

$\bar{q} = 0.643$

HANDOUT #7 7/28/76 *

BINARY MULTIPLICATIVE CHANNEL



WE ASSUME THE BINARY (MULTIPICATION) OPERATIONS

$$0 = 1 \times 0 = 0 \times 1 = 0 \times 0 \quad , \quad 1 = 1 \times 1$$

a. SINCE THE BMC IS (ASSUMED) NOISELESS, THE CONDITIONAL PROBABILITY MATRIX, $P[b/a,c]$, IS

A		b	
a	c	0	1
0	0	1	0
0	1	1	0
1	0	1	0
1	1	0	1

$P(B/A)$

WE HAVE HERE SPECIFIED THE BINARY OUTPUT ALPHABET $b = \{0, 1\}$

b. WE NOW WISH TO DEVELOP AN EXPRESSION FOR THE TRANSINFORMATION OF THE BMC. THAT IS, SPECIFY $I(A;B)$.*

ASSUME

$$P(a=0) = P(c=0) = p$$
$$p(a=1) = P(c=1) = q$$

EVENTS a & c ARE FURTHEMORE ASSUMED TO BE STATISTICALLY INDEPENDENT SO THAT:

$$p(a=0, c=1) = p(a=1, c=0) = pq$$
$$p(a=0, c=0) = p^2$$
$$p(a=1, c=1) = q^2$$

* FOLLOWING NOTATION IN PROB. 5-9 OF TEXT, A IS HERE ASSUMED TO BE COMPRISED OF THE (IND) EVENTS a & c.

IT FOLLOWS THAT

$$-H(A) = 2pq \lg pq + p^2 \lg p^2 + q^2 \lg q^2 \qquad ①$$

FROM THE INPUT PROBABILITIES AND THE CONDITIONAL PROB. MATRIX, WE WRITE $P(A;B) = P(B/A)\,P(A)$:

$$\begin{array}{cc|cc} & & b & \\ a & c & 0 & 1 \\ \hline 0 & 0 & p^2 & 0 \\ 0 & 1 & pq & 0 \\ 1 & 0 & pq & 0 \\ 1 & 1 & 0 & q^2 \end{array} \quad \Leftarrow P(A;B)$$

THUS

$$\begin{aligned} -H(A;B) &= p^2 \lg p^2 + q^2 \lg q^2 + 2pq \lg pq \\ &= -H(A) \end{aligned} \qquad ②$$

SINCE $p(B) = \Sigma_A\, p(A;B)$, WE HAVE

b	P(b)
0	$p^2 + 2pq$
1	q^2

AND

$$\begin{aligned} -H(B) = {} & (p^2 + 2pq) \lg (p^2 + 2pq) \\ & + q^2 \lg q^2 \end{aligned} \qquad ③$$

NOW

$$I(A;B) = H(A) + H(B) - H(A;B)$$

BUT, FROM ②, $H(A;B) = H(A)$. THUS

$$I(A;B) = H(B)$$
$$= -(p^2+2pq)\,lg\,(p^2+2pq)$$
$$- q^2 lg\, q^2 \qquad ④$$

WHERE $q = 1-p$. SIMPLIFYING:

$$I(A;B) = -[p^2+2p(1-p)]\,lg\,[p^2+2p(1-p)]$$
$$- (1-p)^2 lg (1-p)^2$$
$$= -[p^2+2p-2p^2]\,lg\,[p^2+2p-2p^2]$$
$$-2(1-p)^2 lg (1-p)$$
$$= (p^2-2p)\,lg(2p-p^2) - 2(1-p)^2 lg(1-p)$$
$$= -p(2-p)\,lg\,p(2-p) - 2(1-p)^2 lg(1-p) \qquad ⑤$$

THIS LOOKS LIKE A GOOD FINAL RESULT.

C. WE WISH TO FIND NOW THE CHANNEL CAPACITY $C = Max\, I(A;B)$. NOW, FROM ④:

$$I(A;B) = H(B)$$
$$= H[p^2+2pq, q^2]$$

WE HAVE SEEN THAT THE MAXIMUM VALUE OF THE ENTROPY OF TWO (DISJOINT) EVENTS IS 1 BIT. THIS OCCURS WHEN THE EVENTS ARE EQUALLY PROBABLE. THUS, WE REQUIRE

$$p^2 + 2pq = q^2 \qquad \Longrightarrow$$

SOLVING FOR p:

$$p^2 + 2p(1-p) - (1-p)^2 = 0$$

$$p^2 + 2p - 2p^2 - (1 - 2p + p^2) = 0$$

$$-p^2 + 2p - 1 + 2p - p^2 = 0$$

$$-2p^2 + 4p - 1 = 0$$

$$p = \frac{-4 \pm \sqrt{16 - 8}}{-4}$$

$$= 1 \pm \sqrt{8}/4$$

$$= 1 \pm \sqrt{2}/2$$

$$= 1 \pm 1/\sqrt{2}$$

SINCE $0 < p < 1$, WE HAVE

$$p = 1 - 1/\sqrt{2} \cong 0.293$$

$$\Rightarrow q = 1 - p \cong \frac{1}{\sqrt{2}} = 0.707$$

AND, AGAIN, THE CHANNEL CAPACITY IS

$$C = 1 \text{ BIT}$$

BY USING THESE INPUT PROBABILITIES, WE INSURE THAT, ON THE AVERAGE, AN EQUAL NUMBER OF 1'S & 0'S WILL BE "RECEIVED."

8-4-76 *

4-6 p.91 (TEXT)

OUR SOURCE AND ASSOCIATED PROB.'S ARE

$$S = \begin{cases} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 \\ 1/3 & 1/3 & 1/9 & 1/9 & 1/27 & 1/27 & 1/27 \end{cases}$$

a. LETS FIRST COMPUTE $H_3(S)$, SINCE ITS THE EASIEST:

$$H_3(S) = \tfrac{1}{3}\lg_3 3 + \tfrac{1}{3}\lg_3 3 + \tfrac{1}{9}\lg_3 9 + \tfrac{1}{9}\lg_3 9 + \tfrac{1}{27}\lg_3 27 + \tfrac{1}{27}\lg_3 27 + \tfrac{1}{27}\lg_3 27$$

$$= \tfrac{2}{3}\lg_3 3 + \tfrac{2}{9}\lg_3 3^2 + \tfrac{3}{27}\lg_3 3^3$$

$$= \tfrac{2}{3}(1) + \tfrac{2}{9}(2) + \tfrac{1}{9}(3)$$

$$= 13/9$$

ALSO, SINCE ALL PROBABILITIES ARE OF THE FORM 3^{-l_i}, WE CAN EXPECT A COMPACT CODE HERE.

NOW

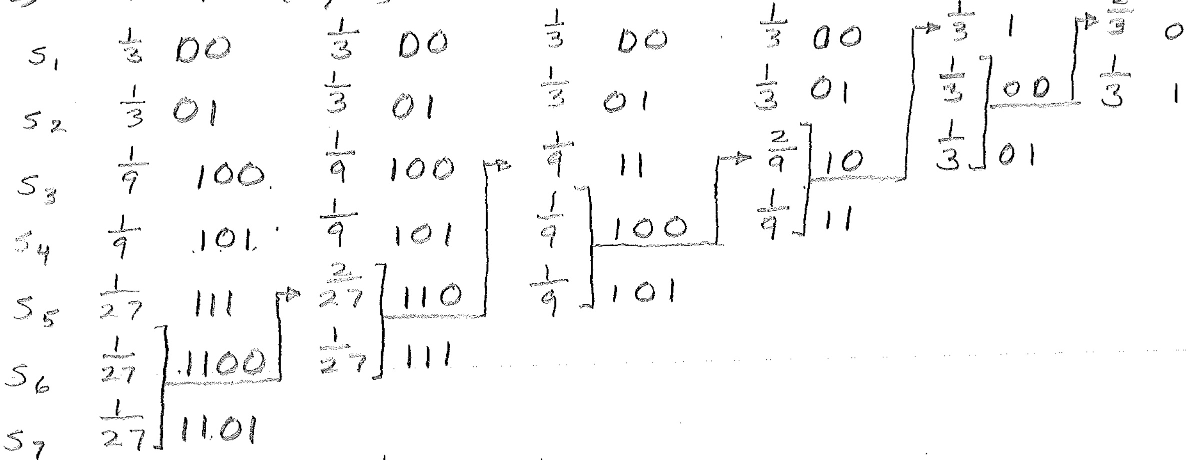
$$H_2(S) = \frac{H_3(S)}{\lg_3 2}$$

BUT

$$\lg_3 2 = \frac{\ln 2}{\ln 3}$$

$$\Rightarrow H_2(S) = \frac{13}{9}\;\frac{\ln 3}{\ln 2} = 2.289390279$$

(bc) FOR $X = \{0, 1\}$ (HUFFMAN CODING)



$$\bar{L} = 4\cdot\frac{1}{3} + 6\cdot\frac{1}{9} + 11\cdot\frac{1}{27}$$
$$= \frac{1}{27}\left[36 + 18 + 11\right]$$
$$= {}^{65}/_{27} = 2.40740741 > H_2(S) \text{ AS EXPECTED.}$$

FOR $X = \{0, 1, 2\}$, WE MAY WRITE $p_i = 2^{-l_i}$

$\Rightarrow l_1 = l_2 = 1 \quad l_2 = l_3 = 2 \quad l_4 = l_5 = l_6 = 3$

A CODE (INSTANT), OBEYING THE PREFIX PROPERTY, IS

s_1	$\frac{1}{3}$	0
s_2	$\frac{1}{3}$	1
s_3	$\frac{1}{9}$	20
s_4	$\frac{1}{9}$	21
s_5	$\frac{1}{27}$	220
s_6	$\frac{1}{27}$	221
s_7	$\frac{1}{27}$	222

DUE TO PROBABILITY STRUCTURE, WE ARE ASSURED THAT THIS CODE IS COMPACT. NOW

$$\bar{L} = 2\left(\frac{1}{3}\right) + 4\left(\frac{1}{9}\right) + 9\left(\frac{1}{27}\right)$$
$$= \frac{2}{3} + \frac{4}{9} + \frac{9}{27}$$
$$= \frac{1}{9}\left[6 + 4 + 3\right]$$
$$= {}^{13}/_{9} = H_3(S) \text{ AS EXPECTED}$$

8-5-76 *

WE WISH TO REPEAT HERE THE PROOF THAT HUFFMAN CODING YIELDS COMPACT CODES (pp 52-3 OF TEXT - SEC 4-7) FOR THE CASE OF A BINARY ALPHABET.

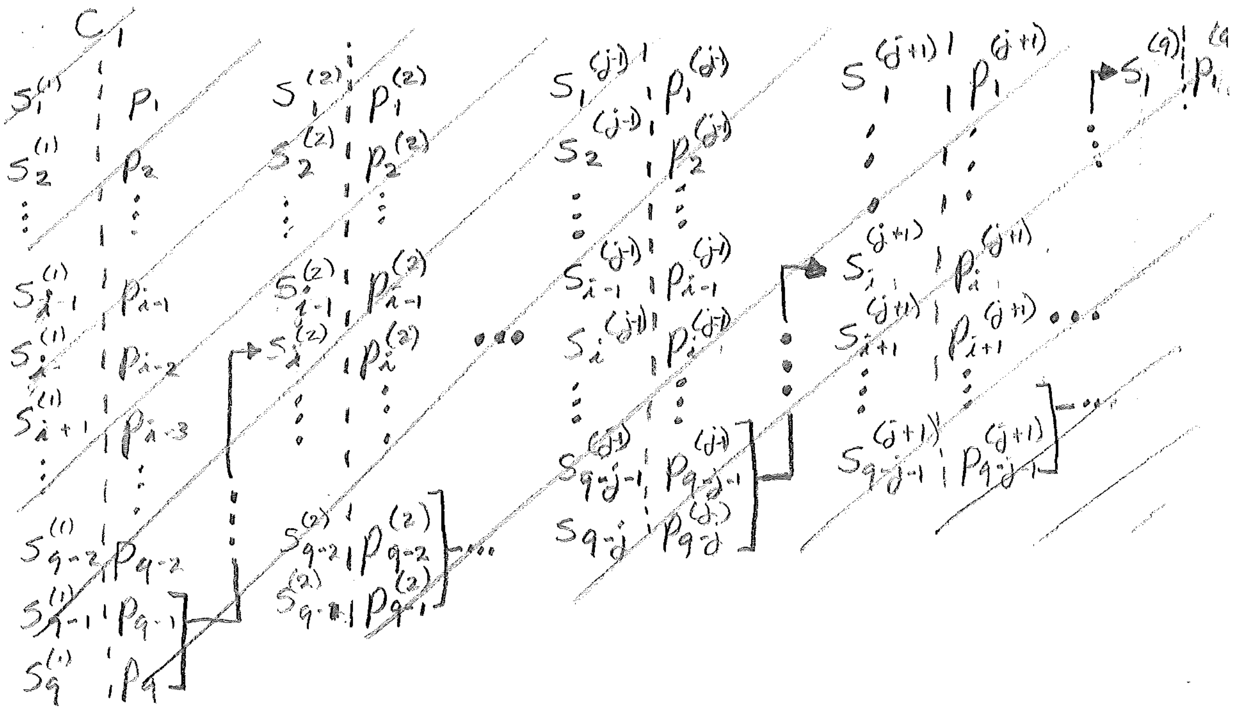
LETS BEGIN BY ESTABLISHING NOMENCLATURE. WE ASSUME q SOURCE SYMBOLS THE i^{TH} OF WHICH HAS PROBABILITY p_i:

$$S:\begin{Bmatrix} s_1 & s_2 & \dots & s_i & \dots & s_q \\ p_1 & p_2 & \dots & p_i & \dots & p_q \end{Bmatrix}$$

WITHOUT LOSS OF GENERALITY, LET

$$p_1 \geq p_2 \geq p_3 \geq \dots \geq p_i \geq \dots \geq p_q$$

WE WON'T GO THROUGH AN EXPLANATION OF HUFFMAN CODING PROCEDURES. THIS IS EXPLAINED IN SEC 4-6 OF THE TEXT. LET'S GENERALIZE THE TABLE RESULTING FROM HUFFMAN CODING FOR THE ABOVE SOURCE ⟹



CONSIDER, THEN, THE AVERAGE WORD LENGTH FOR THE (ASSUMED) COMPACT CODE) FOR C_j :

$$\overline{L_j} = \sum_{k=1}^{q-j} \ell_k^{(j)} P_k^{(j)} \qquad ①$$

WITHOUT LOSS OF GENERALITY, ASSUME THAT

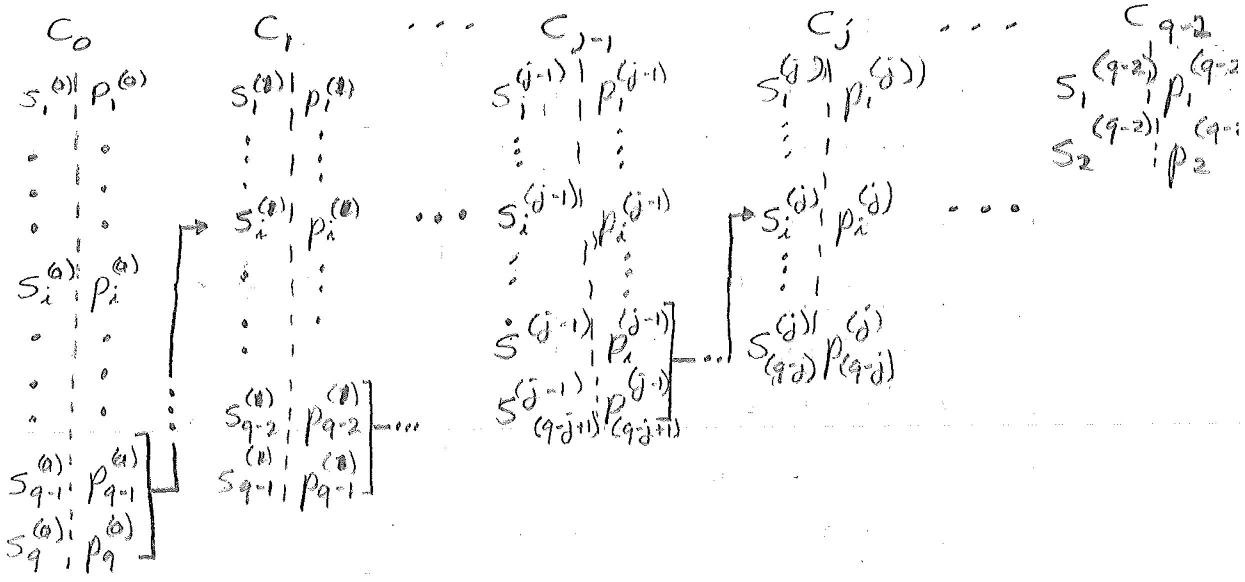
$$P_1^{(k)} \geq P_2^{(k)} \geq \ldots \geq P_i^{(k)} \geq \ldots \geq P_{q-k}^{(k)} \qquad \forall \; k$$

THUS, FOLLOWING THE HUFFMAN CODING PROCEDURE, WE WILL TAKE $P_{q-j}^{(j)}$ AND $P_{q-j-1}^{(j)}$ (DENOTED IN TEXT BY $P_{\alpha 0}$ AND $P_{\alpha 1}$) AND COMBINE THEM, ADDING A ZERO TO THE FIRST CODE WORD AND A ONE TO THE SECOND. THUS:

$$\overline{L_{j-1}} = \sum_{k=1}^{q-j+1} \ell_k^{(j-1)} P_k^{(j-1)}$$

$$= \overline{L_j} + P_{q-j}^{(j)} \times (1 \text{ BINIT}) + P_{q-j-1}^{(j)} \times (1 \text{ BINIT})$$

$$= \overline{L_j} + P_{q-j}^{(j)} + P_{q-j-1}^{(j)} \qquad ②$$



FOR q SYMBOLS, THE HUFFMAN CODE WILL GENERATE q-1 REDUCTIONS AS SHOWN. THE ZEROTH REDUCTION, C_0, CORRESPONDS HERE TO THE GIVEN SOURCE ALPHABET. THE C_j TH CODE CONSISTS OF AN ALPHABET OF q-j SYMBOLS, $S_i^{(j)}$, THE i TH OF WHICH HAS PROBABILITY OF $p_i^{(j)}$ AND SYMBOL LENGTH $l_i^{(j)}$. THE WORD LENGTH, $l_i^{(j)}$, ARISES FROM HUFFMAN CODING PROCEDURE. WE KNOW THAT, FOR C_{q-1} (ONE SYMBOL), WE HAVE A COMPACT CODE. WE MAY THUS PROVE THAT THE HUFFMAN CODE IS COMPACT FOR ALL C_j IF WE ASSUME THAT IT IS TRUE FOR SOME C_j AND SHOW THAT COMPACTNESS FOLLOWS FOR C_{j-1}. THIS CONSTITUTES PROOF BY INDUCTION.

TO PROVE THAT C_{j-1} IS COMPACT GIVEN THAT C_j IS, WE WILL ASSUME THE CONTRARY AND SHOW A CONTRADICTION. ASSUME $\exists$ A CODE $\tilde{C}_{j-1}$ WITH AVERAGE LENGTH

$$\tilde{L}_{j-1} < L_{j-1}$$

ASSUME THAT THE CODE CONSIST OF WORD LENGTHS $\tilde{l}_1, \tilde{l}_2, \ldots, \tilde{l}_{q-j+1} \ni$

$$\tilde{l}_1 \leq \tilde{l}_2 \leq \ldots \leq \tilde{l}_{q-j+1}$$

WE NOW ARGUE THAT ONE OF THE CODE WORDS, SAY CORRESPONDING TO $\tilde{l}_i$, MUST DIFFER FROM ANOTHER (ADJACENT) CODE WORD, SAY CORRESPONDING TO $\tilde{l}_{i+1}$ ONLY IN THE LAST DIGIT. OTHERWISE, WE COULD DROP THE LAST DIGITS OF EACH WORD, RETAIN THE PREFIX PROPERTY, AND REDUCE $\tilde{L}_{j-1}$. (ie, $\tilde{C}_{j-1}$ WOULD NOT BE COMPACT). IT FOLLOWS, THEN, THAT WE CAN CONSTRUCT A CODE $\tilde{C}_j$ BY COMBINING THE CODE WORDS CORRESPONDING TO $\tilde{l}_i$ AND $\tilde{l}_{i+1}$ (ie, KEEPING THEIR $\tilde{l}_i - 1 = \tilde{l}_{i+1} - 1$ FIRST BINITS WHICH ARE EQUIVALENT) AND KEEPING ALL OTHER CODE WORDS THE SAME. THEN, IF $\tilde{L}_j$ IS THE AVERAGE WORD LENGTH OF

CODE C_j, THEN

$$\tilde{L}_j = \tilde{L}_{j-1} - \tilde{P}_i - \tilde{P}_{i-1}$$

WHERE $\tilde{P}_i$ & $\tilde{P}_{i-1}$ CORRESPOND TO THE SYMBOLS ASSOCIATED WITH WORD LENGTHS $\tilde{l}_i$ AND $\tilde{l}_{i+1}$. WE MAY REWRITE THIS AS

$$\tilde{L}_{j-1} = \tilde{L}_j + \tilde{P}_i + \tilde{P}_{i+1} \qquad ③$$

BUT WE HAVE ESTABLISHED, FROM CONSIDERATIONS PREVIOUS TO ② THAT

$$\min\left(\tilde{P}_i + \tilde{P}_{i+1}\right) = P^{(j)}_{q-j} + P^{(j)}_{q-j-1} \qquad ④$$

FURTHERMORE, SINCE WE HAVE ASSUMED THAT C_j IS COMPACT, AND HAVE FURTHERMORE SPECIFIED THAT $\tilde{C}_j$ BE COMPACT, THEN $\tilde{L}_j = L_j$. FROM ②, ③, AND ④, IT THEN FOLLOWS THAT

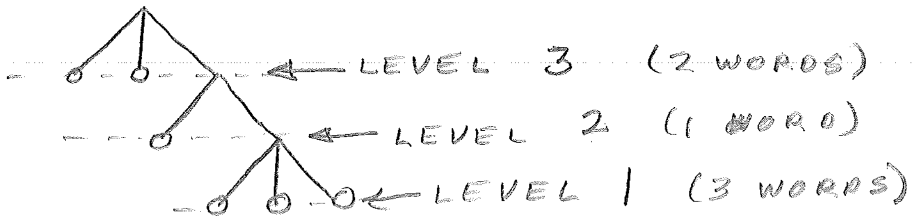
$$\tilde{L}_{j-1} \geq L_{j-1}$$

THIS IS A CONTERDICTION TO OUR ASSUMPTION THAT $\tilde{L}_{j-1} < L_{j-1}$ AND THE PROOF IS COMPLETE.

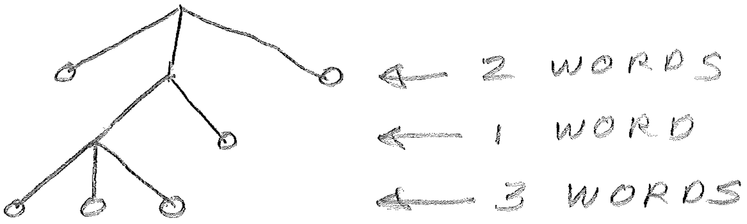
8-5-76*

4-13 p.92 (TEXT)

IN WRITING THESE "TREES", WE SHALL USE THE IDEA OF WORDS PER LEVEL. FOR EXAMPLE:



HERE LEVEL 3 HAS 2 WORDS LEVEL 2 HAS ONE WORD, AND LEVEL 1 HAS 3. BY SPECIFYING WORDS PER LEVEL, WE SPECIFY THE TREE. FOR EXAMPLE

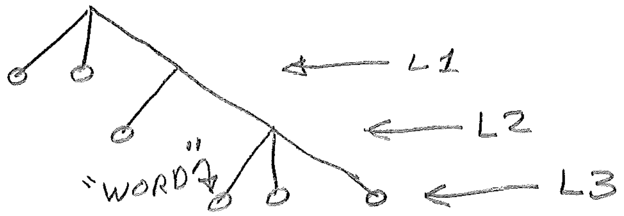


THIS TREE WILL GIVE THE SAME CODE AS THE FIRST TREE*, BUT COULD TRICK THE EYE INTO THINKING ITS DIFFERENT. AS SUCH, THE WORD-LEVEL ASSOCIATION GIVES A GOOD ORDERING PROCESS.

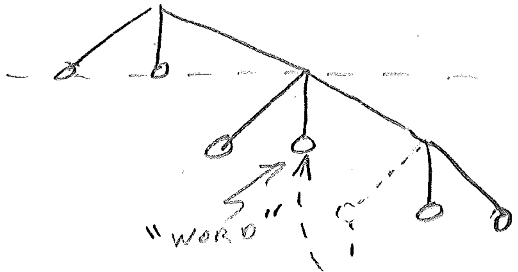
---

* IN THE SENSE THAT CORRESPONDING CODE WORD LENGTHS ARE THE SAME.

A NECESSARY CONDITION FOR A TREE TO REPRESENT A COMPACT CODE IS THAT ALL OF IT'S UPPER LEVELS BE FILLED. FOR EXAMPLE, CONSIDER THE "NON-COMPACT" TREE ON THE PREVIOUS PAGE REDRAWN HERE:



L2 IS NOT "FILLED" FOR THE CASE OF TRINARY CODING. WE COULD TAKE THE "WORD" SPECIFIED IN L3 AND PLACE IT IN L2:



NOW ALL UPPER LEVELS ARE FILLED IN THE SENSE THAT A HORIZONTAL LINE DRAWN THOUGH A LEVEL WILL INTERSECT IN AN INTEGRAL MULTIPLE OF THREE POINTS.

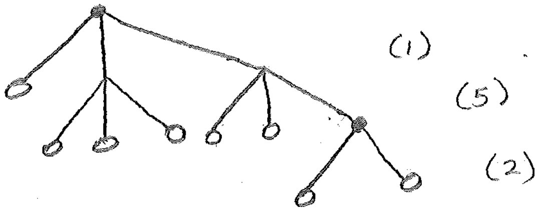
ALSO, NOTE THAT FOR r=3, WE MUST ADD, AT MOST, ONE DUMMY SYMBOL FOR HUFFMAN CODING. NOW, (INCLUDING THE DUMMY) THE LAST THREE CODE

WORDS WILL HAVE EQUAL LENGTH. THUS, IN THE TREE DIAGRAMS, WE MUST HAVE A MINIMUM OF <u>TWO</u> WORDS IN THE TREE DIAGRAM IN THE LOWEST LEVEL. THIS IS SOMEWHAT INTUITIVE.

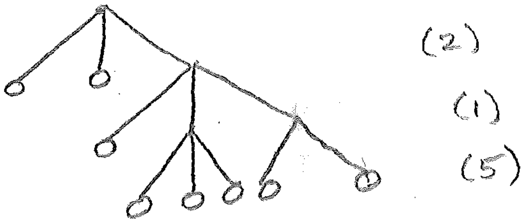
⟹

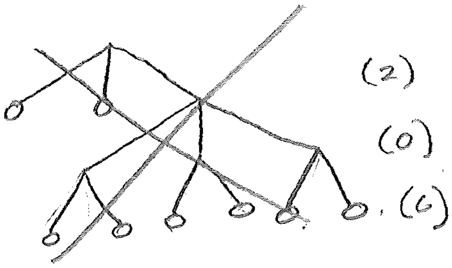
(2) $r = 3$ $q = 8$

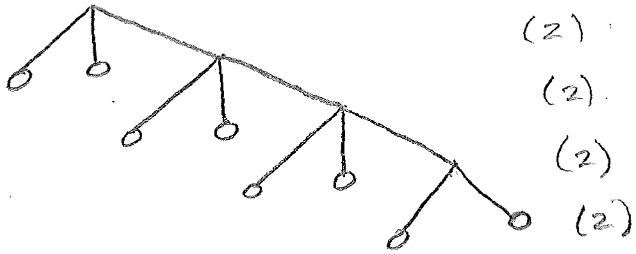
ONE WORD IN TOP LEVEL



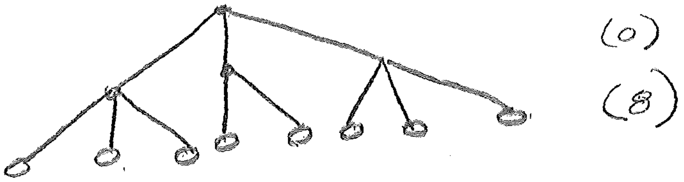
TWO WORDS IN TOP LEVEL







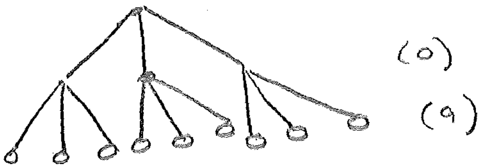
ZERO WORDS AT TOP LEVEL



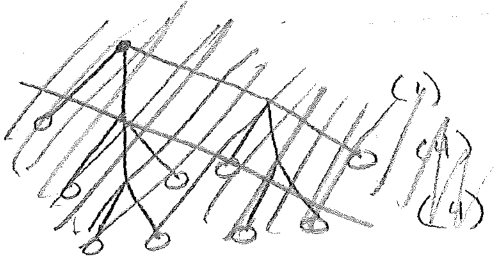
LOOKS LIKE, FOR $r = 3$, $q = 8$, THERE'S 4 POSSIBLE CONPACT CODES

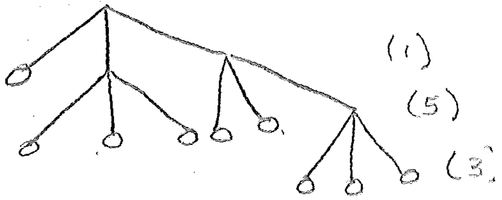
(b) $r=3$ $q=9$

ZERO ON TOP

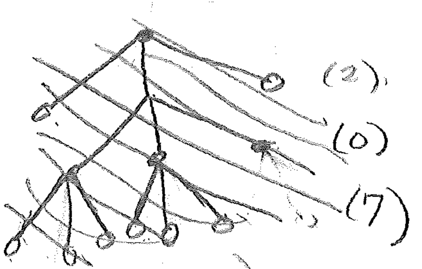


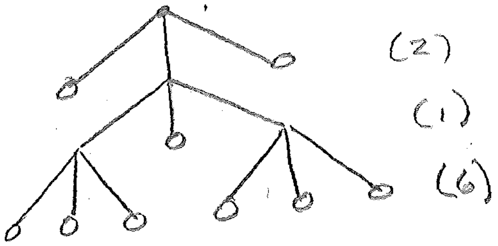
ONE ON TOP

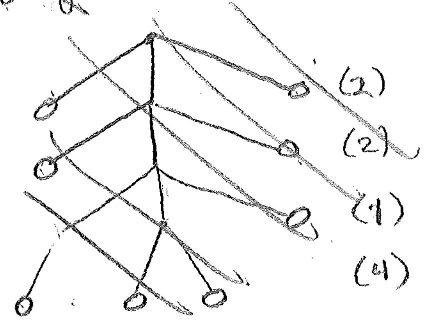


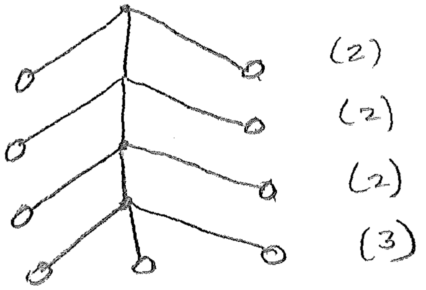


TWO ON TOP









LOOKS LIKE, FOR THIS CASE, THERE ARE ALSO 4 POSSIBLE COMPACT CODES.

WE WISH TO SHOW THAT, FOR HUFFMAN CODING, THAT

$$H(S) \leq \bar{L} \leq \bar{H}(S) - 2P_{min} + 1 \quad ①$$

WE WILL HAVE TO ASSUME BINARY CODING. WE HAVE ALREARY ESTABLISHED (IN SEC. 4-1 OF TEXT) THAT

$$H(S) \leq \bar{L} \quad ②$$

THUS THE UPPER BOUND IN EQ.① WILL BE OF PRIME INTEREST.

WE WILL UTILIZE THE HUFFMAN CODING PROCEDURE IN OUR PROOF. OUR CODING TABLE WILL LOOK LIKE

$$\begin{array}{cccccc}
 & S & S' & & & q=2 \\
 & p_1 & p_1' & \cdots & & p_1'' \\
\cdots & p_2 & p_2' & \cdots & & p_2'' \\
 & p_3 & p_3' & & \cdot & \\
 & \vdots & \vdots & \cdot & & \\
 & p_{q-1} & p_{q-1}' & & & \\
 & p_q & & & &
\end{array}$$

THE CODE ASSOCIATED WITH S IS C. DENOTE THE SYMBOLS IN S & S' BY $\{S_i\}$ & $\{S_i'\}$ AND THE CORRESPONDING WORD LENGTHS BY l_i & l_i'

ALSO, $p_{i+1} \le p_i$ AND $p'_{i+1} \le p'_i$.
WE DENOTE THE FIRST REDUCTION OF C BY C'. THIS IS FORMED BY TAKING THE SMALLEST TWO PROB. ASSOCIATED WITH C (SPECIFICALLY p_q AND p_{q-1}) AND FORMING A NEW SYMBOL s'_i WITH WORD LENGTH $l'_i = l_q - 1 = l_{q-1} - 1$. (WE ARE ASSURED THAT S_q AND S_{q-1} HAVE EQUAL WORD LENGTHS FROM Eq. 4-27 ON pg. 83).

OUR PROOF IS BASED ON INDUCTION. FOR $q = 2$ (ie $S = \{s_1, s_2\}$) WE WRITE

$$0 \le H(s) \le 1 \quad \text{(BITS)}$$
$$p_{\text{min.}} \le 1/2 \qquad ⑤$$
$$\bar{L} = 1 \text{ BINIT}$$

① IS OBVIOUSLY SATISFIED FOR THESE VALUES.

TO COMPLETE THE PROOF, WE WILL ASSUME ① IS TRUE FOR SOME HUFFMAN DERIVED CODE, C', THEN, SHOW AS A CONSEQUENCE, IT IS TRUE FOR CODE C. LET US DENOTE THE SYMBOLS FOR C' BY

$$S' = \{ s'_1, s'_2, \ldots, s'_{q-1}) \qquad ⑥$$

AGAIN, THEIR PROB'S ARE ARRANGED $\ni$

$$p'_1 \ge p'_2 \ge \ldots \ge p'_{q-2} \ge p'_{q-1} \qquad ⑦$$

AND C BY

$$S = \{s_1, s_2, \cdots, s_q) \quad (8)$$

$$p_1 \geq p_2 \geq \cdots \geq p_{q-2} \geq p_{q-1} \geq p_q \quad (9)$$

NOW, IF THE ENTROPY OF CODE C IS H AND THE ENTROPY OF C' IS H', THEN, FOLLOWING HUFFMAN CODING PROCEDURE AND UTILIZING THE "ADDITIVE PROPERTY OF ENTROPY" GIVES

$$H = H' + (p_{q-1} + p_q) H_{q,q-1} \quad (10)$$

WHERE

$$H_{q,q-1} \triangleq H\left(\frac{p_{q-1}}{p_q + p_{q-1}}, \frac{p_q}{p_q + p_{q-1}}\right) = \frac{p_{q-1}}{p_q + p_{q-1}} \lg \frac{p_q + p_{q-1}}{p_{q-1}} + \frac{p_q}{p_q + p_{q-1}} \lg \frac{p_q + p_{q-1}}{p_q} \quad (11)$$

~~WE MAY REWRITE (10) AS~~

$$\cancel{H = H' + (p_{q-1} + p_q) H_{q,q-1}} \quad \cancel{(12)}$$

FOLLOWING THE SPIRIT OF INDUCTIVE PROOF, WE ASSUME ① IS TRUE FOR CODE C'. THAT IS

$$H' \leq L' \leq H' + 1 - 2p'_{min} \tag{13}$$

WHERE L' IS THE AVERAGE WORD LENGTH ASSOCIATED WITH C' AND

$$p'_{min} = min(p'_{q-1}, p'_{q-2}) = p'_{q-1} \tag{14}$$

IF L IS THE AVERAGE WORD LENGTH ASSOCIATED WITH C, THEN (FROM TEXT Eq 4-25)

$$L = L' + p_q + p_{q-1} \tag{15}$$

SUBSTITUTING THIS INTO ⑬

$$H' \leq L - (p_q + p_{q-1}) \leq H' + 1 - 2p'_{min} \tag{16}$$

SUBSTITUTING ⑩:

$$H' \leq L - (p_q + p_{q-1}) \leq H - (p_{q-1} + p_q)H_{q,q-1} + 1 - 2p'_{min} \tag{17}$$

WE KNOW (FROM SEC 4-1) THAT

$$L \geq H \tag{18}$$

SO THAT ⑰ MAY BE WRITTEN AS

$$H \le L \le H - (p_{q-1} + p_q) H_{q,q-1} + (p_q + p_{q-1}) + 1 - 2p'_{\text{MIN}}$$

OR

$$H \le L \le H + \{1 - H_{q,q-1}\}(p_{q-1} + p_q) + 1 - 2p'_{\min} \qquad (19)$$

DEFINE

$$p_{\min} = \min[p_{q-1}, p_q] = p_q \qquad (20)$$

OBVIOUSLY

$$2p_{\min} \le p_{q-1} + p_q \qquad (21)$$

SINCE

$$1 - H_{q,q-1} \ge 0 \qquad (22)$$

WE CAN REWRITE ⑲ AS

$$H \le L \le H + \{1 - H_{q,q-1}) 2p_{\min} + 1 - 2p'_{\min} \qquad (22a)$$

NOW, IN ORDER TO PROVE ①, WE MUST SHOW THAT, (WITH REFERENCE TO (22a))

$$H + (1 - H_{q,q-1}) 2p_{\min} + 1 - 2p'_{\min} \overset{?}{\ge} H + 1 - 2p_{\min} \qquad (22b)$$

OR, EQUIVALENTLY

$$2(2 - H_{q,q-1}) p_{\min} - 2p'_{\min} \overset{?}{\ge} 0 \qquad (23)$$

FROM HUFFMAN PROCEDURE, AND DUE TO THE PROBABILITY ORDERING, IT IS OBVIOUS THAT

$$p'_{min} \geq p_{min} \qquad ㉔$$

NOW THE MAXIMUM VALUE THE L.H.S. OF ㉓ CAN ACHEIVE IS WHEN $H_{q,q-1} = 0$ AT WHICH TIME IT IS EQUAL TO

$$2p_{min} - 2p'_{min}$$

OBVIOUSLY, $p'_{min} \geq p_{min}$. THUS

$$2p_{min} - 2p'_{min} \leq 0 \qquad ㉕$$

~~THIS COMPLETES THE PROOF. IN SUMMARY, WE HAVE SHOWN FROM ㉕ THAT ㉓ AND (22b) ARE TRUE. SINCE WE HAVE PROVEN ① FOR C_2 (ie q=2), ASSUMED IT FOR $C' = C_{q-1}$ AND SHOWN AS A CONSEQUENCE, THAT ① IS TRUE FOR $C = C_q$, IT FOLLOWS THAT ① IS TRUE FOR ALL CODES GENERATED BY THE HUFFMAN CODING TECHNIQUE (WHICH ARE, OF COURSE, COMPACT)~~

ERGO, WE HAVE SHOWN THAT ㉓ AND ㉔ ARE TRUE. REWRITE (22b):

$$H + (1 - H_{q,q-1})\, 2p_{min} + 1 - 2p'_{min} \le H + 1 - 2p_{min}$$

THEN, IT FOLLOWS FROM (22a) THAT

$$H \le L \le H + 1 - 2p_{min} \qquad ㉖$$

AND THE PROOF IS COMPLETE.

IN SUMMARY, WE HAVE SHOWED THAT, IN A HUFFMAN CODING SCHEME, ALL RESULTING CODES SATISFY ㉖. WE DID THIS INDUCTIVELY BY SHOWING THAT ㉖ IS TRUE FOR $n=2$ AND BY NEXT SHOWING THAT, BY ASSUMING ㉖ FOR $n=q-1$, ㉖, AS A CONSEQUENCE, ALSO GAVE CORRECT BOUNDS FOR $n=q$.

8-12-76

WE WISH TO CODE 7-4-1776 ALA HAMMING USING SINGLE ERROR CORRECTING CODE NOW

$(1776)_{10} = (11011110000)_2 \Rightarrow 11$ BINITS

WE WILL SEPARATELY ENCODE 7, 4, AND 1776 (AS INSTRUCTED) USING $m = 11$ FOR EACH. (IT SHOULD BE NOTED THAT THIS IS <u>NOT</u> AN OPTIMAL CODING SCHEME FOR CODING A DATE IN TERMS OF WORD LENGTH).

THE EXPANSION OF 7 IS

00000000111

FOR $m = 11$, $K = 4$. THUS, SET UP THE TABLE:

15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
0	0	0	0	0	0	0	0	0	1	1	0	1	0	0
							P_3				P_2		P_1	P_0

WE'LL USE EVEN 1 BIT PARITY. NOW

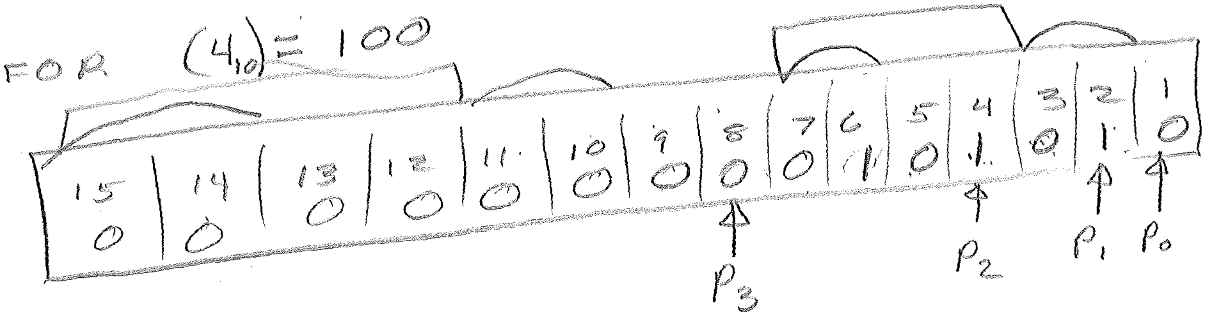
$P_0 \oplus 1 \oplus 1 \oplus 0 \oplus 0 \oplus 0 \oplus 0 \oplus 0 = 0 \Rightarrow P_0 = 0$

$P_1 \oplus 1 \oplus 1 \oplus 0 \oplus 0 \oplus 0 \oplus 0 \oplus 0 = 0 \Rightarrow P_1 = 0$

$P_2 \oplus 1 \oplus 1 \oplus 0 \oplus 0 \oplus 0 \oplus 0 \oplus 0 = 0 \Rightarrow P_2 = 0$

$P_3 \oplus 0 \ldots \oplus 0 = 0 \Rightarrow P_3 = 0$

FOR $(4_{10}) = 100$



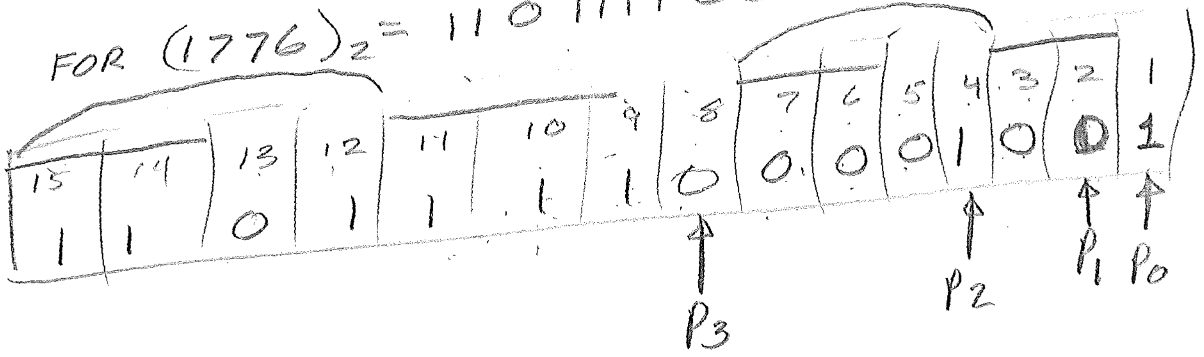
$p_0 \oplus 0 \oplus \ldots \oplus 0 = 0 \Rightarrow p_0 = 0$

$p_1 \oplus 0 \oplus 1 \oplus 0 \ldots \oplus 0 = 0 \Rightarrow p_1 = 1$

$p_2 \oplus 0 \oplus 1 \oplus 0 \ldots \oplus 0 = 0 \Rightarrow p_2 = 1$

$p_3 \oplus \ldots \oplus 0 \Rightarrow p_3 = 0$

FOR $(1776)_2 = 110\ 1111\ 0000$



$p_0 \oplus 0 \oplus 0 \oplus 0 \oplus 1 \oplus 1 \oplus 0 \oplus 1 = 0 \Rightarrow p_0 = 1$

$p_1 \oplus 0 \oplus 0 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 = 0 \Rightarrow p_1 = D$

$p_2 \oplus 0 \oplus 0 \oplus 0 \oplus 1 \oplus 0 \oplus 1 \oplus 1 = 0 \Rightarrow p_2 = 1$

$p_3 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 0 \oplus 1 \oplus 1 = 0 \Rightarrow p_3 = 0$

STRINGING THESE TOGETHER GIVES

0000 00000 11 0100 000000000

10101011 01111 0000 1 011

I HAVE READ
"MATHEMATICAL THEORY
OF COMMUNICATION"
PARTS I THRU V
by CLAUDE SHANNON

COMPUTER PROGRAM
DONE WITH WES REDUS

H.O #9

Given the Fano Bound

$$H(X/Y) \leq H\left[p(e), 1-p(e)\right] + p(e)\log(s-1)$$

and the fact

$$I(X_1, X_2, \ldots X_n \mid Y_1, Y_2 \ldots Y_n) \leq \sum_{i=1}^{n} I(X_i/Y_i)$$

with equality iff Y_i's are independent

prove that,

(1) $p(e)$, the prob. of error satisfies the relation $\log s \leq \dfrac{nC + \log 2}{1 - \overline{p(e)}}$

(2) if $s \geq 2^{n(c+\delta)}$, $\delta > 0$, then

$$\overline{p(e)} \geqslant 1 - \frac{c+\frac{1}{n}}{} \longrightarrow 1 - \frac{c}{c+\delta} > 0$$

where C = channel capacity

R is a + number being the source uncertainty

$\overline{p(e)}$ = average prob. of error.

s = no. of symbols in the input alphabet

(3) Use (2) to show that for $R > C$, no sequence of codes $\left[[2^{nR}], n \right]$ can have an average probability which $\rightarrow 0$ as $n \rightarrow \infty$; hence that no sequence of codes $\left[[2^{nR}], n, \lambda_n \right]$ can exist with $\lim_{n \rightarrow \infty} \lambda_n = 0$

Note: The code designation $[A, B, c_1]$ means that

$A = S$ is the no. of input n-sequences

$B = n$ is the number n, the extension number of primary (0,1) alphabet

c_1 is the max. prob. of error. (like ϵ)

7-13-76 (TUES)

HOUSEKEEPING

J.C. PRABHAKAR EE104 (AFTERNOON)

5375 INFORMATION THEORY

TEXT: INFORMATION THEORY & CODING

by: NORMAN ABRAMSON (U. of HAWAII)

McGRAW HILL, 1963

OTHER REFERENCES

1. "AN INTRODUCTION TO INFO. THEORY" (READABLE)
   by F.M. REZE (McGRAW HILL, 1968)
2. "INFORMATION THEORY" (EXCELLENT)
   by ROBERT ASH (INTERSCIENCE, WILEY)
3. "INFORMATION AND CODING THEORY" (FAIRLY GOOD)
   by F.M. INGELS (INTEXT)

4-6. RECENT (1959-1974) IEEE JOURNAL PAPERS.

COURSE CONTENT:

1. PROBABILITY REVIEW
   - NOTIONS OF PROBABILITY
   - PROBABILITY MEASURE
   - MARGINAL, CONDITIONAL & JOINT PROBABILITY
   - STATISTICAL INDEPENDENCE OF EVENTS
   - PROBABILITY DENSITY (& DISTRIBUTION) FUNCTIONS
2. "SOURCES" OF INFORMATION
   - UNCERTAINTY MEASURE
   - SOURCE "ENTROPY"
   - JOINT, MARGINAL & CONDITIONAL ENTROPY

3. "CLASSICAL" CODING PROCEDURES
   - COST
   - NOISELESS CODING
   - SHANNON'S 1ST THEOREM
   - CHANNELS & THEIR CHARACTERISTICS
4. DECIPHERABLE & UNIQUELY DECIPHERABLE CODING SCHEMES
   - CORRECTING CODES
5. "MODERN CODING SCHEMES
   - CODING THEOREMS
   - PARITY CHECKING CODES
   - GROUP CODES
   - BOSE CHAUDBURY CODING SCHEMES (READING ASSIGNMENT)
   - THE MATHEMATICS OF CODES.

GRADING:
- 3 QUIZES (CLASSTIME : 2 HRS) (60%)
- 10-15 HOMEWORK STARRED PROBLEMS } 40%
- POSSIBLY A FINAL }

ALSO
- SUGGESTED HOMEWORK PROBLEMS

7-14-76

READ THIS PAPER:

"THE MATHEMATICAL THEORY OF COMMUNICATION"

by CLAUDE SHANNON & W. WEAVER

IN BELL SYSTEM TECHNICAL JOURNAL

27 379-423, 623-656 (1948)

ALSO READ BEFORE 7-19-76 (MON) pg. 1-10 TEXT

I. PROBABILITY THEORY: SOME STATEMENTS AND RELATIONS

(a) FREQUENCY OF EVENTS APPROACH:

FORM A BASIC EXPERIMENT

e.g. PULLING OUT A SPADE FROM A PACK OF 52

IN GENERAL, LET $n(x_k)$ = NUMBER OF TIMES AN EVENT x_k OCCURS & N = TOTAL # OF EVENTS.

THEN, ONE SAYS

$$P[x_k \text{ WILL OCCUR}] = \frac{n(x_k)}{N}$$

AS N GETS BIG.

IT IS CLEAR THAT $0 \le n(x_k) \le N$

$$0 \le n(x_k)/N \le 1$$

$$\Rightarrow 0 \le \lim_{N \to \infty} \frac{n(x_k)}{N} \le 1$$

$$0 \le P(x_k) \le 1$$

$\therefore P(x_k)$ IS SINGLE VALUED & REAL

$P(x_k) = 0 \Rightarrow$ IMPOSSIBLE EVENT

$P(x_k) = 1 \Rightarrow$ CERTAIN EVENT

(b) THE PROBABILITY MEASURE APPROACH (AXIOMATIC APPROACH)

THE PROB. MEASURE IS A SPECIFIC TYPE OF FUNCTION (IN THE FRAMEWORK OF MEASURE THEORY) WHICH CAN BE ASSOCIATED WITH SETS. HERE, EACH POSSIBLE OUTCOME OF AN EXPERIMENT CORRESPONDS TO A POINT, a_K, IN A SAMPLE SPACE. THEN, $m(a_K)$ (m FOR "MEASURE") IS A REAL & SINGLE VALUED FUNCTION CALLED THE PROBABILITY MEASURE AND THE PROB. MEASURE ON AN EVENT = Σ OF PROB. MEASURE OF POSSIBLE OUTCOMES $\{a_K\}$ THAT MAKE UP THAT EVENT.

EX. FOR A PACK OF CARDS, $m(a_K) = \frac{1}{52}$

P[CHOOSING A SPADE] = $^{13}/_{52} = \frac{1}{4}$

* TWO EVENTS ARE DISJOINT IF THEY CONTAIN NO OUTCOMES IN COMMON. i.e. "CANNOT HAPPEN SIMULTANEOUSLY"

ON THIS BASIS

(1) $0 \leq m(a_K)$

(2) $m[A \cup B] = m(A) + m(B)$ IF A & B ARE DISJOINT

THIS IS CALLED THE "ADDITIVITY PROPERTY" OF THE MEASURE.

ALSO, $m(X) = 0$ ONLY IF $X = \phi$ (NULL EVENT)

$m(X) = 1$ ONLY IF $X = U$ (UNIVERSAL SET)

ALSO (1) $m(A) \leq m(B)$ IF $A \subset B$

"$\subset$" ≡ "IS A SUBSET OF".

(2) $m(A) = m(B) - m(B-A)$ IF $A \subset B$

(3) $m(A') = m(\overline{A}) = m(U-A)$

$= m(U) - m(A) = 1 - m(A)$

(4) $m(A \cup B) = m[(A - AB) \cup B]$

($AB \equiv A$ INTERSECT B)

$m(A \cup B) = m(A - AB) - m(B)$

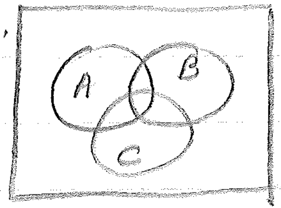
$= m(A) - m(AB) + m(B)$

THIS IS THE GENERAL ADDITIVE LAW OF PROBABILITY MEASURE.

(5) $m[A \cup B \cup C] = m(A) + m(B) + m(C)$

$- m(AB) - m(CA) - m(BC)$

$+ m(ABC)$



EXTENSIONS TO MORE THAN 3 SUBSETS IS OBVIOUS

ONE MAY SUMMARIZE AS FOLLOWS:

THE PROB. IS A PROB. FUNCTION ON A SAMPLE SPACE S WITH THE FOLLOWING AXIOMS →

AXIOM 1: $P(A)$ IS A REAL NUMBER $\ni P(A) \geq 0$ $\forall$ EVENT $A \in S$

AXIOM 2: $P(S) = 1$

AXIOM 3: IF $S_1, S_2, \ldots$ IS A SEQUENCE OF MUTUALLY EXCLUSIVE (DISJOINT) EVENTS IN S, THES $S_i \cap S_j = \phi$ $\forall$ $i \neq j = 1, 2, \ldots$, THEN ADDITIVE LAW HOLDS:

$$P[S_1 \cup S_2 \cup S_3 \cup \ldots] = P[S_1] + P[S_2] + \ldots$$

THESE AXIOMS RESULT IN

THEOREM 1: LET S BE A SAMPLE SPACE & P A PROBABILITY (MEASURE) FUNCTION ON S. THEN PROB THAT EVENT A DOES NOT OCCUR $= 1 - P(A)$ i.e. $P(\bar{A}) = P(A') = 1 - P(A)$

THEOREM 2: IF S IS A S.S. WITH PROB MEASURE P, THEN $0 \leq P(A) \leq 1$ $\forall A \in S$

THEOREM 3: IF S IS A S.S. WITH PROB. MEASURE P & IF S_0 IS A NULL SET, THEN $P(S_0) = 0$.

* MARGINAL, JOINT, & CONDITIONAL PROBABILITIES.

LET S BE A SAMPLE SPACE WITH n POINTS WITH A PROBABILITY (MEASURE) = $1/n$. PARTITION S INTO r DISJOINT SUBSETS $A_1, A_2, A_3, \cdots, A_R$. ALSO PARTITION S INTO s DISJOINT SUBSETS $B_1, B_2, \cdots, B_s$.

	B_1	B_2	B_3	$\cdots$	B_s
A_1	n_{11}	n_{12}	n_{13}		n_{1s}
A_2	n_{21}	n_{22}	n_{23}		n_{2s}
A_3					
$\vdots$					
A_r	n_{r1}	n_{r2}	n_{r3}		n_{rs}

$\ni n_{ij}$ = # OF OUTCOMES WITH ATTRIBUTE A_i AND B_j

$$\therefore \sum_{i,j} n_{ij} = n$$

THEN THE PROB. OF THE EVENT A_1 AND B_3

$$= P(A_1, B_3) = P[A_1 \cap B_3] = n_{13}/n$$

THIS IS THE JOINT PROBABILITY. HERE ONE LOOKS AT TWO OR MORE ATTRIBUTES AT A TIME

IF ONE IS INTERESTED IN ONLY ONE ATTRIBUTE A_2, THEN

$$P[A_2] = \sum_{j=1}^{s} n_{2j} \Big/ n = \sum_{j=1}^{s} P(A_2, B_j)$$

IN GENERAL,

$$P[A_i] = \sum_j P[A_i, B_j]$$

AND

$$P[B_j] = \sum_i P[A_i, B_j]$$

PROBS. SUCH AS $P(A_i)$ OR $P[B_j)$ ARE MARGINAL PROBS.

FROM THE POINT OF VIEW OF SET THEORY:

$$A_i = (A_i \cap B_1) \cup (A_i \cap B_2) \cup (A_i \cap B_3) \cup \ldots (A_i \cap B_5)$$

SINCE $(A_i \cap B_j) \cap (A_i \cap B_{j'}) = \phi$

WHERE $j \neq j'$

THEN, FROM AXIOM 3:

$$P(A_i) = P(A_i \cap B_1) + P(A_i \cap B_3) + \ldots + P(A_i \cap B_5)$$

$$= \sum_j P(A_i, B_j)$$

$$= \sum_{j=1}^{5} n_{ij}/n$$

ALSO

$$P(A_i, C_K) = \sum_j P(A_i, B_j, C_K)$$

ALSO

$$P(C_K) = \sum_j \sum_i P(A_i, B_j, C_K)$$

* CONDITIONAL PROBABILITY

HERE, THE OUTPUT IS EXAMINED FOR ONE ATTRIBUTE KNOWING (A-PRIORI) THAT THE OTHER ATTRIBUTE HAS ALREADY RESULTED.

e.g. LET THE GIVEN (A PRIORI) ATTRIBUTE BE B_3. QUESTION: KNOWING THIS, WHAT IS THE PROBABILITY THAT IT IS ALSO A_2? FROM THE CHART, THE # OF OUTCOMES FOR A GIVEN $B_3 = \sum_{i=1}^{r} n_{i3}$ THE NUMBER OF DESIRED OUTCOMES OUT OF THESE IS n_{23}.

$$\therefore \quad P(A_2/B_3) = n_{23} \Big/ \sum_{i=1}^{r} n_{i3}$$

OR, IN GENERAL

$$P(A_i/B_j) = n_{ij} \Big/ \sum_{i=1}^{r} n_{ij}$$

$$= \frac{n_{ij}/n}{\sum_{i=1}^{r} n_{ij}/n}$$

COND.

$$\boxed{\Rightarrow P(A_i/B_j) = \frac{P[A_i, B_j]}{P[B_j]}} \begin{matrix} \leftarrow \text{JOINT} \\ \leftarrow \text{MAGRINAL} \end{matrix}$$

THIS RELATION IS SOMETIMES REFERED TO AS MULTIPLICATIVE LAW OF PROBABILITY MEASURE:

$$P(A_i, B_j) = p(B_j)\, P(A_i/B_j)$$

FOR THREE ATTRIBUTES:

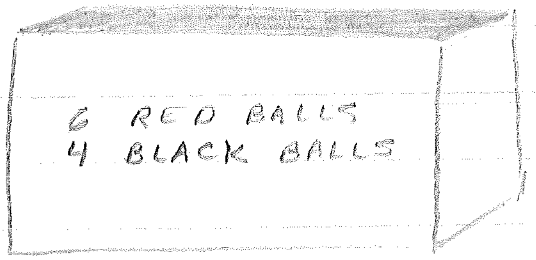
$P[A_i, B_j / C_K] = P[A_i, B_j, C_K] / P[C_K]$

$P[A_i | B_i, C_K] = P[A_i, B_j, C_K] / P[B_i, C_K]$

$\Longrightarrow$ HOMEWORK:

1. WRITE OUT $P(A_i, B_j, C_K)$ IN TERMS OF VARIOUS CONDITIONAL & MARGINAL PROBABILITIES.

2.



2 BALLS DRAWN W/O REPLACEMENT

FIND $P[2^{nd} \text{ RED} / 1^{ST} \text{ IS RED}]$

(VERIFY RIGOUROUSLY)

$P(A, B) =$ JOINT

$=$ RELAVENT OUTCOMES / POSSIBLE OUTCOME

$= {}^6C_2 / {}^{10}C_2 = 1/3$

$P[A/B] = P[A,B]/P[B] = \frac{1}{3} / \frac{6}{10} = \frac{5}{9}$

7-15-76 (WED)

STATISTICAL INDEPENDENCE

$P(B/A) = P[A,B] / P[A]$

IF $P[B/A] = P[B]$

THEN $P[B/A] = P(A,B)/P(A) = P(B)$

THUS $P(A,B) = P(A)P(B)$

THEN EVENTS A & B ARE SAID TO BE STATISTICALLY INDEPENDENT

$$\therefore P(A,B) = P(A/B)P(B) = P(B/A)P(A)$$
$$= P(A)P(B) = P(B)P(A)$$

HOWEVER, WHEN MORE THAN TWO EVENTS ARE INVOLVED, ADDITIONAL INFORMATION IS NEEDED TO ESTABLISH STATISTICAL IND. OF THESE EVENTS.

e.g. CONSIDER AN EXPERIMENT WITH FOUR MUTUALLY EXCLUSIVE OUTCOMES A_1, A_2, A_3, A_4 EACH WITH PROB. $\frac{1}{4}$

DEFINE: $B_j \ni$ $B_1 = (A_1 \text{ OR } A_2)$
$B_2 = (A_1 \text{ OR } A_3)$
$B_3 = (A_1 \text{ OR } A_4)$

NOW $P(B_1) = P(B_2) = P(B_3) = \frac{1}{2}$

NEXT, LOOK AT JOINT PROBABILITIES

PAIRWISE : $P(B_1, B_2)$, DENOTES "AND"

$$P(B_1, B_2) = P(A_1) = \frac{1}{4}$$
$$= P(B_1)P(B_2)$$

$\therefore B_1$ AND B_2 ARE STAT. IND. $\Longrightarrow$

SAME HOLDS FOR $P(B_1, B_3)$ & $P(B_2, B_3)$.
THUS, PAIRWISE EVENTS ARE STAT. IND.

HOWEVER, CONSIDER

$$P(B_3 / B_1, B_2) = 1 \neq P(B_3) = \frac{1}{2}.$$

HENCE, WE NEED TO CHANGE S.I. IDEA FOR N EVENTS;

DEF: N EVENTS ARE SAID TO BE S.I. IF FOR ALL COMBINATIONS $1 \leq i < j < k, \ldots \leq N$ IF THE FOLLOWING RELATIONSHIPS HOLD;

$$P(A_i, A_j) = P(A_i) P(A_j) \quad \leftarrow \text{PAIRWISE INDEP.}$$

$$P(A_i, A_j, A_k) = P(A_i)\, P(A_j)\, P(A_k)$$

$$\vdots \qquad\qquad \vdots \qquad \vdots \qquad \vdots$$

$$P(A_i, A_j, A_k, \ldots A_N) = P(A_i) P(A_j) P(A_k) \ldots P(A_N)$$

** RANDOM VARIABLE (A MATTER OF ASSOCIATING REAL #'S WITH OUTCOME OF EXPERIMENTS)

A RIGOUROUS DEFN IS:
A REAL VALUED FUNCTION $X(S)$ DEFINED ON A SAMPLE SPACE, S, IS A RANDOM VARIABLE IF FOR EVERY REAL # a, THE SET OF POINTS FOR WHICH $X(S) \leq a$ IS ONE OF THE CLASS OF ADMISSIBLE SETS FOR WHICH A PROB. IS DEFINED.

RANDOM #'S CAN BE

(1) DISCRETE WHEN THE NUMBER OF OUTCOMES OF AN EXPERIMENT IS FINITE OR COUNTABLY INFINATE

(2) CONTINUOUS RANDOM VARIABLE (REQUIRE pdf OR CDF FOR DESCRIPTION)

A FUNCTION OF A RANDOM VARIABLE IS A RANDOM VARIABLE.

* PROBABILITY DISTRIBUTION FUNCTION IS MERELY A PROBABILITY THAT A R.V. $X(x)$ IS BOUNDED BY AN ARBITRARILY CHOSEN REAL # $\mathbb{X}$.

ie, $P[x \leq \mathbb{X}] \triangleq$ PDF OF THE R.V. X.

$$P[x \leq \infty] = 1$$

$$P[x < -\infty] = 0$$

$$P[x \leq b] - P[x \leq a] = P[a < x \leq b] \text{ IF } b > a$$

THUS, THE PDF IS A NON-DECREASING FUNCTION OF X.

JOINT PDF (PROB. DISTRIBUTION FUNCTION)

$$\triangleq P[x \leq \mathbb{X}, \underbrace{\;Y \leq \mathbb{Y}}_{\text{"," $\equiv$ "AND"}}]$$

= PROBABILITY THAT THE SAMPLE POINT IS IN THE APPROPRIATE QUADRANT.

OBVIOUSLY:

$$P[x \leq -\infty, Y \leq \mathbb{Y}] = 0$$

$$P[x \leq \infty, Y \leq \infty] = 1$$

(THIS ENCOMPASES THE ENTIRE $\mathbb{XY}$ PLANE OR SAMPLE SPACE)

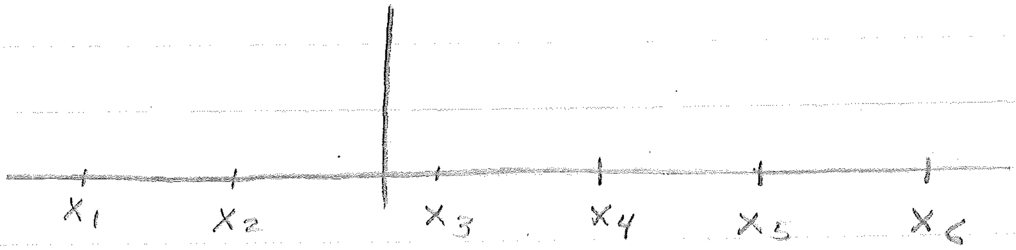
MARGINALLY:

$$P[x \leq \mathbb{X}, Y \leq \infty] = P[x \leq \mathbb{X}]$$

$$P[x \leq \infty, Y \leq \mathbb{Y}] = P[Y \leq \mathbb{Y}]$$

THESE PDF'S ARE CALLED "MARGINAL" PDF'S.

e.g. CONSIDER A DISCRETE r.v. X WHICH CAN TAKE ON 6 VALUES x_i, $i = 1-6$.

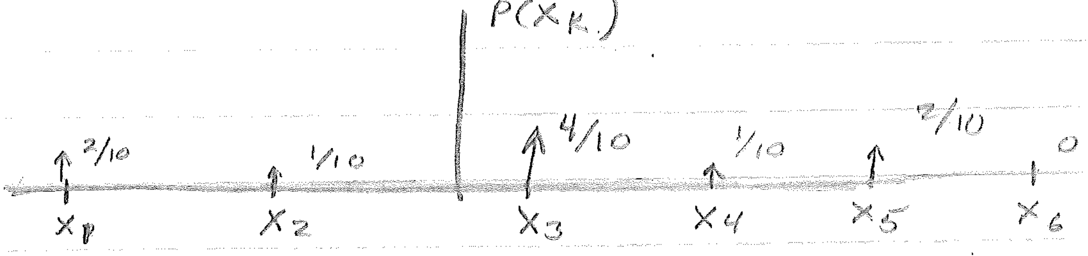


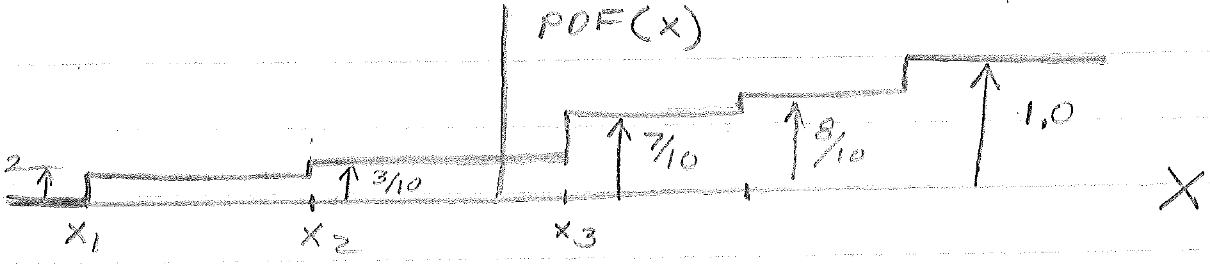
WE HAVE PROBABILITIES:

$P[x_1] = 0.2$ $\quad$ $P[x_4]\ 0.1$

$P[x_2] = 0.1$ $\quad$ $P[x_5] = 0.2$

$P[x_3] = 0.4$ $\quad$ $P[x_6] = 0$

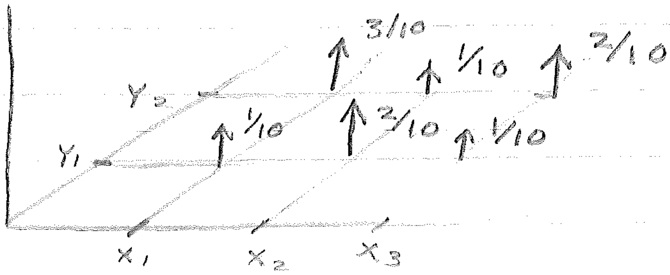




CLEARLY: $\sum_{x_k \le X} P(x_k) = P[x \le X]$

AND $P[x \le \infty] = \sum_{x_k} P[x_k] = 1$

CONSIDER JOINT DISCRETE SITUATION.
LET PROB. PLOT BE AS FOLLOWS:



⟹ HOMEWORK: SKETCH THE JOINT CUMMULATIVE DISTRIBUTION FUNCTION.

MARGINALLY: $p(X_K) = \sum_m P(X_K, Y_m)$

$p(Y_m) = \sum_K P(X_K, Y_m)$

AND $p(X_K, Y_m) = P(Y_m / X_K) P(X_K)$

$= P(X_K / Y_m) P(Y_m)$

THUS $\sum_K P(X_K / Y_m) = \sum_m P(Y_m / X_K) = 1$

A CONTINUOUS RANDOM VARIABLE IS A R.V. FOR WHICH THE PDF IS EVERYWHERE CONTINUOUS. (THAT IS, THE R.V. MAY TAKE ON A CONTINUUM OF VALUE $-\infty \leq x \leq \infty$.) MOREOVER, IF THE PDF IS ALSO DIFFERENTIABLE WITH A CONTINUOUS DERIVATIVE. THUS, ITS PROB. DENSITY FUNCTION (pdf)

$$p(X) \triangleq \frac{d}{dx} PDF = \frac{d}{dX} P(x \leq X)$$

$$= \lim_{\Delta x \to 0} \frac{P[x \leq X] - P[x \leq X - \Delta x]}{\Delta x}$$

THUS, $p(X)\Delta x = P[x - \Delta x < X < x]$

ie, PROB. THAT x IS IN A CERTAIN RANGE = pdf AT x TIMES THE RANGE FOR "SMALL" RANGE.

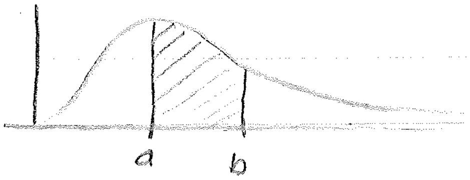
THUS (1) $p(x) \geq 0$

(2) $P[a \leq x \leq b] = \int_a^b p(x)\,dx$

(3) $\int_{-\infty}^{\infty} p(x)\,dx = 1$

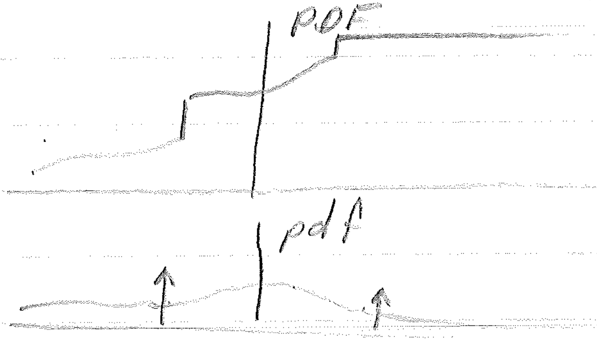
(1), (2), & (3) CONSTITUTE THE PROPERTIES OF A pdf OF A CONTINUOUS R.V.

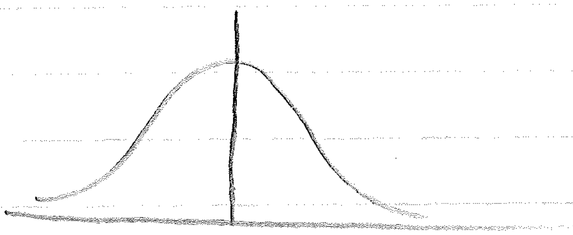
ALSO: $p(x = x_0) = 0$ FOR CONTINUOUS R.V.



← RAYLEIGH DISTRIBUTIO

CONSIDER THE DISTRIBUTION:





← GAUSSIAN OR NORMAL pdf

* ⇒ (7-16-76 HOMEWORK HANDOUT)

7-16-76

* JOINT PROBABILITY DENSITY FUNCTION.

IF IN THE TWO-DIMENSIONAL SAMPLE SPACE, THE JOINT PDF = $P[x \le X, y \le Y]$ IS EVERYWHERE CONTINUOUS AND POSSESSES A MIXED CONTINUOUS MIXED 2nd DERIVATIVE EVERYWHERE, THEN

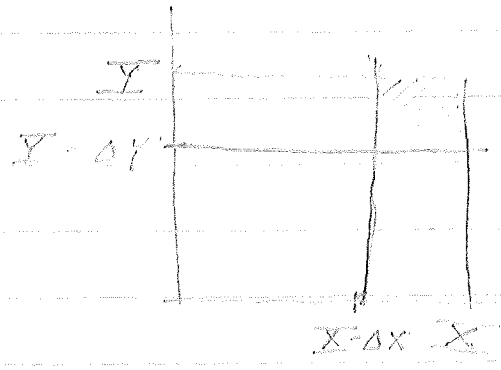
$$\text{JOINT pdf} = \frac{\delta^2}{\delta X \, \delta Y} P[x \le X, y \le Y]$$

THEN $P[x \le X, y \le Y]$

$$= \int_{-\infty}^{X} \int_{-\infty}^{Y} p(x, y) \, dx \, dy$$

THUS,

$$p[X, Y] = \lim_{\substack{\Delta X \to 0 \\ \Delta Y \to 0}} \frac{P[x \le X, y \le Y] - P[x \le X - \Delta X, y \le Y] - P[x \le X, y \le Y - \Delta Y] + P[x \le X - \Delta X, y \le Y - \Delta Y]}{\Delta X \, \Delta Y}$$



IN THE DIFFERENTIAL FORM:

$$p(X, Y) \, dx \, dy = P[X - dy < x < X; Y - \Delta Y < y < Y]$$

RECALL THE SIMILAR RELATION FOR SINGLE VARIABLE;

$p(X, Y)$ IS CALLED THE JOINT pdf

AGAIN, SINCE THE JOINT PDF IS A NON-NEGATIVE NONDECREASING FUNCTION OF ITS ARGUMENT: $p(X,Y) \geq 0$ AND THE PROB THAT A SAMPLE POINT S FALLS IN A REGION $\mathcal{R}$ OF OUR SS IS

$$P[s \in R] = \iint_{\mathcal{R}} p(x,y)\, dx, dy$$

AND THUS

$$\int_{-\infty}^{\infty}\int p(x,y)\, dx\, dy = 1$$

AND, AS BEFORE, MARGINALLY

$$\left.\begin{aligned} \int_{-\infty}^{\infty} p(X,y)\, dy &= p(X) \\ \int_{-\infty}^{\infty} p(x,Y)\, dx &= p(Y) \end{aligned}\right)$$

AND, ALSO

$$\int_{-\infty}^{\infty}\int_{-\infty}^{X} p(x,y)\, dx\, dy = P[x \leq X]$$

$$\int_{-\infty}^{Y}\int_{-\infty}^{\infty} p(x,y)\, dx\, dy = P[y \leq Y]$$

CONDITIONAL pdf

HERE, WHAT IS THE PROB THAT THE R.V. IS $< Y$ SUBJECT TO THE HYPOTHESIS THAT A SECOND R.V

$\mathbb{X} - \Delta x < x < \mathbb{X}$?

$$\therefore P[y < \mathbb{Y} / \mathbb{X} - \Delta x < x < \mathbb{X}]$$

$$= \frac{P[\mathbb{X} - \Delta x < x < \mathbb{X}, y < \mathbb{Y}]}{P[\mathbb{X} - \Delta x < x < \mathbb{X}]}$$

$$= \frac{\int_{-\infty}^{Y} \int_{\mathbb{X}-\Delta x}^{X} p(x,y)\,dx\,dy}{\int_{\mathbb{X}-\Delta x}^{X} p(x)\,dx}$$

NOTE: TO FIND pdf, WE MUST REQUIRE THAT $p(x,y)$ BE DIFFERENTIABLE

NOW WE REWRITE THE CONDITIONAL PDF AS

$$P[y \leq \mathbb{Y}/x] = \frac{\int_{-\infty}^{\mathbb{Y}} p(\mathbb{X},y)\,dy \cdot \Delta x}{p(\mathbb{X})\Delta x}$$

$$= \int_{-\infty}^{Y} p[\mathbb{X},y]\,dy \Big/ p[\mathbb{X}]$$

DIFFERENTIATING W.R.T. Y:

$$p(y/x) = p[\mathbb{X},\mathbb{Y}] \big/ p(\mathbb{X})$$

ALSO, AS BEFORE $p(\mathbb{Y}/\mathbb{X}) \geq 0$

$$P[a < y < b / \mathbb{X}] = \int_a^b p(y/\mathbb{X})\,dy$$

& OF COURSE

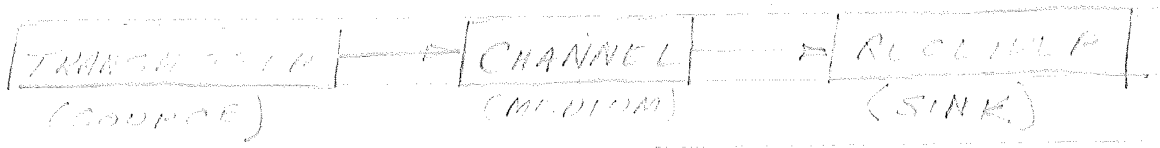
$$\int_{-\infty}^{\infty} p(y/\mathbb{X})\,dy = 1$$

ON TO INFO. THEORY:

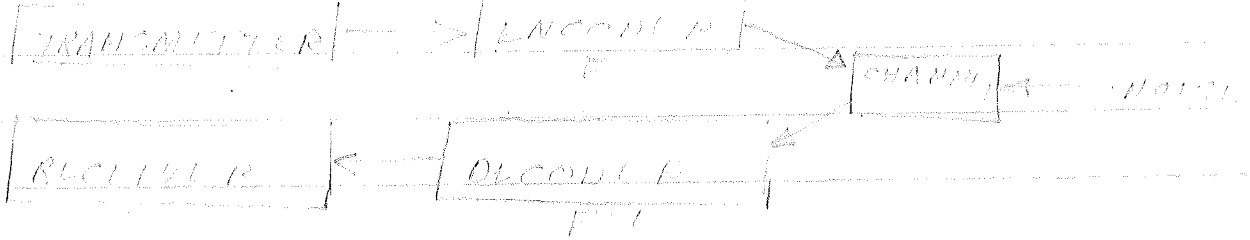
OF INTEREST TO COMPUTATION ! NSR.

COMMUNICATION PROCESS: FLOW OF SOME INFORMATION BEARING COMMODITY.

A SIMPLE MODEL IS:



A MORE GENERAL MODEL IS



ENCODER PERFORMS A 1 TO 1 MAPPING F(I).

DECODER, F^{-1}, GIVES $F^{-1}F(I) = I$

THE SOURCE MAY SELECT MESSAGE ON A PROBABALISTIC BASIS.

WE SEEK TO ASK AND ANSWER THE QUESTIONS:

(1) IF THE SOURCE IS NON DETERMINISTIC, HOW DOES ONE MEASURE INFORMATION AND WHAT UNIT SHALL WE PRESCRIBE TO IT.

(2) HAVING DEFINED THIS UNIT WHAT IS THE RATE AT WHICH THE SOURCE SUPPLIES INFORMATION.

(3) WHAT IS A CHANNEL? HOW DO WE CHARACTERIZE IT.

(4) GIVEN A SOURCE & CHANNEL, WHAT IS COMBINED RATE OF INFO. TRANS.

(5) WHAT EFFECT DOES "NOISE" HAVE ON THE PERFORMANCE OF THE CHANNEL?
(ANSWERED BY SHANNON'S THEOREM)

EXAMPLE:

A CATALOG WITH n MODELS DESIGNATED $(x_1, x_2, x_3, \dots, x_n)$ IN m COLORS $(c_1, c_2, \dots, c_m)$. THE TOTAL # OF ARTICLES = $m \times n$.

A VERY KEY STATEMENT: THE DESIRED AMOUNT OF INF $I(x_k)$ ASSOCIATED WITH THE SELECTION OF A PARTICULAR MODEL x_k SHOULD BE SOME FUNCTION OF THE PROB. WITH WHICH THAT MODEL x_k "OCCURS" IN THE CATALOG. [RECOGNIZED BY HARTLEY (1928)]

i.e., $I(x_k) = f[P(x_k)]$

IF ALL MODELS ARE EQUALLY LIKELY, THEN $P(x_k) = \frac{1}{n} \; \forall n \Rightarrow I_1(x_k) = f(1/n)$

NEXT, EXECUTE A CHOICE W.R.T. COLOR

$I_2(c_j) = f(1/m)$

IF THE CHOICE OF MODEL & COLOR IS EXECUTED INDEPENDENTLY, THEN THE INFORMATION ASSOCIATED WITH A SPECIFIC COLOR IS: $I_1(x_k) + I_2(c_j) = f(\frac{1}{n}) + f(\frac{1}{m})$

BUT WHAT IS THE INFO ASSOCIATED WITH EXTRACTION x_k AND c_j: $f(1/nm)$. CLEARLY:

$$f\left(\frac{1}{nm}\right) = f\left(\frac{1}{n}\right) + f\left(\frac{1}{m}\right)$$

THERE ARE MANY FUNCTIONS $f(\cdot)$ WHICH WILL SATISFY

$$f\left(\tfrac{1}{nm}\right) = f\left(\tfrac{1}{n}\right) + f\left(\tfrac{1}{m}\right)$$

a. LET $f = -\log P[X_k] = -\log \frac{1}{n} = \log n$

b. f = # OF FACTORS IN THE DECOMPOSITION OF $\frac{1}{x}$ IN PRODUCT OF PRIMES (w/o 1)

EXAMPLE : $n = 18 = 2 \cdot 3 \cdot 3 \Rightarrow f\left(\tfrac{1}{18}\right) = 3$

$n = 8 = 2 \cdot 2 \cdot 2 \Rightarrow f\left(\tfrac{1}{8}\right) = 3$

$\therefore f\left(\tfrac{1}{144}\right) = 6$

7/19/76

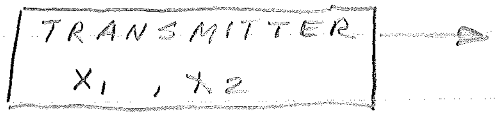
$-\log p_1 =$

SELF INFORMATION OF THE SYMBOL X, WHOSE PROB OF OCCURENCE WAS p_1

= UNCERTAINTY ASSOCIATED WITH x_1

= THE AMMOUNT OF INF. ASSOCIATED WITH THE SELECTION OF X_1

IF, FOR A GIVEN SOURCE, THERE ARE ONLY 2 SYMBOLS X_1, X_2 EACH EQUALLY LIKELY:



$$p(x_1) = \tfrac{1}{2} \rightarrow -\log\tfrac{1}{2} = \log 2$$
$$= \log_2 2 = 1 \text{ BIT OF INFO.}$$

$$p(x_2) = 1 \text{ BIT}$$

THE AVERAGE UNCERTAINTY ASSOCIATED WITH THE ENTIRE SOURCE = AVE $[-\log p_i]$, NOTING THAT EACH SYMBOL x_i OCCUR WITH PROB p_i

ASIDE: EXPECTED VALUES (AVERAGE)

CONSIDER $X_1, X_2, \ldots, X_n$

$P(x_1), P(x_2), \ldots, P(x_n)$

THEN $\overline{X} \triangleq \sum_{i=1}^{n} x_i \, p(x_i)$

= STATISTICAL AVERAGE

$$E[-\log p_i] = -\sum_{i=1}^{n} p_i \log p_i$$

$= \text{ENTROPY} = \text{AVERAGE INFO}$

$= H(p_1, p_2, \ldots, p_n) \geq 0$

USING DEF: $0 < p_i \leq 1 \ \forall i$ (ie $(0,1]$)

CONSIDER

$$\boxed{\begin{array}{l} x_1, x_2, \ldots\ldots x_{10} \\ 0, 1, 2, 3 \ldots \end{array}} \qquad p(x_i) = \frac{1}{10}$$

$$I = -\log \frac{1}{10}$$

$$= \log_{10} 10$$

$$= 1 \text{ HARTLEY}$$

$$\begin{cases} \text{BASE } 2 \Longrightarrow \text{BIT} \\ \text{BASE } 10 \Longrightarrow \text{HARTLEY} \\ \text{BASE } e \Longrightarrow \text{NAT} \end{cases}$$

1 HARTLEY = 3.32 BITS

1 NAT = 1.44 BITS ETC

THIS IS FROM:

$\log_a x = \log_b x / \log_b a$ ← TEST KNOW

CONSIDER

$S_1 : [A_1, A_2] \Rightarrow \left(\frac{1}{256}, \frac{255}{256}\right)$

$S_2 : [B_1, B_2] \Rightarrow \left(\frac{1}{2}, \frac{1}{2}\right)$

$S_3 : [C_1, C_2] \Rightarrow \left(\frac{7}{16}, \frac{9}{16}\right)$

DEFN: IF $\sum_{1} p_i = 1$

$\ni$ # OF SYMBOLS IS FINITE

THEN SOURCE IS CALLED

"A DISCRETE COMPLETE SOURCE"

$$H_1(\quad) = \frac{1}{256}\log 256 + \frac{255}{256}\log\frac{256}{255}$$
$$= \frac{1}{256}\log 2^8 + \frac{255}{256}\left[\log 2^8 - \log 255\right.$$
$$= \frac{8}{256} + \frac{255}{256}\left[8 - 7.994353\right]$$
$$= \overset{\text{BITS}}{0.0369} \ll 1$$

$$H_2(\quad) = 1 \text{ BIT}$$
$$H_3(\quad) = \frac{7}{16}\log\frac{16}{7} + \frac{9}{16}\log\frac{16}{9}$$
$$= 0.989 \text{ BITS.}$$

THE MATHEMATICS OF H

THE REQUIREMENTS ON H

(1) CONTINUITY WITH REGARD TO p_i

(2) SYMMETRY i.e. $H(p_1, p_2, \ldots, p_k, \ldots p_n) = P(p_3, p_2, p_1, \ldots p_k, \ldots p_n, p_{n-1})$

(3) EXTREMAL PROPERTY

$H(p_1, p_2, \ldots, p_n)$ IS MAXIMUM WHEN $p_i = \frac{1}{n} \forall i$

(4) IF ONE OF THE EVENTS IS DISECTED INTO m SUBEVENTS, THE RESULTING UNCERTAINTY SHOULD BE LARGER THAN THE ORIGINAL AVERAGE.

IN FACT:

$$H(p_1, p_2, \ldots, p_{n-1}; q_1, q_2, \ldots q_m)$$
$$= H(p_1, p_2, \ldots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \ldots, \frac{q_m}{p_n}\right)$$

CHECK OF SHANNON'S SUGGESTED ENTROPY FUNCTION

(1) $H(p_1, p_2, \dots p_n) = -\sum_{k=1}^{n} p_k \ln p_k$

(2) SYMMETRY (OBVIOUS)

(3) MAXIMALITY (OR EXTREMALITY) OF H

$$\frac{dH}{\delta p_k} = \sum_{i=1}^{n} \frac{\delta H}{\delta p_i} \frac{\delta p_i}{\delta p_k}$$

$$= -\frac{d}{dp_k} p_k \log p_k \cdot \frac{dp_k}{dp_k}$$

$$-\frac{1}{dp_n}(p_n \log p_n) \frac{dp_k}{dp_n} + 0$$

ALSO, DUE TO COMPLETENESS,

$$p_n = 1 - (p_1 + p_2 + \dots + p_{n-1})$$

$$\therefore \frac{dH}{dp_k} = -(\log_2 e + \log p_k)$$

$$+ (\log_2 e + \log p_n)$$

IF THE ORIGINAL LOG IS BASE e

$$\frac{dH}{dp_k} = \log p_n - \log p_k = 0 \quad \text{FOR MAX OR MIN}$$

$$\Rightarrow p_n = p_k$$

FOR MAXIMALITY, CONSIDER $H[1, 0, 0 \dots 0] = 0$

$\therefore$ HERE $H\left[\frac{1}{n}, \frac{1}{n}, \dots \frac{1}{n}\right]$ IS MAX

(4) $H[p_1, p_2, \dots, p_{n-1}, q_1, q_2, \dots q_m]$

$$= -[p_1 \ln p_1 + \dots + p_{n-1} \ln p_{n-1}]$$

$$-[q_1 \ln q_1 + \dots + q_m \ln q_m]$$

$$= -\sum_{i=1}^{n-1} p_i \ln p_i - \sum_{k=1}^{m} q_k \ln q_k$$

$$= -\sum_{i=1}^{n} p_i \ln p_i + p_n \ln p_n - \sum_{k=1}^{m} q_k \ln q_k$$

$$= H(p_1, p_2, \dots, p_n) + p_n \sum_{k=1}^{m} \frac{q_k}{p_n} \ln p_n - p_n \sum_{k=1}^{m} \frac{q_k}{p_n} \ln q_k$$

$$= H(p_1, p_2, \dots, p_n) + p_n \sum_{k=1}^{m} \frac{q_k}{p_n} \ln\left(\frac{p_n}{q_k}\right)$$

$$= H(p_1, p_2, \dots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \dots, \frac{q_k}{p_n}\right)$$

= ADDITIVITY PROPERTY

ET.

9-20-76 (TUES)

ADDITIVITY PROPERTY

①: $X : \{x_1, x_2, x_3)$ $P: (\frac{1}{5}, \frac{4}{15}, \frac{8}{15})$, $\Sigma P = 1$

A DISCRETE & COMPLETE SCHEME

WE WILL LOOK AT

1) $H(\frac{1}{5}, \frac{4}{15}, \frac{8}{15})$

2) $X: \{x_1, x_2 \cup x_3)$ & $P(\frac{1}{5}, \frac{12}{15})$

3) $X: \{ \frac{x_2}{x_2 \cup x_3}, \frac{x_3}{x_2 \cup x_3} \}$

$P: \{ \frac{4/15}{12/15}, \frac{8/15}{12/12} \} \Rightarrow (\frac{1}{3}, \frac{2}{3})$

QUESTION

RELATE $H(\frac{1}{5}, \frac{4}{5}, \frac{8}{15})$ TO

$H(\frac{1}{5}, \frac{4}{5})$ & $H(\frac{1}{3}, \frac{2}{3})$

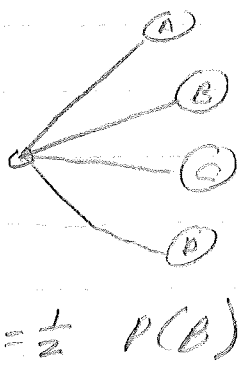
$H(\frac{1}{5}, \frac{4}{15}, \frac{8}{15}) = \frac{1}{5} \ln 5 + \frac{4}{15} \ln \frac{15}{4} + \frac{8}{15} \ln \frac{15}{8}$

$H(\frac{1}{5}, \frac{4}{5}) = \frac{1}{5} \ln 5 + \frac{4}{5} \ln 5/4$

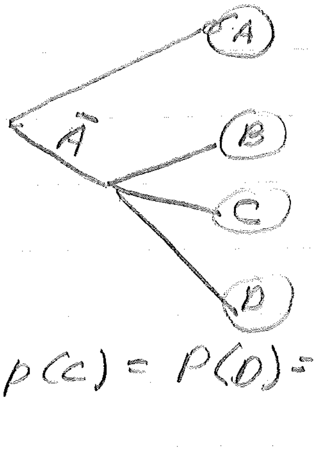
ETC

⟹ ② DEMONSTRATE THE FACT THAT THE AVERAGE UNCERTAINTY OF A SYSTEM IS NOT AFFECTED BY THE ARRANGEMENT OF THE EVENTS AS LONG AS THE INDIVIDUAL PROBS. ARE UNAFFECTED.

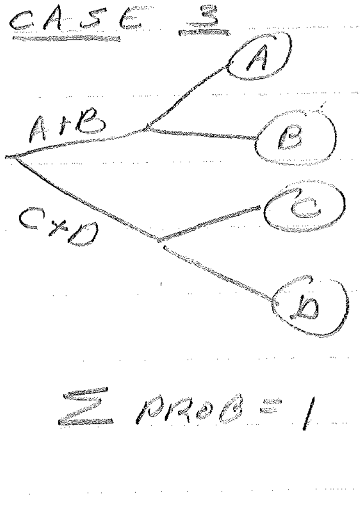
CASE 1



CASE 2



CASE 3



$P(A) = \frac{1}{2}$ $P(B) = \frac{1}{4}$ $p(C) = P(D) = \frac{1}{2}$ Σ PROB $= 1$

FOR CASE 1:

$$H(\ ) = \frac{1}{2}\ln 2 + \frac{1}{4}\ln 4 + \frac{1}{8}\ln 8 + \frac{1}{8}\ln 8$$
$$= \left[\frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{3}{8}\right] \text{BITS}$$
$$= 1\tfrac{3}{4} \text{ BITS}$$

FOR CASE 2

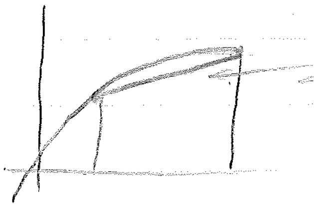
$$H_2(\ ) = -p(A)\ln p(A) - [1-P(A)]\ln[1-P(A)]$$
$$- [1-P(A)]\left[\frac{P(B)}{1-P(A)}\ln\frac{P(B)}{1-P(A)} + \frac{P(C)}{1-P(A)}\ln\frac{P(C)}{1-P(A)} + \frac{P(D)}{1-P(A)}\ln\frac{P(D)}{1-P(A)}\right]$$
$$= \frac{1}{2}\ln 2 + \frac{1}{2}\ln 2$$
$$+ \frac{1}{2}\left[\frac{1}{4}\ln 4 + \frac{1}{8}\ln 8 + \frac{1}{8}\ln 8\right]$$
$$= 1 + \frac{1}{2}\left[\frac{1}{2} + \frac{6}{8}\right]$$
$$= 1\tfrac{3}{4} \text{ BITS}$$

FOR CASE 3

$$H_3' = -(P_A+P_B)\ln(P_A+P_B) - (P_C+P_D)\ln(P_C+P_D)$$
$$+ (P_A+P_B)\left[\frac{-P_A}{P_A+P_B}\ln\frac{P_B}{P_A+P_B} - \frac{P_B}{P_A+P_B}\ln\frac{P_B}{P_A+P_B}\right]$$
$$+ (P_C+P_D)\left[\frac{-P_C}{P_C+P_d}\ln\frac{P_C}{P_C+P_D} - \frac{P_D}{P_C+P_D}\ln\frac{P_D}{P_C+P_d}\right]$$
$$= 1\,3/4 \text{ BITS}$$

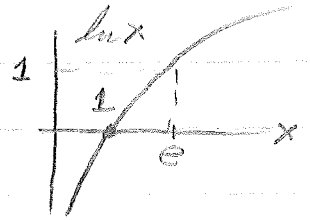
** LEMMAS PERTAINING TO H FUNCTION

LEMMA 1: THE LOG FUNCTION IS A CONVEX (UPWARD) FUNCTION:



CONVEX IF CORD IS ALWAYS BELOW CURVE

CONSIDER:



A NECESSARY & SUFFICIENT CONDITION FOR A CONVEX FUNCTION IS THAT

$$d^2y/dx^2 \le 0$$

NOW $\frac{d}{dx}\ln x = \frac{1}{x}$

$$\frac{d^2}{dx^2}\ln x = -\frac{1}{x^2} \; ; \; 0 < x < \infty$$

ALSO, FOR ANY CONVEX FUNCTION:

$$\frac{1}{2}\left[f(x_1) + f(x_2)\right] \le f\left(\frac{x_1+x_2}{2}\right)$$

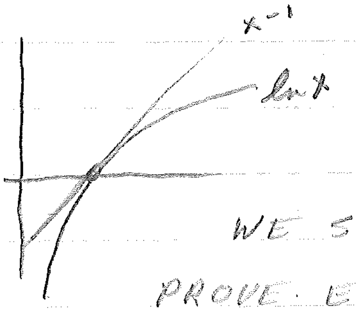
ASSUME $x_1, x_2 > 0$ SO THAT

$$\frac{1}{2}\left[\ln x_1 + \ln x_2\right] \overset{?}{\le} \ln\left(\frac{x_1+x_2}{2}\right)$$

$$\ln\sqrt{x_1x_2} \overset{?}{\le} \ln\frac{x_1+x_2}{2}$$

$$\underbrace{\sqrt{x_1x_2}}_{\text{GEOMETRIC MEAN}} \le \underbrace{\frac{x_1+x_2}{2}}_{\text{ARITHMATIC MEAN}} \leftarrow \text{TRUE F} \; x_1, x_2 > 0$$

LEMMA 2: $\ln x \le x-1$



OR

$\ln x \ln_2 e$

$= \ln_2 x \le x-1 \ln_2 e$

WE SHALL USE THIS LEMMA TO PROVE EXTREMAL PROPERTY OF H

LET $X = \{x_1, x_2, \dots, x_m\}$

$P(x_1) \ne P(x_2) \ne \dots \ne P(x_m)$

WE WISH TO PROVE

$$H(x) \le -m\left(\frac{1}{m}\ln\frac{1}{m}\right)$$

$$\le \ln\frac{1}{m}$$

BE DEFINITION

$$H(x) - \ln m = \sum_{i=1}^{m} p_i \ln\frac{1}{p_i} + \ln\frac{1}{m}$$

$$= \sum_{i=1}^{m} p_i \ln\frac{1}{p_i} + \Big(\overbrace{\sum_{i=1}^{m} p_i}^{=1}\Big)\ln\frac{1}{m}$$

$$= \sum_{i=1}^{m} p_i \ln\frac{1}{m p_i}$$

$$\le \sum_{i=1}^{m} p_i\left(\frac{1}{m p_i} - 1\right)\ln_2 e \quad \left(\text{BY LEMMA 2}\right)$$

$$\le \sum_{i=1}^{m}\left(\frac{1}{m} - p_i\right) \le 0$$

HENCE

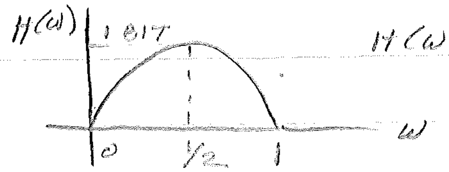
$H(x) \le \ln m \Rightarrow$ EXTREMA PROPERTY 3

$H(S)$: $S(x_1, x_2)$ $p:(w, \bar{w})$

$H(S) = -(w \ln w + \bar{w} \ln \bar{w}) \Leftarrow$ FIXED FOR FIXED w

$H(w)$ = ENTROPY FUNCTION

<u>COR</u>: THE ENTROPY FUNCTION IS CONVEX



$H(\omega) = -(\omega \ln \omega + \bar{\omega} \ln \bar{\omega})$

⇒ CONSIDER 3 SOURCES EACH WITH 2 SYMBOLS

$S_1(A,B)$ $\quad P_1 = (\frac{1}{3}, \frac{2}{3})$

$S_2(C,D)$ $\quad P_2 : (\frac{1}{4}, \frac{3}{4})$

$S_3(E,F)$ $\quad P_3 : (\frac{7}{24}, \frac{17}{24})$ ⟸ AVE OF S_1 & S_2

SHOW THAT

$$\frac{1}{2}[H(X_1) + H(X_2)] \le H\left(\frac{X_1 + X_2}{2}\right)$$

$$H(X_1) = \frac{1}{3}\ln 3 + \frac{2}{3}[\ln 3 - \ln 2] = 0.88$$

$$H(X_2) = \frac{1}{4}\ln 4 + \frac{3}{4}(\ln 4 - \ln 3) = 0.81$$

$$H(X_3) = \frac{7}{24}\ln\frac{24}{7} + \frac{17}{24}\ln\frac{24}{17} =$$

LEMMA 3: LET $X_1, X_2, \ldots X_q$ & $Y_1, Y_2, \ldots Y_q$ BE TWO SETS OF COMPLETE SCHEMES.

ie $\sum_1^q X_i = \sum_{i=1}^q Y_i = 1$

THEN

$$\sum_{i=1}^q X_i \ln \frac{1}{X_i} \le \sum_{i=1}^q X_i \ln 1/Y_i$$

NOTICE THAT $\log_a X = \log_b X / \log_b a$ (H.W).

$\therefore \sum_{i=1}^q X_i \ln_2 \frac{Y_i}{X_i} = \frac{1}{\ln 2} \sum_{i=1}^q X_i \ln Y_i/X_i$

ALSO $X_i \ln Y_i/X_i \le X_i/\ln_2 (Y_i/X_i - 1)$ (FROM LEMMA)

$$\Rightarrow \sum_i X_i \ln Y_i/X_i \le \frac{1}{\ln 2}\sum X_i (Y_i/X_i - 1)$$

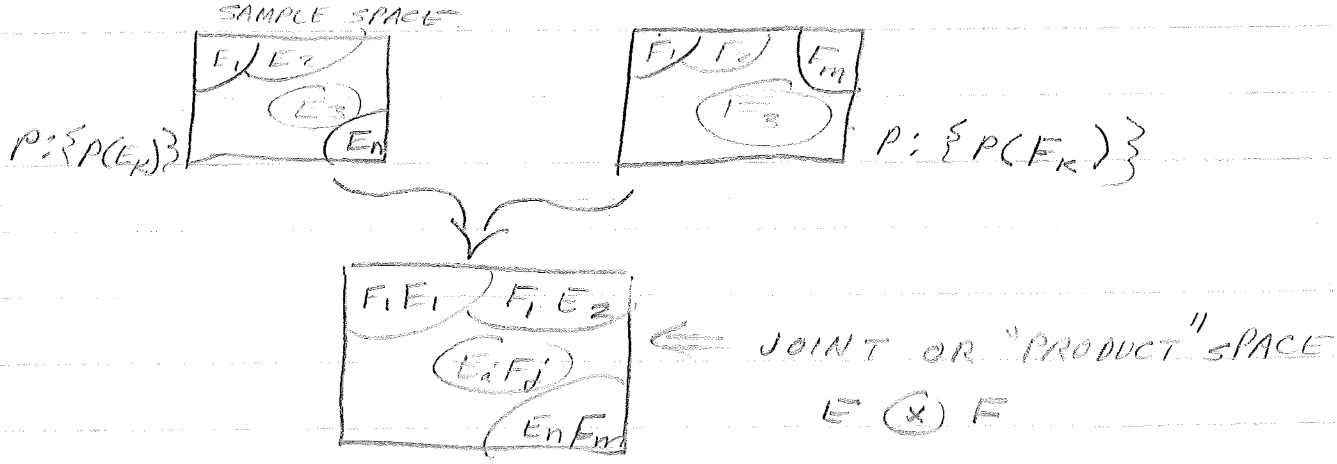
$$\le \frac{1}{\ln 2}\left[\sum_i Y_i - \sum X_i\right] = 0$$

$\therefore \sum_{i=1}^q X_i \ln_2 Y_i/X_i \le 0$

OR

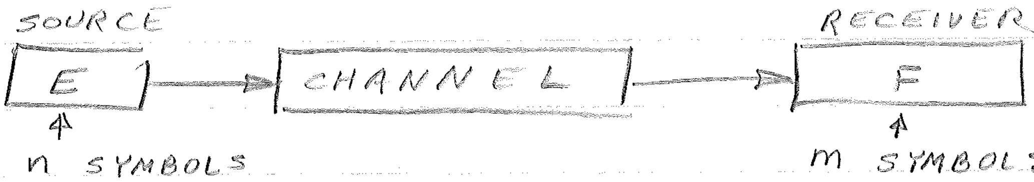
$$\sum_i X_i \ln \frac{1}{X_i} \le \sum_i X_i \ln \frac{1}{Y_i}$$

LEMMA 4: RELATES TO JOINT ENTROPY FUNCTIONS



$$P: \begin{bmatrix} E_1F_1 & E_1F_2 & \cdots & E_1F_m \\ E_2F_1 & & & \\ \vdots & & & \\ E_nF_1 & \cdots & & E_nF_m \end{bmatrix} \Longleftarrow$$

"JOINT PROBABILITY MATRIX"



IT IS CLEAR THAT

$$P[E_1] = P[E_1F_1 \cup E_1F_2 \cup \ldots \cup E_1F_m]$$

$$= \sum_{i=1}^{m} P[E_1 F_i]$$

$$P[F_2] = \sum_{i=1}^{n} F[E_i, F_2]$$

} SAME AS MARGINAL PROB

ON THIS BASIS, IT IS REASONABLE TO DEFINE TH

JOINT ENTROPY OF $\{EF\}$ SPACE AS

$$H(X,Y) \overset{\Delta}{=} -\sum_i \sum_j P(x_i, y_j) \ln p(x_i, y_j)$$

OVER THE ENTIRE PLANE

7/21/76 (WED)

$$H(X,Y) = -\sum_{i=1}^{M} \sum_{j=1}^{L} p(x_i, y_j) \ln p(x_i, y_i)$$

$X = \{x_1, x_2, \ldots, x_M\}$

$Y = \{y_1, y_2, \ldots, y_L\}$

AND $p(x_i, y_j) \triangleq P[X_i = x_i, Y = y_i]$, $i = 1, 2, \ldots M$

$j = 1, 2, \ldots L$

THUS, THERE ARE ML OUTCOMES OF INTEREST

LEMMA 4 (WILL CONNECT THE INDIVIDUAL ENTROPIES $H(X)$, $H(Y)$ WITH JOINT $H(X,Y)$ AS DEFINED ABOVE): $H(X,Y) \le H(X) + H(Y)$

NOTE THAT $p(x_i) = \sum_{j=1}^{L} p(x_i, y_j)$

$p(y_j) = \sum_{i=1}^{M} p(x_i, y_j)$

$$\therefore H(X) = -\sum_{i=1}^{M} p(x_i) \ln p(x_i)$$

$$= -\sum_{i=1}^{M} \sum_{j=1}^{L} p(x_i, y_i) \ln p(x_i)$$

$$H(Y) = -\sum_{j=1}^{L} p(y_j) \ln p(y_j)$$

$$= -\sum_{i=1}^{M} \sum_{j=1}^{L} p(x_i, y_j) \ln p(y_j)$$

$$H(X) + H(Y) = -\sum_{i=1}^{M} \sum_{j=1}^{L} p(x_i, y_j) \ln p(x_i) p(y_j)$$

$$= -\sum_{i=1}^{M} \sum_{j=1}^{L} p(x_i, y_j) \ln q_{ij} \ni q_{ij} = p(x_i) p(y_j)$$

$\ni \sum q_{ij} = 1$

ALSO $H(X,Y) = -\sum_{i=1}^{M}\sum_{j=1}^{L} p_{ij} \ln p_{ij}$

$\ni \; p_{ij} \triangleq p(x_i, y_j)$

$$
\begin{aligned}
\therefore \; H(X,Y) &\le -\sum_{i=1}^{M}\sum_{j=1}^{L} p_{ij} \ln q_{ij} \\
& \qquad\quad \shortparallel \\
&\le -\sum_{i=1}^{M}\sum_{j=1}^{L} p_{ij} \left[\ln p(x_i) + \ln p(y_j) \right] \\
& \qquad\quad \shortparallel \\
&\le -\sum_{i=1}^{L} p(x_i) \ln p(x_i) - \sum_{j=1}^{M} p(y_j) \ln p(y_j) \\
& \qquad\quad \shortparallel \\
&\le H(X) + H(Y))
\end{aligned}
$$

WHICH WAS TO BE SHOWN. EQUALITY WILL OCCUR WHEN X & Y ARE STATISTICALLY INDEPEND.

COR. 1: $H(X_1, X_2, \ldots X_n)$
$\le H(X_1) + H(X_2) + \ldots + H(X_n)$

EQUALITY PREVAILING WHEN X_i'S ARE STATISTICALLY INDEPENDENT.

COR. 2: $H(X_1, X_2, \ldots X_n, Y_1, Y_2, \ldots Y_m)$
$\le H(X_1, X_2, \ldots X_n) + H(Y_1, Y_2, \ldots, Y_m)$

EQUALITY PREVAILING WHEN THE RANDOM VECTOR $(X_1, X_2, \ldots X_N)$ IS STATISTICALLY IND. OF THE RANDOM VECTOR $(Y_1, Y_2, \ldots, Y_m)$

* CONDITIONAL ENTROPY

CONSIDER $H[Y/X=x_i]$:

$$H[Y/X=x_i] = -\sum_{j=1}^{L} p(Y_j/X_i) \ln p(Y_j/X_i)$$

NEXT, WE DEFINE

$$\begin{aligned} H[Y/X] &= \overline{H[Y/X=x_i]} \\ &= p(X_1)H(Y/X=X_1) + p(X_2)H(Y/X=X_2) \\ &\quad + \dots + p(X_M)H(Y/X_M) \\ &= \sum_{i=1}^{M} p(X_i)\, H(Y/X_i) \\ &= -\sum_{i=1}^{M} \sum_{j=1}^{L} p(Y_j/X_i)p(X_i) \ln p(Y_i/X_i) \\ &= -\sum_{i=1}^{M} \sum_{j=1}^{L} p(X_i, Y_i) \ln p(Y_j/X_i) \end{aligned}$$

AS AN EXTENSION, MAKE SURE THAT THE FOLLOWING IS TRUE

(1) $H(Y,Z/X) = -\sum_{ijk} p(X_i,Y_j,Z_k) \ln p(Y_j,Z_k/X_i)$

(2) $H(Z/X,Y) = -\sum_{ijk} p(X_i,Y_j,Z_k) \ln p(Z_k/X_i,Y_j)$

⇒ (3) $H(Y_1 \dots Y_m / X_1, X_2 \dots X_n)$ = HOMEWORK)

LEMMA 5: (RELATES $H(X)$ $H(Y)$, $H(X,Y)$, $H(Y/X)$, $H(X,Y)$). WE SHALL SHOW

$$H(X,Y) = H(X) + H(Y/X)$$
$$= H(Y) + H(X/Y)$$

CONSIDER

$$H(X,Y) = -\sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i, y_j) \ln p(x_i, y_j)$$
$$= -\sum_i \sum_j p(x_i, y_j) \ln p(x_i) p(y_j/x_i)$$
$$= -\sum_i \sum_j p(x_i, y_j) \ln p(x_i)$$
$$- \sum_i \sum_j p(x_i, y_j) \ln p(y_j/x_i)$$
$$= H(X) + H(X/Y)]$$ END!

(HOMEWORK: @ HOME, SHOW $H(Y) + H(X/Y) = H(X,Y)$)

AS AN EXTENSION,

$$H(X,Y,Z) = H(X) + H(Y/X) + H(Z/X,Y)$$
$$= H(X) + H((Y,Z)/X)$$

⟹ IN GENERAL (HOMEWORK):

$$H(X_1, X_2, \ldots X_n, Y_1, Y_2, \ldots Y_m)$$
$$= H(X_1, X_2, \ldots X_n) + H(\quad ? \quad)$$

LEMMA 6: (RELATES THE CONDITIONAL $H(Y/X)$ TO $H(Y)$)

$H(Y/X) \leq H(Y)$ WITH EQUALITY WHEN X & Y ARE INDEPENDENT.

FROM BEFORE:

$$H(X,Y) = H(X) + H(Y/X) \leftarrow \text{LEMMA 5}$$
$$\leq H(X) + H(Y) \leftarrow \text{LEMMA 4}$$
$$\therefore \quad H(Y/X) \leq H(Y)$$

⟹ HOMEWORK

⟹* ✓ TEXT 2-3 p41 / 2-4* p42 / 2-14 p44

1. FROM ENGLE'S BOOK

LET $X = \{x_1, x_2)$ $\quad p: (\frac{1}{4}, \frac{3}{4})$

LET $Y = \{Y_1, Y_2, Y_3)$

$$\Rightarrow \begin{array}{l|l} p(Y_1/X_1) = \frac{1}{4} & p(Y_1/X_2) = \frac{1}{10} \\ p(Y_2/X_1) = 0.35 & p(Y_2/X_2) = \frac{7}{10} \\ p(Y_3/X_1) = {}^4/_{10} & p(Y_3/X_2) = \frac{2}{10} \end{array}$$

FIND $H(X)$, $H(Y)$, $H(X,Y)$, $H(Y/X)$, $H(X/Y)$ AND VERIFY THE RELATIONS FROM LEMMA'S 4, 5, 6.

ASH 1.3. SUPPOSE THAT IN A CERTAIN CITY $\frac{3}{4}$ OF THE HIGH SCHOOL PASS & $\frac{1}{4}$ FAIL. OF THOSE WHO PASS, 10% OWN CARS, WHILE 50% OF THE FAILING STUDENTS OWN CARS. ALL OF THE CAR OWNING STUDENTS BELONG TO FRATS, WHILE 40% OF THOSE

WHO DO NOT OWN CARS BUT PASSED, AS WELL 40% OF THOSE WHO DO NOT OWN CARS BUT FAIL, BELONG TO FRATS

(a) HOW MUCH INFO IS CONVEYED ABOUT A STUDENT'S ACADEMIC STANDING BY SPECIFYING WHETHER OR NOT HE OWNS A CAR.

(b) ... BY SPECIFYING WHETHER OR NOT HE BELONGS TO A FRAT.

(c) IF A STUDENT'S ACADEMIC STANDING, CAR OWNING STATUS, & FRAT STATUS ARE TRANSMITTED BY 3 SUCCESSIVE BINARY DIGITS, HOW MUCH INFO IS CONVEYED BY EACH DIGIT.

ASH 1-4. ESTABLISH THE FOLLOWING

a) $H(Y,Z/X) \leq H(Y/X) + H(Z/X)$

WITH EQUALITY IFF

$p(Y_j, Z_k/X_i) = P(Y_i/X_i)\, p(Z_k/X_i)$

FOR ALL i, j, k

b) $H(Y,Z/X) = H(Y/X) + H(Z/X,Y)$

c) $H(Z/X,Y) \leq H(Z/X)$

WITH EQUALITY IFF

$p(y_j, z_k/x_i) = p(y_j/x_i)\, p(z_k/x_i)\ \forall\, i, j, k$

** AN EXTENDED SOURCE: SOURCE EXTENSION

$\boxed{0,1}$ ◄ IF WE CAN ONLY SEND A DIGIT FOR INFO, WE ONLY HAVE <u>2</u> MESSAGES.

FOR 2 @ A TIME (WORD LENGTH: 2) WE GET 2^2 MESSAGES

FOR WORD LENGTH = n, WE GET 2^n MESSAGES

QUESTION: WHAT IS ENTROPY OF THE n^{TH} EXTENSION OF A SOURCE S WHICH HAS q SYMBOLS

$$\begin{cases} S = [s_1, s_2, \ldots, s_q] \\ p: [p(s_1), p(s_2) \ldots p(s_q)] \end{cases}$$

$$\begin{cases} S^n = [\sigma_1, \sigma_2, \ldots, \sigma_{q^n}] \\ p: [p(\sigma_1), p(\sigma_2), \ldots, p(\sigma_{q^n})] \end{cases}$$

WE WILL SHOW

$H[S^n] = nH(S)$ IF SOURCE HAS <u>NO</u> MEMORY

IF THE SUCCESIVE SYMBOLS EMMITED FROM S^n ARE STATISTICALLY INDEPENDENT, THE INFORMATION SOURCE IS SAID TO BE A ZERO MEMORY SOURCE.

S WILL ALSO BE A ZERO MEMORY SOURCE.

7/22/76 (THURS)

$S: [s_1, s_2, \ldots, s_q]$

$S^n: [\sigma_1, \sigma_2, \ldots, s_{q^n}]$

$H(S) = -\sum_i P(s_i) \log P(s_i)$

$H^n(S) = +\sum_{S^n} p(\sigma_i) \ln 1/p(\sigma_i)$

$$\sum_{S^n} = \sum_{i=1}^{q^n} = \sum_{q^n}$$

$\sum_{S_n}$ = n SUMMATIONS EACH OVER S

SINCE THE SOURCE HAS NO MEMORY:

$P[\sigma_i] = P_{i_1} P_{i_2} P_{i_3} \ldots P_{i_n}$

∴ $i \Rightarrow i^{th}$ MESSAGE ; n = # SYMBOLS IN A MESSAGE

(1) IT NEEDS TO BE SHOWN THAT S^n IS COMPLETE

$\sum_{S_n} p(\sigma_i) = \sum_{S_n} p_{i_1} p_{i_2} p_{i_3} \cdots p_{i_n}$

$$= \sum_{i_1=1}^{q} p_{i_1} \sum_{i_2=1}^{q} p_{i_2} \cdots \sum_{i_n=1}^{q} p_{i_n} = 1$$

(2) $H^n(S) = \sum_{S_n} p(\sigma_i) \log 1/p(\sigma_i)$

$$= \sum_{S_n} P(\sigma_i) \log \frac{1}{p_{i_1} p_{i_2} \cdots p_{i_n}}$$

$$= \sum_{S_n} P(\sigma_i) \log \frac{1}{p_{i_1}} + \sum_{S_n} p(\sigma_i) \log \frac{1}{p_{i_2}} + \ldots + \sum_{S_n} P(\sigma_i) \log \frac{1}{p_{i_n}}$$

CONSIDER THE FIRST TERM

$\sum_{S_n} P(\sigma_i) \log 1/p_{i_1}$

$$= \sum_{S_n} p_{i_1} p_{i_2} \ldots p_{i_n} \log 1/p_{i_1}$$

$$= \sum_{i_1=1}^{q} p_{i_1} \log \frac{1}{p_{i_1}} \sum_{i_2=1}^{q} p_{i_2} \sum_{i_3=1}^{q} p_{i_3} \cdots \sum_{i_n=1}^{q} p_{i_n}$$

$$= \sum_{i_1=1}^{q} p_{i_1} \log \frac{1}{p_{i_1}} = H(S)$$

CLEARLY $H^n(S) = n H(S)$

EXAMPLE (FROM BOOK)

$S: (s_1, s_2, s_3) \Rightarrow P = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$

$$\begin{cases} S^2: (s_1 s_1, s_1 s_2, s_1 s_3, s_2 s_1, s_2 s_2, s_2 s_3, s_3 s_1, s_3 s_2, s_3 s_3) \\ P: (\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}) \end{cases}$$

$$H^2 = \frac{1}{4} \log_2 4 + 4 \times \frac{1}{8} \log 8 + 4 \frac{1}{16} \log 16$$

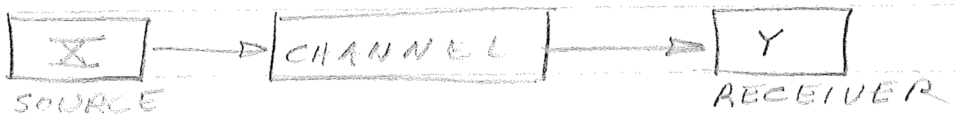
$$= \frac{1}{4} \times 2 + \frac{1}{2} \times 3 + \frac{1}{4} 4 = 3 \text{ BITS}$$

$H(S) = 1\frac{1}{2}$ BITS

⇒ HOMEWORK: TRY $n = 3$

CHANNELS

$H(X)$ = AVERAGE INFORMATION PER CHARACTER (SYMBOL) AT THE SOURCE



$H(Y)$ = AVE INF. PER CHARACTER @ RECEIVER

$H(X,Y)$ = AVE INF. PER PAIR OF CHARACTERS i.e., AVE UNCERTAINTY OF THE COMMUNICATION SYSTEM AS A WHOLE

$H(Y/X) = \overline{H(Y/X = x_i)}$ = A MEASURE OF INFORMATION ABOUT THE RECEIVING PORT KNOWING WHAT WAS TRANSMITTED. IT IS MEASURE OF "NOISE" OR "ERROR" IN THE CHANNEL

$H(X/Y) = \overline{H(X/Y = y_i)}$ = EQUIVOCATION OF THE CHANNEL & IS A MEASURE OF THE RECOVERY OF THE INPUT KNOWING THE OUTPUT

CHANNEL CHARACTERIZATION

(1) A SOURCE WITH A PDF

(2) A JOINT PROB. MATRIX, WHICH DESCRIBES THE INPUT-OUTPUT RELATIONS ON A PROBABALISTIC MEASURE

$$\begin{array}{ccccccc} & & Y_1 & Y_2 & \cdots & Y_m & \Leftarrow \text{RECEIVED} \\ \text{TRANSMITTED} & \left\{ \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ X_n \end{array} \right. & \begin{array}{c} p(x_1,Y_1) \\ p(x_2,Y_1) \\ \\ p(x_n,Y_1) \end{array} & \begin{array}{c} p(x_1,Y_2) \\ p(x_2,Y_2) \\ \\ p(x_n,Y_2) \end{array} & \begin{array}{c} \cdots \\ \\ \\ \cdots \end{array} & \begin{array}{c} p(x_1,Y_m) \\ p(x_2,Y_m) \\ \vdots \\ p(x_n,Y_m) \end{array} & \end{array}$$

THIS, THE JOINT PROBABILITY MATRIX, COMPLETELY DESCRIBES THE CHANNEL. FROM IT, WE CAN FIND $p(x_k)$ & $p(Y_k)$ & THUS $H(x)$ & $H(Y)$. CAN ALSO GET CONDITIONAL PROBABILITIES & MATRICES.

CONDITIONAL PROB MATRIX

$$\begin{array}{ccccc} & Y_1 & Y_2 & \cdots & Y_m \\ x_1 & p(x_1/Y_1) & & & \\ y_2 & \vdots & & & \\ \vdots & & & & \\ X_n & & & & \end{array}$$

⟹ HOMEWORK: WORK OUT CONDITIONAL MATRICES IN TERMS OF JOINT MATRIX.

TWO SIMPLE CHANNELS

(1) DISCRETE NOISE FREE CHANNEL

$$P(X,Y)\quad \begin{array}{c|cccc} & x_1 & x_2 & \cdots & x_n \\ \hline Y_1 & p(x_1,Y_1) & 0 & & 0 \\ Y_2 & 0 & p(x_2,Y_2) & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ Y_n & 0 & 0 & & p(x_n,Y_n) \end{array} \Leftarrow n \times n \text{ DIAGONAL MATRIX}$$

CONDITIONAL MATRIX WILL BE A UNIT MATRIX

ALSO $H(X,Y) = H(X) + H(Y)$

$H(X/Y) = H(Y/X) = 0$

(2) A DISCRETE CHANNEL WITH INDEPENDENT INPUT/OUTPUT

$$P(X,Y): \quad \begin{array}{c|cccc} & Y_1 & Y_2 & \cdots & Y_m \\ \hline x_1 & p_1 & p_1 & & p_1 \\ x_2 & p_2 & p_2 & & p_2 \\ \vdots & & & & \\ x_n & p_n & p_n & & p_n \end{array}$$

$$\sum_{i=1}^{m} p(x_i) = 1/m \qquad p(x_i) = mp_i$$

JOINT:

$p(x_i, Y_j) = p_{ij} = P(x_i)P(Y_j) = P_i$

CONDITIONAL: $p(x_i/Y_j) = p(x_i) = mp_i$

$p(Y_j/x_i) = P(Y_j) = 1/m$

ENTROPIES: $H(X,Y) = -\sum_i mp_i \ln p_i \Leftarrow$ FROM DOUBLE SUM

$H(X) = -\sum_{i=1}^{n} mp_i \ln mp_i = -m\sum_i p_i \ln p_i - \ln m$

$H(Y) = m \cdot \frac{1}{m} \ln m = \ln m$

$H(X/Y) = -\sum\sum p_i \ln mp_i = -m\sum p_i \ln mp_i = H(X)$

$H(Y/X) = H(Y)$

EXAMPLE: THE JOINT PROB. MATRIX WITH

$S:(x_1, x_2 \ldots X_5)$ $R:(Y_1, Y_2, Y_3, Y_4)$

	Y_1	Y_2	Y_3	Y_4	$P(x_i)$
X_1	1/4	0	0	0	0.25
X_2	1/10	3/10	0	0	.40
X_3	0	0.05	0.10	0	.15
X_4	0	0	0.05	0.1	.15
X_5	0	0	0.05	0	.05
	0.35	.35	.2	.1	Σ 1

CONDITIONAL $P[X/Y]$

	Y_1	Y_2	Y_3	Y_4
X_1	$\frac{.25}{.35}$	–	–	–
X_2	$\frac{10}{35}$	$\frac{30}{35}$	0	0
X_3	0	$\frac{5}{35}$	$\frac{1}{2}$	0
X_4	0	0	5/20	1
X_5	0	0	1/4	0

CONDITIONAL $P[Y/X]$

	X_1	X_2	X_3	X_4	X_5
Y_1	1				
Y_2					
Y_3					
Y_4					

→ HOMEWORK: FIGGER OUT

$H(X), H(Y), H(X,Y), H(X/Y), H(Y/X)$

& VERIFY $H(X,Y) < H(X) + H(Y)$

2.665 2.066 1.856

—

7/23/76 FIRST TEST

7/26/76 (MON)

⟹ HOMEWORK

1. PROB 5-4 p. 142
2. USING THE DEFN. OF MUTUAL INFORMATION, SHOW THAT

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$
$$= H(X) - H(X/Y)$$
$$= H(Y) - H(Y/X)$$

3. WILL BE ASSIGNED TUESDAY OR WED.

PART II OF OUR WORK

1. THE CHANNEL
2. "CLASSICAL" CODING TECHNIQUES

$$\xrightarrow{X:(x_1,x_2,\dots x_n)}\boxed{Ch: P[x_i, y_i]}\xrightarrow{Y:(y_1,y_2,\dots y_m)}$$

(1) LET US DEFINE A NEW ELEMENT, NAMELY MUTUAL INFORMATION:

$$I(x_i;y_j) \overset{\Delta}{=} \lg_2 p(x_i/y_i) - \lg_2 p(x_i)$$
$$= \lg\, p(x_i/y_i)/p(x_i)$$
$$= \lg\, p(x_i, y_i)/p(x_i)p(y_i)$$

COMPARE WITH SELF INFORMATION:

$$I(x_i) = -\lg p(x_i)$$

(2) THE A'PRIORI KNOWLEDGE THAT x_i IS BEING TRANSMITTED

= Σ PROB (x_i IS BEING XMITTED <u>AND</u> IS BEING RECEIVED AS <u>ANY</u> y_j) = $p(x_i)$

THE POSTERIOR KNOWLEDGE OF THE OBSERVER IS BASED ON THE CONDITIONAL PROB OF x_i BEING TRANSMITTED GIVEN THAT A PARTICULAR y_j IS RECEIVED = $p(x_i/y_j)$

THE DIFFERENCE (INFORMATION THEORY WISE) IS THE GAIN IN INFORMATION

= lg RATIO OF TWO PROB

$$= \lg\, p(x_i/y_i)/p(x_i)$$

IT CAN BE SHOWN

(1) $I(x_i;y_j)$ IS CONTINUOUS

(2) $I(x_i;y_j) = I(y_j;x_i)$ (SYMMETRY)

(3) SELF INFO: $I(x_i;x_i) = -\lg 1/p(x_i)$

$$\left\{\begin{aligned} I(x_i) &= I(x_i;x_i) \geq I(x_i;y_j) \\ I(y_j) &= I(y_j,y_j) \geq I(x_i;y_j) \end{aligned}\right.$$

(4) LET US AVERAGE THIS "GAIN" IN INFO OVER THE ENTIRE SET OF (x_i, y_j) PAIRS:

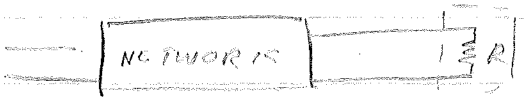
$$\begin{aligned} I(X;Y) &= \overline{I(x_i,y_j)} \\ &= \Sigma_i \Sigma_j \, p(x_i;y_j)\, I(x_i;y_j) \\ &= \Sigma_i \Sigma_j \, p(x_i,y_j) \lg p(x_i/y_j)/p(x_i) \end{aligned}$$

IF WE GO THRU THE SUMMATIONS, ONE WOULD SEE THAT

$$\left.\begin{aligned} I(X,Y) &= H(X)+H(Y)-H(X,Y) \\ &= H(X)-H(X/Y) \\ &= H(Y)-H(Y/X) \end{aligned}\right\} \text{p 108-110 OF TEXT}$$

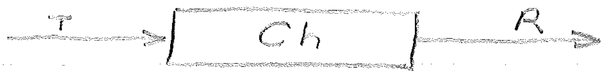
THUS, ON AN AVERAGE, THE OBSERVANCE OF ANY Y, PROVIDES US WITH $I(X;Y)$ BITS OF INF

CONSIDER:



TO MAX PWR WHEN R IS MATCHED

COMPARE WITH



WE WANNA MAXIMIZE $I(X,Y)$.

THE MAX OF $I(X,Y)$ IS CHANNEL CAPACITY

SHANNON'S DEFINITION OF CHANNEL CAPACITY

$$C \triangleq \mathrm{Max}\, I(X,Y)$$
$$= \mathrm{Max}\,[H(X) - H(X/Y)]$$

EXAMPLE (INDEPENDENT INPUT/OUTPUT)

$$\begin{array}{c|cc|c} & Y_1 & Y_2 & \\ \hline X_1 & 1/4 & 1/4 & 1/2 \\ X_2 & 1/4 & 1/4 & 1/2 \\ \hline & 1/2 & 1/2 & \end{array}$$

$H(X) = 1$ BIT
$H(Y) = 1$ BIT
$H(X,Y) = 2$ BITS

$$\therefore I(X,Y) = H(X) + H(Y) - H(X,Y)$$
$$= 0$$ (INTUITIVE! INPUT & OUTPUT ARE INDEPENDENT)

IF EACH SYMBOL TAKES t SEC TO PROPAGATE, THEN THE RATE OF INFORMATION TRANSMISSION OF INFO. CAN BE WRITTEN AS

$$C_t = C/t \text{ BITS/SEC}.$$

FOR A NOISE FREE SYSTEM:

$$C_t = C/t = \lg n/t \ \frac{\text{BITS}}{\text{SEC}}$$

IF THE CHANNEL IS NOISY, THE DIFFERENCE TWIXT MAX POSSIBLE VALUE OF $I(X;Y)$ (CHANNEL CAPACITY) & THE ACTUAL RATE IS CALLED THE ABSOLUTE

REDUNDANCY $= C - I(X,Y)$ → GENERAL
$= \lg n - H(X)$ → NOISE FREE

AND THE RELATIVE REDUNDANCY
$= [\lg n - H(X)]/\lg n$ ← GENERAL
$= \left[1 - \frac{H(X)}{\lg n}\right]$ ← NOISE FREE

THE EFFICIENCY OF THE CHANNEL:

$$\eta = I(X;Y)/\lg n = H(X)/\lg n$$
$$= 1 - \text{RELATIVE REDUNDANCY}$$

IF A PARTICULAR SYMBOL X_i TAKES t_i SEC TO PROPAGATE THRU A NOISELESS CHANNEL:

$$R_t = \frac{H(X)}{\sum p(x_i) t_i} = -\frac{\sum_i p(x_i) \lg p(x_i)}{\sum p(x_i) t_i}$$

7/27/76 (TUES)

INFORMATION CHANNEL

· DEFN: AN INF. CHANNEL IS DESCRIBED BY GIVING AN ALPHEBET

$A\ ;\ \{a_i\}\quad i = 1, 2, \dots, r$ ⇐ INPUT ALPHEBET

$B\ ;\ \{b_i\}\quad j = 1, 2, \dots, s$ ⇐ OUTPUT ALPH.

AND A SET OF CONDITIONAL PROBABILITIES

$P(b_j/a_i)\ \forall\ i,j$ WHEN $p(b_j/a_i)$ = CONDITIONAL PROB WITH WHICH b_j WAS RECEIVED IF INPUT SYMBOL a_i WAS SENT

$$\begin{array}{ccccc} & b_1 & b_2 & \dots & b_s \\ a_1 & P(b_1/a_1) & P(b_2/a_1) & & P(a_1/b_s) \\ a_2 & P(b_1/a_2) & & & \\ \vdots & & & & \\ a_r & P(b_1/a_r) & & & P(a_r/b_s) \end{array}$$

DEFINE: $P[b_j/a_i] = P_{ij}$

$$\begin{array}{cccc} & b_1 & b_2 & b_s \\ a_1 & p_{11} & p_{12} & p_{1s} \\ a_2 & p_{21} & p_{22} & \\ \vdots & & & \\ a_r & p_{r1} & p_{r2} & p_{rs} \end{array}$$

CLEARLY:

$$\Sigma_{ij} p_{ij} = 1 \quad \forall \; i = 1,2,\dots r$$

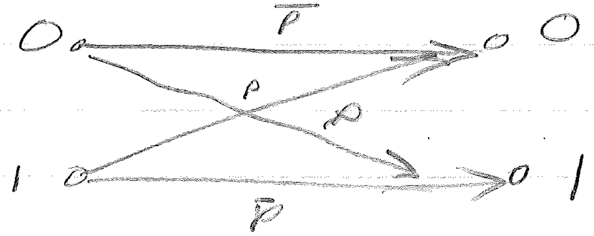
SINCE, IF a_i IS SENT IT MUST BE RECEIVED AS SOME b_j

DEFN: A BINARY SYMMETRIC CHANNEL (BSC) IS "DESCRIBED" BY

$$\begin{vmatrix} \bar{p} & p \\ p & \bar{p} \end{vmatrix}$$

WHERE $\bar{p} = P(0/0) = P(1/1) = P$[CORRECT XMISSION

$p = P(0/1) = P(1/0) = P$[ERROR



AS WITH A SOURCE A, THERE IS AN EXTENSION OF CHANNEL, NAMELY n^{TH} EXTENSION OF THE CHANNEL.

DEF: CONSIDER AN INFO CHANNEL WITH INPUT ALPHEBET $A = \{a_i\}; i = 1,2,\dots,r$ & OUTPUT $B = \{b_j\}; j = 1,2,\dots s$ AND A PRIMARY CHANNEL DESCRIBED BY THE COND. PROB. MATRIX:

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1s} \\ \vdots & & \vdots \\ p_{r1} & \cdots & p_{rs} \end{bmatrix}$$

(CONT ⟶)

NEXT, CONSIDER THE SOURCE A^n, THE n^{TH} EXTENSION OF A (MEMORYLESS) WHERE $A^n = \{\alpha_i\}$, $i = 1, \ldots, r^n$

ALSO, THE RECEIVER ALPHABET

$$B^n = \{\beta_j\} \quad j = 1, \ldots, s^n$$

THE CHANNEL MATRIX WILL THEN BE

$$\Pi = \begin{bmatrix} \Pi_1 & \Pi_2 & \ldots & \Pi_{1s^n} \\ \vdots & & & \\ \Pi_{r^n 1} & & \ldots & \Pi_{r^n s^n} \end{bmatrix}$$

WHERE EACH INPUT, α_i, DEFINED ABOVE, CONSISTS OF A SEQUENCE OF n PRIMARY SYMBOLS $(a_{i1}, a_{i2}, \ldots a_{in})$. SIMILARLY, FOR $\beta_j : \{b_{j1}, b_{j2} \ldots b_{jn})$ AND EACH $\Pi_{ji} = P(\beta_j / \alpha_i)$ = PRODUCT q CORRESPONDING ELEMENTARY PROBS (DUE TO LACK OF MEMORY)

EX. 2nd EXTENSION OF $P = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & \bar{p} & p \\ 1 & p & p \end{array}$ IS

$$\begin{array}{ccccc} & 00 & 01 & 10 & 11 \\ 00 & \bar{p}^2 & \bar{p}p & \bar{p}p & p^2 \\ 01 & \bar{p}p & \bar{p}^2 & p^2 & \bar{p}p \\ 10 & \bar{p}p & p^2 & \bar{p}^2 & \bar{p}p \\ 11 & p^2 & \bar{p}p & \bar{p}p & \bar{p}^2 \end{array}$$

$$\equiv \begin{bmatrix} \bar{p}P & pP \\ pP & \bar{p}P \end{bmatrix}$$ = "KRONECKER SQUARE" FOR $n = 2$

IF ONE KNOWS $P = \begin{bmatrix} p_{11} & \cdots & p_{1s} \\ \vdots & & \vdots \\ p_{r1} & \cdots & p_{rs} \end{bmatrix}$

AND $P(a_i)$ ∀ i (INPUT ALPH), THEN IT HAS BEEN SEEN THAT $P(b_j)$ CAN BE COMPUTED ∀ j. (PROB. 3 ON TEST 1) THIS LEADS TO $\{P(a_i, b_j)\}$. THUS, ONE CAN DETERMINE THE "BACKWARD" COND. PROB. MATRIX FROM THE "FORWARD" COND. PROB. MATRIX.

$\therefore$ FORWARD $\Rightarrow P(b_j/a_i)$

BACKWORD $\Rightarrow P(a_i/b_j)$

NOTICE THAT $\{P(a_i)\}$ YIELDS

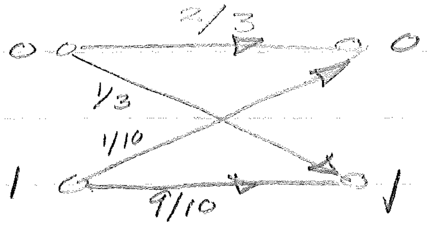
$$H(A) = -\Sigma_i P(a_i) \lg P(a_i)$$

= APRIORI ENTROPY

ALSO $H(A/b_j) \triangleq \Sigma_{\underset{(i)}{A}} P(a_i/b_j) \lg P(a_i/b_j)$

$\triangleq$ A POSTERIORI ENTROPY

<u>EX</u> CONSIDER A NOISY BINARY CHANNEL



(NOTE: NOT SYMMETRIC)

$P(\text{COND}) = \begin{bmatrix} 2/3 & 1/3 \\ 1/10 & 9/10 \end{bmatrix}$ $\quad P(a=0) = 3/4$

$P(a=1) = 1/4$

THIS IS ALL THE INFO WE NEED TO CHARACTERIZE THE CHANNEL. $\Rightarrow$

P[0 IS RECEIVED @ OUTPUT]

$$= \frac{3}{4} \times \frac{2}{3} + \frac{1}{4} \times \frac{1}{10} = \frac{21}{40}$$

$P(a=0) \times P(b=0/a=0) + P(a=1)P(b=0/a=1)$

P[1 IS RECEIVED @ OUTPUT]

$$= \frac{3}{4} \times \frac{1}{3} + \frac{1}{4} \times \frac{4}{10} = \frac{19}{40}$$

$$P[a=0/b=0] = \frac{P[a=0,b=0]}{P(b=0)}$$

$$= \frac{P[b=0/a=0]\,P[a=0]}{P[b=0]}$$

$$= \frac{3}{4} \times \frac{2}{3} \Big/ \frac{21}{40} = \frac{20}{21}$$

$$P[b=1, a=1] = P[b=1/a=1]\,P[a=1]$$

$$= 9/10 \times 1/4 \big/ 19/40 = \frac{9}{14}$$

$$P[a=1/b=0] = \frac{P[b=0, a=1]}{P[b=0]} = \frac{1}{21}$$

$$P[a=0/b=1] = \frac{10}{19}$$

FINDING APRIORI & APOSTERIORI PROBABILITIES.

$$H(A) = \frac{3}{4} \lg \frac{4}{3} + \frac{1}{4} \ln 4$$

$$= \frac{3}{4} \lg 4 + \frac{1}{4} \lg 4 - \frac{3}{4} \lg 3$$

$$= 2 - 1.585 \left(\frac{3}{4}\right) = 0.811 \text{ BITS}$$

$$H(A/b=0) = -p(a=0/b=0) \lg p(a=0/b=0)$$

$$- p(1/0) \lg p(1/0)$$

$$= \frac{20}{21} \lg \frac{21}{20} + \frac{1}{21} \lg 21$$

$$= \lg 21 - \frac{20}{21} \lg 20$$

$$= (4.392) - \frac{20}{21}(4.322) = 0.276 \text{ BITS}$$

$$H(A/b=1) = -P(0/1) \lg P(0/1)$$

$$- P(1/1) \lg P(1/1)$$

$$= \frac{10}{19} \lg \frac{19}{10} + \frac{9}{19} \lg \frac{19}{9}$$

$$= 0.998 \text{ BITS}$$

A UNIFORM CHANNEL: A CHANNEL [DESCRIBED BY ITS FORWARD COND. MATRIX] IS SAID TO BE UNIFORM IF THE ELEMENTS IN EVERY ROW & EVERY COLUMN OF ITS MATRIX CONSIST OF AN ARBITRARY PERMUTATION OF THE TERMS IN THE FIRST ROW.

EX:

$$\begin{array}{ccccc} & b_1 & b_2 & b_3 & b_4 \\ a_1 & \frac{1}{3} & 1/3 & 1/6 & 1/6 \\ a_2 & \frac{1}{6} & 1/6 & 1/3 & 1/3 \end{array}$$

C = CHANNEL CAPACITY = $\mathrm{Max}\ I(X;Y)$

= TRANS INFO = MUTUAL INFO

$= \mathrm{Max}\left[H(X) - H(X/Y)\right]$

$= \mathrm{Max}\left[H(Y) - H(Y/X)\right]$

LET $p(y_j/x_i) = \alpha_{ij}$

$p(x_i, y_j) = a_i \alpha_{ij} \ni P(x_i) = a_i \ \forall i$

$$P = (a)\overset{(b)}{\begin{bmatrix} p_{11} & \cdots & p_{1s} \\ \vdots & & \vdots \\ p_{r1} & \cdots & p_{rs} \end{bmatrix}}$$

$$H(Y/x_i) = -\sum_{j=1}^{m} p(y_i/x_i) \lg p(y_j/x_i)$$

THUS, FOR A UNIFORM CHANNEL

$$H(Y/x_i) = h = \text{CONST.}$$

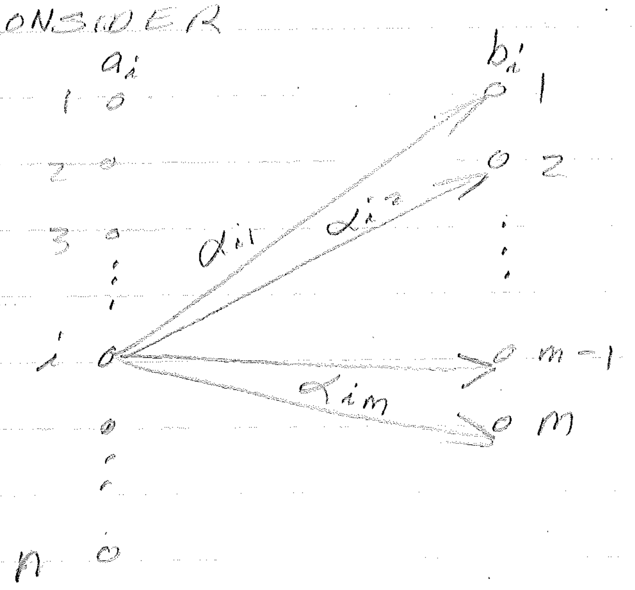
AND

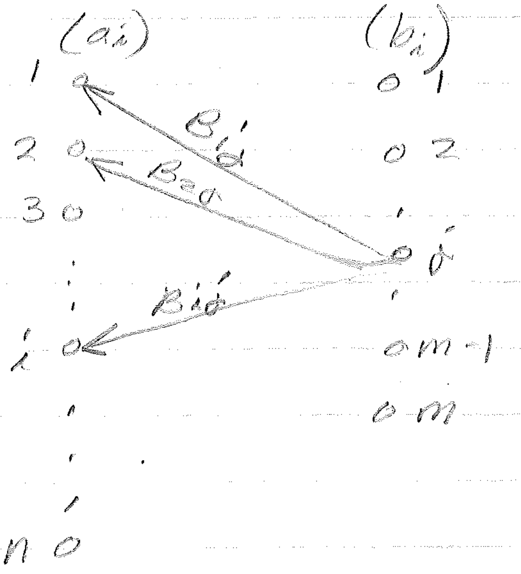
$$H(Y/X) = \sum_{i=1}^{m} a_i h = h$$

CONSIDER, THEN

$$C = \text{Max}\left[H(Y) - h\right]$$
$$= \lg s - h$$

CONSIDER

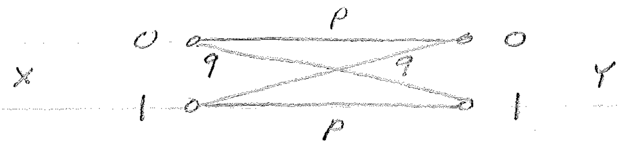




$$C = \lg r - h' = \lg s - h$$

7/28/76 (WED)

BSC (BINARY SYMMETRIC CHANNEL)



$P(0)=\alpha \quad , P(1)=1-\alpha$

$P(0/0)=P(1/1)=p$

$P(0/1)=P(1/0)=q$

$H(X)=H(\alpha, 1-\alpha)$

$= \alpha \lg \frac{1}{\alpha} + (1-\alpha) \lg \frac{1}{1-\alpha}$

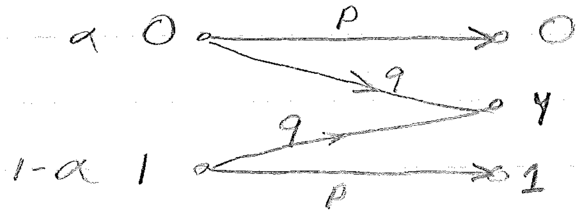
$H(Y/X)= -p \lg p - q \lg q$

$I(X;Y)= H(Y) - H(Y/X)$

$= H(Y)+p \lg p+q \lg q$

$Max[I(X;Y)] = 1+p \lg_2 p+q \lg_2 q$ BITS

BEC (BINARY ERASURE CHANNEL)



$p+q=1$

$$P(Y/X)= \begin{array}{c|ccc} & 0 & Y & 1 \\ 0 & p & q & 0 \\ 1 & 0 & q & p \end{array}$$

$P(0)=\alpha \qquad P(1)=1-\alpha=\bar{\alpha}$

$H(X)= \alpha \lg \frac{1}{\alpha} + \bar{\alpha} \lg \frac{1}{\bar{\alpha}}$

$H(X/Y)$ (WE FIRST NEED $P(X/Y)$) $\longrightarrow$

$$
P(X,Y) = \begin{array}{c} \\ \\ \\ \\ \end{array}\begin{array}{ccc} 0 & Y & 1 \\ \left|\begin{array}{c} \alpha p \\ 0 \end{array}\right. & \begin{array}{c} \alpha q \\ \bar{\alpha} q \end{array} & \left.\begin{array}{c} 0 \\ \bar{\alpha} p \end{array}\right| \\ \alpha p & q & \bar{\alpha} p \Leftarrow P(Y = Y_i) \end{array}
$$

$$
P(X/Y) = \left| \begin{array}{ccc} \frac{\alpha p}{\alpha p} & \frac{\alpha q}{q} & 0 \\ 0 & \frac{(1-\alpha) q}{q} & \frac{(1-\alpha) p}{(1-\alpha) p} \end{array} \right|
$$

$$
= \left| \begin{array}{ccc} 1 & \alpha & 0 \\ 0 & \bar{\alpha} & 1 \end{array} \right]
$$

NOW, FIND

$$
H(X/Y) = \left| \begin{array}{l} \alpha p \lg 1 + \alpha q \lg \alpha \\ \alpha q \lg 0 + 0 - \bar{\alpha} \lg q + \bar{\alpha} p \lg 1 \end{array} \right]
$$

$$= q\,H(x)$$

$$= \bar{p}\;H(x)$$

$$\therefore\; I(x;Y) = H(X) - (1-p)\,H(x)$$

$$= p\,H(x)$$

MAXIMIZE $\{H(x)\} = 1$ BIT

$\therefore\; C = \text{Max}\, I(x;Y) = p$ BITS

SECOND EXTENSION OF BSC

CLEARLY, IF 0,1 COMPRISE THE INPUT ALPHEBET FOR THE PRIMARY CHANNEL, THEN 00, 01, 10, 11 COMPRISE THE SYMBOLS FOR INPUT (& OUTPUT) OF THE SECOND EXTENSION. ASSUME THE CHANNEL HAS ZERO MEMORY:

$$U = [\overset{x_1 x_2}{00}, \overset{x_1 x_2}{01}, \overset{x_1 x_2}{10}, \overset{x_1 x_2}{11}]$$

$$V = [\underset{Y_1 Y_2}{00}, \underset{Y_1 Y_2}{01}, \underset{Y_1 Y_2}{10}, \underset{Y_1 Y_2}{11}]$$

$$P(X_1, X_2) = P(X)P(X_2) = P(U)$$

$$P(Y_1, Y_2) = P(Y_1)P(Y_2) = P(V)$$

$$P(V/U) = P(Y_1Y_2/X_1X_2) = P(Y_1/X_1)\,P(Y_2/X_2)$$

$$P(U/V) = P(X_1/Y_1)\,P(X_2/Y_2)$$

$$\begin{aligned} P(U,V) &= P(U)P(V/U) \\ &= P(X_1)P(X_2)P(Y_1/X_1)P(Y_2/X_2) \\ &= P(X_1,Y_1)\,P(X_2,Y_2) \end{aligned}$$

SOURCE ENTROPY:

$$\begin{aligned} H^2(X) = H(U) &= 2H(X) \\ &= -2[\alpha \lg \alpha + \bar{\alpha} \lg \bar{\alpha}] \\ &= H(X_1) + H(X_2) \end{aligned}$$

$$H(U,V) = H(X_1, Y_1) + H(X_2, Y_2)$$

$$H(V/U) = H(Y_1/X_1) + H(Y_2/X_2)$$

$$H(U/V) = H(X_1/Y_1) + H(X_2/Y_2)$$

$$\begin{aligned} I(U;V) &= H(U) - H(U/V) = H(X_1) - H(X_1/Y_1) + H(X_2) - H(X_2/Y_2) \\ &= 2\,I(X;Y) \end{aligned}$$

SEE HAND-OUTS ③ ~~& 4~~

*⟹ HOMEWORK: USE (THIS) GRAPHICAL TECHNIQUE FOR THE FOLLOWING CHANNELS:

(a) $\begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$ (b) $\begin{bmatrix} 9/10 & 1/10 \\ 2/10 & 8/10 \end{bmatrix}$

IF YOU NEED TO, ASSUME APPROPRIATE NUMBERS FOR INPUT &/OR OUTPUT PROBABILITIES

ANALYTICAL TECHNIQUE FOR DETERMINING CAPACITY OF BINARY CHANNEL

[S. MUROGA, "ON THE CAPACITY OF A DISCRETE CHANNEL" J. PHY. SOC. JAP 8, 484, 1953

CONSIDER

① $p_{11}Q_1 + p_{12}Q_2 = p_{11}\lg p_{11} + p_{12}\lg p_{12} = {}^{-}H(\varphi_{11})$

$$\begin{array}{cc} Y \to p_1' & p_2' \\ \left[\begin{array}{c} p_{11} \\ p_{21} \end{array}\right. & \left.\begin{array}{c} p_{12} \\ p_{22} \end{array}\right] \end{array}$$

ALSO, LET

② $p_{21}Q_1 + p_{22}\lg Q_2 = p_{21}\lg p_{21} + p_{22}\lg p_{22}) = {}^{-}H(\varphi_{22})$

$I(X;Y) = H(Y) - H(Y/X)$

③ $= -(p_1'\lg p_1' + p_2'\lg p_2')$
$+ p_1(p_{11}\lg p_{11} + p_{12}\lg p_{12})$
$+ p_2(p_{21}\lg p_{21} + p_{22}\lg p_{22})$

① & ② CAN BE WRITTEN

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = -\begin{bmatrix} H(p_{11}) \\ H(p_{22}) \end{bmatrix}$$

SHOW ⟹ USE THIS SOLN TO WRITE:

④ $$I(X;Y) = -(p_1' \lg p_1' + p_2' \lg p_2') + p_1' Q_1 + p_2' Q_2$$

WE WANNA MAXIMIZE WITH RESPECT TO p_1' AND p_2' AND RESORT TO THE LAGRANGE FUNCTIONAL:

LAGRANGE MULTIPLIER

⑤ $$U = -(p_1' \lg p_1' + p_2' \lg p_2') + p_1' Q_1 + p_2' Q_2 + \mu (p_1' + p_2')$$

$$\frac{dU}{dp_1} = -(\lg p_1' + \lg_2 e) + Q_1 + \mu = 0$$

$$\frac{dU}{dp_2} = -(\lg_2 e + \lg p_2') + Q_2 + \mu = 0$$

SOLVE THESE FOR μ:

$$\mu = -Q_1 + (\lg_2 e + \lg p_1')$$
$$\mu = -Q_2 + (\lg_2 e + \lg p_2')$$

USE THESE VALUES OF μ IN ⑤ ONE BY ONE. THUS

CHANNEL CAPACITY $= C = \text{Max } I(X;Y) = Q_1 - \lg p_1'$

⑥ $= Q_2 - \lg p_2'$

WHICH SAYS THAT

$$p_i' = 2^{Q_i - C} \quad i = 1,2$$

OR $$C = \lg\left[2^{Q_1} + 2^{Q_2}\right]$$ BITS

WE MAY FIND Q_1 & Q_2

1. BY SOLVING MATRIX @ TOP OF -63-
2. CURVES ON HANDOUT # 4 (7/28/76)

EX. LET $p_{11} = p_{12} = \frac{1}{2}$ $p_{21} = \frac{1}{4}$ $p_{22} = 3/4$

$$\begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$$

FROM CHART, $Q_1 \simeq -1.4$ $\quad$ $\left[\begin{matrix} Q_1 = -1.378 \\ Q_2 = -0.622 \end{matrix}\right]$

$Q_2 \simeq -0.6$

⟹ AS HOMEWORK, WORK BY MATRIX INVERSION

NOW, FROM CHART, $C \doteq 0.06$ BITS

⟹ AS HOMEWORK, VERIFY THAT $C = 0.0485$

K ⟹ USE MUROGA'S TECHNIQUE TO FIND

$$\begin{bmatrix} 3/4 & 1/8 & 1/8 & 0 \\ 1/8 & 3/4 & 0 & 1/8 \\ 1/8 & 1/8 & 3/4 & 0 \\ 0 & 0 & 1/4 & 3/4 \end{bmatrix}$$

7/29/76 (WED)

COMMENTS ON MUROGA'S TECHNIQUE

(1) FOR A m×m MATRIX, ONE NEEDS TO SET UP m SIMULTANEOUS EQUATIONS ;

$$p_{11} Q_1 + \dots + p_{1m} Q_m = \sum_{j=1}^{m} p_{1j} \lg p_{1j}$$

$$\vdots \qquad\qquad \vdots$$

$$p_{m1} Q_1 + \dots + p_{mm} Q_m = \sum_{j=1}^{m} p_{mj} \lg p_{mj}$$

THEN

$$I(X,Y) = -\sum_{i=1}^{m} p_i' \lg p_i' + \sum_{i=1}^{m} p_i' Q_i$$

∋ m = # OF INPUT <u>AND</u> OUTPUT SYMBOLS

THEN

$$C = \lg \sum_{i=1}^{m} 2^{Q_i}$$ FOR THE BINARY ELEMENT SOURCE (0, 1)

(2) IT IS CLEAR THAT THE CHANNEL MATRIX HAS TO BE SQUARE

(3) IT IS CONCEIVABLE THAT THE INPUT PROB'S CORRESPONDING TO C DETERMINED BY THE TECHNIQE MAY NOT MEET THE REQUIREMENT

$$0 \le p_i \le 1 \qquad \sum_i p_i = 1$$

SO <u>WATCH</u> <u>OUT</u>

(4) SILVERMAN, CHANG & LOEB SOLVED THIS PROBLEM IN MID 50'S

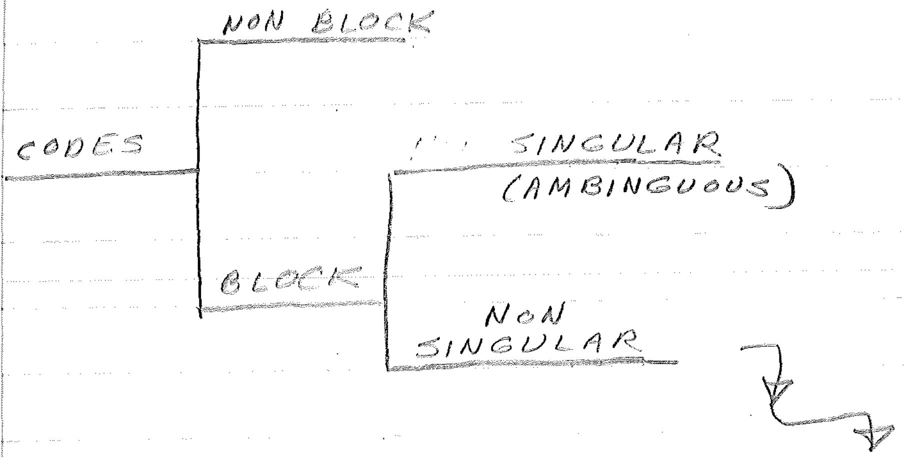
CODES & THEIR "ELEMENTARY" PROPERTIES
(MATERIAL FROM TEXT, p46 ON):
READ p. 51

DEFN: LET THE SET OF SYMBOLS COMPRISING A GIVEN ALPHEBET BE CALLED $S; \{s_1, s_2, \ldots, s_q\}$. THEN A CODE IS A MAPPING (TRANSFORMATION) OF ALL POSSIBLE SEQUENCES OF THE SYMBOLS OF S INTO OTHER SEQUENCES OF SOME OTHER ALPHEBET $X = \{x_1, x_2, \ldots x_r\}$.
S IS CALLED SOURCE ALPHEBET
X IS CALLED CODE ALPHEBET
EX: $S = \{0, 1, 2, \ldots, 9)$ $X = (0, 1)$

DEFN: A BLOCK CODE IS ONE WHICH MAPS EACH OF THE SYMBOLS OF THE SOURCE ALPHEBET S INTO A FIXED SEQUENCE OF SYMBOLS OF X (CODE ALPHEBET). SUCH FIXED SEQUENCES IS CALLED CODE WORDS. A COLLECTION OF CODE WORDS IS A CODE BOOK.

EX:

s_1	0
s_2	11
s_3	00
s_4	11



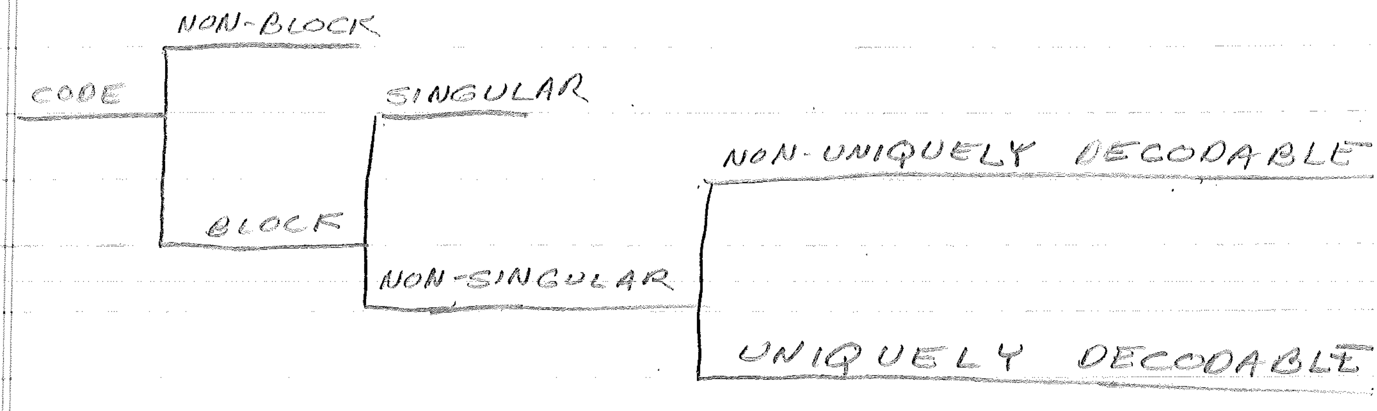
DEFN: A BLOCK CODE IS NON-SINGULAR IF ALL THE WORDS OF THE CODE ARE DISTINCT. OTHERWISE, THE BLOCK CODE IS SINGULAR

Ex:

S_1	0
S_2	11
S_3	00
S_4	01

BUT, SUPPOSE WE RECEIVE 0011, COULD BE $S_1 S_1 S_2$ OR $S_3 S_4$

THIS CODE IS NON-SINGULAR IN THE "SMALL", BUT SINGULAR IN THE LARGE.



CODE EXTENSION: LET A GIVEN BLOCK CODE MAP SYMBOLS FROM S INTO <u>FIXED SEQUENCES</u> OF SYMBOLS FOR X. S, ITSELF, WE HAVE SEEN, CAN BE AN EXTENSION OF AN OTHER ELEMENTARY SOURCE. THUS, TO AN EXTENSION OF A SOURCE, MUST CORRESPOND TO AN EXTENSION OF THE CODE.

DEF: THE n^{TH} EXTENSION OF A BLOCK CODE WHICH MAPS SYMBOLS s_i INTO WORDS X_i IS THE BLOCK CODE WHICH MAPS SEQUENCES OF <u>SOURCE</u> SYMBOLS $(s_{1i}\ s_{i2}, \dots, s_{in})$ INTO SEQUENCES OF CODE WORDS $(X_{i1}, X_{i2}, \dots, X_{in})$

EX	s_1	0	$s_1 s_2$	00
	s_2	11	$s_1 s_2$	011
	s_3	00	⋮	
	s_4	01		

DEF: A BLOCK CODE IS DECODABLE IF THE n^{TH} EXTENSION OF THE CODE IS NON-SINGULAR FOR EACH FINITE n.

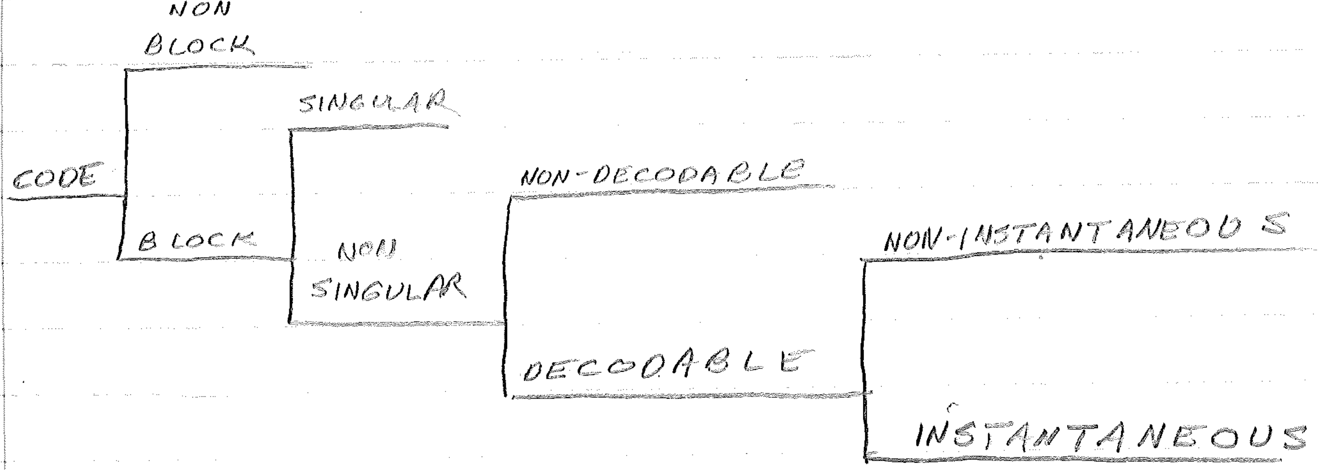
EX.

	A	B	C
	CODES		
S_1	00	10	0
S_2	01	10	01
S_3	10	110	011
S_4	11	1110	0111

A, B, & C ARE COMPLETELY DECODABLE

B IS SOMETIMES CALLED "COMMA" CODES

C IS UNIQUELY DECODABLE, BUT IS A "DELAY" CODE OR A "NON-INSTANTANEOUS" CODE. THAT IS, YA ALWAYS GOTTA LOOK @ NEXT SYMBOL



DEF: A UNIQUELY DECODABLE CODE IS INSTANEOUS IF IT IS POSSIBLE TO DECODE EACH WORD IN A SEQUENCE W/O REFERENCE TO THE SUCCEEDING CODE SYMBOL. TO TEST FOR INSTANTANEOUSLY, WE DEFINE A PREFIX OF A CODE WORD.

DEF: LET $X_i = X_{i_1}, X_{i_2}, \ldots, X_{i_n}$ BE A WORD. THE SEQUENCE $(X_{i_1}, X_{i_2}, \ldots, X_{i_j})$ $j \leq n$ IS CALLED A PREFIX OF X_i

EX. $X_i = 0111$

PREFIXES: 0, 01, 011, 0111

<u>PREFIX PROPERTY</u>

A NECESSARY & SUFFICIENT CONDITION FOR A CODE TO BE INSTANTANEOUS IS NO COMPLETE WORD IN A CODE BOOK BE A PREFIX OF SOME OTHER CODE WORD.

HOW TO CONSTRUCT INSTANTANEOUS CODES:
LET THE ENCODING ALPHEBET BE (0,1).
SAY THE CODEBOOK HAS 5 WORDS
ARISING FROM

s_1	0	(1) ⟵ $l_i,\ i=1,5$
s_2	10	(2)
s_3	110	(3)
s_4	1110	(4)
s_5	1111	(4)

ANOTHER IS

		l_i
s_1	00	(2)
s_2	01	(2)
s_3	10	(2)
s_4	110	(3)
s_5	111	(3)

7-29 ⟹ HOMEWORK → START WITH 10 $\frac{1}{1}$ BUILD

7/30/76 (FRI)

CODE → BLOCK → NON-SINGULAR → UNIQUELY DECODABLE → INSTANT.

A NEC. & SUFFICIENT CONDITION FOR INSTANT CODE → MEET THE PREFIX PROPERTY

KRAFT (1949)

LET US SEEK TO BUILD ANI INSTANT CODE FROM

$S: \{s_1, s_2, \ldots s_q\}$ & A CODE ALPHEBET

$X: \{x_1, x_2, \ldots, x_r\}$

WE WILL CONSIDER $X: \{0,1\}$

∋ $X_1, X_2, \ldots, X_q$ ARE WORDS OUT OF THE ALPHEBET X WITH THE CORRESPONDING LENGTHS $l_1, l_2, \ldots, l_q$. IT IS CONCEIVABLE THAT SOME SEQUENCES OF X_i HAVE THE SAME l_i.

l_i = LENGTH OF CODE WORD = # DIGITS IN X_i

KRAFT'S INEQUALITY

A NECESSARY & SUFFICIENT CONDITION *FOR THE EXISTANCE OF* INST. CODES WITH WORD LENGTH $l_1, l_2, \ldots, l_q$ IS $\sum_{i=1}^{q} r^{-l_i} \leq 1$ ∋ r = # OF DIGITS IN THE ENCODING ALPHABET. FOR $r = 2$

$$\sum_{i=1}^{q} 2^{-l} \leq 1$$

NOTE: THIS DOES *NOT* ASSURE THE CODE IS INSTANTANEOUS.

LOOK AT SOME CODES $(r=2)$

SOURCE SYMBOL	A	B	C	D	E
S_1	(2) 00	(1) 0	(1) 0	(1) 0	(1) 0
S_2	(2) 01	(3) 100	(2) 10	(3) 100	(2) 10
S_3	(2) 10	(3) 110	(3) 110	(3) 110	(3) 110
S_4	(2) 11	(3) 111	(3) 111	(2) 11	(2) 11

A: $\sum_{i=1}^{4} 2^{-2} = 1$: A IS ALSO INSTANT

B: $2^{-1}+2^{-3}+2^{-3}+2^{-3}=\frac{7}{8}$: B IS ALSO INSTANT

C: $2^{-1}+2^{-2}+2^{-3}+2^{-3} \leq 1$: C " " "

D: $2^{-1}+2^{-3}+2^{-3}+2^{-2} \leq 1$ BUT D IS <u>NOT</u> INSTANT

E: $2^{-1}+2^{-2}+2^{-3}+2^{-2} > 1$

$\Rightarrow$ THIS CODE <u>CANNOT</u> BE INSTANTANEOUS

EXAMPLE USING KRAFT INEQUALITY:

CODE THE DECIMAL ALPHABET 0, 1, ..., 9 BY USING THE CODING ALPHEBET (0,1) $(r=2)$

$0 \rightarrow 0$

$1 \rightarrow 10$

$$\left.\begin{matrix} 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{matrix}\right\} \text{REQUIRE } l_i \text{ FOR THE REST HAVE EQUAL LENGTH. WE SEEK AN INSTANTANEOUS CODE}$$

→

USING KRAFT'S INEQUALITY, WE REQUIRE

$$2^{-1} + 2^{-2} + 8 \times 2^{-l} \leq 1$$

$$2^{-l} \leq \frac{1}{32}$$

$$\Rightarrow \frac{1}{2^l} \leq \frac{1}{32} \Longrightarrow l \geq 5\text{, SO LET } l = 5$$

LETS BUILD IT

0	0	5	11 011
1	10	6	11 100
2	11 000	7	11 101
3	11 001	8	11 110
4	11 010	9	11 111

PROOF OF KRAFT'S INEQUALITY

SUFFICIENCY: $\sum_{i=1}^{q} r^{-l_i} \leq 1$ ① AND WE NEED TO SHOW THAT THE SELECTION OF l_i BASED ON THIS EQUALITY WILL GENERATE INSTANTANEOUS CODES. LET n_j = # OF CODE WORDS WITH LENGTH j $\ni$ ~~max~~ $j = l$. & THE CODES $X_1, X_2, \ldots X_q$ HAVE LENGTH $l_1, l_2, \ldots, l_q$. THEN ② $q = \sum_{i=1}^{l} n_i$ $\ni$ l is LONGEST LENGTH CODE IN THE BOOK. ① CAN THEN BE WRITTEN AS

$$\sum_{i=1}^{l} n_i r^{-i} \leq 1$$

(CONVINCE YERSELF BY OPENING SUMMATION ON LEFT & COUNT TERMS LIKE $\frac{1}{r}, \frac{1}{r^2}, \ldots, \frac{1}{r^3}, \ldots$)

$\Longrightarrow$

EXPANDING THIS:

$$\Sigma_{i=1}^{\ell} n_i r^{-i} = n_1 r^{-1} + n_2 r^{-2} + \ldots + n_{\ell-1} r^{-(\ell-1)} + n_\ell r^{-\ell}$$
$$\leq 1$$

$$\Longrightarrow n_\ell r^{-\ell} \leq 1 - n_1 r^{-1} - n_2 r^{-2} - \ldots - n_{\ell-1} r^{-(\ell-1)}$$

OR $n_\ell \leq r^\ell - n_1 r^{\ell-1} - n_2 r^{\ell-2} - \ldots - n_{\ell-1} r^1$

$$\Rightarrow 0 \leq r^\ell - n_1 r^{\ell-1} - n_2 r^{\ell-2} - \ldots - n_{\ell-1} r^1$$

$$\therefore n_{\ell-1} r^1 \leq r^\ell - n_1 r^{\ell-1} - n_2 r^{\ell-1} - \ldots - n_{\ell-2} r^2$$

$$\Longrightarrow n_{\ell-1} \leq r^{\ell-1} - n_1 r^{\ell-2} - n_2 r^{\ell-3} - \ldots - n_{\ell-2} r$$

$$\vdots$$

$$n_3 \leq r^3 - n_1 r^2 - n_2 r$$
$$n_2 \leq r^2 - n_1 r$$
$$n_1 \leq r$$

THE CLAIM IS THAT THESE n_i ARE ADEQUATE TO BUILD AN INSTANT CODE.

$n_1 \leq r \Rightarrow$ FOR r ELEMENS, YOU CAN ONLY BUILD, AT MOST, n_1, r element INSTANT CODE $\Rightarrow r - n$ ARE LEFT TO BUILD OTHER WORDS OF THE BOOK.

$n_2 \leq r(r - n_1)$ = # OF WAYS WE CAN BUILD INST. CODES OF LENGTH TWO.

FIRST TWO LOCATIONS HAVE BEEN TAKEN UP. $(r^2 - n_1 r - n_2) r$ = # OF WAYS INSTANTANEOUS CODES OF LENGTH $n_3 = 3$.

$\Rightarrow$ HOMEWORK 3-2, 3-3

READ "NOTES" p. 61-62

8/2/76

KRAFT'S INEQUALITY IS ALSO CALLED THE NOISELESS CODING THEOREM

WE HAVE YET TO PROVE "NECESSITY" OF KRAFT'S INEQUALITY. LET'S DO SO:

WE HAVE NOTED THAT THE TWO MESSAGES X_i & X_k CAN HAVE THE SAME LENGTH. (WE ARE TRYING TO SHOW, THAT, GIVEN AN INSTANT CODE, THEN KRAFT'S INEQUALITY HOLDS). LET n_i, AS BEFORE, = # OF ENCODED MESSAGES OF LENGTH l_i. THEN $n_i \le r \ni r$ = # OF DIGITS IN ENCODING ALPHABET. THEN THE NUMBER OF ENCODED MESSAGES OF LENGTH 2, n_2 CANNOT BE $> (r - n_1) r$.

ALONG THE SAME LINES

$$n_3 < [(r-n_1)r - n_2]r$$
$$\le r^3 - n_1 r^2 - n_2 r$$

PROCEEDING LIKEWISE:

n_m = # OF ENCODED WORDS WITH LENGTH m

$$\le r^m - n_1 r^{m-1} - n_2 r^{m-2} - \dots - n_{m-1} r$$

$$0 \le r^m - n_1 r^{m-1} - n_2 r^{m-2} - \dots - n_{m-1} r - n_m$$

$$0 \le 1 - n_1 r^{-1} - n_2 r^{-2} - \dots - n_{m-1} r^{-m+1} - n_m r^{-m} \Longrightarrow$$

$$\therefore \sum_{i=1}^{m} n_i r^{-i} \le 1$$

IF m IS THE # OF DIGITS IN THE LARGEST WORD, THE $m = \ell$

$$\therefore \sum_{i=1}^{\ell} n_i r^{-i} \leq 1$$

BUT, AS WE <u>SAID</u>* ON FRIDAY:

$$\sum_{i=1}^{\ell} n_i r^{-i} = \sum_{j=1}^{q} r^{-l_j}$$

$$\therefore \sum_{i=1}^{q} r^{l_i} \leq 1$$ (WHICH IS K)

* WILL NOW SHOW THAT $\sum_{i=1}^{\ell} n_i r^{-i} = \sum_{j=1}^{q} r^{j}$

LET $X : \{x_1, x_2, \ldots x_7\}$

LET $l_1 = 2 \quad l_2 = 2 \quad l_3 = 3 \quad l_4 = 3$

$l_5 = 3 \quad l_6 = 4 \quad l_7 = 5$

l_i IS THE LENGTH OF X_i

$n_1 = 0 \quad n_2 = 2 \quad n_3 = 3 \quad n_4 = 1 \quad n_5 = 1$

n_j IS THE NUMBER OF WORDS WITH LENGTH $l_i = j$

$\ell =$ LENGTH OF LONGEST WORD $= 5$

COMPUTING $(q = 7)$

$$① \sum_{i=1}^{\ell} n_i r^{-i} = 2\frac{1}{r^2} + 3\frac{1}{r^3} + 1\frac{1}{r^4} + 1\frac{1}{r^5}$$

$$= \frac{1}{r^2} + \frac{1}{r^2} + \frac{1}{r^3} + \frac{1}{r^3} + \frac{1}{r^3} + \frac{1}{r^4} + \frac{1}{r^5}$$

$$② \sum_{j=1}^{q} r^{-l_j} = \frac{1}{r^2} + \frac{1}{r^2} + \frac{1}{r^3} + \frac{1}{r^3} + \frac{1}{r^3} + \frac{1}{r^4} + \frac{1}{r^5}$$

IN OTHERWORDS, ① IS A # OF STRINGS N_K EACH OF THE TYPE $1/r^K$ HAVING K TERMS IN EACH STRING.

** McMILLAN'S INEQUALITY
(FOR UNIQUELY DECODABLE CODES
i.e NON-SINGULAR)

NOTE THAT

(1) ALL INSTANTANEOUS CODES ARE UNIQUELY DECODABLE. HENCE, IF KRAFT'S INEQUALITY IS SUFFICIENT FOR INST. CODE IS SUFFICIENT FOR U.D.

(2) THE QUESTION IS: IS KRAFT A NECESSARY CONDITION FOR U.D. CODES. McMILLAN SAYS <u>YES</u>. i.e, GIVEN A UNIQUELY DECODABLE CODE, IS SATISFIES KRAFTS INEQUALITY. (McMILLAN'S INEQUALITY (1956))

CONSIDER $\left(\sum_{i=1}^{q} r^{-\ell_i}\right)^n \ni$
n IS REAL, n & ℓ_i IS AN INTEGER.

$$\left(\sum_{i=1}^{q} r^{-\ell_i}\right)^n = \left(r^{-\ell_1} + r^{-\ell_2} + \dots + r^{-\ell_q}\right)^n$$

$$\overset{\text{OF THE}}{\underset{\text{TYPE}}{=}} \sum r^{-\ell_1-\ell_2-\ell_3-\dots-\ell_q} = r^{-k}$$

$$\ni k \triangleq \ell_{i_1} + \ell_{i_2} + \ell_{i_3} + \dots + \ell_{i_n}$$

IF ℓ = LENGTH OF LONGEST WORD:

$$n \le k \le n\ell$$

LET, AS BEFORE, N_k = # OF TERMS OF THE FORM r^{-k}. THEN FROM BEFORE

$$\left(\sum_{i=1}^{q} r^{-l_i}\right)^n = \sum_{k=n}^{nl} N_k r^{-k} \Leftarrow \text{JUST A REORDERING}$$

WHERE $N_k \overset{\text{ALSO}}{=}$ # OF STRINGS OF n CODE WORDS THAT CAN BE FOUND SO THAT EACH STRIP IS OF LENGTH k.

OBVIOUSLY, $N_k \le r^k$

$$\left(\sum_{i=1}^{q} r^{-l_i}\right)^n = \sum_{k=n}^{nl} N_k r^{-k}$$

$$\le \sum_{k=n}^{nl} r^k r^{-k}$$

$$\le \sum_{k=n}^{nl} 1$$

$$\le nl - n + 1 \le nl$$

THIS GIVES

$$\left(\sum_{i=1}^{q} r^{-l_i}\right)^n \le nl$$

RECALL, $x^n \le nl \Rightarrow x \le 1$

THEN THIS GIVES McMILLAN'S INEQUALITY:

$$\sum_{i=1}^{q} r^{-l_i} \le 1$$

THIS IS KRAFTS INEQUALITY AS APPLIED TO UNIQUELY DECODABLE CODES.

EX: LET THE # OF WORDS WITH EQUAL LENGTHS BE

$[n_1, n_2, n_3, n_4] = [0, 3, 0, 5]$

ie # WORDS LENGTH 1 = 0

" " 2 = 3

" " 3 = 0

" " 4 = 5

$\Rightarrow$ THE CODE BOOK HAS 8 WORDS.

QUESTION: CAN ONE BUILD AN INSTANT CODE WITH BINARY ALPHABET (0,1). THEN

$$\sum_{i=0}^{4} n_i r^{-i} \leq 1$$

$$\sum_{i=0}^{4} n_i r^{-i} = \sum_{i=1}^{8} r^{-l_i} \leq 1$$

$$\sum_{i=0}^{4} n_i r^{-i} = \frac{3}{r^2} + \frac{5}{r^4} \leq 1$$

NOT SATISFIED FOR $r = 2$. GOTTA LOOK FOR ANOTHER ALPHABET

$3/r^2 + 5/r^4 \leq 1$

$r^2 + 5 \leq r^4$

$r^4 - 3r^2 - 5 \geq 0$

FOR EQUALITY:

$r^2 = \frac{3 \pm \sqrt{29}}{2} = 4.2, \ \frac{3-5.4}{2} < 0$

$r \simeq 2.1$

$\Longrightarrow r = 3$

LET, THEN, THE ALPHABET BE $(0,1,2)$
A SCHEME FOR $n_i = \{0,3,0,5\}$ IS

00	1000	2000
01	1001	2222
02	1002	

/ $\Longrightarrow$ AS HOMEWORK, $\overset{6/3}{}$ DEVISE TWO INST. CODING SCHEMES WITH $r = (0,1,2)$ FOR THIS SAME EXAMPLE.

(EX): LET THE MESSAGE ENSEMBLE $X = \{x_1, x_2, x_3\}$. SHOW THAT ALL POSSIBLE SETS OF BINARY CODES WITH PREFIX PROPERTY (INST. CODES) CAN BE ENCODED IN WORDS NOT MORE THAN THREE DIGITS LONG. i.e. $n_i \leq 3$.

NOW

$$\sum_{i=1}^{\ell} n_i r^{-i} \leq 1$$

$$\ell_1 2^{-1} + \ell_2 2^{-2} + \ell_3 2^{-3} \leq 1$$

ALSO, $\ell_1 + \ell_2 + \ell_3 = q = 3$

$$\Rightarrow 4\ell_1 + 2\ell_2 + \ell_3 \leq 8$$

$$\ell_1 + \ell_2 + \ell_3 = 0$$

$\Longrightarrow$

l_1	l_2	l_3
1	1	1
1	2	0
1	0	2
0	3	0
0	2	1
0	1	2
0	0	3

$\Rightarrow$ WE HAVE TO SATISFY

$$4l_3 + 2l_3 + l_3 \leq 8$$

AND $l_1 + l_2 + l_3 = 3$

ALL THESE WILL WORK

✓ $\Longrightarrow$ CHECK THEM

SHANNON'S FIRST THEOREM

DEFN: AVERAGE LENGTH OF A CODING SCHEME:

$$\overline{L} = \sum_i p(x_i)\, l_i$$

$$\ni \quad X = \begin{Bmatrix} x_1, x_2, \ldots, x_q \\ l_1, l_2, \ldots, l_q \end{Bmatrix}$$

WE SEEK TO KEEP $\overline{L}$ AS LOW AS POSSIBLE

IT TURNS OUT

$$\overline{L} \geq H(x)/\lg r$$

8/3/76

SHANNONS FIRST THEOREM (NOISELESS CASE)

(a) THE AVERAGE LENGTH OF A U.D. CODING SCHEME CANNOT BE REDUCED BELOW $H(x)/\lg_2 r$.
(= H(x) FOR BINARY ALPHABET)

$$\bar{L} \triangleq \sum_{i=1}^{q} p(x_i) l_i$$

q = TOTAL # OF MESSAGES
l_i = LENGTH OF i^{TH} MESSAGE
x_i = PROB OF OCCURANCE OF x_i MESSAG

PROOF: CONSIDER p_i & q_i

$$\Rightarrow \sum_i p_i = \sum q_i = 1$$

WE HAVE PROVED THAT

$$-\sum_{i=1}^{n} q_i \lg q_i \leq -\sum_{i=1}^{q} q_i \lg p_i$$

LET $q_i = r^{-l_i} / \sum_{i=1}^{n} r^{-l_i}$

IT IS CLEAR $\sum_i q_i = 1$

$$\therefore -\sum_i^n p(x_i) \lg p(x_i) = H(x)$$

$$\leq -\sum_{i=1}^{q} P(x_i) \lg \frac{2^{-l_i}}{\sum_{i=1}^{q} r^{-l_i}}$$

$$\leq -\sum_{i=1}^{q} p(x_i) \lg r^{-l_i} + \sum_{i=1}^{q} p(x_i) \lg \sum_{k=1}^{q} r^{-l_k}$$

$$\leq \sum_{i=1}^{q} l_i p(x_i) \lg r + \sum_{i=1}^{q} p(x_i) \lg \sum_{k=1}^{q} r^{-l_k}$$

$$\leq \bar{L} \lg r + \sum_{i=1}^{q} p(x_i) \lg \sum_{k=1}^{q} r^{-l_k}$$

BUT $\sum_{k=1}^{q} r^{-l_k} < 1$ FOR U.D.C.

OR $H(x) \leq \bar{L} \lg r$

$$\therefore \bar{L} \geq H(x)/\lg r$$

ie, FOR UNIQUELY DECIFERABLE CODES THE AVERAGE LENGTH CANNOT BE VALUED BELOW $H(x)/\lg r$

IF THE RESTRICTION ON THE CODE BEING U.D. IS RELAXED, A LOWER $\bar{L}$ MAY BE ACHEIVED.

IN CASE THE LOWEST $\bar{L} = L_0$ IS NOT ATTAINABE, THE ATTAINABLE LOW AQUIRED AS UNDER:

CONSIDER A WORD (MESSAGE) OF LENGTH $l_k \ni$

① $-\lg p(x_k) / \lg r \le l_k \le \frac{-\lg p(x_k)}{r} + 1$

AVERAGE LHS OF INEQUALITY OVER $p(x_i)\ \forall i$:

$$H(x) \lg r \le \bar{L} \le 1 + H(x)/\lg r$$

NOW, LET $r=2$ & RAISE THE LHS OF ①:

$$[p(x_k)]^{-1} \le 2^{l_k}$$

OR $p(x_k) \ge 2^{-l_k}$

WHICH SATISFIES KRAFT. THUS, UNIQUELY DECODABLE CODES EXIST WHICH FOLLOW THE PATTERN OF PROBABILITIES GIVEN BY

$p(x_1) = \frac{1}{2}, p(x_2) = \frac{1}{2}, p(x_3) = 1/8\ \ p(x_4) = 1/8$

IF THE RESTRICTION OF U.D. IS ELIMINATED, A LOWER $\bar{L}$ IS POSSIBL

CONSIDER

$X_1 = 1 \quad p_1 = .4 \quad H(X) \cong 0.46$

$X_2 = 0 \quad p_2 = .4 \quad \bar{L} = 1.4$

$X_3 = 100 \quad p_3 = .3 \quad H(X)/\lg_{10} 2 = 1.53$

> 1.4

EX

$\mathcal{X} = \{ x_1, x_2, x_3, x_4, x_5, x_6 \}$

$P : \; 2^{-1} \; 2^{-2} \; 2^{-4} \; 2^{-4} \; 2^{-4} \; 2^{-4} \quad \Rightarrow \sum_i P(x_i) = 1$

	C_1	C_2	C_3	C_4	C_5	C_6
x_1	0	1	0	111	1	0
x_2	10	011	10	110	01	01
x_3	110	010	110	101	0011	011
x_4	1110	001	1110	100	0010	0111
x_5	1011	000	11110	011	0001	01111
x_6	1101	110	111110	010	0000	011111

(1) WHICH ONE OF THESE SCHEMES IN U.D.?

C_1: BLOWS KRAFT INEQ. (NOT INSTANT OR U.D)

C_2: $\frac{1}{2} + \frac{5}{8} = \frac{9}{8} \Rightarrow$ BLOWS KRAFT (" " " ")

C_3: $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64} = \frac{63}{64}$

$\Rightarrow$ THIS CODE <u>IS</u> U.D.

C_4: OK

C_5: $\frac{1}{2} + \frac{1}{2} + 4\,\frac{1}{16} = 1 \Rightarrow$ O.K

C_6: $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64} = \frac{63}{64} \Rightarrow$ OK

(2) WHICH ONES OF U.D. CODES OBSERVE THE PREFIX PROPERTY?

ONLY C3, C4, C5

(3) FIND $\overline{L}$ FOR EACH UDC:

$\overline{L_3} = 1 \cdot \frac{1}{2} + \frac{2}{4} + \frac{3}{16} + \frac{4}{16} + \frac{5}{16} + \frac{6}{16} = \frac{21}{8}$

$\overline{L_4} = 3$

$\overline{L_5} = 2$

$\overline{L_6} = 21/8$

(4) DOES ANY OF THESE CODES REACH THE LOWEST $\bar{L} \triangleq L_0$?

WELL, $\bar{L} \geq H(x)/\lg r$ (BY SHANNON'S FIRST)

$$H(x) = \tfrac{1}{2}\lg 2 + \tfrac{1}{4}\lg 4 + \tfrac{1}{4}\lg 8$$

$$= \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{3}{4} = 1\tfrac{3}{4} \text{ BITS}$$

BEST CODE IS C_5 (ie, MOST EFFICIENT)

** SPECIFIC CODING PROCEDURES

(1) SHANNON'S BINARY CODING PROCEDURE

WE HAVE SEEN THAT IF $p(x_k) \geq 2^{-\ell_k}$ THEN IT IS POSSIBLE TO ACHEIVE A REALISTIC LOW $\bar{L}$.

THUS $H(X) \leq \bar{L} \leq H(X) + 1$

STEPS

(a) ARRANGE THE ENSEMBLE IN DECREASING ORDER OF PROBS

$$p_1 \leq p_2 \leq p_3 \leq \ldots \leq p_q$$

(b) COMPUTE α_i'S $\ni$

$\alpha_1 = 0$ ALWAYS

$\alpha_2 = P(x_1)$

$\alpha_3 = p(x_2) + P(x_1) = P(x_2) + \alpha_2$

$\alpha_4 = P(x_3) + P(x_2) + P(x_1) = P(x_3) + \alpha_3$

$\vdots$

$\alpha_{q+1} = P(x_q) + p(x_{q-1}) + \ldots + P(x_2) + P(x_1) = \underline{1}$

$\Rightarrow$

(c) DETERMINE THE SET OF INTEGERS (THE SMALLEST) WHICH SATISFIES

$2^{l_i} p(x_i) \geq 1$ (AS PER THE THEM)

(d) EXPAND EACH α_i (WHICH IS IN DECIMAL FORM) INTO BINARY NOTATION UP TO l_i PLACES & NO FURTHER:

EX $X : \{x_1\ x_2\ x_3\ x_4\}$
$\{\frac{4}{10}\ \frac{3}{10}\ \frac{2}{10}\ \frac{1}{10}\}$

$\alpha_1 = 0$

$\alpha_2 = 4/10$

$\alpha_3 = 7/10$

$\alpha_4 = 9/10$

$l_1 : 2^{l_1}\ 4/10 \geq 1.0 \Rightarrow l_1 = 2$

$l_2 : 2^{l_2}\ 3/10 \geq 1 \Rightarrow l_2 = 2$

$l_3 : 2^{l_3}\ 2/10 \geq 1 \Rightarrow l_3 = 3$

$l_4 : 2^{l_4}\ 1/10 \Rightarrow l_4 = 4$

$\alpha_1 = 00 \quad \Rightarrow \quad x_1 = 00 \quad p_1 = .4$

$\alpha_2 = (0.4)_{10} = (.01)_2 \quad x_2 = 01 \quad p_2 = .3$

$\alpha_3 = (0.7)_{10} = (.101)_2 \quad x_3 = 101 \quad p_3 = 0.2$

$\alpha_4 = (0.9)_{10} = (.1110) \quad x_4 = 1110 \quad p_4 = 0.1$

$\Rightarrow$ COMPUTE $\bar{L}$ & $H(x)$ & COMPARE

8-4-76

→ HOMEWORK 4-7✓ 4-11✓

4-6*✓

1. CONSTRUCT A HUFFMANN CODE FOR THE FOLLOWING SYMBOLS. COMPARE ITS AVERAGE LENGTH, $\overline{L}$, WITH THE AVERAGE UNCERTAINTY $H(x)$.

$$\begin{Bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} & x_{11} & x_{12} & x_{13} \\ 0.2 & .18 & .10 & .10 & .10 & .061 & .059 & .04 & .04 & .04 & .04 & 0.03 & 0.01 \end{Bmatrix}$$

HUFFMAN CODES CALLED MINIMUM REDUNDANCY CODE

2. DERIVE A SHANNON CODE FOR

$$X = \begin{Bmatrix} x_1 & x_2 \\ 9/10 & 1/10 \end{Bmatrix}$$ AND FIND

THE CODES, $\overline{L}$, AND CORRESPONDING EFFICIENCIES FOR 2^{nd} & 3^{d} EXTENTION

⟹ * 3. PROVE, THAT FOR A HUFFMAN CODE:

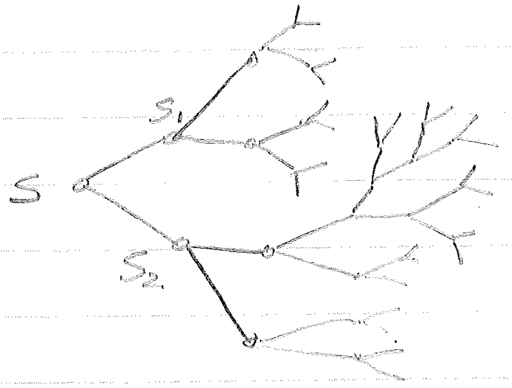
$$H(x) \leq \overline{L} \leq \overline{H}(x) + 1 - 2p_{min}$$

$> p_{min} = P[$LEAST PROBABLE MESSAGE IN THE ENSEMBLE.

SHANNON - FANO CODING SCHEME

SEPARABLE BINARY CODES:

DEALS WITH BINARY CODE TREES:



IF $X = \left\{ \begin{matrix} x_1 & x_2, & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{matrix} \right\}$

THEN IT IS DESIREABLE TO ASSOCIATE A SEQUENCE C_K OF $(0,1)$ OF UNSPECIFIED LENGTH ℓ_K FOR X_K SUCH THAT

(1) THE PREFIX PROPERTY IS OBSERVED

(2) THE XMISSION OF THE ENCODED MESSAGES IS "REASONABLY" EFFICIENT ie, 1 & 0 APPEAR INDEPENDENTLY AND WITH (ALMOST) EQUAL PROBS.

LET X:

x_1	0.25	00	(2)	
x_2	0.25	01	(2)	
x_3	.125	100	(3)	CUT 1
x_4	.125	101	(3)	#3
x_5	.0625	1100	(4)	CUT 2
x_6	.0625	1101	(4)	#4
x_7	.0625	1110	(4)	#3
x_8	.0625	1111	(4)	#4

NOTE: PROB.'S GIVEN IN THIS EXAMPLE ARE THE TYPE $P(x_a) = 2^{-\ell_k}$ IN SUCH A CASE, $H(x) = \bar{L}$ AND NO BETTER CODING CAN BE ACHIEVED (FROM SHANNON'S FIRST). IF PROBABILITIES DO NOT FALL IN THIS PATTERN, THE CODE CAN NOT BE OPTIMUM USING SHANNON-FANO.

Ex:

X	P		
x_1	0.49	0	
x_2	.14	100	#1 #3
x_3	.14	101	#2
x_4	.07	1100	4
x_5	.07	1101	#3
x_6	.04	1110	4
x_7	.02	11110	
x_8	.02	111110	
x_9	.01	111111	

CONSIDER ALPHABET BE 0, 1, 2

X	P			
X_1	$3/8$	.375	0	#1
X_2	$1/6$	.167	1 0	#2
X_3	$1/8$	.125	1 1	#1
X_4	$1/8$	.125	2 0	#2
X_5	$1/8$	.125	2 1	#2
X_6	$1/12$	.083	2 2	

$p(0) = \frac{16}{39}$ $p(1) = 13/39$ $p(2) = 10/39$

THEN FIND $H(X)$, $\bar{L}$, & η

$$\eta = H(x)/\bar{L}\,\lg_2 3$$

SHANNON'S FIRST THEOREM, PART (b)
(OR SHANNON'S NOISELESS CODING THEOREM)

- WE HAVE SEEN THAT, IF l_i IS CHOSEN $\ni \lg_r 1/P_i \leq l_i \leq \lg_r 1/P_i + 1$

THEN: $1/P_i \leq r^{l_i}$

$$P_i \geq r^{-l_i}$$

AND ALSO $H_r(S) \leq \bar{L} < H_r(S) + 1$

IF THE SOURCE WERE EXTENDED (w/o MEMORY), THEN

$$H_r(S^n) \leq \overline{L_n} < H_r(S^n) + 1$$

$$\therefore \quad n H_1(S) \leq \bar{L}_n < n H_r(S) + 1$$

$$\ni \quad \bar{L} = \sum_{i=1}^{n} P(\sigma_i)\lambda_i$$

$$\ni p(\sigma_i) = p(x_1)\ldots p(x_n) \text{ ETC.} \longrightarrow$$

$$H_r(S) \le \overline{L_n}/n < H_r(S) + \frac{1}{n}$$

AS $n \to \infty$, WE GET

$$\lim_{n\to\infty} \overline{L_n}/n = H_r(S) \Longleftarrow \left\{ \begin{array}{l} \text{SHANNONS } 1^{ST} \\ \text{PART } 1 \end{array} \right.$$

$\frac{\overline{L_n}}{n} \triangleq$ AVERAGE # OF CODE SYMBOLS USED PER SAMPOL SYMBOLS FROM THE ORIGINAL SOURCE.

NOTES:

① THE PRICE ONE PAYS FOR LARGE EXTENSIONS IS THE COMPLEXITY OF THE CODING

② EVEN THOUGH, ASYMTOTICALLY, ONE APPROACHES THE ABOVE LIMIT, THE PROCEDURE DOES NOT NECESSARILY YIELD A MONOTONICALLY INCREASING IMPROVEMENT (OR EFFICIENCY)

ie $\overline{L_n}/n \overset{\text{MAY BE}}{\ge} \overline{L_{n-1}}/n-1$ & POSSIBLY

BUT, YOU ALWAYS IMPROVE EFFICIENCY BY INCREASING n.

⟹ ③ HOMEWORK: PUT EXAMPLE ON p. 76 OF THE TEXT IN THE FRAMEWORK OF $n=2$ AND SEE IF $\overline{L_2}/2$ IS OPTIMIZED BY THE USE OF EQN. 4-10 p.72 (WHICH IS $\lg_r 1/p_i \le l_i \le \lg_r 1/p_i + 1$

HUFFMAN'S MINIMUM REDUNDANCY (OPTIMAL, OR MAX EFFICIENCY) SEPARABLE (INST) CODES.

FROM: D.A. HUFFMAN; "A METHOD FOR CONSTRUCTING MINIMUM REDUNDANCY CODES" PROC. IRE, Sept 52

RESULTS:

(1) FOR AN OPTIMUM CODING, THE LARGER CODE WORD SHOULD CORRESPOND TO A MESSAGE OF LOWER PROB. THUS, ONE SHOULD LIST THE MESSAGES IN ORDER OF INCREASING PROBABILITY:

$$P(x_1) \geq P(x_2) \ldots \geq P(x_q)$$

$$\Rightarrow L(x_1) \leq L(x_2) \ldots \leq L(x_q)$$

(2) THE LENGTH OF $l(x_{q-1}) = l(x_q)$ WHERE q = # OF MESSAGES TO BE ENCODED AND THAT, FOR AN OPTIMUM CODE, n_0, THE # OF LEAST PROBABLE MESSAGES OF EQUAL LENGTH IS GIVEN BY $q - n_0 / r-1$ = INTEGER

8-5-76 (THURS)

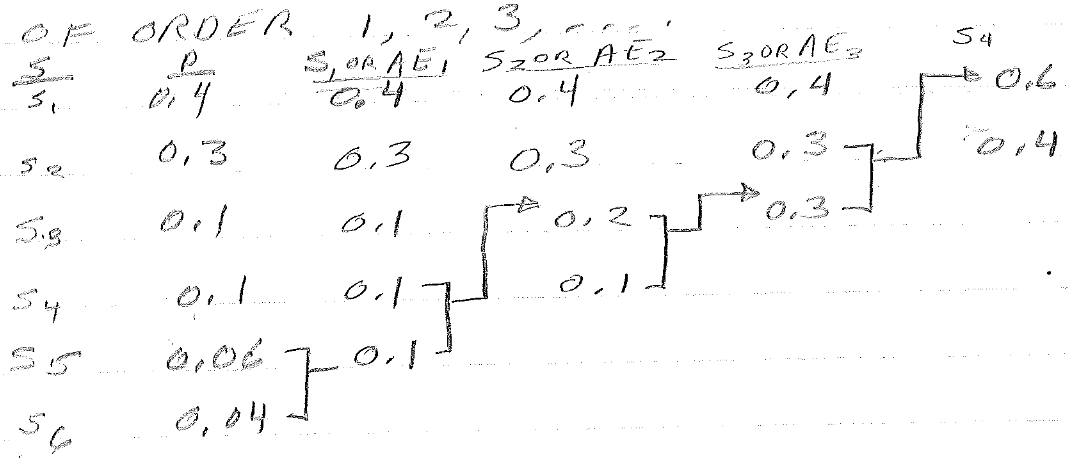
HUFFMAN CODES (CONT.)

(1) ....

(2) ....

THIS STATEMENT SUGGESTS THAT THE TOTAL # OF MESSAGES $\overset{\text{MUST}}{=} r + \alpha(r-1)$ WHERE α IS AN INTEGER

(3) SOURCE REDUCTION (AUXILERY ENSEMBLE) OF ORDER 1, 2, 3, ...



S	P	S_1 OR AE_1	S_2 OR AE_2	S_3 OR AE_3	S_4
S_1	0.4	0.4	0.4	0.4	0.6
S_2	0.3	0.3	0.3	0.3	0.4
S_3	0.1	0.1	0.2	0.3	
S_4	0.1	0.1	0.1		
S_5	0.06	0.1			
S_6	0.04				

IF r = # SYMBOLS IN ENCODING ALPHABET = 2.

WORK BACKWORDS

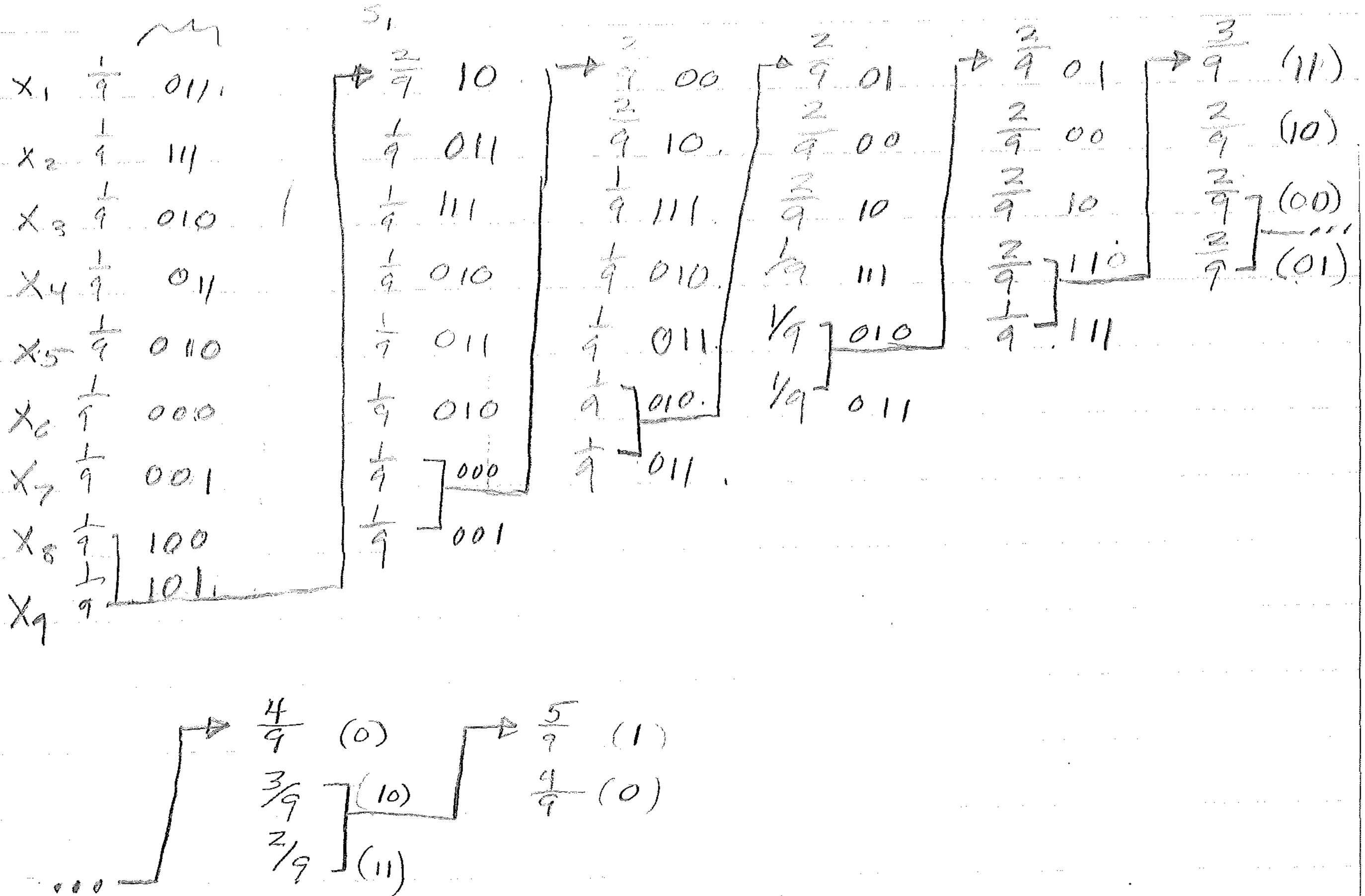
(4) CONSIDER THE J^{TH} REDUCTION S_j. ONE OF ITS MESSAGES HAS ORIGINATED BY THE MERGER OF 2 MESSAGES FROM S_{j-1}. LET THAT MESSAGE BE S_α. LET THE MERGING MESSAGES OF S_{j-1} BE S_{α_0} & S_{α_1}. OTHER MESSAGES OF S_j CAME DIRECTLY FROM S_{j-1} (W/O MERGING)

NOW, ASSIGN TO EACH MESSAGE

S_{j-1} (EXCEPT S_{α_0} & S_{α_1} WHICH ARE MERGING) THE CODE WORDS USED BY CORRESPONDING MESSAGES OF S_j. THE CODE WORDS S_{α_0} & S_{α_1} ARE FORMED BY ADDING 0 & 1 RESPECTIVELY TO THE CODE WORD USED FOR S_z.

1-5 → HW: CODE REDUCTION SAMPLE ON PREVIOUS PAGE.

Ex: $X = x_1 \ldots x_9$ & $P(x_i) = \frac{1}{9} \; \forall i$



ALSO FIND $\bar{I}$, η, $H(X)$ & REDUNDANCY

$\bar{I} = 3.22$ BITS, $\eta = 95\%$, $H(X) = 2 \lg 3 = 3.06$

NOTES:

(1) REPLACEMENT OF EQUAL PROBABILITIES IN ANY S_j IS ARBITRARY & WILL GENERATE CODES OF THE SAME $\bar{L}$ (JUST A REORDERING OF 0's & 1's)

(2) WHEN $r \neq 2$ $r > 2$, THEN THE # OF MESSAGES, q, TO BE ENCODED MUST BE $\ni$ $q = r + (r-1)\alpha \ni \alpha$ IS AN INTEGER. FOR $r=0$, <u>ANY</u> q WILL do: $q = 2+\alpha$

FOR, SAY, $r = 2$

$$q = r + (r-1)\alpha = \frac{q-4}{3}$$

$$\Rightarrow q = 7, 10, 13, 16 \text{ ETC.}$$

EXAMPLE: THE CODE FOR THE ENSEMBLE (p 84), $\hat{q} = 11$. THUS $q \overset{\text{MUST}}{=} 13$. THUS, SET $p(X_{12}) = p(X_{13}) = 0$

(3) A COMPACT CODE (DEFINED ON p. 66 OF TEXT) MAY BE MANY U.D. CODES WITH MINIMUM $\bar{L}$.

EX: THE CODE ON p80, FIG. 4.3 CAN GENERATE l_i : 1, 2, 4, 4, 4, 4

or l_i : 1, 2, 3, 4, 5, 5

<u>BOTH</u> OF THESE COMPACT CODES HAVE $\bar{L} = 2.2$ BINITS.

⇒ HW* (a) PRODUCE THE PROOF OF MINIMUM REDUNDANCY (COMPACTNESS) OF HUFFMAN CODES (p. 82-83 OF TEXT)
(b) 4-13 ON p. 92 TEXT WITH $q=8, 9$ ONLY (BASED ON NOTE 2).
USE TREES.

8-6-75 (FRI)

- SECOND TEST -

8-9-76 (MON) ([illegible])

NOISY CHANNELS

→ ERROR CORRECTING CODE

→ P[MAKING AN ERROR] = $P(e)$ OR $P[E]$

→ FANO BOUND [RELATE $H(A/B)$ BOUNDED BY FANO BOUND]

→ SHANNON'S (CELEBRATED) SECOND THEOREM

→ HAMMING'S SINGLE & DOUBLE ERROR CORRECTING CODED

*→ CODE THESE DATES:

7-4-1776

8-15-1947 (INDIA'S IND. TEST)

WITH SIMPLE (ERROR CORRECTING) CAPABILITY

ONE MAY DECIDE THE NATURE OF SYMBOL SENT BY THE OBSERVATION OF THE RECEIVED SIGNAL. CLEARLY, THERE'S AN ELEMENT OF ERROR INVOLVED IN THE ASSOCIATED PROB. (MAKING THAT ERROR) WE EVOLVE FOR DECISION RULES & TEST THEIR VALIDITY FOR A MIN $P(E)$.

RULE 1 : $p[a^*/b_j]$

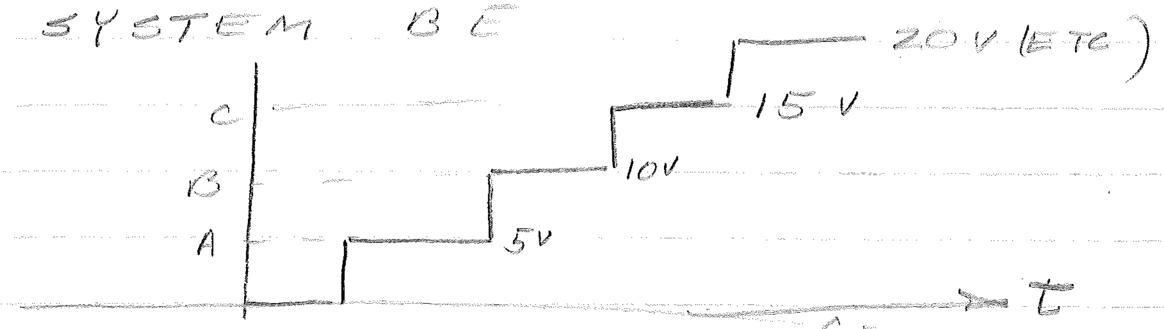
$$\begin{bmatrix} p(a_1/b_1) \\ p(a_2/b_1) \\ \vdots \\ p(a_N/b_1) \end{bmatrix}$$

PICK $p(a^*/b_j)$

$$\ni \; p(a^*/b_j) \geq p(a_i/b_j) \; \forall i \qquad ①$$

A JUDGEMENT BASED ON THIS RULE IS THE "IDEAL OBSERVER DECISION RULE"

EX: LET AN AM PULSE MOD^n SYSTEM BE



THE CHANNEL ERROR (ie AVERAGE OVER THE ENTIRE RECEIVED OUTPUT)

$= \text{AVERAGE } P(E/b_j) \text{ OVER ALL } b_j$

$= \sum_j p(E/b_j)\,p(b_j)$

$\ni \; p(E/b_j) \triangleq P[\text{MAKING AN ERROR WHEN } b_j \text{ IS RECEIVED}$

ALSO, $p(a^*/b_j) + \sum_{\substack{i \\ a^* \neq a_i}} p(a_i/b_j) = 1$

$\therefore \; p(E/b_j)$

$$= \sum_{\substack{\text{REMAINING} \\ i}} p(a_i/b_j)$$

$$= 1 - p(a^*/b_j) \qquad ②$$

ERGO, THE AVERAGE ERROR ASSOCIATED WITH THE JUDGEMENT THAT a^* WAS RECEIVED (BY RULE 1)

$$= P(E/b_j)_{a^*}$$

$$= 1 - P(a^*/b_j) \leq 1 - P(a_i/b_j)$$

FROM ② ON PREVIOUS PAGE:

$$1 - p(a_i/b_j) = P(E/b_j)_{a_i} \quad ③$$

BY BAYES RULE:

$$p(a^*/b_j) = \frac{p(a^*)\,p(b_j/a^*)}{p(b_j)}$$

$$\geq \frac{p(a_i)\,p(b_j/a_i)}{p(b_j)}$$

(FROM ①)

IF ALL INPUTS ARE EQUALLY LIKELY;

$$P(a_i) = 1/q$$

WHERE q = # OF INPUT SYMBOLS, THEN

$$p(a^*/b_j) = \frac{P(b_j/a^*)}{q\,p(b_j)}$$

& ③ BECOMES

$$P(E/b_j)_{a^*} = 1 - \frac{p(b_j/a^*)}{q\,P(b_j)} \quad ④$$

AVERAGING ④,

$$P(E) = \Sigma_j \; p(b_j) \left[1 - \frac{p(b_j/a^*)}{q P(b_j)} \right]$$

$$= 1 - \frac{1}{q} \Sigma_j \; p(b_j/a^*).$$

$= P$[CHANNEL ERROR BASED ON IDEAL OBSERVOR SCHEME]

$\neq 0$ GENERALLY.

EX

CONSIDER THE ABOVE FOUR SYMBOLS BEING INJECTED INTO THE NOISY CHANNEL & LET THEM ALL BE EQALLY LIKELY.

$$p(A) = P(B) = p(C) = P(C) = \frac{1}{4}$$

ALSO, LET THE NOISE BE ∋

$$p(\text{NOISE} < 0) = 0.$$

$$p[0 \leq \text{NOISE} \leq 2.5] = 3/4$$

$$p[2.5 \leq \text{NOISE} \leq 7.5] = 3/16$$

$$p[7.5 \leq \text{NOISE} \leq 12.7] = 1/16$$

$$P[\text{NOISE} > 12.5] = 0$$

RULE 2 (THE MAXIMUM LIKELIHOOD RULE)

$$p(b_j/a^*) \geq p(b_j/a_i)$$

	$p(b_j/a_i)$	b_1	b_2	b_3	b_4
$\frac{1}{4}$	a_1	$3/4$	$3/16$	$1/16$	0
$\frac{1}{4}$	a_2	0	$3/4$	$3/16$	$\frac{1}{16}$
$\frac{1}{4}$	a_3	0	0	$3/4$	$1/4$
$\frac{1}{4}$	a_4	0	0	0	1

LET US CALL THIS $P(B/A)$.
NOW, $P(E)$ FOR ENTIRE CHANNEL IS

$$P(E) = 1 - \frac{1}{q}\sum_{j=1}^{4} p(b_j/a^*)$$

$$P(b_1/a_1) = P(b_1/a^*) = 3/4$$

RECEIV	ASSOC.
b_1	a_1
b_2	a_2
b_3	a_3
b_4	a_4

$$\Rightarrow P(E) = 1 - \frac{1}{4}\left[\left(\frac{3}{4} + \frac{3}{4} + \frac{3}{4} + 1\right)\right] = 3/16$$

WHEN THE INPUT SYMBOLS TO THE CHANNEL ARE EQUALLY LIKELY, THEN $P(E)$ IS THE SAME BASED ON RULE 1 OR RULE 2.

EXAMPLE:

	b_1	b_2	b_3	b_4
a_1	3/4	3/16	1/16	0
a_2	0	1/4	3/4	0
a_3	0	3/4	0	1/4
a_4	0	0	1/4	3/4

b_j	a
b_1	a_1
b_2	a_3
b_3	a_2
b_4	a_4

$$P(E) = 1 - \frac{1}{4}\left[\frac{3}{4} + \frac{3}{4} + \frac{3}{4} + \frac{3}{4}\right]$$
$$= \frac{4}{16}$$ (A BIT LARGER)

NOTE: THE PROBABILITY OF ERROR IS A FUNCTION OF THE CHANNEL. SINCE THE CHANNEL IS CHARACTERIZED BY $I(X;Y)$ AND C, ONE WOULD EXPECT P(E) TO BE SOME FUNCTION OF THE CHANNEL CAPACITY. THE IDEA LEADS TO SHANNON'S SECOND THEOREM -

THUS, THERE MUST BE A CONNECTION TWIXT P(E) AND THE RATE INFO. IS SENT ACCROSS THE CHANEL. THIS "RATE" IS CHARACTERIZED BY $H(A/B)$ (THE EQUIVOCATION)

THE FANO BOUND EXPLORES THAT CONNECTION.

$$p[E/b_j] = 1 - P(a_i/b_j) \quad ①$$

$$= \sum_{i \neq j} p(a_i/b_j)$$

THE CONDITIONAL ENTROPY

$$H(A/B) \triangleq \sum_j H(A/b_j)\, P(b_j) \quad ②$$

$$\Rightarrow H(A/b_j) = -\sum_i p(a_i/b_j) \lg p(a_i/b_j)$$

$$= -P(a_j/b_j) \lg P(a_j/b_j) \quad ③$$

$$-\sum_{i \neq j} p(a_i/b_j) \lg p(a_i/b_j)$$

$$= -[1 - P(E/b_j) \lg \{1 - p(E/b_j)\}]$$

$$- \sum_{i \neq j} p(a_i/b_j) \lg p(a_i/b_j) \quad ⑤$$

ADD & SUBTRACT $p(E/b_j) \lg p(E/b_j)$ TO ⑤

$$\Rightarrow -[1 - p(E/b_j) \lg \{1 - p(E/b_j)\} + p(E/b_j) \lg P(E/b_j)]$$

$$- P(E/b_j) \Big(\underbrace{\sum_{i \neq j}}_{(m-1)\ \text{SYMBOLS}} P(a_i/b_j) / P(E/b_j) \lg \frac{p(a_i/b_j)}{P(E/b_j)} \Big)$$

$\uparrow$ MAXIMUM OF $\lg(m-1)$

$$\therefore H(A/b_j) \leq -[1 - P(E/b_j)] \lg \{1 - P(E/b_j)\}$$

$$- P(E/b_j) \lg P(E/b_j)$$

$$- P(E/b_j) \lg (m-1)$$

TURNS OUT:

$$H(A/B) = H[P(E/B)] + P(E) \lg(m-1)$$

THIS IS THE "FANO BOUND"

8-10-76

GRADES ON TEST #2

$^{-}9$, $^{-}12$, $^{-}14$, $^{-}14$, $^{-}17$, $^{-}19$, $^{-}21$, $^{-}22$, $^{-}25$, $^{-}29$

8-11-76

1) FANO BOUND

2) SHANNON'S SECOND THEOREM (NOISY CHANNEL)

3) ERROR-CORRECTING CODE - HAMMING CODE

FANO BOUND EQUIVOCATION

$$H(A/B) \leq H[p(E)] + p(E)\,\lg(m-1)$$

$\Rightarrow m =$ # OF OUTPUT SYMBOLS

- $H(A/B)$, THE EQUIVOCATION, IS
  - A MEASURE OF THE INFORMATION LOST THRU THE CHANNEL
  - AVERAGE ADDITIONAL INFO. NEEDED TO DETERMINE WHICH a_i WAS SENT.
- $H(P(E))$ IS THE AVERAGE AMOUNT OF INFO REQUIRED TO FIGGER OUT IF AN ERROR HAS OCCURED.

$$H[P(E)] = H[P(E), P(\bar{E})]$$
$$= p(E)\,\lg P(E) - p(\bar{E})\,\lg P(\bar{E})$$

- $P(E)\lg M-1$ MAY BE INTERPRETED AS THE ADDITIONAL INFO NEEDED TO FIGGER OUT WHICH OF THE REMAINING $(m-1)$ SYMBOLS HAS BEEN ERONEOUSLY RECEIVED.

WE ALSO KNOW THAT $H(A/B)$ IS RELATED TO $I(A;B)$ WHICH WHEN MAXIMIZED, IS THE CHANNEL CAPACITY. THEN FANO BOUND SUGGESTS A CONNECTION TWIXT $P(E)$ AND THE CHANNEL CAPACITY.

SHANNON'S SECOND THEM
(GENERAL STATEMENT)

IT IS POSSIBLE TO XMIT INFO WITH AS SMALL A $P(E)$ AS WE DESIRE, IF WE TRANSMINT INFORMATION AT A RATE < CH. CAP. (RATE IN THE SENSE OF BITS/SYMBOL, OR BITS/SEC, OR I/n)
ie, KEEP $H(\text{SOURCE}) <$ CHANNEL CAPACITY

SOME MATH PRELIMINARIES

- STERLING'S FORMULA

$$n! \simeq 2\pi e^{-n} n^{n+\frac{1}{2}} \quad \text{FOR } n \text{ LARGE}$$

ONWARD →

SHANNON'S SECOND THEORM FOR NOISY B.S.C.
(USING INGEL'S SIMPLIFIED PROOF)

LET THE SOURCE HAVE m MESSAGES: $(s_1, s_2 \ldots s_m)$ ASSUMED XMITTED INDEPENDENTLY AND EQUALLY LIKELY. EACH MESSAGE, s_r, MAY BE A SEQUENCE OF a_k's $\ni s_r = a_1 a_2 \ldots a_p \ni$ THE LENGTH q OF THE MESSAGE S_r MAY BE VERY LARGE. THEN THE SAME ENTROPY $= H(S) = \lg m$. ALSO, LET US USE SHANNON'S RANDOM CODING TECHNIQUE. i.e., IF EACH SEQUENCE IS n DIGITS LONG, $\exists$ 2^n SEQUENCES, ASSIGN ANY OF THE ABOVE S_i's TO ANY ONE OF THE 2^n POSSIBLE SEQUENCE EACH OF LENGTH n. CLEARLY

$$2^n \geq m \Rightarrow n \geq \lg_2 m$$

AND THE PROB THAT A PARTICULAR n DIGIT CODE WORD (MESSAGE) OUT OF A POSSIBE 2^n CODE WORDS WILL BE CHOSEN WHEN a_i WAS SENT TO REPRESENT A SPECIFIC S_i OUT OF m POSSIBLE MESSAGES $= m/2^n$

LET THE CHANNEL BE BSC ∋

$$P(y_k/x_j) \Longrightarrow \begin{bmatrix} p & \bar{p} \\ \bar{p} & p \end{bmatrix}$$

$p = P$[CORRECT XMISSION]

$\bar{p} = 1 - p = P$(INCOR. XMISSION]

THE CRITERION OF DECISION:

FIND THE SAME MESSAGE WHICH DIFFERS FROM THE RECEIVED MESSAGE IN THE LEAST # OF BINARY DIGITS.

[IDEA OF HAMMING DISTANCE]

FROM QUIZ 1, THE PROB OF MAKING r ERRORS IN A STRING OF n DIGITS, AS PER THE ABOVE ASSIGNMENT OF

PROB ERR = $({}^{n}C_{n-1})\, p^{n-r} q^{r};\ r = 0, 1 \ldots n$

$$\Rightarrow\ {}^{n}C_{n-r} = {}^{n}C_{r} = \frac{n!}{(n-r)!(r!)}$$

THE EXPECTED VALUE OF r (IN A SEQUENCY OF n SYMBOLS)

$$= \bar{r} = \sum_{r=0}^{n} r \binom{n}{n-r} p^{n-r} q^{r}$$

$$= \sum_{r=1}^{n} r \binom{n}{r} p^{n-r} q^{r}$$

$$= nq = n(1-p)$$

DUE TO RANDOM ASSIGNMENT, THE # OF CODE WORDS THAT DIFFER FROM A RECEIVED SYSTEM IS EXACTLY r DIGITS IS $\binom{n}{r}$ AND:

M = TOTAL # OF CODE WORDS THAT DIFFER FROM A RECEIVED SEQUENCE BY nq SYMBOLS

$$= \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{nq}$$

$$= \sum_{i=0}^{rq} \binom{n}{i} \quad \Longleftarrow$$

UTILIZE STERLING'S FORMULA:

FOR n 'BIG', $n! \simeq \sqrt{2\pi}\, e^{-n} e^{-n} n^{n+\frac{1}{2}}$

FOR $q < \frac{1}{2}$ (WHICH IS TRUE FOR A WORTHWILE CHANNEL $(q = \bar{p} = 1-q)$.

THEN, $\binom{n}{nq-1} < \binom{n}{nq}$

$\binom{n}{nq-2} < \binom{n}{nq-1}$

$$\therefore M \le (nq+1)\binom{n}{nq} = \frac{(nq+1)(n!)}{(nq)!\,(n-nq)!}$$

INVOKING STERLING:

$$M \le \sqrt{\frac{(nq+1)^2}{2\pi\, npq}}\; q^{-nq}\, p^{-np}$$

NOW OF THE M SEQUENCES THAT, ON THE AVERAGE, CAN BE CONSIRED AS POSSIBLE MESSAGES THAT WE SENT, ONLY ONE IS CORRECT AND (M-1) ARE WITH ERRORS. SINCE WE HAVE IN POSSIBLE MESSAGE & 2^n POSSIBLE CODEWORDS. ∴ THE

P[OF AN n-DIGIT SEQUENCE SELECTED @ RANDOM CORRESPONDS TO ONE OF THE m MESSAGES]

$$= m/2^n$$

SINCE WE GOT M TOTAL POSSIBLE RECEIVED SEQUENCES EACH OF WHICH HAS A PROBABILITY OF CORESPONDENCE TO ARE OF THE RANDOMLY PICKED n-DIGIT SEQUENCE.

∴ THE EXPECTED # OF MESSAGES THAT COULD BE CHANGED BY XMISSION ERROR (AND BE CONFUSED ?) AS CORRECT

$$\therefore N = \frac{Mm}{2^n} \le m\, 2^{-n} \sqrt{\frac{(nq+1)^2}{2\pi npq}}\, q^{-np}\, p^{-np}$$

$$\le m \sqrt{\frac{(nq+1)^2}{2\pi npq}}\, 2^{-n(1+p\lg p \cdot q\lg q)}$$

BOILS DOWN TO $H(x) = C - \frac{\lg n}{n}$

WHEN THIS HAPPENS

$$N \leq m \sqrt{\frac{(nq+1)^2}{2\pi npq}}\; 2^{-nC}$$

LET $m = 2^{nC}/n$

THEN AS $n \Rightarrow \infty$

THEN $N \Rightarrow 0$

① IT IS A CLEAR INDICATION THAT $P(E) \Rightarrow 0$ AS $n \Rightarrow \infty$.

② ALSO. $H(X) \Rightarrow C$ AN $n \Rightarrow \infty$.

FOR A MEMORYLESS CHANNEL WITH A CAPACITY C AND A DISCRETE SOURCE WITH ENTROPY H AND A # $\epsilon > 0$ IFF $0 < H < C$, IT IS POSSIBLE TO ENCODE SEQUENCES OF M-SOURCE SYMBOLS IN CODE-CODES OF LENGTH n DIGITS FOR TRANSMISSION OVER THE CHANNEL $\ni$

$$P[\text{SUCH A SEQUENCE WILL BE INCORRECTLY RECEIVED}] < \epsilon$$

IF n IS CHOSEN LARGE ENOUGH $\ni$ $m \leq 2^{nH}$. UNDER SUCH CONDITIONS, m XMITTED SEQUENCES, $(u_1, u_2 \ldots u_m)$ WILL BE RECEIVES AS $(v_1, v_2 \ldots v_m)$ $\ni$

$p(u_0^* / v_0^*) \geq 1 - \epsilon$ $\ni$ "$*$" DENOTES ASSUMED SENT SYMBOL.

THE CONVERSE

IF WE TRY TO SEND INFO THRO THE CHANNEL AT A RATE $H(x) > C$, THEN IT IS NOT POSSIBLE TO ENCODE THE MESSAGE ALPHABET SO THAT DETECTION CAN BE ACCOMPLISHED WITH ARBITRARILY SMALL PROB. OF ERROR. THE XMISSION WILL STILL TAKE PLACE, BUT THE $P(E)$ WILL BE $> \epsilon$.

NOTE: LIKE KRAFT & McMILLAN, SHANNON'S SECOND THEOREM DOES NOT GIVE SPECIFIC CODING TECHNIQUES BUT MERELY SUGGESTS SELECTION OF m RELATIVE TO C & H.

8-12-76 (THURS)

CHANGE LAST "*" INTO

* ⟹ "*" 7-4-1776

* ⟹ "*" 8-15-1974 ← HAMMING'S SECOND ERROR CORRECTING CODE

IF 11TH DIGIT FROM THE RIGHT HAS BEEN RECEIVED ERRONEOUSLY, DETECT & CORRECT CODE. (12$\overset{\downarrow}{7}$)

* ⟹ HANDOUT

ABOVE PROBLEMS DUE IN CLASS ON 8/16/76

CORRECTIONS IN SHANNON'S SECOND THEM

1 $m \geq 2^{nH}$ FOR THE PROOF FOR GENERAL CHANNEL

m = # OF MESSAGES

ERROR DETECTION

(I) HAMMING DISTANCE BETWEEN TWO MESSAGES A & B.

$\triangleq$ THE # OF DIGITS IN WHICH A & B DIFFER

EX: (1) A : 100

B 101 $\longrightarrow d(A,B)=1$

(2) A $\boxed{0}01$ $\longrightarrow d(A,B)=2$

B 100

PROBERTIES OF d:

$d(A,B)$ CAN BE $= d(B,C)$

$d(A,B)=0$ IFF $A=B$

$d(A,B)+d(A,C) \geq d(A,C)$

$d(A,B)=d(B,A)$

(II) PARITY CHECK

TWO CASES: $\begin{cases} \text{EVEN 1} & \text{EVEN 0} \\ \text{ODD 1} & \text{ODD 0} \end{cases}$

01100111110011

OF 1'S = 9 $\Longrightarrow$ THIS MESSAGE DOES NOT SATISFY EVEN 1 PARITY

OF 0'S = 5 $\Longrightarrow$ EVEN 0'S <u>NOT</u> SAT

ODD 1'S SATISFIED

ODD 0'S "

WE WILL WORK ALL OUR PROBLEMS IN EVEN 1 PARITY.

THIS CHECK CAN BE USED TO DETECT ERRORS.

EX: LET M = # OF MESSAGE BITS = 7, AND LET THE MESSAGE BE

01 0111 0

USING EVEN 1 PARITY ⇒ 0 SINCE 41 1'S MAKE UP EVEN PARITY CHECK. THUS, WE SEND 01011100, AN 8 BIT WORD, &

IF THE RECEIVED MESSAGE HAS PARITY, WE SAY THAT NO ODD # SINGLE ERROR HAS OCCURED (1, 3, 5 ... ERRORS HAVE OCCURED & ARE DETECTED).

III. LET THE CODE BOOK CONSIST OF MESSAGES BE

A = 000

B = 111

$d(A; B)$

LET THE FOLLOWING BE RECEIVED:

$001 \xrightarrow{\text{ASSIGN TO}} 000$

$010 \longrightarrow 000$

$100 \longrightarrow 001$

$011 \longrightarrow 111$

$101 \longrightarrow 111$

$110 \longrightarrow 111$

THUS, WE CAN SAY THAT THE DISTANCE d IN ⓔ, SUCH ERROR CAN BE DETECTING

DISTANCE TWIXT WORDS	CODING TYPE
1	NO DETECTON OR CORRECTION.
2	1 ERR & DESCRIBED
3	1 " " CORRECTIVE
4	1 " " CORRECT & DETECT 2 ERRORS
5	2 ERR CORRECTED (AND, OF COURCE DETECTED)
6	— DETECTED
7	3 —

IN GENERAL, $d(A,B) \geq 2q+1$ IN ORDER TO CORRECT q ERRORS IN A STRING OF n SYMBOLS. IT IS CLEAR FROM THE SIMPLE EXAMPE ~~THAT~~ A FINITE # (>1) OF MESSAGES (EACH n SYMBOLS LONG) GOTTA BE RESERVED FOR EACH MESSAGE (n DIGIT LONG) SENT.

IV. THE MAX # OF WORDS (MESSAGES) IN A CODE BOOK ∋ n = LENGTH OF EACH METHOD & THE BINARY ALPHABET (0,1) AND THE # OF ERRORS THAT CAN BE TOLERATED = q.

WE HAVE SEEN (WED) THAT M = # OF WORDS OF LENGTH n THAT DIFFER FOR A SPECIFIC SQUARE (OF LENGTH n) BY q DIGITS OR LESS

$$= \sum_{i=0}^{q} \binom{n}{i}$$

ALL OF THE M SEQUENCES MUST TO ASSIGNED TO ONE CODE WORD OF THE CODE BOOK.

IF ∃ r WORDS IN THE CODE BOOK, WE MUST HAVE rM POSSIBLE WORDS @ THE OUTPUT. $\therefore rM \leq 2^n$

OR, $r \leq 2^n / \sum_{i=1}^{q} \binom{n}{i}$

THIS IS A NECESSARY CONDITION ON r:

IS IT SUFFICIEN(?)

—No!

EX: $n=4$ & $q=1$

$r \le \sum_{i=1}^{q} \binom{4}{i} \le 3.2$

TRY $r=3$

$r=3$. ie, IF THIS IS A SUFFICIENT CONDITION, WE SHOULD TO BE ABLE TO PICK OUT 3 WORDS OUT OF A POSSIBLE 16 (EACH 4 DIGITS LONG).

d(ANY PAIR = 0) = (AT LEAST 3)

<u>BUT</u> IT DON WORK. THUS, THIS IS <u>NOT</u> A SUFFICIENT CRITERIA.

HAMMING'S PAPER: <u>ERROR</u> <u>CORRECTION</u>

(R.W. HAMMING BSTJ VOL 29, 1950, p147-50)

(1) HAMMING INTRODUCED REDUNDANT DIGITS (K IN #) TO A MESSAGE (m DIGITS LONG) $\ni$ THE # OF DIGITS IN A XMITED MESSAGE $= n = m+k$ & CALLS K BITS THE PARITY CHECK BIT

e.g. $(1854)_2 = 11100111110$, $m=11$

(2) THE K BITS ARE APPROPRIATELY POSITIONED $\ni$ THEY SATISFY CERTAIN (EVEN) PARITY 1 CHECKS AND ALSO GIVE THE DECIMAL LOCATION OF THE SINGLE ERROR. THE K BIT CHECKING # MUST HAVE ENOUGH POSSIBLE "STATES" TO

IDENTIFY IF AN ERROR HAS OCCURE IN ANY LOCATION OR NO ERROR @ ALL. K BINITS HAVE 2^K STATES. THUS

$$2^K \geq m+k+1$$ FOR NO ERROR

THUS, FOR A GIVEN m, K CAN BE EVALUATED. IF m=12, K=4

⟹ HOMEWORK

m	$K \ni 2^K \leq m+k+1$	n
1		
⋮		
10		

(3) WHERE DO WE PLACE K PARITY DIGITS IN n LOCATIONS? PLACE ONE OF THE PARITY BITS, SAY P_0, IN THE FIRST (OR LEAST SIGNIFICANT OR RIGHTMOST) POSITION. PUT ANOTHER, P_1, NEXT TO LEAST SIGNIFICANT. PLACE P_j IN THE 2^j TH POSITION FROM THE RIGHT

(4) P_0 WILL BE THE ONLY PARITY BIT WHICH WILL PARTICIPATE IN THE EVEN 1 CHECK FOR LOCATION 1, 3, 5, 7, 9, 11, ...

P_1 ... FOR 2, 3, 6, 7, 10, 11, 14, 15 ...

P_2 ... " 4, 5, 6, 7, 12, 13, 14, 15 ...

P_3 ... " 8, 9, 10, 11, 12, 13, 14, 15 ...

⋮

BIT POSITION	BINARY EQUIV	P_0 CHECKS	P_1	P_2	P_3
1	0001	✓			
2	0010		✓		
3	0011	✓	✓		
4	0100			✓	
5	0101	✓		✓	
6	0110		✓	✓	
7	0111	✓	✓	✓	
8	1000				✓
9	1001	✓			✓
10	1010		✓		✓
11	1011	✓	✓		✓
12	1100			✓	✓
13	1101	✓		✓	✓
14	1110		✓	✓	✓
15	1111	✓	✓	✓	✓
⋮	$P_3 P_2 P_1 P_0$				

THUS, SINCE EACH P_i OCCURS ONCE IN A PARITY CHECK, THAT P_i CAN BE DETERMINED EASILY.

⑤ AFTER CONDING (ENCODING) TRANSMIT M+1 DIGITS AS ONE MESSAGE, APPLY THE K PARITY CHECKS AND EVOLVE THE BINARY WORD $(P_K \ldots P_1 P_0)$ AND READ IT IN DECIMAL NOTATION. IF IT IS $\neq 0$, IT IS THE LOCATION OF THE FAULTY DIGIT. INVERT IT.

EX

IT IS CLEAR THAT FOR $m=11$, $k=4$.
LETS CODE $(1854)_2$

						0 ↓								
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
							P_3				P_2		P_1	P_0

$(1854)_2 = 11100111110$

RUN PARITY (EVEN 1) ON

$P_0 : P_0 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \oplus 0 \oplus 1 \oplus 1 \Rightarrow P_0 = 1$

$P_1 : P_1 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \oplus 0 \oplus 1 \oplus 1 \Rightarrow P_1 = \underline{1}$

$P_2 : P_2 \oplus 1 \oplus 1 \oplus 1 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \Rightarrow P_2 = 0$

$P_3 : P_3 \oplus 1 \oplus 1 \oplus 0 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \Rightarrow P_3 = 1$

NOTE 1: IN SOLVING 7-4-1776
WE GOT M=11, K=4 FOR 1776
FOR 7 $\underbrace{M=3,\ K=3}_{\text{NO GOOD. USE } M=4}$ FOR 7

∴ CHOOSE M=11 UNIFORMLY
FOR ALL PARTS OF THE
DATE

NOTE 2: HAMMING CODE ALSO
CORRECTS AN ERROR IN
A PARITY BIT.

8-12-76

1 - FOR THE SAKE OF UNIFORMITY

$7-4-76$

$m = 11, 11, 11$

$n = 15, 15, 15$

2 - THE HAMMING CODE WILL CORRECT THE ERRORS IN LOCATIONS OCCUPIED BY PARITY DIGITS

3 - THE "FIGURE OF MERIT" FOR A B.S.C AND A HAMMING CODE

$$\left.\begin{array}{l} P(\text{ERROR}) = P \\ P(\text{NO ERROR}) = \bar{P} \end{array}\right\} \text{FOR A SINGLE DIGIT}$$

IF n IS THE # OF DIGITS IN A CODE WORD $\ni np \ll 1$ $(n = 10, \frac{1}{100} = p)$, THEN

P[RECEIVING AN INCORRECT WORD TEN DIGITS LONG W/O HAMMING CODE] =

$$= \binom{n}{1} p^1 (\bar{p})^{n-1} + \binom{n}{2} p^2 \bar{p}^{n-2} \ldots$$

$$= 1 - (1-p)^n$$

$$\approx np$$

THE P[RECEIVING AN INCORRECT WORD AFTER APPLYING HAMMINGS SINGLE ERROR CORRECTING CODE]

$$= \binom{n}{2} p^2 (1-p)^{n-2} \ldots \binom{n}{n} p^n \bar{p}^0$$

$$= 1 - (1-p)^n - np(1-p)^{n-1}$$

$$= \frac{n}{2}(n-1) p^2 + \ldots$$

$$< np$$

ie, THE HAMMING CODING HAS REDUCED THE PROB (RECEPTION OF AN INCORRECT WORD.

ON THIS BASIS

$$\text{FIGURE OF MERIT} \triangleq \frac{1-(1-p)^n}{1-(1-p)^n - np(1-p)^{n-1}}$$

FOR MESSAGES n DIGITS LONG

4. LET N = # OF MESSAGES TO BE ENCODED. THE # OF INFORMATION DIGITS m IS THE SMALLEST DIGIT THAT IS LARGER THAN $\lg N$. NOTE THAT $2^m \leq 2^n/n+1$. THUS, A TABULATION CAN BE DRAWN:

$N \rightarrow$	4	8	16	32	64	128	256	512	1024
$m \rightarrow$	2	3	4	5	6	7	8	9	10
$n \rightarrow$	5	6	7	9	10	11	12	13	14
PARITY $k \rightarrow$	3	3	3	4	4	4	4	4	4

** GROUP CODES.

GROUP, RING, FIELD, INTEGRAL DOMAIN, LINEAR ALGEBRA (MATRICES), GEOMETRY & TOPOLOGY

A GROUP IS A SET DEFINED ON ONLY ONE OPERATION (WITH INVERSE)

(1) MUST BE CLOSED WITH REGARD TO THE OPERATION (CLOSURE PROPERTY)

$u, a, b \in G \Rightarrow a \odot b = c \in G$

(2) ASSOCIATIVE LAW

IF $a, b \in G$, THEN

$(a \odot b) \odot c = a \odot (b \odot c)$

(3) IDENTITY ELEMENTS u

$\ni \quad u \odot a = a$

(4) EACH ELEMENT MUST HAVE AN INVERSE $\ni \; a \oplus a^{-1} = u$

EX: +, AND − #'S FORM A GROUP UNDER THE OPERATION OF +

DO THEY FORM A GROUP UNDER MULT? <u>NO</u>. THE GROUP OF FRACTIONAL #'S ≠ 0 FORM A GROUP UNDER MULTIPICATION

IT IS POSSIBLE FOR TWO NUMBERS TO FORM A GROUP (0, 1), THE OPERATION BEING HALF ADDING. THAT IS

$0 \oplus 0 = 0$
$0 \oplus 1 = 1$
$1 \oplus 0 = 1$
$1 \oplus 1 = 0$

IDENTITY: 0 = U
INVERSE: EACH ITS OWN INVERSE

A FINITE GROUP IS A GROUP WITH ONLY A FINITE # OF ELEMENTS AND AN OPERATION. THE NUMBER OF ELEMENTS IS A FINITE GROUP IS ITS ORDER.

A SUBGROUP OF A GROUP G IS A SET OF ELLMENTS FROM G WHICH, BY THEMSELVES, OBSERVE ALL OF THE PROPERTIES OF THE GROUP. THE IDENTITY MUST ALSO BELONG TO THE SUBGROUP.

e.g. $G(\ldots -40, -39 \ldots 0, 1, 2, \ldots 40)$
A SUBGROUP IS
$(-14, -7, 0, 7, 14, 21 \ldots)$ ETC
(UNDER, OF COURSE, ADDITION)

A COSET: LET A GROUP G (DEFINED UNDER MULT) BE $(g_1 g_2, \ldots g_m, s_1 \ldots s_n)$ AND A SUBGROUP $\{s_1 \ldots s_n)$

$$\begin{array}{lllll}
S \rightarrow s_1 & s_2 & s_3 & s_4 & s_n \\
g_1 s_1 = g_1 & g_1 s_2 & g_1 s_3 & g_1 s_4 & g_1 s_n \\
g_2 s_1 = g_2 & g_2 s_2 & g_2 s_3 & g_2 s_4 & g_2 s_n \\
\vdots & \vdots & \vdots & \vdots & \\
g_m s_1 & g_m s_2 & g_m s_3 & g_m s_4 & g_m s_n
\end{array}$$

THIS IS CALLED A COSET ARRAY WITH THE FIRST COL (LHS) IS "COSET LEADER" EACH ELEMENT OF A COSET (A ROW) IS FORMED BY THE MULTIPLICATION OF LEADERS WITH S_i^{s}. IF THE LEADERS ARE TO THE LEFT, THE LEADERS ARE CALLED LEFT COSETS (w 1ST ELEMENT

ON THE LEFT AS THE COSET LEADER.

PROPERTIES

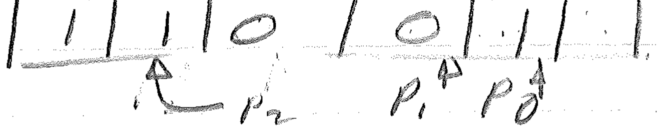
① TWO ELEMENTS, g & g' OF A GROUP G ARE IN THE SAME LEFT COSET OF A SUBGROUP S IFF $g^{-1}g'$ IS AN ELEMENT OF S.

② EVERY ELEMENT OF G IS IN ONE & ONLY ONE COSET OF A SUBGROUP S.

eg. LET $m=2$ FOR A HAMMING (SINGLE ERROR CORRECTIN CODE)

($2^k \geq n+1+k \Rightarrow k=3$)

$\therefore$ $n=n+k=5$ AND WE SHOULD HAVE 2^5 POSSIBLE WORDS. THE # OF MESSAGES CANNOT BE MORE THAN $2^m = 2^2 = 4$,



THE FOUR WORDS XMITTED ARE

$S_1 = 00000$

$S_2 = 00111$

$S_3 = 11001$

$S_4 = 11110$

LET THESE FORM THE "SUBSET" AND LET US GENERATE THE LEFT COSETS:

$S_1 = 00000$	$S_2 = 00111$	$S_3 = 11001$	$S_4 = 11110$
10000	10111	01001	01110
01000	01111	10001	10110
00100	00011	11101	11010
00010	00101	11011	11100
00001	00110	11000	11111
01100	01011	10101	10010
01010	01101	10011	10100

WORDS WITH LEAST # OF 1'S

½ ADDER $S_1 \oplus S_2 \Rightarrow \{S_2\}$

AN ARRAY FORMED WITH COSET LEADERS WITH A MINIMUM HAMMING RATE IS CALLED A STANDARD ARRAY. NOTE THAT NONE OF THE 2^5 WORDS ARE REPEATED.

* ⟹ 6-1 p 184

USE HAMMING'S SINGLE ERROR CORRECTING CODE FOR # BINITS/MESSAGE = 5 = m, K = ? [ONLY 2 CASES]. FOR BOTH CODES, COMPUTE P[INCORRECTLY ENCODING A WORD SENT W/O ERROR CORRECTION]. ASSUME BSC WITH P[MAKING AN ERROR] = 1/

ıave been prov

ormers for tran
have specified, fo
ers be ertib

ounded; the eart
in the initial stag
loped.
t-up areas is reso
ngeover to neutra
ıvestigation. Th
owever, as eart
ine poles and in
lines have to b
ervative German

ted, with the ex
may be grounded

of Hamburg will
, a provision that
through to the
hould contribute
supply.

at the IEEE Winter
970.

rungssich eit von
t, no. 293

earned the Dipl.
electrical-power
m the Technical
tuttgart, Federal
ermany, in 1958
joined the Ham
ectricitäts-Werke
e evaluator. In
erred to the util
Department and
the development
age system. In
gencies of over
development of
ms. Mr. Meinert
that is respons
uclear plant.

welt joined the
Electricitäts
mburg, Federal
rmany, after re
. Ing. degree in
eering from the
ersity of Berlin
1957. His duties
have involved
low-voltage in
etwork, the up
V und uno
nent o pply
orking on the
nning, reliabil
member of the

rk for Hamburg

Coding and its application in space communications

Once regarded as purely academic, coding theory has turned out to be eminently practical for the modern applications of space channels

G. David Forney, Jr. *Codex Corporation*

Between 1948—when Shannon first proposed his basic theorems on information theory—and the start of the space age, little practical application developed from the lessons of coding theory. This article presents an overview of the Shannon theorem, interesting practical codes, and their application to the space channel. It turns out that a simple encoder in combination with a decoder of modest complexity placed into an uncoded communications system can increase the data rate by a factor of four or more depending on the coding scheme and the allowable error rate. Use of a convolutional code with sequential decoding has proved to be the outstanding scheme for these applications. It appears that, in the future, coding will find a place in most new digital space communication systems.

Coding theory has a history no doubt unique among engineering disciplines: the ultimate theorems came first, practical applications later. For many years after Shannon's announcement of the basic theorems of information theory in 1948, the absence of any actual realization of the exciting improvements promised by the theory was a source of some embarassment to workers in the field. A standard feature of IEEE Conventions in this period was a session entitled "Progress in Information Theory," or something similar, in which the talks purporting to show that the theory was approaching practical application tended instead to confirm the prejudices of practical men that information theory would do nothing for them.

In retrospect, there were two principal reasons for this lag. First, Shannon's coding theorems were existence theorems, which showed that within a large class of coding schemes there existed some schemes—nearly all, actually—that could give arbitrarily low error rates at any information rate up to a critical rate called channel capacity. The theorems gave no clue to the actual construction of such schemes, however, and the search for coding techniques capable of remotely approaching the theoretical capacity proved so difficult that a folk theorem was proposed: "All codes are good, except those we can think of."

Second, the channels of practical interest—telephone lines, cable, microwave, troposcatter, and HF radio—proved not to have anything like the statistical regularity assumed in the proof of the coding theorems. In fact, most theorems are based on the assumption of statistical independence in the noise affecting each transmitted symbol, whereas on the channels just cited disturbances tend to be manifested in bursts spanning many bits. This is to say nothing of other anomalies that arise in practice, such as a channel described at a recent information theory symposium as "a very good channel, with errors predominantly due to a noisy Coke machine near the receiver."

Over the past decade, the situation has improved tremendously. The problem of finding workable coding schemes has been recognized to be fundamentally a problem of finding decoders of reasonable complexity. The solution has been sought in considering classes of

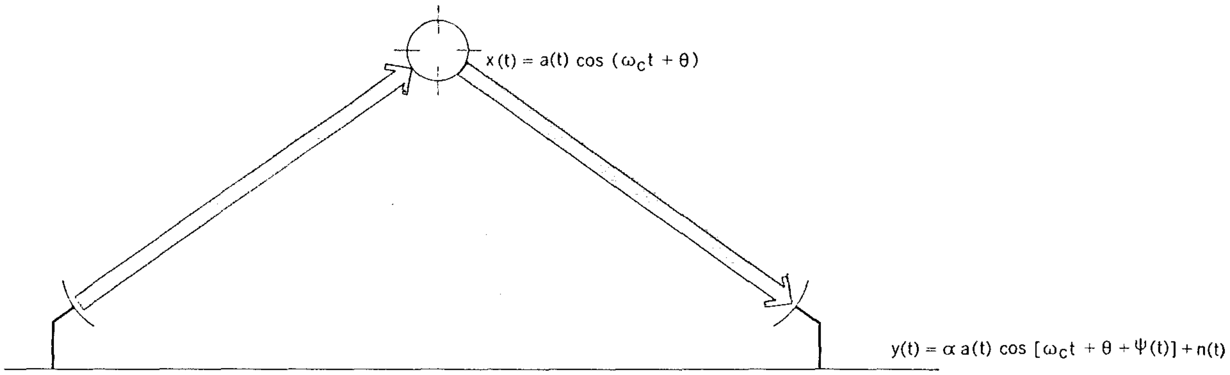
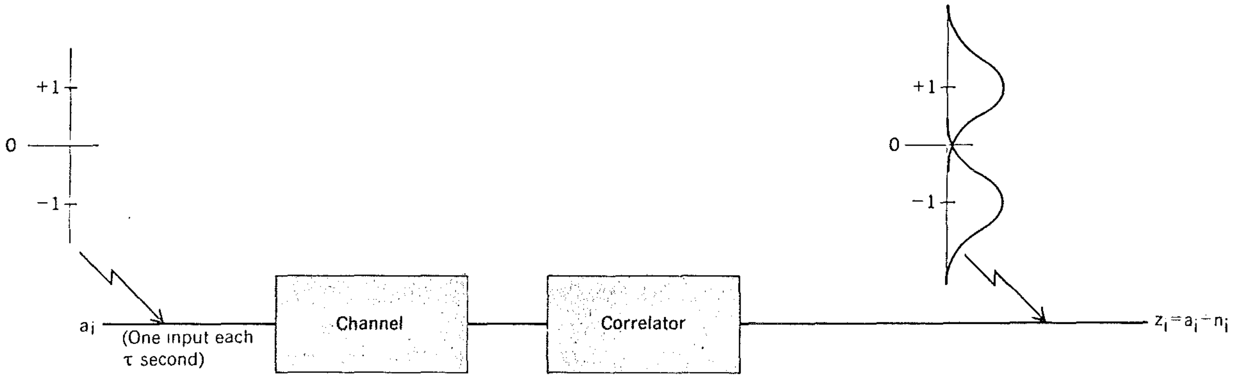


FIGURE 1. Amplitude modulation on a satellite channel.

FIGURE 2. Discrete-time channel model.



codes so structured that efficient decoding becomes feasible (but not so much structured that the codes themselves are no good). The most popular approach has been to use the structures of abstract algebra to generate classes of good, decodable block codes. A second approach uses linear sequential circuits to generate a class of codes that are called convolutional; at least for the applications to be discussed here, convolutional codes seem to have better balance between structure and randomness than is capable with the perhaps too-structured block codes.

A second major development of the last decade has been the emergence of the space channel into practical importance, both in the requirements of NASA for efficient transmission from deep-space probes, and in the proliferation of earth-orbiting communications satellites. The remarkable characteristic of the space channel is that, within the sensitivity of tests performed to date, it appears to be accurately modeled as a white-Gaussian-noise channel. Anyone who has ever taken a statistical subject knows that white Gaussian noise is the archetype of statistically regular, nonbursty noise, and as such is the theorist's dream. Consequently, in considering possible schemes for the space channel, one may use the most profound theorems, the most subtle analyses, and the most accurate simulations. One is also able to propose the most sophisticated and powerful decoding procedures, and predict performance to the accuracy of a fraction of a decibel. The initial successes of coding on the space channel have led to its incorporation in all space-system designs (of which the author is aware) in the last two years or so. For this reason, as well as the pedagogical neatness of the white-Gaussian-noise channel, this article uses the space channel for both orientation and motivation. We shall say little about the literally more mundane channels mentioned earlier, for although applications of coding have also been increasing in those environments, the schemes used are much more *ad hoc*, and more than qualitative predictions about behavior on real channels rarely can be made.

The space channel

The model of the space channel that we shall use reflects all the significant characteristics of the channel, without some details important only in practice; it is illustrated in Fig. 1. An amplitude-modulated carrier

$$x(t) = a(t)\cos(\omega_c t + \theta)$$

is generated aboard a satellite and transmitted to an earth antenna. (Frequency and phase modulation are also used, but not as often as AM, and offer no advantage in principle.) The model still applies when the signal actually originates at another ground station and the satellite is only a repeater, since the power available on the ground is so much greater than that aboard the satellite that the uplink may be considered perfect in most cases. The received signal

$$y(t) = \alpha a(t)\cos[\omega_c t + \theta + \psi(t)] + n(t)$$

is subject to several principal disturbances:

1. Simple attenuation α due to distance (assumed perfectly linear). The received signal power is denoted P.

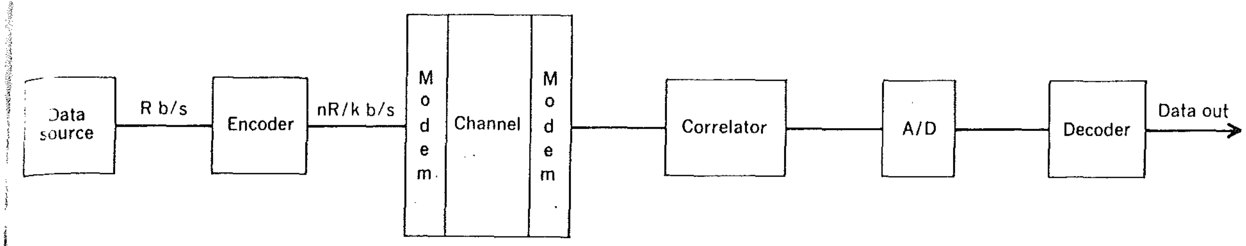


FIGURE 3. System block diagram.

2. Additive white Gaussian noise $n(t)$ arising in the receiver front end, with single-sided spectral density N_o.

3. Phase variations $\psi(t)$ due to imperfect tracking, uncompensated Doppler shifts, an unstable carrier oscillator, and so forth. In the applications with which the author is familiar, with the carrier ω_c in S band, the phase variations are the only important departure from the ideal white-Gaussian-noise model, and make themselves felt at low data rates by frustrating perfectly coherent demodulation. On a NASA mission with a terminal of the Goldstone type, phase variations can be kept to a few hertz or less, and are unimportant unless the bit rate is of the order of 10 bits per second or less. However, in some military applications where the receiver is aboard a plane, ship, jeep, or other moving platform, "low" data rates may be as high as 75 to 2400 b/s. We shall assume hereafter that we are at high enough rates that essentially perfect phase tracking and coherent demodulation can be maintained.

It will also be assumed that the information to be transmitted is already in digital form, leaving totally aside the kind of coding (source coding) that is concerned with efficient representation of the information in bits. (The gains from efficient source coding may be expected to equal or exceed those claimed in the following for efficient channel coding. The best techniques of the infant field of data compression are, however, even more *ad hoc* than those for channel coding on bursty channels.) The information rate will be denoted as R b/s.

When a communications system can pass R information bits per second over a white Gaussian channel on which the received power is P and the noise density N_o, with some acceptable quality, we say that the system is operating at a *signal-to-noise ratio per information bit* $E_b/N_o = P/N_oR$. This dimensionless parameter then serves as a figure of merit for different coding and modulation schemes. Note that it incorporates any effective power loss due to coding redundancy. A system designer who simply wants to select a communications scheme to get the most data rate for a given power and receiver noise temperature, or to use the least power for a fixed data rate, will pick the scheme that can operate at the lowest E_b/N_o with adequate quality (if he can possibly afford it).

An appropriate modulation technique, and the only one we shall consider, is pure time-discrete, N-level amplitude modulation. By this we mean that the modulating waveform $a(t)$ can only change at discrete intervals τ seconds apart, and during any τ-second period, sometimes called a baud, it can take on one of N discrete values, usually equally spaced. We let a_i be the value in the ith interval. If N is a power of two, say 2^m, then the *signaling rate* is $1/\tau$ symbols (bauds) per second, and the *transmitted rate* m/τ bits per second. Ideally, the bandwidth occupied is $W = \frac{1}{2}\tau$ hertz, but this is only an approximation (and a lower bound) to the practical bandwidth. By far the most common scheme of this class is the binary ($N = 2$) case, with $a(t) = \pm 1$; this is commonly called PSK or phase-shift keying, the terminology arising from a viewpoint in which $a(t)$ has constant magnitude 1 and the phase θ is modulated to the two values $\pm \pi/2$.

With white Gaussian noise, and perfect phase tracking, it is appropriate to use a correlation or matched filter receiver. Mathematically, in the ith baud such a receiver forms the integral

$$z_i = \int_{i\tau}^{i\tau+\tau} y(t) \cos\,[\omega_c t + \theta + \psi(t)]\, dt$$

It is easily shown that $z_i = a_i + n_i$, where a_i is the modulation amplitude (scaled) in the ith baud and n_i is the noise, a Gaussian random variable centered on 0 and independent from baud to baud. (This assumes perfect synchronization of the timing intervals, which can be approached as closely as desired in practice.) Furthermore, no information is lost in the correlation operation, in the sense that any decision on what was sent that is based on the correlator outputs z_i can be just as good as the information based on the complete received waveform. Thus we have replaced our continuous-time model with a discrete-time model, illustrated in Fig. 2 for PSK. Every τ seconds, a level a_i (one of N) is sent, and a correlator output z_i is received.

In the absence of coding, a *hard decision* is made on the correlator output as to which level was actually sent. For example, with binary PSK, a positive z_i leads to a decision of +1, and negative to −1. With coding, it is usually desirable to keep an indication of how reliable the decision was; this can range from establishing a null zone around 0, which is treated as no decision or an *erasure*, to retaining essentially all the information in the correlator output by sufficient finely quantized analog-to-digital conversion (normally three bits), called a *soft* (or *quantized*) *decision*. Schematically, any of these possibilities will be represented by a box following the correlator output labeled A/D.

We can now lay out the complete block diagram of a system that includes coding (Fig. 3). Information bits arrive at a rate of R b/s. An encoder of code rate k/n inserts $n - k$ redundant bits for every k information bits, giving a transmitted bit rate of nR/k b/s. These bits are taken m per baud into the modulator; at the receiver, a noisy correlator output is developed for each baud and A/D converted. The resulting hard decisions, soft decisions, or whatever, enter the decoder, which uses the redundancy in the data as well as (with soft decisions) the reliability of the received information to estimate which information bits were actually sent. When the signal-to-noise ratio is specified, this is a well-defined mathematical model, and it makes sense to ask the question: How much information can we transmit through this channel, and what do we put in the encoder and decoder boxes to do it? The surprising fact upon which we commented at the beginning of this article is that the answer to the first question was announced long before anyone had the remotest idea how to answer the second.

Channel-capacity statements

Shannon's original work[1] showed that the capacity of the communication system blocked out in Fig. 3 is

$$C = \tfrac{1}{2}\log_2(1 + P/N_oW) \quad \text{bits/baud}$$

$$\text{or } W\log_2(1 + P/N_oW) \quad \text{bits/second}$$

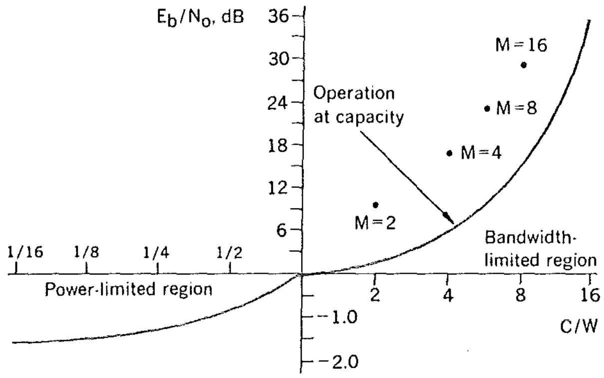
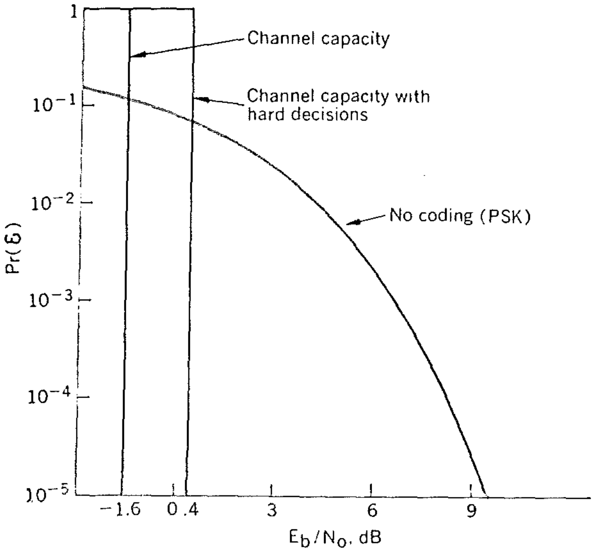


FIGURE 4. E_b/N_o as a function of C/W in bandwidth-limited and power-limited regions (note change of scale), with operation at capacity contrasted with M-level amplitude modulation [Pr (ε) = 10^{-5}].

FIGURE 5. Bit-error probability as a function of signal-to-noise ratio per information bit for situations involving no coding and coding at capacity.



where P is the received signal power, N_o the single-sided noise spectral density, and W the nominal bandwidth $\frac{1}{2}\tau$. Shannon showed that whenever the information rate R is less than C, then there exists some coding and modulation scheme with as low a decoded error probability as you like; whereas if $R > C$, then the error probability cannot approach zero and more coding generally only makes things worse. Finally, it can be shown that the same results apply when the special modulation assumptions of Fig. 3 are removed, and any signaling scheme whatsoever is allowed.

At one time, this classic formula fell into disrepute, after it had been used loosely by all sorts of coarse fellows who applied it promiscuously to channels not remotely characterized by the white-Gaussian-noise model. With the advent of the space channel, however, it is time to rehabilitate it for the insight it provides.

Suppose we could actually transmit at capacity; the signal-to-noise ratio per information bit would then be $E_b/N_o = P/N_oC$. The number of bits per cycle of bandwidth under the same conditions would be C/W. The capacity formula is usefully rewritten as a relation between these two dimensionless parameters:

$$C/W = \log_2[1 + (P/N_oC)(C/W)]$$

This relation is plotted in Fig. 4. We see that, for a fixed power-to-noise ratio P/N_o, more and more efficient communication is possible as the bandwidth is increased, and that with no bandwidth limitations, E_b/N_o approaches a limit of ln 2 ($\approx$0.69, or -1.6 dB), called the Shannon limit. To date, space communication has been characterized by severe power limitation and bandwidth to burn, so that this so-called power-limited case has been the regime of interest. We note that, although the E_b/N_o limit is reached only for infinite bandwidth, at ½ bit per cycle of bandwidth (or a code rate of about ¼ with PSK) we are practically there.

Let us now see what coding has to offer in the power-limited case. Figure 5 is a more standard curve of error probability versus E_b/N_o in decibels. The no-coding curve is that for ideal PSK, which is representative of what was in fact used in the years B.C. (before coding), as in the Mariner '64 system that returned the first pictures from Mars. We see that an E_b/N_o of 6.8 dB is required to obtain a bit error probability of 10^{-3} and 9.6 dB to obtain 10^{-5}. On the other hand, the capacity theorem promises essentially zero error probability whenever E_b/N_o exceeds -1.6 dB. This means that potential coding gains of 8 to 11 dB (a factor of 6 to 12) are possible, which is rather exciting in an environment where the cost of a decibel is frequently measured in millions of dollars. Since, in the power-limited region, R is directly proportional to P, this gain may be taken either as reduced power or as increased data rate.

Another curve of parenthetical interest is included in Fig. 5, the capacity curve when the A/D box of Fig. 3 makes hard decisions. It turns out that this costs a factor of $\pi/2$ or 2 dB. We remark on this loss here because it seems to be one of the universal constants of nature: regardless of the coding scheme, use of hard decisions rather than soft in the power-limited region always costs about 2 dB.

The situation is quite different when the channel is bandwidth-limited rather than power-limited. The following simple argument shows that, in this region, coding no longer offers such dramatic gains. Referring back to the capacity formula, we see that for $P/N_0W \gg 1$, with fixed N_0 and W, each increase by a factor of four in P leads to an increase of 1 bit/baud in channel capacity. On the other hand, consider what is required to increase the transmission rate in conventional multilevel amplitude modulation by 1 bit/baud. To double the number of signal levels while maintaining the same level separation and therefore the same probability of error requires increasing the amplitude span of the levels by a factor of two, as in Fig. 6, or the average power P by a factor of about four (this rapidly becomes exact as $N = 2^m$ increases). Thus, if R_{AM} is the rate achievable with amplitude modulation and C the capacity for some power P, then as P increases by k factors of four, we have

$$P \to 4^k P$$

$$R_{AM} \to R_{AM} + k$$

$$C \to C + k$$

$$\frac{R_{AM}}{C} \to \frac{R_{AM} + k}{C + k} \to 1 \quad \text{as } k \to \infty$$

Thus we can nearly achieve capacity without coding as we get deeper into the bandwidth-limited region. In Fig. 4, we plotted the first few AM points for $\Pr(\mathcal{E}) = 10^{-5}$ to show how rapidly R_{AM} approaches C. It therefore may be anticipated that as communications satellites achieve greater and greater effective radiated power the attractiveness of coding will diminish. One also suspects that this argument partially explains why, despite the fact that much outstanding early work on coding, including Shannon's, came out of the Bell Telephone Laboratories, to date there has been negligible operational use of coding on telephone circuits, which are engineered to be high signal-to-noise ratio, narrow-bandwidth lines. Comsat, by inheriting telephone-type tariffs that require its bandwidth to be offered in narrow slices, has been hobbled in the same way.

Maximum-length shift-register codes

In the remaining sections, we will discuss different types of codes and decoding methods, in an attempt to give an impressionistic feel for what they involve, with particular reference to performance on the power-limited space channel. We begin with block codes, which were the first to be studied and have the most well-developed theory. The maximum-length shift-register (or pseudorandom or simplex) codes are a class of codes that make a good introduction to algebraic block codes. Their properties are interesting and easy to derive, and serve as an easy entrée to the mysteries of finite fields, upon which further developments in block codes depend. Furthermore, they are actually useful in space applications and in noncoding areas as well. The number and quality of the pictures of Mars returned from the recent Mariner probes depended on the use of codes like these.

Consider first a digital feedback circuit such as the one depicted in Fig. 7; i.e., an m-bit shift register whose serial input is the modulo-2 (exclusive-or) sum of two or more of the bits in the shift register. In Fig. 7, $m = 4$ and the two bits are the rightmost b_1 and the leftmost b_4, so that the input b_{in} is expressed mathematically as

$$b_{in} = b_1 + b_4 \quad \text{modulo 2} \tag{1a}$$

or, using the notation $\oplus$ for modulo-2 addition,

$$b_{in} = b_1 \oplus b_4 \tag{1b}$$

When we say "shift register," we imply that whenever the circuit is pulsed by a clock pulse (not shown), b_{in} enters the left end, all other bits shift one place to the right, and the rightmost bit b_1 is lost.

It is well to be absolutely solid on the properties of modulo-2 arithmetic before striding off into the woods of algebraic coding theory.* Only two quantities occur in the arithmetic, 0 and 1. They may be added and multiplied as though they were ordinary integers, except that $1 \oplus 1 = 0$. This leads to the curious property that any number (0 or 1) added to itself in this arithmetic "cancels," i.e., equals zero, so that each number can be regarded as the negative of itself, and addition and subtraction are indistinguishable. (For example, if $a = b \oplus c$, then $b =$

I. Modulo-2 arithmetic

Addition

+	0	1
0	0	1
1	1	0

Multiplication

×	0	1
0	0	0
1	0	1

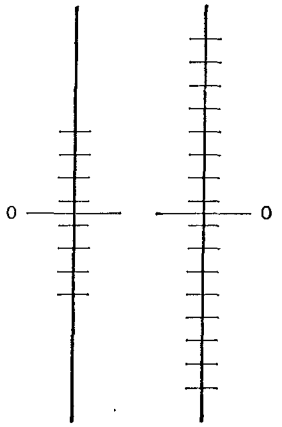
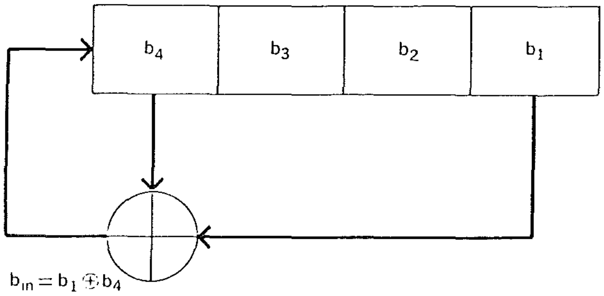


FIGURE 6. Doubling the number of levels with the same level spacing requires quadrupling the power in pulse amplitude modulation.

FIGURE 7. Maximum-length shift-register sequence generator with m = 4 stages.



---

* In general, the operations of modulo-N arithmetic (N equal to any integer) are the same as those of ordinary arithmetic after every number is reduced to its remainder when divided by N. For example, 8 modulo 3 is 2.

$a \ominus c = a \oplus c$.) Addition and multiplication tables are given explicitly in Table I. It is easy to verify that all the ordinary rules of arithmetic—i.e., $a + b + c = c + b + a$, $a(b + c) = ab + ac$, etc.—apply in modulo-2 arithmetic, so that we can manipulate symbolic expressions freely, just as though they involved ordinary numbers, with the additional rule that $a + a = 0$.

Return now to the feedback circuit of Fig. 7. What happens when it is shifted a number of times? The answer clearly depends on what its initial contents are. If all stages initially contain zeros, then the input will be zero, so that a shift will leave the register in the all-zero state. There are 15 other initial states; if we pick one of them, say 0001, and use Eq. (1), we find that 15 shifts cycle the register through all nonzero states and return the register to the starting point. The state diagram is shown in Fig. 8; it consists of two cycles: the one-state all-zero cycle, and the 15-state nonzero cycle. The name "maximum-length shift register" is given to this circuit since, given that 0000 must go to 0000, the 15-state cycle is the maximum length possible.

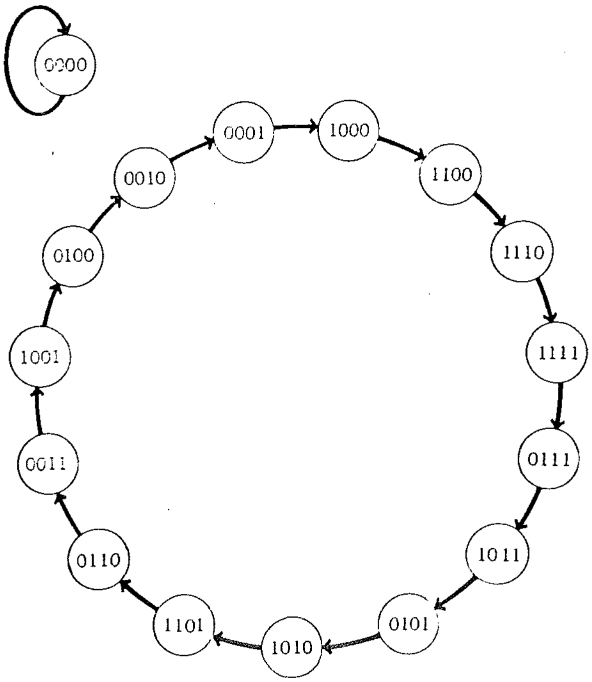
It is a nontrivial result of algebra that for any number of stages m we can always find a circuit like Fig. 7 with a state diagram like Fig. 8. The input is always a modulo-2 sum of certain stages of the register, so the all-zero state always gives a zero input, and the zero state always goes into the zero state on a shift. The remaining $M - 1$ states form a maximum-length cycle, where $M = 2^m$. Table II specifies input connections to the modulo-2 adder that will give a maximum-length shift register for $1 \leq m \leq 34$.

A block code using the circuit of Fig. 7 as an encoder operates as follows: The message to be transmitted, assumed to be a sequence of bits, is segregated into 4-bit segments. Each segment is loaded into the 4-bit shift register, and the register is shifted 15 times. The 15 bits coming out of the rightmost stage of the register are transmitted as a block, or code word. Table III gives the 15-bit code words corresponding to each 4-bit information segment.

This code is called a (15, 4) code, since code words have 15 bits for each 4 information bits. By using registers of different lengths m, we can create $(M - 1, m)$ codes. Since $M = 2^m$, as m gets large, the ratio of information bits to transmitted bits (the code rate) becomes very small, which limits the usefulness of these codes for coding purposes; in other applications, however, the fact that a very long nonrepeating sequence can be generated with a short register is the feature of interest.

We can quickly determine some properties of the $(M - 1)$-bit sequences generated by these registers. First, the bits in these sequences are the rightmost bits of the $M - 1$ nonzero state sequences of length m. Since exactly half of all m-bit sequences end in "1," precisely $M/2$ 1's occur in any maximum-length sequence (for example, 8 bits out of the 15 in the sequence of the example). In a long sequence, if we look at the output at a random time, the probability of seeing a "1" is $(M/2)/(M - 1)$, or just

FIGURE 8. State diagram of feedback circuit in Fig. 7.



II. Connections for MLSR generators

m	Stages Connected to Modulo-2 Adder	m	Stages Connected to Modulo-2 Adder
1	1	18	1, 12
2	1, 2	19	1, 15, 18, 19
3	1, 3	20	1, 18
4	1, 4	21	1, 20
5	1, 4	22	1, 22
6	1, 6	23	1, 19
7	1, 7	24	1, 18, 23, 24
8	1, 5, 6, 7	25	1, 23
9	1, 6	26	1, 21, 25, 26
10	1, 8	27	1, 23, 26, 27
11	1, 10	28	1, 26
12	1, 7, 9, 12	29	1, 28
13	1, 10, 11, 13	30	1, 8, 29, 30
14	1, 5, 9, 14	31	1, 29
15	1, 15	32	1, 11, 31, 32
16	1, 5, 14, 16	33	1, 21
17	1, 15	34	1, 8, 33, 34

III. Code words in a (15, 4) code

Information Bits	Code Word
0000	000000000000000
0001	000111101011001
1000	100011110101100
0100	010001111010110
0010	001000111101011
1001	100100011110101
1101	110010001111010
0110	011001000111101
1011	101100100011110
0101	010110010001111
1010	101011001000111
1101	110101100100011
1110	111010110011001
1111	111101011001000
0111	011110101100100
0011	001111010110010

bout ½. Furthermore, since all m-bit sequences except ne all-zero sequence occur somewhere in the maximum-ngth sequence, the probability of seeing a "1" given any $- 1$ or fewer preceding bits is still nearly one half. hese and other statistical properties make a maximum-ngth sequence difficult to distinguish from a sequence enerated truly randomly, as by flipping a coin, yet these equences are easy to generate and repeatable. Thus they re commonly used to generate pseudorandom bits.

The class of maximum-length shift-register codes is representative of the major classes of algebraic block odes, in that such codes have the properties of being

1. Systematic; that is, the information bits are transmitted unchanged as part of the code word. In the example (Table III), the first four bits of each code word are the information bits.
2. A parity-check code; that is, each of the noninformation (parity) bits is a parity check on (modulo-2 sum of) certain information bits. This can be proved inductively; for example, in the example code, the fifth bit is the modulo-2 sum of the first and fourth; the sixth is the sum of the second and fifth, but this is the same as the second plus the first plus the fourth; in general, the nth bit is some modulo-2 sum of previous bits, which are themselves each modulo-2 sums of information bits, so the nth bit is also some modulo-2 sum of information bits. (In fact, in the maximum-length shift-register codes, the parity bits consist of all possible different parity checks on the information bits.)
3. Cyclic; that is, the end-around shift of any code word is another code word.

The parity-check property can be used to prove the most important single result concerning parity-check codes (the group property), which is that if we form the modulo-2 sum of two code words, we get another code word.

The modulo-2 sum of two n-bit code words is defined as the bit-by-bit modulo-2 sum; that is, if x_i and y_i, $1 \leq i \leq n$ are the bits in the two original code words, then the bits z_i in their sum are

$$z_i = x_i \oplus y_i$$

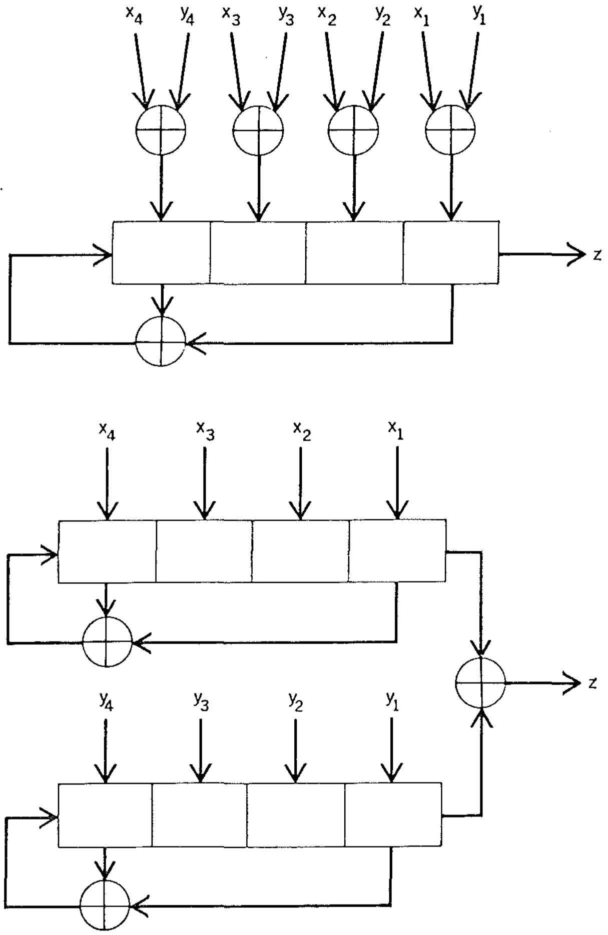
Thus the information bits in z are the modulo-2 sum of the information bits in x and y. The parity bits in z are what we get when we put the modulo-2 sum of the information bits in x and y into our 4-bit register and shift 15 times; it is not hard to see that they are the modulo-2 sum of the parity bits in x and y, since the shift-register connection is itself a modulo-2 sum. In other words, the two circuits in Fig. 9 have identical outputs.

This can be verified also by taking any two of the words in Table III and forming their modulo-2 sum; the result will be another one of the cyclic shifts of the basic sequence.

The group property gives immediate answers to questions about distance or correlation between code words. The distance (Hamming distance) between two code words is defined as the number of places in which they differ. If we form the modulo-2 sum of two code words, the resulting word will have zeros in the positions in which the two code words agree, and ones where they differ; thus the distance between two code words is exactly the number of ones in their sum. But, from the group property, their sum is another code word; and in the maximum-length shift-register codes all words have the same number of ones,* namely $M/2$ (eight in our example). Thus the distance between any two words in these codes is $M/2$, or about half the code length.

The equidistant property of maximum-length shift-register codes makes them an optimum solution to the following problem in signal design: How can one construct M equal-energy signals to minimize the cross-correlation between any two signals, with no bandwidth limitations? Let us suppose that a code word is sent by PSK, so that a 0 is sent as a baud of amplitude -1 and a 1 as amplitude $+1$. The M-code words then correspond to M vectors in $M - 1$ dimensions, all of equal energy (autocorrelation) $M - 1$. The cross-correlation (inner product) of any two vectors is a sum of baud-by-baud correlations, equal to $+1$ if the vectors agree in that place, and -1 if they disagree. But we have just proved that the Hamming distance between any two code words is $M/2$, so that any two vectors disagree in $M/2$ places and agree in the remaining $M/2 - 1$. Consequently, any two vectors are anticorrelated with cross-correlation -1. This implies that as vectors in $(M - 1)$-space, the code words form a geometrical object called a simplex, which is universally believed (though it has never quite been proved) to be the distribution of equal-energy signals in signal space that minimizes the probability of in-

FIGURE 9. Two equivalent linear circuits.



---

* Except the all-zero word, of course; this is the code word we get when we sum any code word with itself.

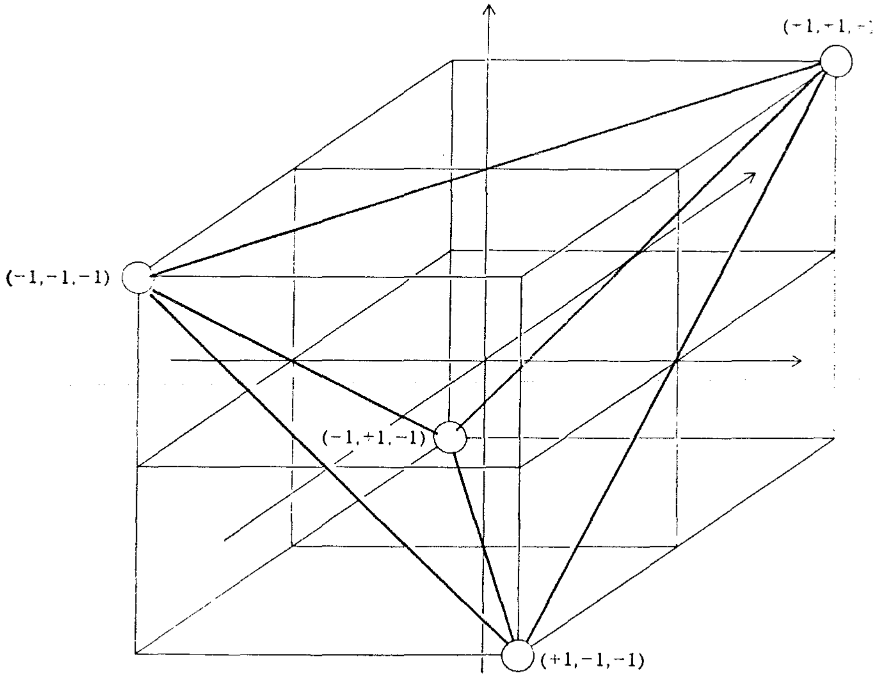
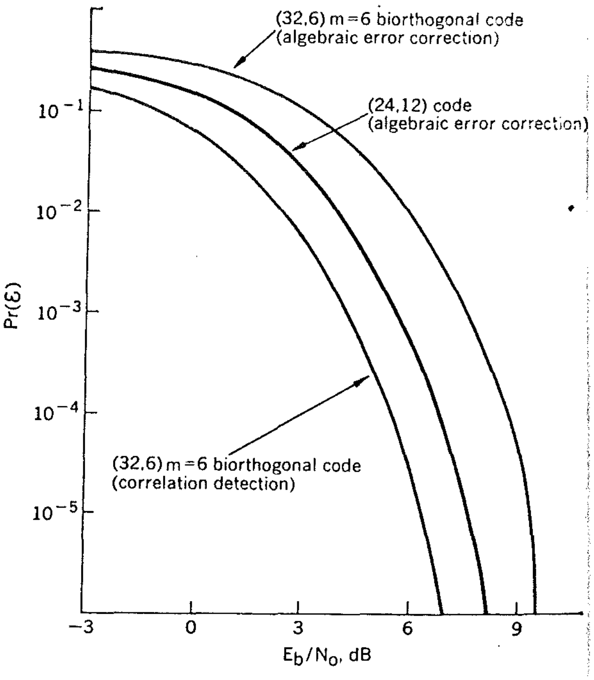


FIGURE 10. Simplex (tetrahedron) formed by m = 2, (3, 2) code in three dimensions.

FIGURE 11. Performance of various block codes.



correct detection. Figure 10 shows the simplex corresponding to the $m = 2$ maximum-length shift-register code, which takes the form of a tetrahedron in three dimensions. Here is an intriguing contact between algebraic coding theory and the geometry of N dimensions.

Suppose now that we use such a binary code with PSK modulation; how shall we decode it at the receiver? As in Fig. 3, we assume that we start with the $M - 1$ correlator outputs z_i that correspond to the $M - 1$ bauds required to send a code word. For definiteness, we use the code of our example in which $m = 4$ and $M - 1 = 15$. Here we shall see a distinction between the viewpoints of the signal designer and of the algebraic coding theorist. The signal designer would take the attitude that what we have here is a way of sending one of 16 signals through a white Gaussian channel, where each possible signal is made up of 15 binary chips, and thus is a vector in 15 dimensions. As in the pure binary case, the optimum detection method is to correlate the received signal against all the 16 possible transmitted signals, which can be done by simply summing the correlator outputs z_i multiplied by ± 1 according to the code word amplitude in the corresponding baud. Thus 16 computations followed by a selection of the largest correlation must be performed. (It turns out that the correlations can be done simultaneously in a special-purpose computer—called the "Green machine" at Jet Propulsion Laboratory[2]—as an M-point fast Hadamard transform, which is structurally very similar to a fast Fourier transform.) The computational load remains manageable for m less than eight or so, which is also where the bandwidth occupied by these codes begins to be absurdly large. A modified (biorthogonal) $m = 6$ code was used in the Mariner '69 expedition; its performance curve is shown in Fig. 11.[3]

An alternate approach is usually taken by the algebraic coding theorist. The first step is to make a hard decision on each correlator output to obtain a 15-bit digital word called the received word. Now we are back in the realm of modulo-2 arithmetic, Hamming distance, and so forth. In the hard-decision process, a number of bit errors will usually be made. From the distance properties of the original code, one can determine that if fewer than some maximum number of errors occur, then correct decoding is guaranteed. In the example, where the Hamming distance between any two words is eight, it is easy to see that if three errors occur in the reception of any code word, then the received word will differ in three places from the correct word, but in at least five places from any other word, so that in principle decoding should be correct. In

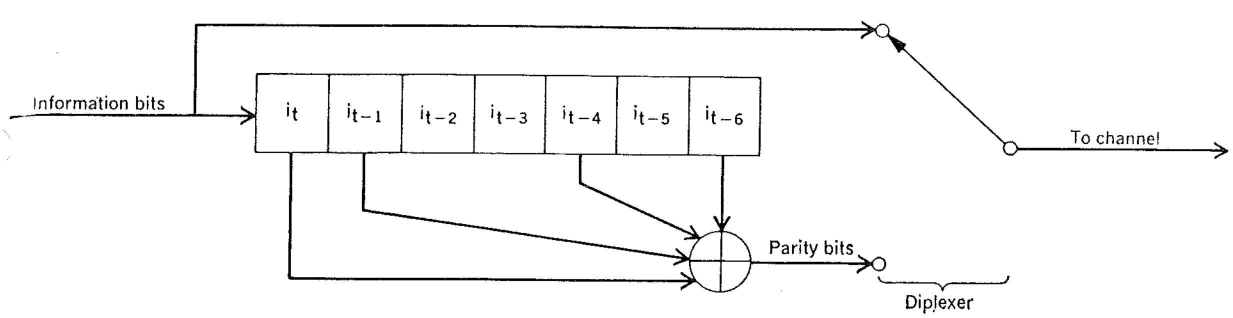


FIGURE 12. Convolutional encoder.

general, a number of bit errors equal to the greatest integer less than half the Hamming distance is guaranteed to be correctable.

One decoding method suitable for the (15, 4) example code is permutation decoding.[4] From the distance properties of this code, we know that if we can find some code word within Hamming distance three of the received word, then we should assume that word was sent, since all other code words must be at least distance five from the received word. To find such a word, we can start by simply reencoding the four received information bits, and checking whether the reencoded parity bits agree with the received parity bits in all but three or fewer places. If so, we are done. If not, then because of the cyclic property of the code, we can take any other four consecutive received bits, treat them as information bits, and generate the rest of the cycle (end-around) that makes up the code. (If this is not immediately clear, try taking any four consecutive positions of any code word in Table III, loading them into the encoder shift register, and shifting 15 times to generate the whole code word, starting at that point and cycling around past the beginning.) If there are actually three or fewer bit errors, at least one set of four consecutive positions will be received correctly, so by taking each set of four in turn, reencoding, and comparing, we will eventually find the correct code word. (This cyclic permutation scheme is also used to correct burst errors, since a correctable burst of errors will not affect at least one set of k consecutive bits.[5])

The performance of hard decisions followed by algebraic error correction is also shown in Fig. 11, for the same (32, 6), distance-16 code as in the correlation detection curve. We see that it is more than 3 dB worse for the lower error probabilities. It might therefore seem that the correlation technique is the better one; however, algebraic decoding remains feasible for much longer code lengths and numbers of information bits, where correlation detection is computationally infeasible. A curve for the (24, 12) (minimum distance eight) extended Golay code is also shown in Fig. 11; with longer codes, the hard-decision disadvantage can eventually be overcome.

Ideally, one would like a scheme whose computational complexity was like that of the algebraic decoding schemes, but would make use of all the information in the correlator output and thus achieve performance like that of correlation detection. At least two approaches (orthogonal equation decoding[6,7] and generalized minimum-distance decoding[8]) with these features are known, but they have not been extensively studied due to the existence of superior convolutional coding schemes (to be described in the next section).

Although we have studied only the maximum-length shift-register codes here, more advanced algebraic block codes involve quite similar ideas. Peterson[9] and Berlekamp[10] are the standard references of the field.

Convolutional codes

Historically, the coding world has been divided between block-code people and convolutional-code people. Although relations between these groups are perfectly amicable, block-code types tend to harp on the relatively primitive theoretical understanding and development of convolutional codes vis-à-vis block codes, whereas convolutional-code types point out that in all respects in which convolutional codes can be compared with block codes they are essentially as good in theory, and in some major respects better, while in practice they are typically simpler. The correctness of both these viewpoints will be illustrated in this section. Whereas we have considered an infinite class of good block codes, we cannot now consider such a class of convolutional codes, since classes of reasonably good codes in the block-code sense are unknown. Instead we shall consider a simple typical code and some reasonable ways of decoding it. The best of these methods will be seen to give better performance on the space channel than any block-code techniques.

Consider the linear sequential circuit illustrated in Fig. 12. Like the maximum-length shift-register generator of Fig. 7, it consists of a shift register and a modulo-2 adder connected to several shift-register stages. In this case, however, information bits are continuously entered into the left end of the register, and for each new information bit a parity bit (a parity check on the current bit and three of those in the past) is computed according to the formula

$$p_t = i_t \oplus i_{t-1} \oplus i_{t-4} \oplus i_{t-6}$$

Information and parity bits are transmitted alternately over the channel. The code generated by this encoder is called a rate-½ convolutional code: rate ½ because there are two transmitted bits for every information bit, convolutional because the parity sequence is the convolution of the information sequence with the impulse response 1,1,0,0,1,0,1, modulo-2. Like the block codes considered earlier, the code is systematic (information bits are transmitted), and is a parity-check code; therefore, it has the group property (the modulo-2 sum of two encoded sequences is the encoded sequence corresponding to the modulo-2 sum of the information sequences).

We shall now suppose that the encoded sequence is sent over a binary channel and that hard decisions are made at the receiver output. How do we decode? First, the decoder must establish which received bits are information and which parity, but as there are only two possibilities, trial and error is a feasible procedure. (For block codes, the comparable problem involves a choice between n phases, where n is the block length, and some special synchronization means may be required.) This done, we shall let the decoder form *syndromes*, which are defined as follows:

Take the received information sequence, and from it recompute the parity sequence with an encoder identical to that of Fig. 12. Compare these recomputed parity bits with the parity bits actually received; the outputs from the comparator (another modulo-2 adder) are called the syndromes (see Fig. 13). (The syndrome idea is equally useful with block codes.)

It is evident that if no errors occur in transmission over the channel, the recomputed parity bits will equal the received parity bits and all syndromes will be zero. On the other hand, if an isolated error occurs in the parity sequence, then a single syndrome will be equal to one at the time of the error. If an isolated error occurs in the information sequence, then the syndromes will equal one at all times when the incorrect bit is at a tapped stage of the shift register, so the syndrome sequence will be 1,1,0,0,1, 0,1,0,0 . . . , starting at the time of the error. The syndrome pattern for more than one error is just the linear superposition (modulo-2) of the syndrome patterns for each of the individual errors. Thus do the syndromes indicate the nature of the disease.

An obvious technique for correcting single isolated errors now suggests itself. Such an error will manifest itself as a syndrome pattern of 1100101 or 1000000, depending on whether it is in an information or a parity bit. The first time we see a 1 in the syndrome sequence, we know that an error has occurred; the value of the following syndrome tells us whether it was an information or parity error. Since only information errors need be corrected, an AND gate looking for two successive syndrome "ones" suffices, as illustrated in Fig. 14(A).

One can correct double errors with the hardly more complicated circuit of Fig. 14(B). Here the syndromes are fed into a 7-stage shift register; a threshold circuit fires if three or four of four selected places contain ones. The selected places are those that would contain ones if there were only a single information error. A single parity error, in addition, can only disturb one input to the threshold circuit; similarly, it can be verified that with this particular code a second information error can only interfere with one input, so that if only two errors occur the threshold circuit will certainly fire at the right time. On the other hand, it can also be verified that under the assumption of only two errors the circuit will never fire at the wrong time. Finally, the complement line is included to take out the effect of a corrected error in those syndrome bits that were inverted by it, so that the decoder can handle all error patterns that do not have more than two errors in any seven consecutive pairs of received bits.

Both these decoders are examples of threshold decoders[7] (working on a self-orthogonal code[11]). Threshold

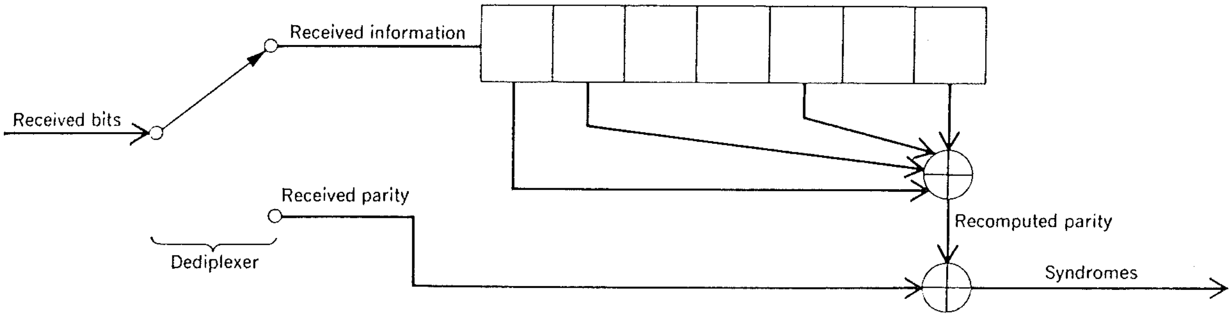
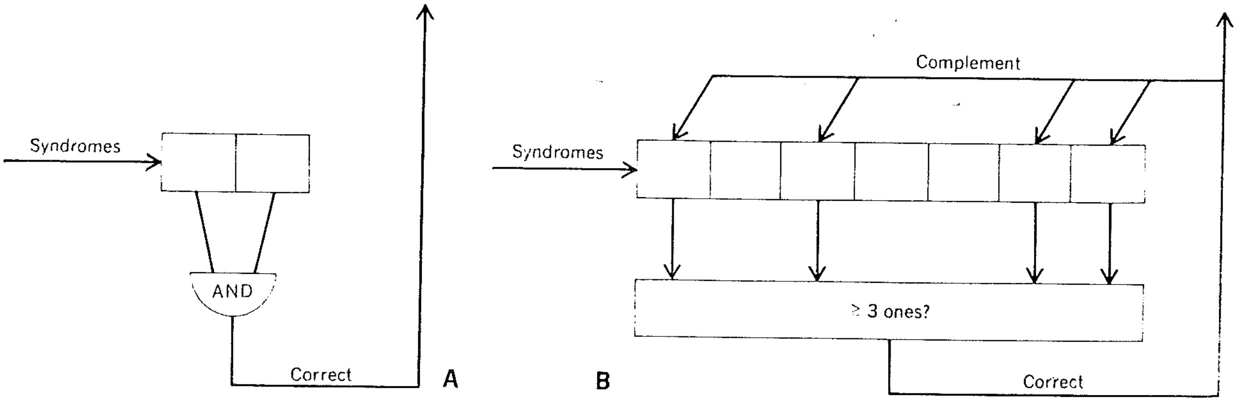


FIGURE 13. Syndrome formation at the receiver.

FIGURE 14. Simple single- and double-error-correcting threshold decoders.



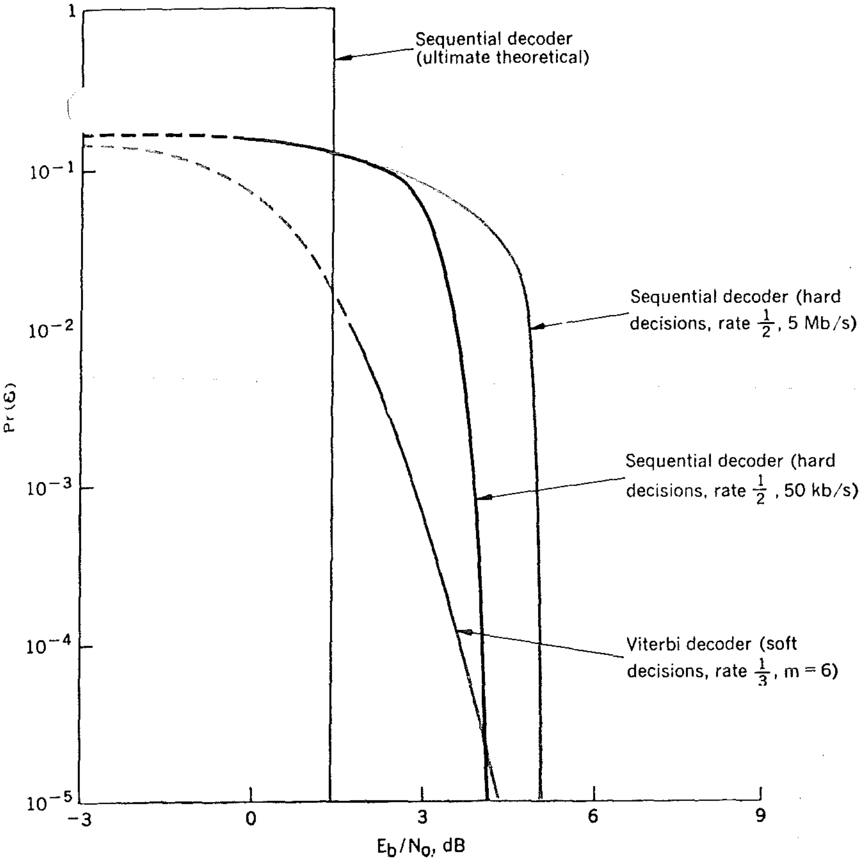


FIGURE 15. Performance of convolutional decoders.

decoding is an extremely simple technique that applies to many short codes correcting a few errors, and that is easily extended to correct bursts of errors. Its efficiency diminishes as the number of errors to be corrected becomes large, and for this reason it is not an outstanding performer on the space channel. With hard decisions, the performance of the three-error-correcting (24, 12) convolutional code (shift register length 12) is about the same (to within 0.2 dB) as that of the (24, 12) block code of Fig. 11.

Sequential decoding was invented by Wozencraft[12] in about 1957. Through a decade of improvement, analysis, and development, it has become the best-performing practical technique known for memoryless channels like the space channel, and will probably be the general-purpose workhorse for these channels in the future. Like much else in the convolutional-coding domain, it is hard to explain and analyze, but relatively easy to implement. Very crudely, a sequential decoder works by generating hypotheses about what information sequence was actually sent until it finds some that are reasonably consistent with what was received. It does this by a backward and forward search through the received data (or through syndromes). It starts by going forward, generating a sequence of hypotheses about what was sent. It checks what [illegible] received against what would have been transmitted given the hypotheses, and according to the goodness of the agreement updates a measure of its happiness called the metric. As long as it is happy, it goes forward; when it becomes unhappy, it turns back and starts changing hypotheses one by one until it can go forward happily again. A simple set of rules for doing this is called the Fano algorithm.[13–15]

It is evident even from this meager description that sequential decoding involves a trial-and-error search of variable duration. When reception is perfect, the decoder's first guess is always correct, and only one "computation" (generation of a hypothesis) is required per bit. The more noise, the more hypotheses must be generated, up to literally millions to decode a single short segment. Because of the variability of the computational load, buffer storage of the received data must be provided to permit long searches. Whenever this buffer overflows, the decoder must jump ahead and get restarted, leaving a section of data undecoded. This overflow event therefore leads to a burst of output errors; its frequency generally dominates the probability of decoding error, since the code can be made long enough that the probability the decoder is actually happy with incorrect hypotheses can be made negligible.

Sequential decoding is outstandingly adaptable; it can work with soft or hard decisions and PSK, or with any modulation and detection scheme. In the four implementations for the space channel to date, the Lincoln Experimental Terminal decoder[16] works with 16-ary frequency-hopping modulation and incoherent list detection; the NASA Ames decoder for the Pioneer satellites[17] and the JPL general-purpose decoder[18] work with PSK and soft (eight-level) decisions; and the Codex decoder, built for the U.S. Army Satellite Communication Agency,[19] works with PSK (or DPSK or QPSK) and hard decisions, the choice in every case being based on system considerations. Sequential decoding can even make efficient use of known redundancies in the data, as was done for some preexisting parity checks in the Pioneer data format. The one thing a sequential decoder cannot tolerate is bursts of errors, which will cause excessive computation; therefore, it cannot be applied without modification to any channel but the space channel.

The performance of sequential decoding depends both on the modulation and detection scheme with which it is used, and on the data rate relative to the internal computation rate. The theoretical limit of any sequential decoder on a white Gaussian channel is $E_b/N_o = 1.4$ dB, exactly 3 dB above the Shannon limit; this limit can be approached with PSK, soft decisions, and low-rate codes. The simplest possible sequential decoder working with rate-½ codes, PSK, and hard decisions has a theoretical limit of $E_b/N_o = 4.5$ dB; 2 dB of this loss is due to hard decisions, 1 dB to the choice of rate ½ rather than a lower rate. Actual performance depends on the data rate as well as the error rate desired, although the curves are very steep; Fig. 15 shows measured curves at 50 kb/s and 5 Mb/s for the Codex decoder,[20] which has an internal computation rate of 13.3 Mb/s.

Somehow the idea that sequential decoding is complicated to implement has achieved considerable circulation. This is undoubtedly partly due to the difficulty of the literature. Also, the first sequential decoder (SECO[21]), built at Lincoln Laboratory for telephone lines with the technology of an earlier day, was an undoubted monster, due in part to large amounts of auxiliary equipment such as equalizers. It should be emphasized that three of the four implementations just mentioned involve only a drawer of electronics with a core memory system for the buffer storage; the fourth, the Pioneer system, was actually done in software because of the low maximum bit rate (512 b/s).

We conclude by mentioning two more classes of schemes of current interest. One, the Viterbi algorithm,[22] performs optimum correlation detection of short convolutional codes much as the Green machine does of block codes. Figure 15 shows the performance of this algorithm[23] with soft decisions when the decoding complexity is comparable to that of the $m = 6$ block decoder of Fig. 11; performance is uniformly superior. This algorithm is competitive in performance with sequential decoding for moderate error rates, but cannot achieve very low error rates efficiently. On the other hand, it can be implemented in a highly parallel pipe-lined decoder capable of extremely high speeds (tens of megabits) where sequential decoders become uneconomic. It therefore may find application in high-data-rate systems with modest error requirements, such as digitized television.

The second class represents attempts to bridge the final 3-dB gap between the sequential decoding limit and the Shannon limit by combining sequential decoding with algebraic block code constraints. Recent unpublished work of Jelinek gives promise of performances between 1 and 2 dB from the Shannon limit without excessive computation. At the moment, all schemes in this class seem most suited for software implementation, and will probably be used only for low-data-rate applications where the ultimate in efficiency is desired, as in deep-space probes.

Thus do we near practical achievement of the goal set by Shannon 20 years ago.

REFERENCES

1. Shannon, C. E., "A mathematical theory of communication," *Bell System Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

2. Green, R. R., "A serial orthogonal decoder," Jet Propulsion Lab. Space Programs Summary 37–39, vol. 4, pp. 247–252, 1966.

3. *Digital Communications with Space Applications*, appendix 4 S. Golomb, ed. Englewood Cliffs, N.J.: Prentice-Hall, 1964.

4. MacWilliams, J., "Permutation decoding of systematic codes," *Bell System Tech. J.*, vol. 43, pp. 485–506, 1964.

5. Gallager, R. G., *Information Theory and Reliable Communication.* New York: Wiley, 1968, pp. 291–297.

6. Gallager, R. G., *Low-Density Parity-Check Codes.* Cambridge, Mass.: M.I.T. Press, 1963, pp. 42–52.

7. Massey, J. L., *Threshold Decoding.* Cambridge: M.I.T Press, 1963, pp. 59–63.

8. Forney, G. D., "Generalized minimum distance decoding," *IEEE Trans. Information Theory*, vol. IT-12, pp. 125–131, Apr. 1966.

9. Peterson, W. W., *Error-Correcting Codes.* Cambridge: M.I.T. Press, 1961.

10. Berlekamp, E. R., *Algebraic Coding Theory.* New York: McGraw-Hill, 1968.

11. Robinson, J. P., and Bernstein, A. J., "A class of binary recurrent codes with limited error propagation," *IEEE Trans. Information Theory*, vol. IT-13, pp. 106–113, Jan. 1967.

12. Wozencraft, J. M., and Reiffen, B., *Sequential Decoding.* Cambridge: M.I.T. Press, 1961.

13. Fano, R. M., "A heuristic discussion of probabilistic decoding," *IEEE Trans. Information Theory*, vol. IT-9, pp. 64–74, Apr. 1963.

14. Wozencraft, J. M., and Jacobs, I. M., *Principles of Communication Engineering.* New York: Wiley, 1965.

15. Gallager, R. G., *op. cit.*, pp. 263–286.

16. Lebow, I. L., and McHugh, P. G., "A sequential decoding technique and its realization in the Lincoln Experimental Terminal," *IEEE Trans. Communication Technology*, vol. COM-15, pp. 477–491, Aug. 1967.

17. Lumb, D. R., "Test and preliminary flight results on the sequential decoding of convolutionally encoded data from Pioneer IX," 1969 IEEE Internat'l Communications Conf. Record, Boulder, Colo., pp. 39/1–8.

18. Lushbaugh, W., "Multiple-mission sequential decoder," Jet Propulsion Lab. Space Programs Summary 37–58, vol. 2, pp. 33–36, 1969.

19. Forney, G. D., Jr., and Langelier, R. M., "A high-speed sequential decoder for satellite communications," *1969 IEEE Internat'l Communications Conf. Record*, Boulder, Colo., pp. 39/9–17.

20. "High-speed sequential decoder," Codex Corp. final rept., Contract DAAB07-69-C-0051, U.S. Army Satellite Communication Agency, Ft. Monmouth, N.J., June 6, 1969.

21. Lebow, I. L., *et al.*, "Application of sequential decoding to high-rate data communication on a telephone line," *IEEE Trans. Information Theory*, vol. IT-9, pp. 260–269, Apr. 1963.

22. Viterbi, A. J., "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Information Theory*, vol. IT-13, pp. 260–269, Apr. 1967.

23. Heller, J., "Improved performance of short constraint length convolutional codes," Jet Propulsion Lab. Space Programs Summary 37–56, vol. 3, pp. 83–84, 1969.

G. David Forney, Jr. (M) received the B.S.E. degree from Princeton in 1961, and the M.S. and Sc.D. degrees from M.I.T. in 1963 and 1965, respectively. He has been with Codex Corporation, Watertown, Mass., since 1965, and now serves as director of research. Responsible for company efforts in space communications, he has also been involved in the design and development of products in the areas of coding, multiplexing, and modems. Dr. Forney is the author of several journal articles and a book entitled "Concatenated Codes"; although the book is concerned with block codes, he is usually identified as a convolutional-code type. He is currently vice chairman of the Boston IEEE Information Theory Group Chapter and is also a member of the AAAS.

Codes correcteurs d'erreurs

par A. Hocquenghem,
Professeur au Conservatoire des Arts et Métiers,
Ingénieur conseil à la S.E.A.

Généralisant un travail de Hamming, l'auteur construit des codes permettant de corriger k erreurs dans une transmission de digits binaires.

The paper is a generalization of Hamming's work. The author gives a coding system available to correct k errors in a transmission of binary digits.

Eine Arbeit von Hamming verallgemeinernd, entwickelt der Autor Kodes die es ermöglichen bei Ubertragung binärer bits k Fehler zu korrigieren.

Обобщая работу Хамминга, автор предлагает коды, которые дают возможность исправлять k ошибок в передаче двоичных цифр.

1. Introduction.

Introduisons dans un système de transmission un mot, constitué par un nombre de n chiffres binaires :

$$a_1 \, a_2 \, \ldots\ldots\ldots \, a_n$$

Le mot reçu peut différer du mot initial par un certain nombre d'erreurs (certains chiffres a_i étant altérés en $1 - a_i$). Pour essayer de détecter et de corriger ces erreurs, on n'utilise que m chiffres du mot comme support de l'information, les chiffres restant appelés chiffres de test devant servir à la vérification du mot après la transmission. Donner une loi de détermination de ces chiffres de test en fonction des m chiffres d'information de façon à pouvoir détecter — ou corriger — un nombre maximum k d'erreurs, c'est former un code détecteur — ou correcteur — de k erreurs.

L'exemple le plus simple est le code détecteur d'une erreur. Dans ce cas $m = n - 1$, et on choisit le chiffre de test de façon que le nombre total de chiffres 1 du mot soit pair. La vérification du mot consiste alors en un test de parité.

Hamming (*Bell System Technical Journal*, 1950) a donné la loi de formation d'un code correcteur d'une erreur. Le nombre de chiffres de test est l'entier N déterminé par les inégalités

$$(11) \qquad \mathrm{Log}_2(1+n) \leqq \mathrm{N} < 1 + \mathrm{Log}_2(1+n)$$

Dans le cas général d'un code correcteur de k erreurs, le nombre de configurations d'erreurs possibles est :

$$\mathrm{H} = 1 + \mathrm{C}_n^1 + \mathrm{C}_n^2 + \ldots + \mathrm{C}_n^k$$

Par suite le code le plus économique utiliserait un nombre de chiffres de test égal à l'entier immédiatement supérieur à $\mathrm{Log}_2\mathrm{H}$. A part le code de Hamming, on n'a pu construire de tels codes. Ceux que nous proposons utilisent un nombre de chiffres de test égal à

$$n - m = kN$$

La différence

$$k\mathrm{N} - \mathrm{Log}_2\mathrm{H}$$

est de l'ordre de $\mathrm{Log}_2(k!)$, donc assez faible pour que ces codes soient satisfaisants.

Après avoir défini un anneau dans lequel nous ferons nos calculs, nous exposerons le code de Hamming sous cette optique, puis les principes de formation des codes qui nous conduiront à une détermination quasi-expérimentale et à une détermination systématique de ces codes. Nous terminerons par un exemple de code correcteur de 2 erreurs.

2. Définition de l'anneau $\mathcal{A}$

Les éléments de l'anneau $\mathcal{A}$ sont les nombres entiers écrits en numération binaire.

A chaque élément de l'anneau $\mathcal{A}$ nous faisons correspondre un polynôme ayant comme coefficients les chiffres de l'élément. Le polynôme est alors défini sur le corps de caractéristique 2.

Toute opération sur les éléments de $\mathcal{A}$ sera faite sur les polynômes correspondants — au cours de ces opérations tout coefficient pair sera remplacé par O, tout coefficient impair par 1. Le résultat sera un polynôme auquel correspondra un élément de l'anneau $\mathcal{A}$.

On a donc toutes les opérations habituelles sur les nombres entiers — afin d'éviter toute ambiguïté, toutes les expressions calculées selon ces règles seront suivies de l'indication $(\mathcal{A})$.

Exemples :

Addition : $101 + 111 = 10 \quad (\mathcal{A})$
Multiplication : $101 \times 111 = 11.011 \quad (\mathcal{A})$
Puissance : $101^2 = 10.001 \quad (\mathcal{A})$
Division : $1.101 = 111 \times 10 + 11 \quad (\mathcal{A})$
En particulier : $p + p = 0,\ (p+q)^2 = p^2 + q^2 \quad (\mathcal{A})$

Lorsque le polynôme sera irréductible sur le corps de caractéristique 2, nous dirons que le nombre correspondant est irréductible (il n'admet pas, dans l'anneau $\mathcal{A}$, d'autre diviseur que lui-même et l'unité).

On peut classer évidemment les nombres dans l'anneau $\mathcal{A}$ par ordre de grandeur, mais beaucoup plus important est le nombre de chiffres. On démontre que parmi les nombres ayant un nombre de chiffres donné, il existe toujours un nombre irréductible.

Etant donné un mot écrit en binaire

$$a_1 \, a_2 \ldots a_n$$

nous attacherons à chaque indice i un nombre p_i de l'anneau $\mathcal{A}$) et au mot lui-même nous attacherons le nombre

$$\mathrm{T} = a_1 p_1 + a_2 p_2 + \ldots + a_n p_n \qquad (\mathcal{A})$$

C'est la considération du nombre T qui, grâce à un choix convenable des nombres p_i nous permettra de corriger les erreurs éventuelles.

3. Code de Hamming.

Nous retrouvons le code de Hamming en faisant

$$p_i = i$$

Les chiffres de test sont les chiffres du mot d'indices

$1, 2, 2^2, \ldots, 2^{N-1}$ (N défini par les inégalités 11).

L'information sera portée par les chiffres

$$a_3 \, a_5 \, a_6 \, a_7 \, a_9 \ldots a_n$$

On détermine les chiffres de test par la condition

$$\mathrm{T} = \Sigma \, p_i a_i = 0$$

condition qui s'écrit ici

$$a_1 + 2a_2 + 4a_4 + \ldots + 2^{N-1} a_{2^{N-1}} = 3a_3 + 5a_5 + 6a_6 + \ldots + na_n \qquad (\mathcal{A})$$

Le second membre est un nombre binaire connu d'au plus N chiffres. L'égalité détermine donc parfaitement les valeurs des chiffres de test.

Si, après transmission, il n'y a pas d'erreur, on retrouvera $\mathrm{T} = 0$.

S'il y a une erreur, portant par exemple sur le chiffre a_α remplacé par $(1 - a_\alpha)$, le nombre T prendra la valeur :

Reprinted with permission from *Chiffres,* vol. 2, pp. 147-156, 1959.

$$T = a_1 + 2a_2 + \ldots + \alpha(1 - \alpha_\alpha) + \ldots + na_n = \alpha \qquad (\mathcal{A})$$

la valeur de T sera l'indice du chiffre erronné.

S'il y a deux erreurs, portant sur les chiffres d'indice α et β, T prendra la valeur :

$$T = \alpha + \beta \neq 0 \qquad (\mathcal{A})$$

S'il y a plus de deux erreurs, T pourrait être nul. Le code obtenu est donc correcteur d'une erreur, détecteur de deux erreurs.

Il est commode, pour automatiser le contrôle, de supposer les nombres p disposés en matrice. Par exemple pour $n = 7$, on aura la matrice

$$\begin{vmatrix} 0001111 \\ 0110011 \\ 1010101 \end{vmatrix}$$

Aux nombres

$$0010110 \qquad \text{et} \qquad 0010010$$

correspondront les matrices

$$\begin{vmatrix} 0000110 \\ 0010010 \\ 0010100 \end{vmatrix} \qquad \text{et} \qquad \begin{vmatrix} 0000010 \\ 0010010 \\ 0010000 \end{vmatrix}$$

Le nombre T s'obtient en faisant suivre chaque ligne de la matrice de son chiffre de parité (§ 1).

On obtient ici :

$$\begin{vmatrix} 0 \\ 0 \\ 0 \end{vmatrix} \qquad \text{et} \qquad \begin{vmatrix} 1 \\ 0 \\ 1 \end{vmatrix} = 5$$

Le premier nombre est correct, le 5ᵉ chiffre du second nombre est faux.

4. Principe d'un code correcteur de k erreurs.

Voyons maintenant à quelles conditions doivent satisfaire les nombres p pour que le calcul de T permette de corriger k erreurs.

Nous supposerons que les chiffres de test sont en nombre suffisant pour que, connaissant les chiffres d'information, on puisse réaliser la condition

(41) $$T = \Sigma a_i p_i = 0 \qquad (\mathcal{A})$$

Si après transmission, les chiffres de rang

$$\alpha_1, \alpha_2 \ldots, \alpha_j \qquad (j \leq k)$$

sont erronés, le nombre T calculé sur le mot déformé prendra la valeur :

$$T = p_{\alpha_1} + p_{\alpha_2} + \ldots + p_{\alpha_j} \qquad (\mathcal{A})$$

Il faut que le nombre ainsi trouvé soit caractéristique des rangs $\alpha_1, \alpha_2, \ldots, \alpha_j$, c'est-à-dire que :

$$p_{\alpha_1} + p_{\alpha_2} + \ldots + p_{\alpha_j} \neq p_{\alpha'_1} + p_{\alpha'_2} + \ldots + p_{\alpha'_{j'}} \qquad (\mathcal{A})$$

lorsque

$$j \leq k, \quad j' \leq k$$

et les deux ensembles

$$(\alpha_1, \alpha_2, \ldots, \alpha_j), \qquad (\alpha'_1, \alpha'_2, \ldots, \alpha'_j),$$

non identiques.

Cette condition peut encore s'écrire :

(42) $$p_{\lambda_1} + p_{\lambda_2} + \ldots + p_{\lambda_l} \neq 0 \qquad (\mathcal{A})$$

lorsque $l \leq 2k$

et des $\lambda_1, \lambda_2, \ldots, \lambda_l$ étant tous différents.

On devra donc choisir les nombres p tels que l'addition, dans l'anneau $\mathcal{A}$, d'au plus $2k$ de ces nombres donne un résultat non nul.

Une fois déterminé un ensemble de n nombres p, il faudra choisir les chiffres de test. Il est commode pour cela de remplacer l'ensemble obtenu par un autre ensemble de n nombres mais contenant les puissances successives de 2 :

$$1, 2, 2^2, \ldots, 2^{K-1}$$

K désignant le nombre de chiffres du plus grand nombre p obtenu.

Disposons pour cela les nombres p en une matrice M de n colonnes et K lignes (K < n), chaque nombre p étant donc représenté par une colonne

$$p_i = \begin{vmatrix} \varpi_i^K \\ \vdots \\ \varpi_i^2 \\ \varpi_i^1 \end{vmatrix}$$

Si le rang de cette matrice (dans l'anneau $\mathcal{A}$) est $K' < K$, c'est que $K - K'$ lignes de cette matrice sont des combinaisons linéaires des K' lignes restantes. Si l'on supprime ces $K - K'$ lignes, on obtiendra une matrice M′ de nombres p' qui vérifieront encore la condition (42).

Ceci étant, nous pourrons extraire de la matrice M′ une matrice carrée Δ de K' lignes dont le déterminant calculé dans l'anneau $\mathcal{A}$ ne sera pas nul. On aura donc

$$\det \Delta = 1$$

puisque les seules valeurs possibles sont 0 ou 1. En multipliant la matrice M′ par Δ^{-1}, on obtiendra la matrice M″ formée de nombres p'' tels que

$$p''_i = \begin{vmatrix} \varpi_i''^{K'} \\ \vdots \\ \varpi_i''^{1} \end{vmatrix} = \Delta^{-1} \begin{vmatrix} \varpi_i'^{K'} \\ \vdots \\ \varpi_i'^{1} \end{vmatrix}$$

et par suite les nombres p''_i vérifieront encore la condition (42). De plus, la matrice M″ contiendra à ce moment la matrice $\Delta^{-1} \times \Delta$, c'est-à-dire la matrice unité, donc l'ensemble des p'' contiendra les puissances successives de 2 :

$$1, 2, 2^2, \ldots, 2^{K'-1}$$

Les indices correspondants seront pris comme chiffres de test et la condition (41) déterminera ces chiffres en fonction des chiffres d'information par égalité de deux nombres binaires de K' chiffres.

Tout le problème se ramène donc à construire des ensembles de nombres p satisfaisant à la condition (42).

5. Formation de proche en proche d'une suite de nombres p.

Prenons d'abord :

$$p_1 = 1, \; p_2 = 2, \; p_3 = 2^2, \; \ldots, \; p_{2k} = 2^{2k-1}$$

puis :

$$p_{2k+1} = 2^{2k} - 1$$

$$p_{2k+2} = 2^{2k}$$

Ces nombres satisfont déjà aux conditions (42). Pour prolonger cette suite dans l'ordre des p croissants, supposons être arrivé au nombre p_i de l chiffres. Considérons l'ensemble des nombres p_1 à p_i et de leurs sommes dans l'anneau $\mathcal{A}$ par groupes de 2, 3, … $(2k - 1)$. Tous les nombres obtenus ont au plus l chiffres.

S'il existe un nombre non contenu dans l'ensemble ainsi formé et compris entre p_i et 2^l, ce nombre sera pris pour valeur de p_{i+1} (s'il y a plusieurs nombres on choisira évidemment le plus petit). Sinon on prendra $p_{i+1} = 2^l$.

On peut ainsi continuer pas à pas jusqu'à l'obtention des n nombres p. Si p_n a K chiffres, les nombres

$$1, 2, 2^2, \ldots, 2^{K-1}$$

seront inclus dans la suite des p. La suite sera donc directement utilisable pour former un code. Il restera à établir le tableau de correspondance entre les H valeurs de la somme

$$p_{\alpha_1} + p_{\alpha_2} + \ldots + p_{\alpha_j} \qquad (\mathcal{A}) \qquad (j \leq k)$$

et la valeur des indices $\alpha_1, \alpha_2, \ldots, \alpha_j$.

Le procédé ainsi défini est assez long à exploiter. Cependant, pour des valeurs raisonnables de n et k, il ne dépasse pas les possibilités d'une calculatrice de moyenne puissance.

La détermination à priori du nombre K de chiffres de test paraît assez difficile. Aussi allons-nous exposer un procédé plus systématique de recherche des nombres p.

6. Formation systématique des nombres p.

La théorie des congruences, si utilisée dans les preuves des opérations arithmétiques, va nous fournir un mode de calcul des nombres p. Désignons par ϱ un nombre irréductible de $N + 1$ chiffres et par q' le reste de la division dans l'anneau $\mathcal{A}$ d'un nombre q par ϱ. Le nombre q' aura au maximum N chiffres.

Nous poserons alors :

$$p_i = i + 2^{N}(i^3)' + 2^{2N}(i^5)' + \ldots + 2^{(k-1)N}(i^{2k-1})' \qquad (\mathcal{A})$$

$$(i = 1, 2, \ldots, n)$$

c'est-à-dire que le nombre p_i est formé de la juxtaposition des restes successifs de la division par ϱ des puissances impaires dans l'anneau $\mathfrak{A}$ du nombre i. Nous allons montrer que ces nombres p_i satisfont à la condition (42).

En effet, supposons :

$$(61) \qquad p_{\lambda_1} + p_{\lambda_2} + \dots + p_{\lambda_l} = 0 \qquad (\mathfrak{A}) \qquad (l \leqq 2k)$$

Cela entraînerait :

$$(62) \qquad S_1 = S'_3 = S'_5 = \dots = S'_{2k-1} = 0$$

en posant :

$$S_i = (\lambda_1)^i + (\lambda_2)^i + \dots + (\lambda_l)^i \qquad (\mathfrak{A})$$

et

$$S'_i \equiv S_i \pmod{\varrho} \qquad (\mathfrak{A})$$

Or, si nous considérons le produit

$$\Pi(\lambda_i + \lambda_j) \qquad \begin{matrix} i = 2, 3, \dots, l \\ j = 1, 2, \dots, l-1 \end{matrix} \qquad i > j \qquad (\mathfrak{A})$$

ce produit peut s'écrire sous forme d'un déterminant de Van der Monde dont le carré contiendra la ligne

$$S_1 \; S_2 \; S_3 \; \dots\dots \; S_l$$

Comme $S_{2i} = S_i^2$, les conditions (62) entraînent

$$\Pi(\lambda_i + \lambda_j) \equiv 0 \pmod{\varrho} \qquad (\mathfrak{A})$$

Donc un des facteurs, par exemple $\lambda_i + \lambda_j$, serait divisible par ϱ. Comme la somme dans $\mathfrak{A}$ des nombres $\lambda_i + \lambda_j$ a moins de chiffres que le nombre ϱ, il en résulterait

$$\lambda_i + \lambda_j = 0 \qquad \lambda_i = \lambda_j$$

Par suite l'hypothèse (61) ne peut être réalisée que si au moins deux indices étaient égaux.

Donc les nombres p que nous avons formés remplissent la condition (42) et peuvent servir à former un code correcteur de k erreurs. Naturellement on les transformera comme il est indiqué au § 4 pour former une suite contenant des puissances de 2. Dans le cas général, p_i comprenant kN chiffres, il y aura lieu d'utiliser kN chiffres de test.

7. Exemple.

Nous avons formé un code de 15 chiffres correcteur pour 2 erreurs. Ici $N = \mathrm{Log}_2 16 = 4$, il y aura 8 chiffres de test.

En prenant $\varrho = 19 = 10.011$, on calcule aisément les nombres p et la matrice M :

$$\begin{array}{|ccccccccccccccc|}
0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15
\end{array}$$

Cette matrice M est de rang 8, car le déterminant formé avec les colonnes 1, 2, 4, 8, 6, 12, 7, 14 (choisies parce qu'elles présentent le plus de zéros) vaut 1.

La matrice Δ (§ 4) sera formée avec ces colonnes. En l'inversant on trouve la matrice :

$$\Delta^{-1} = \begin{vmatrix}
0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 1 & 0
\end{vmatrix}$$

Le produit $\Delta^{-1}M$ (dans l'anneau $\mathfrak{A}$) donne la matrice définitive :

$$\begin{array}{|ccccccccccccccc|}
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15
\end{array}$$

Les chiffres de rang

1, 2, 4, 6, 7, 8, 12, 14

serviront de chiffres de test, les 7 autres chiffres seront les supports de l'information.

Pour vérifier et corriger un mot on fera la somme

$$T = \Sigma p_i a_i \qquad (\mathfrak{A})$$

Si elle est nulle, il n'y aura pas eu d'altération du mot (ou plus de 4 erreurs). Si T n'est pas nulle, on pourra retrouver les chiffres faux (en admettant qu'il n'y en ait pas plus de 2) en utilisant la table suivante qui donne les valeurs possibles de T suivies entre parenthèses des rangs des chiffres faux.

1 (14) — 2 (12) — 3 (12,14) — 4 (7) — 5 (7,14)
6 (7,12) — 8 (6) — 9 (6,14) — 10 (6,12) — 12 (6,7)
15 (9,10) — 16 (8) — 17 (8,14) — 18 (8,12) — 19 (10,11)
20 (7,8) — 24 (6,8) — 27 (2,5) — 28 (9,11) — 29 (1,13)
32 (4) — 33 (4,14) — 34 (4,12) — 36 (4,7) — 40 (4,6)
41 (5,9) — 42 (5,10) — 44 (3,13) — 46 (2,11) — 48 (4,8)
49 (1,3) — 50 (2,9) — 53 (5,11) — 61 (2,10) — 64 (2)
65 (2,14) — 66 (2,12) — 68 (2,7) — 72 (2,6) — 75 (5,8)
77 (8,13) — 78 (4,11) — 80 (2,8) — 82 (4,9) — 83 (5,6)
85 (6,13) — 88 (3,15) — 89 (5,12) — 90 (5,14) — 91 (5)
93 (4,10) — 96 (2,4) — 98 (8,9) — 102 (6,11) — 103 (5,7)
105 (1,15) — 106 (7,11) — 108 (11,12) — 110 (11) — 111 (11,14)
112 (9,12) — 113 (8,10) — 114 (9) — 115 (9,14) — 116 (13,15)
117 (6,10) — 118 (7,9) — 121 (7,10) — 122 (6,9) — 123 (4,5)
124 (10,14) — 125 (10) — 126 (8,11) — 127 (10,12) — 128 (1)
129 (1,14) — 130 (1,12) — 132 (1,7) — 135 (11,15) — 136 (1,6)
144 (1,8) — 145 (3,4) — 152 (10,15) — 153 (7,13) — 155 (9,15)
156 (13,14) — 157 (13) — 159 (12,13) — 160 (1,4) — 161 (3,8)
169 (2,15) — 176 (3,14) — 177 (3) — 178 (5,15) — 179 (3,12)
181 (3,7) — 185 (3,6) — 189 (4,13) — 192 (1,2) — 195 (3,9)
198 (5,13) — 201 (4,15) — 204 (3,10) — 219 (1,5) — 221 (2,13)
223 (3,11) — 224 (10,13) — 225 (6,15) — 232 (14,15) — 233 (15)
234 (3,5) — 235 (12,15) — 237 (7,15) — 238 (1,11) — 239 (9,13)
241 (2,3) — 242 (1,9) — 243 (11,13) — 249 (8,15) — 253 (1,10)

On remarquera que le nombre T prend 121 valeurs possibles $\left(1 + C_{15}^1 + C_{15}^2\right)$ et qu'on utilise un nombre de 8 chiffres pour l'écrire. Le code utilise un chiffre de test de plus qu'il n'est théoriquement indispensable, mais il n'est pas sûr qu'on puisse construire des codes n'ayant qu'un nombre de chiffres de test strictement égal à l'entier par excès de Log_2 H.

Intro

I ~~a~~ let us introduce a system of Σ u transmission a word constituted by a number of (n) chiefly binary numbers
$a_1 a_2 - - - a_n$

the no received can differ from the initial word by a certain no. of errors. by certain no a_i being alterned ($1 - a_i$). to try to detect & correct these errors one uses only m numbers nos of the word as support of the information ($n - m$)

is no's remaining called no's of tests (test ~~numbers~~) must serve to verification of the word after the transmission

To give a law of determination of these test number in func of m numb of info in such way as to be able to detect — or correct — to a maximum numb K of errors that is to form a code detector or corrector — of K errors,

The most simple EX is the code detector of one error. in this case $m = N = 1$, & one choses the ~~[illegible]~~ test number in such a way that the total number of numbers (sum) 1 of the word would be even.

The verification of the word consists then in a test of parity.
Hamming Bell system Technical J. 1950 gave the law of formation of code corrector of 1 error,
The number of test digit is the entire n determined by the inequalities

(11) Log — — = ═══ – . . .

In the general case of the code corrector of K errors the ~~amount~~ number of config of errors possible is

$H = 1\ 7$ ——— — ———

Thus the most econom. code would use log K! numb of test digit = to the whole immediate superior to $\log_2 H$.

Besides the code of Hamming they have not been able to con~~stzuct~~struct such a code. Those that we propose use a number of test digits = to

$n - m = kN$

& the difference

$kN - \log_2 H$

is to the order of $\log_2(K!)$, ~~close~~ & it is weak enough that these codes would be satisfactory.

After having defined a O or band arising in which we can make are calculations, we will suppose Hammings code under this Optic.

then the principles of formation of codes which will lead us to a determination partially experimental determined + to a systematic determination of these codes.

We will end ~~determin~~ by an example of a corrector code of 2 errors

2 Definition of Ring a

In the elements of the O a are the whole numbers written in binary numeration

To each element of the O a we correspond polynomial having as a coefficient the digits

of the element.
the polynomial is then
defined on the body of the
characteristic 2,

All operations on the elements
of a will be made on the corresponding polynomials
— during the course of
these operations every all even
coefficients will be replaced
by 0,
every uneven coefficient by
1
the result will be a polynomial
in which will correspond
a element of the 0 a.

If in 9 all then usual habitual operation
on the whole numbers —
in order to avoid all ambiguity, all the calculated expressions
(~~figured~~) ~~expressions~~
~~(followed~~

according to these stated rules

will be (followed by the indication (a)).

Examples

add 101 + 111 ≡ 10

When the polynomial will be unreducable, on the body of the [illegible] 2, we will say that the Correspontig number is unreducable (There s not admitted in the OA another divisor than itself + unity)

You can naturally classify the OA by order of greatness, but much more important is the number of digits. It is shown that among the numbers having a given number of digits there always exists unreducable numbers.

Given a written word in binary ($a_1 a_2 \ldots a_n$)

We will attach to each indication a number p_i of the circle a)

and to the word itself we will attach the number

$T = a_1 p_1 + a_2 + p_2 + \ldots$

It is the consideration of the number T which, thanks to a suitable choice of numbers p_i will permit us to correct the eventual errors.

3. Code of Hamming

We find again the code of Hamming in doing $p_i = i$

The test digits are the digits of the word of the indications

$1, 2, 2^2, \ldots 2^{N-1}$ (N defined by the inequalities 1_1)

The information will be carried by the digits $a_3\ a_5\ a_6\ a_7\ a_9 \ldots a_n$

One determines the test digits by the conditions $T = \Sigma p_i a_i = 0$, a condition which is written here

$a_1 + 2a_2 + 4a_4 + \ldots$ (a)

The second member is a binary # known at the most N digits. The equality determines then perfectly the value of the test digits

If after transmission, there is no error, we will find $T = 0$.

If there is an error, ~~carrying~~ on for example the digit a_a replaced by $(1 - a_a)$ the number T will take the value

$$T = a_1 + 2a_2 + \dots \qquad (a)$$

The value of T will be the indication of the digit ~~which~~ is wrong

If there are two errors on the digits of the indication a and β, T will take the value.

$$T = a + \beta \neq 0$$

If there are more than 2 error, T would be nothing, the code obtained is then corrector of one error, detector of two errors. It is useful to automate the control to suppose the #'s disposed as a matrix. For example for N = 7, one would have the matrix

$$\begin{vmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{vmatrix}$$

to the numbers 0010110 and 0010010 will correspond the matrices

$$\left| \quad \right| \text{ and } \left| \quad \right|$$

The number T is obtained in making it follow each line of the matrix of its digits to parity (Σi)

One obtains here $\begin{vmatrix} 0 \\ 0 \\ 0 \end{vmatrix}$ and $\begin{vmatrix} 1 \\ 0 \\ 1 \end{vmatrix} = 5$.

The first # is correct; the fifth digit of the second # is false.

4. Principle of Corrector Code of K errors.

Lets see now to which condition the numbers p must satisfy in order that the calculation of T would permit to correct K errors.

We will suppose that the test digits are ~~in~~ in sufficient # so that knowing the digits of info. one could realize the condition $T = \sum a_1 p_1 = 0$

If after transmission the digits of the line $a_1 a_2 \dots a_j$ $(j \leq K)$ are erroneous the # T calculated on the deformed word will take the value $T = p_{a1} + p_{a2} + \dots + p_{a_j}$ (a)

It is necessary that the number that is found be $\neq$ of the rows $a_1 a_2 \dots a_j$ that is to say

$$p_{a_1} + p_{a_2} + \dots + p_{a_j} \neq p_{a_1} + p_{a_2} + \dots + p_{a_{j'}}$$

when $j \leq k$ + $j' \leq k$

and the two together

$(a_1 a_2 \dots a_j)$ $(a'_1, a'_2 \dots a'_j)$ are not identical.

This condition can be written

(42) ______

When $l \leq 2K$

⑩ and the $\lambda_1 \lambda_2 \ldots \lambda_1$ being all different. One ought to Choose the #'s p in Such a way that the addition, in the ring ⓐ, at the most 2K of these numbers gives a result not null.

One time determined a group of n numbers p, it will be necessary to Choose the test digits. It is helpful then for this to replace the group obtained by another group of n numbers, but Containing the Successive powers of 2.

$$1, 2, 2^2 \ldots 2^{K-1}$$

K designating the number of digits of the greatest number of p obtained.

Place for this the numbess p in a matrix M of n Columns + K lines $(K < n)$ each # p then being represented by a column.

$$p_1 = \Big|\Big|$$

If the row of this matrix in the O a) $K' < K$, that is, $K-K'$ lines of the matrix are ∃ linear combo of K' lines remaining.

If one gets rid of $K-K'$ lines you will obtain a matrix m' of numbers p' which will verify again the condition (42)

This being so we will be able to extract from the matrix of m' a □ matrix Δ of k' lines of which the calculated determinat in the O a will not be null.

One will have then determinat

$\Delta = 1$

Since the only possib values are 0 & 1.

In ~~multiplying~~ nullifying the matrix m' by Δ^{-1}, one will obtain the matrix m'' formed of #'s ρ'' such that $\rho'' =$

$$\Big| \quad \Big| = \Delta^{+} \Big|\Big|$$

and thus the #'s ρ''_i will verify again the condition 42.

and besides, the matrix m'' will contain at nither moment the matrix $\Delta^{-1} \times \Delta$,

12

this is to say the matrix unity, is for the whole of the groups of p'' will contain the successive powers of 2:

$1, 2, 2^2, \ldots \ldots 2^{K^i}-1$

The corresp. indication will be taken as test digits & the condition 41 will determine these digits in function of digits of info by = of binary #'s of K^2 digits.

all the problem comes back to construction the groups of #'s p satisfactory to the condition 42.

Formation nearest nearer to the following (what comes after the #'s p,

Take 1st $p_1 = 1$ $p_2 = 2^2 \ldots p_K = 2^{2K-1}$

Then $P^{*}_{2K} + 1 = 2^{2K} - 1$

$P_{2K} + 2 = 2^{2K}$

These #'s satisfy already are conditions (42).

To eptend this following in the order of P increasing, suppose we arrive at the numb p' ~~a~~ of L digits.

Let us consider the group of #'s P_1 to P_i & of there sums in the 0 a by groups of 2, 3, --- ($2K-1$)

All the numbers obtained are at the most L digits.

If a # exists not contained in the group thus formed & included between P_i & p' this number will be taken P_{i+1} & have the value of $P_i + 1$ (if there are several #'s you will chose naturally the smallest.)

If not you will take $P_i + 1 = 2^i$

(one can thus continue step by step until you obtain n numbers P.

If P to the base n has k digits, the #s $1, 2, 2^2, \dots 2^{k-1}$ will be included in the following of P.

The following will be then directly useable to form a code

It will remain to establish the table of correspon between the # values of the sum

? $P_2 - P + \dots P_j$ — — — —

+ the value of indications
$\delta_1 \quad \delta_2 \quad - \sim \quad \delta_J$

the process thus defined is ~~long~~ long ~~to figure out~~ enough to utilize.

However for the reasonable values of $n + K$, it doesn't surpass the possibilities of the calculator of medium power.

The determination apriori of K of ~~digits~~ test appears difficult enough.

therefore we are going to expose a more systematic process of research for #'s P.

6. Systematic Formation of #'s p.

The theory of Congruencies, so much used in proofs of arithmetic operations, is going to furnish us a method of Calculating of numbers p.

Let us designate by q a number irreducable of N+1 digits & by g' the rest of the division in the @ of a number g by q. The number g' will have the maximum N digits

We will place then:

$$p_1 = 1 + 2^N \left({}_1^{3}\right)' + \underline{\qquad\qquad\qquad}$$

that is to say that the number p_1 is formed by juxtaposition of the successive remains of the division by @ q, by unequal powers @ of # i. We are going to show that these numbers p_i satisfy the condition 42.

In effect suppose:

(61) ...

And that will bring about

(62) ....

in placing: $S_i = \ldots$

and $S'_i = \ldots$

but if we can considered the product

$$\Pi(\lambda_i + \lambda_j) \quad i = 2, 3 \ldots$$
$$j = 1, 2 \ldots$$

This product can be written in the form of a determinant of Van der Monde of which the square will contain the

line
$S_1, S_2 \dots S_T$.

As $S_{21} = S_{12}$, the condition 62 brings about

$$\Pi(A_i + \lambda_j) = 0 \pmod{q}$$

Then one of the factors for example $\lambda_i + \lambda_j$, will be divisible by ~~pa~~ q. As the sum in ② of #'s $\lambda_i + \lambda_j$ has less digits than the number q, it would result that

$$\lambda_i + \lambda_j = 0 \quad \lambda = \lambda_j.$$

Thus the hypotheses 61 cannot be realized unless at least two indices are equal.

Thus the #'s p that we have formed fulfill the condition 42 + can serve to form a code corrector of K errors. Naturally we will transform them as indicated in paragraph 4 to form a following containing the powers of 2. In the general case, p_i including KN digits, there will be (you would use) KN test digits

In taking $q = 19 = 10.011$ one Calculates easily the numbers p' and the matrix M:

$$\left|\qquad\qquad\qquad\qquad\right|$$

1 2 3 ... 15

This matrix M is of a range 8, for the determinant formed from the Columns 1, 2 4 8 6 12 7 14 (Chosen because they have the most zeroes) = 1

The matrix Δ (par. 4) will be formed with these columns). In inverting one finds the matrix.

$$d^{-1} = \left|\qquad\qquad\qquad\right|$$

The product $\Delta^{-1} M$ (in the ⓐ) gives a definitive matrix

$$\left|\qquad\qquad\qquad\qquad\right|$$

1 ... 15

The digits of the ranges
1 2 4 6 7 8 12 14
Will serve as test digits, the seven other digits will be the Supports for the info.

To verify + correct a word, you will make the sum

$$T = \sum p_i^{a_i}$$

If it is null, there will have not been an alteration of the word (or more than 4 errors). If T is not null one will be able to find again the false digits (in admitting that there would not be more than 2) in using ~~the~~ following table which gives the possible values of T followed in parenthesis by the rang of the false digits.

One will remark that the # T takes a 121 possible values $(1 + C_{15}^1 + C_{15}^2)$ and that one uses a number of 8 digits to write it. The code uses one test # more than ~~it~~ is theoretically indispensable, but it is not certain that one can construct codes having only one number of test digits equal to the whole by the excess of $\log_2 H$.