Information Theory
R.J. Marks II Class Notes
Texas Tech University (1976)
TEST #1 PLUG SHEET

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

* MARGINAL PROBABILITY: \( P[A_1] = \Sigma_j P[A_1, E_j] \)

* CONDITIONAL PROBABILITY: \( P[A \mid B] = \frac{P[A \land B]}{P[B]} \)

* JOINT PROBABILITY: \( P[AB] \)

* STATISTICAL IND: \( P[ABC] = P_A P_B P_C \)

* PDF (PROBABILITY DIST) \( f(X) = P[X = x] \)

* MARGINAL PDF: \( f(X) = \int f(X,Y)dY \)

* CONDITIONAL PDF: \( f(X \mid Y) = \frac{f(X,Y)}{P[Y]} \)

* PROBABILITY DENSITY FUNCTION: \( P(X) = \frac{\partial^2}{\partial x \partial y} P[X \leq x, Y \leq y] \)

* JOINT PDF: \( P(X,Y) = \frac{\partial^2}{\partial x \partial y} P[X \leq x, Y \leq y] \)

* MARGINAL PDF: \( P(X) = \int_{-\infty}^{\infty} P(X,Y) \, dY \)

* CONDITIONAL PDF: \( P(Y \mid X) = \frac{P(X,Y)}{P(X)} \)

* INFORMATION MEASURE: \( I[X \mid K] = -\log P[X \mid K] \)

* THE MATHEMATICS OF ENTROPY

* ENTROPY DEFIN: \( H(S) = -\sum_k P[X = k] \log P[X = k] \)

* ADDITIVE PROPERTY: \( H(p_1, ..., p_n) = H(p_1, ..., p_{n-1}) + p_n H(\frac{q_1}{p_n}, ..., \frac{q_n}{p_n}) \)

* LEMMAS \( \frac{1}{2} \) COR. OF H

* CONVEX FUNCTION CRITERION: \( \frac{d^2 f}{dx^2} \leq 0 \quad \frac{1}{2} [f(x_1) + f(x_2)] \leq f(\frac{x_1 + x_2}{2}) \)

* EXTREMA PROPERTY: \( H(x_1, ..., x_m) \leq \ln m \)

* JOINT ENTROPY: \( H(X,Y) = -\sum_i \sum_j P(x_i, y_j) \log p(x_i, y_j) \)

* CONDITIONAL ENTROPY: \( H(X \mid Y) = H(X) + H(Y) \)

* \( H(X \mid Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y) \)

* \( H(X \mid Y) \leq H(Y) \)
*SOURCE EXTENSION*

S has $q^n$ symbols $\Rightarrow S^n$ has $q^n$ symbols

\[ \sum_{s_n} = \frac{q^n}{1} = \frac{q}{2} \cdots \frac{q^n}{n} \]

\[ p(o_i) = p(s_{i_1}) p(s_{i_2}) \cdots p(s_{i_n}) \]

\[ H^n(s) = n H(s) \]

*CHANNEL ENTROPY*

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

CHANNEL EQUIVOCATION: $H(X/Y)$

CONDITIONAL PROBABILITY MATRICES
A. TWO APPROACHES

a. FREQUENCY OF EVENTS APPROACH
   Let \( N(x_k) \) be the number of times an event \( x_k \) occurs.
   \[ N = \text{total number of events} \geq n(x_k) \]
   \[ \Rightarrow P(x_k) = \lim_{n \to \infty} \frac{n(x_k)}{N} \]

b. AXIOMATIC (PROBABILITY MEASURE) APPROACH
   Let \( \Omega_k \) be a point in a sample space.
   \( m[\Omega_k] \) be a real \( \frac{1}{2} \) single valued measure on \( \Omega_k \).
   \[ m[E] = \sum m[\Omega_k \in E] \]
   Two events are disjoint if \( m[A \cup B] = m(A) + m(B) \)
   (This is called "additive property" of the measure)
   \( m(\emptyset) = 0 \) iff \( X = \emptyset \) (null set)
   \( m(U) = 1 \) iff \( X = U \) (universal set)
   Also
   1) \( m(A) \leq m(B) \) if \( A \subseteq B \)
   2) \( m(A) = m(B) - m(B - A) \) if \( A \subseteq B \)
   3) \( m(A) = m(U - A) = 1 - m(A) \)
   4) \( m(A \cup B) = m[A \cup (A \cup B)] \)
      \[ = m(A) + m(B) - m(AB) \]
   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 \( \exists P(A) \geq 0 \) \forall \text{event } A \subseteq \Omega

Axiom 2: \( P(\emptyset) = 1 \)

Axiom 3: Let \( S = \{s_1, s_2, \ldots \} \Rightarrow s_i \cap s_j = \emptyset \ \forall \ i \neq j \)

(\& all \( s_i \) are disjoint) then

\[ P(s_1 \cup s_2 \cup s_3 + \ldots) = P(s_1) + P(s_2) + \ldots \]

Theorem 1: Let \( S \) be a sample space and \( P \) be a probability measure on \( S \), then

\[ P(\overline{A}) = P(A^c) = 1 - P(A) \]

Theorem 2: Let \( S \) be a sample space with probability measure \( P \), then \( \forall P(A) \geq 1 \) \& \( \text{A} \subseteq \Omega \)

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 \( P \) disjoint subsets \( \{A_1, A_2, \ldots, A_p\} \), repartition \( S \) into \( S \) disjoint subsets \( \{B_1, B_2, \ldots, B_s\} \).

The joint probability of events \( A_i \) and \( B_j \) occurring is denoted \( P[A_i, B_j] \)

Given the joint probability, the marginal probabilities are \( P[A_i] = \sum_{j=1}^{n} P[A_i \cap B_j] \)

and \( P[B_j] = \sum_{i=1}^{m} P[A_i \cap B_j] \)

For three disjoint partitions of \( S \):

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

and \( P[C_k] = \sum_{i=1}^{m} \sum_{j=1}^{n} P[A_i, B_j, C_k] \)

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

\[ S: \{A_1, A_2, \ldots, A_r\} \]
\[ \bigcap_{i \neq j} A_i \cap A_j = \emptyset \]

\[ S: \{B_1, B_2, \ldots, B_s\} \]
\[ \bigcap_{i \neq j} B_i \cap B_j = \emptyset \]

\[ A_1 \quad n_1 \quad n_{12} \quad n_{1j} \quad n_{1s} \]

\[ A_2 \quad n_{21} \quad n_{22} \quad n_{2j} \quad n_{2s} \]

\vdots

\[ A_i \quad n_{i1} \quad n_{i2} \quad \ldots \quad n_{ij} \quad n_{is} \]

\vdots

\[ A_r \quad n_{r1} \quad n_{r2} \quad \ldots \quad n_{rj} \quad n_{rs} \]

\[ \text{LET} \quad \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij} = n \]

\[ \Rightarrow P(A_i, B_j) = P(A_i \cap B_j) = \frac{n_{ij}}{n} \quad \text{JOINT} \]

\[ P(A_i) = \sum_{j=1}^{s} P(A_i, B_j) = \frac{1}{n} \sum_{j=1}^{s} n_{ij} \quad \text{MARGINAL} \]

\[ P(B_j) = \sum_{i=1}^{r} P(A_i, B_j) = \frac{1}{n} \sum_{i=1}^{r} n_{ij} \]

\[ P(A_i/B_j) = \frac{n_{ij}}{\sum_{i=1}^{r} n_{ij}} \quad \text{CONDITIONAL} \]

\[ P(B_j/A_i) = \frac{P(A_i, B_j)}{P(A_i)} \]

\[ P(B_j/A_i) = \frac{P[A_i, B_j]}{P[A_i]} \]
D. STATISTICAL INDEPENDENCE

- Two events, $A \neq B$, are statistically independent iff $P[A, B] = P[A]P[B]$
- It follows that $P[A / B] = P[A]
- $N$ events $\{A_1, A_2, ..., A_i, ... A_n\}$ are statistically independent iff $\forall i \neq j \neq k...
- $P[A_i, A_j] = P(A_i)P(A_j)$
- $P[A_i, A_j, A_k] = P(A_i)P(A_j)P(A_k)$
- $P[A_i, A_j, A_k, ..., A_n] = P(A_i)P(A_j)P(A_k)...P(A_n)$

E. RANDOM VARIABLE

- Defined: 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 admissible sets for which a probability is defined.

F. PROBABILITY DISTRIBUTION FUNCTION

- Defined: $P[x \leq x] = PDF$ of R.v. $x$
- $P[1 \leq b] = f[1 \leq a] = P[a < x \leq b] \forall b > a$
- Joint PDF = $P[x \leq X_B, y \leq X]$
- = $P[sample\ point\ is\ in\ appropriate\ quadrant]$
- Marginal PDF's
  $P[x \leq x] = P[x \leq x, y < \infty]$
  $P[y \leq y] = P[y \leq y, x < \infty]$
- Conditional PDF's
  $P[x \leq x / y \leq y] = P[x \leq x, y \leq y] / P[y \leq y]$
  $P[y \leq y / x \leq x] = P[x \leq x, y \leq y] / P[x \leq x]$
G. Continuous R.V. / Probability Density Function (pdf): 

- pdf \( f(x) = p(x) \) (PDF) = \( \frac{d}{dx} P(x) \) 

\[
(pdfs \text{ are right continuous})
\]

\[
\rho(x) \Delta x = P[x - \Delta x < x < x]
\]

- Properties of a pdf for a continuous R.V.:
  1. \( \rho(x) \geq 0 \)
  2. \( P[a \leq x \leq b] = \int_a^b \rho(x) \, dx \)
  3. \( \int_{-\infty}^{\infty} \rho(x) \, dx = 1 \) (also, \( \rho(x_0) = 0 \))

H. Joint, Marginal, \( \frac{1}{2} \) Conditional pdf's:

- Joint pdf \( f(x, y) \) for joint pdf \( f[x, y] \) must be continuous \( \frac{1}{2} \) twice (mixed) differentiable

\[
\text{Joint pdf} f(x, y) = \lim_{\Delta x, \Delta y \to 0} \frac{1}{\Delta x \Delta y} \left[ \rho(x, y) - \rho(x - \Delta x, y - \Delta y) - \rho(x, y - \Delta y) + \rho(x - \Delta x, y) \right]
\]

- Some relationships:

\[
\int_{-\infty}^{\infty} \rho(x, y) \, dx \, dy = 1
\]

Joint marginal density:

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(x, y) \, dx \, dy = P(x \leq x, y \leq y)
\]

- Conditional pdf:

\[
P[y \leq y / x - \Delta x < x < x] = \frac{P[y \leq y / x - \Delta x < x, x \leq y]}{P[x - \Delta x < x < x]}
\]

\[
= \frac{\int_{x - \Delta x}^{y} \int_{-\infty}^{\infty} \rho(x, y) \, dx \, dy}{\int_{x - \Delta x}^{x} \rho(x) \, dx}
\]

Differentiating wrt \( y \):

\[
\rho(y / x) = \frac{\rho([x, y])}{\rho(x)}
\]
 INFORMATION MEASURE

Consider a sample space \( S \) and an event \( x_k \in S \).

We wish to associate an amount of information associated with the occurrence of \( x_k \), \( I(x_k) \).

Hartley recognized that \( I(x_k) = \log \left( \frac{1}{P(x_k)} \right) \).

Intuitively, we would expect

\[
I(x_k, c_j) = I(x_k) + I(c_j) \quad \forall \quad c_j \in S, \quad c_j \perp x_k \text{ ind.}
\]

Shannon proposed the following INFORMATION MEASURE:

\[
I(x_k) = -\log P(x_k)
\]

- A PRIORI UNCERTAINTY ASSOCIATED WITH
  THE OCCURRENCE OF \( x_k \).

- AMOUNT OF INFORMATION ASSOCIATED
  WITH THE OCCURRENCE OF \( x_k \).

- INFORMATION UNITS

  Using \( \log \) base 2, \( I \) is in \( \text{ BITS} \).

  " " " \( e \), " " " \( \text{ NATS} \).

  " " " \( \text{ 10}, " " " \text{ HARTLEYS} \).

  For conversion, use \( \log_a x = \frac{\log_b x}{\log_b a} \).

  1 HARTLEY = 2.32 \text{ BITS}, 1 NAT = 1.44 \text{ BITS}.
THE MATHEMATICS OF ENTROPY

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

\[
H(S) = \sum_{i=1}^{K} p(x_i) I(x_i) = -\sum_{i=1}^{K} p(x_i) \ln p(x_i)
\]

B. SOME PROPERTIES OF ENTROPY:

1. \( H(S) \) IS CONTINUOUS WITH RESPECT TO \( p_i \)

2. \( H(S) \) IS SYMMETRIC. i.e., \( H(p_1, p_2, ..., p_n) = H(p_2, p_1, ..., p_n) \)

3. \( H(S) \) IS MAXIMUM WHEN \( p_i = \frac{1}{n}, i = 1, 2, ..., n \)

PROOF: \[
\frac{dH}{dp_k} = \sum_{i=1}^{n} \frac{\frac{dH}{p_k}}{p_i} \frac{dp_i}{dp_k} = -\frac{1}{p_k} p_k \ln p_k \frac{dp_k}{dp_k} = -\frac{1}{p_k} (p_k \ln p_k) \frac{dp_k}{dp_k} + O
\]

ALSO, SINCE \( p_n = 1 - (p_1 + p_2 + ... + p_{n-1}) \)
WE HAVE \( \frac{dH}{dp_k} = \ln p_n - \ln p_k = 0 \Rightarrow p_n = p_k \)

4. PROPERTY: \( H(p_1, p_2, ..., p_n, q_1, q_2, ..., q_m) = H(p_1, p_2, ..., p_n) + p_n H(q_1/q_n, q_2/q_n, ..., q_m/q_n) \)

WHERE \( p_n = \sum_{i=1}^{n} q_i \)

PROOF: \( H(p_1, p_2, ..., p_n, q_1, q_2, ..., q_m) = -\sum_{i=1}^{n} p_i \ln p_i - \sum_{k=1}^{m} q_k \ln q_k \)

\[
= -\sum_{i=1}^{n} p_i \ln p_i + p_n \sum_{k=1}^{m} q_k \ln q_k - \sum_{k=1}^{m} q_k \ln q_k
\]

\[
= -\sum_{i=1}^{n} p_i \ln p_i + p_n \sum_{k=1}^{m} q_k \ln q_k - \sum_{k=1}^{m} q_k \ln q_k
\]

\[
= H(p_1, p_2, ..., p_n, q_1, q_2, ..., q_m)
\]
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 \( x, \frac{1}{2} x_2 > 0 \)

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

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

   GEOMETRIC MEAN \( \Rightarrow \sqrt{x_1 x_2} \leq \frac{x_1 + x_2}{2} \leq \) ARITHMETIC MEAN

   GEOMETRIC MEAN \( \leq \) ARITHMETIC MEAN, \( \therefore \) LEMMA 1 IS PROVED

B. LEMMA 2: \( \ln x \leq x - 1 \) (OR \( \ln x \ln 2 \leq x \leq (x-1) \ln 2 \))

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

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

   \[ = \sum_{i=1}^{\infty} p_i \ln \frac{1}{p_i} + \frac{\ln m}{m} \sum_{i=1}^{\infty} p_i \]

   \[ \leq \sum_{i=1}^{\infty} p_i \left( \frac{1}{m p_i} - 1 \right) \leq \text{FROM LEMMA 2} \]

   \[ \leq \sum_{i=1}^{\infty} \left( \frac{1}{m} - p_i \right) \leq 1 - 1 = 0 \]

   QED

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

\[ H(\omega) = -\omega \ln \omega - \omega \ln \omega \]
E. **Lemma 3**: Let \( \{x_1, x_2, \ldots, x_q\} \) and \( \{y_1, y_2, \ldots, y_q\} \) be separate disjoint partitions of \( S \). That is,
\[
\frac{q}{\sum_{i=1}^{q} x_i} x_i = \frac{q}{\sum_{j=1}^{q} y_j} y_j = 1.
\]
Then
\[
\frac{\sum_{i=1}^{q} x_i}{\sum_{j=1}^{q} y_j} \ln \frac{x_i}{y_i} \leq \frac{\sum_{i=1}^{q} x_i}{\sum_{j=1}^{q} y_j} \ln \frac{y_i}{x_i} - 1.
\]

**Proof**: \[
\begin{align*}
\frac{\sum_{i=1}^{q} x_i}{\sum_{j=1}^{q} y_j} \ln \frac{x_i}{y_i} & \leq \frac{\sum_{i=1}^{q} x_i}{\sum_{j=1}^{q} y_j} \ln \left( \frac{y_i}{x_i} \right) - 1 \\
& \leq -\sum x_i + \sum y_i.
\end{align*}
\]

or
\[
\frac{\sum_{i=1}^{q} x_i}{\sum_{j=1}^{q} y_j} \ln \frac{x_i}{y_i} - \frac{\sum_{i=1}^{q} x_i}{\sum_{j=1}^{q} y_j} \ln \frac{y_i}{x_i} \leq 0
\]
\[
\Rightarrow \sum x_i \ln \frac{x_i}{y_i} - \sum y_i \ln \frac{y_i}{x_i} \leq 0.
\]

F. **Lemma 4**: Relation of Joint vs Marginal Entropies

\[
\text{Joint entropy: } H(X, Y) \triangleq \sum_{i=1}^{M} \sum_{j=1}^{N} p(x_i, y_j) \ln p(x_i, y_j)
\]
\[
\exists X = \{x_1, x_2, \ldots, x_m\}, Y = \{y_1, y_2, \ldots, y_l\}
\]
\[
H(X, Y) \leq H(X) + H(Y)
\]

**Proof**: \[
H(X) = -\sum_{i=1}^{M} p(x_i) \ln p(x_i)
\]
\[
H(Y) = -\sum_{j=1}^{N} p(y_j) \ln p(y_j)
\]
\[
\therefore H(X) + H(Y) = -\sum_{i=1}^{M} \sum_{j=1}^{N} p(x_i, y_j) \ln p(x_i, y_j) - \sum_{i=1}^{M} p(x_i) \ln p(x_i)
\]
\[
= -\sum_{i=1}^{M} \sum_{j=1}^{N} p(x_i, y_j) \ln q_{ij}
\]
\[
\exists q_{ij} \triangleq p(x_i, y_j) p(y_j) \Rightarrow \sum q_{ij} = 1
\]

Also, let \( p_{ij} = p(x_i, y_j) \)
\[
H(X, Y) = -\sum_{i=1}^{M} \sum_{j=1}^{N} p_{ij} \ln p_{ij}
\]

From Lemma 3:
\[
H(X, Y) = -\sum_{i=1}^{M} \sum_{j=1}^{N} p_{ij} \ln p_{ij} + \sum p_{ij} \ln q_{ij}
\]
\[
\leq -\sum p_{ij} \ln p_{ij} q_{ij}
\]
\[
\leq H(X) + H(Y) \quad \text{QED}
\]
G. Cor F2: \( H(x_1, x_2, ..., x_n) \leq H(x_1) + H(x_2) + ... + H(x_n) \)

Equality if all \( x_i \) are statistically independent.

H. Cor F2: \( H(x_1, x_2, ..., x_n, y_1, y_2, ..., y_m) \leq H(x_1, x_2, ..., x_n) + H(y_1, y_2, ..., y_m) \)

Equality if the random vector \( \{x_1, x_2, ..., x_n\} \) is statistically independent of \( \{y_1, y_2, ..., y_m\} \).

I. Conditional Entropy (defn)

\[
H[\frac{Y}{X = x_i}] = - \sum_{j=1}^{L} p(y_j|x_i) \ln p(y_j|x_i)
\]

\[
H[\frac{Y}{X}] = E[H[\frac{Y}{X = x_i}]] = -\sum_{i=1}^{m} \sum_{j=1}^{L} p(x_i, y_j) \ln p(y_j|x_i)
\]

J. Lemma 5:

\( H(x, y) = H(x) + H(y|x) = H(y) + H(x|y) \)

Proof:

\[
H(x, y) = -\sum_{i=1}^{m} \sum_{j=1}^{L} p(x_i, y_j) \ln p(x_i, y_j)
\]

\[
= -\sum_{i=1}^{m} \sum_{j=1}^{L} p(x_i, y_j) \ln p(x_i) p(y_j|x_i)
\]

\[
= H(x) + H(y|x)
\]

K. Lemma 6:

\( H(y|x) \leq H(y) \)

\[
H(x, y) = H(x) + H(y|x) \leq H(x) + H(y) \rightleftharpoons \text{from Lemma 5}
\]

\[
\leq H(x) + H(y) \rightleftharpoons \text{from Lemma 4}
\]

\[
\Rightarrow H(y|x) \leq H(y)
\]
Source Extension

A. Definition: Utilizing a source of q symbols to generate a symbol words denotes the source as the n\textsuperscript{th} extension.

The n\textsuperscript{th} extension of S has q\textsuperscript{n} symbols.

B. Some Theorems

1. The n\textsuperscript{th} extension of a source is complete

\[ S = \{s_1, s_2, \ldots, s_q \} \Rightarrow (p_1, p_2, \ldots, p_q) \]
\[ S^n = \{\sigma_1, \sigma_2, \ldots, \sigma_q \}^n \Rightarrow (p'_1, p'_2, \ldots, p'_q) \]
\[ \sum_{x} p'_1 = \sum_{x} p(\sigma_1) \Rightarrow p(\sigma_1) = p_{i_1} p_{i_2} \ldots p_{i_n} \]
\[ \sum_{x} p(\sigma_x) = \sum_{x=1}^{n} \frac{q}{d_1} \frac{q}{d_2} \ldots \frac{q}{d_n} p_{i_1} p_{i_2} \ldots p_{i_n} \]

2. \( H^n(S) = n \cdot H(S) \)

Proof: \( H^n(S) = \sum_{x} p(\sigma_x) \cdot \frac{1}{\log_2 p(\sigma_x)} \)
\[ = \sum_{x} p(\sigma_x) \cdot \frac{1}{\log_2 p_{i_1} p_{i_2} \ldots p_{i_n}} \]
\[ = \sum_{x} p(\sigma_x) \cdot \frac{1}{\log_2 p_{i_1}} + \sum_{x} p(\sigma_x) \cdot \frac{1}{\log_2 p_{i_2}} + \ldots + \sum_{x} p(\sigma_x) \cdot \frac{1}{\log_2 p_{i_n}} \]

Consider the k\textsuperscript{th} term:
\[ \sum_{x} p(\sigma_x) \cdot \frac{1}{\log_2 p_{i_k}} = \sum_{x=1}^{\frac{q}{d_k}} \frac{q}{d_1} \frac{q}{d_2} \ldots \frac{q}{d_{k-1}} p_{i_1} p_{i_2} \ldots p_{i_{k-1}} \cdot \frac{1}{\log_2 p_{i_k}} \]
\[ = \sum_{x=1}^{\frac{q}{d_k}} \frac{1}{\log_2 p_{i_k}} \cdot p_{i_1} p_{i_2} \ldots p_{i_{k-1}} \]
\[ \therefore H^n(S) = n \cdot H(S) \]
**CHANNEL ENTROPY**

A. SOURCE \[ x; n \text{ symbols} \] CHANNEL \[ y; m \text{ symbols} \] RECEIVER

- \( H(y) = \text{Ave. Info per source symbol} \)
- \( H(y/x) = \text{Ave. Info per receiver character} \)
- \( H(y/x) = \text{Info about received symbol given transmitted signal (measure of error \& noise)} \)
- \( H(x/y) = \text{Channel equivocation (measure of the recoverability of the input at the receiver)} \)
- \( H(x,y) = \text{Average info per character pair} \)

B. SOURCE CHARACTERIZATION

1. JOINT PROB. MATRIX:

\[
\begin{array}{cccccc}
  & y_j & y_i & y_l & \ldots & y_m \\
  x_1 & p(x_1) & p(x_1y_j) & p(x_1y_i) & p(x_1y_l) & p(x_1y_m) \\
  x_2 & p(x_2) & p(x_2y_j) & p(x_2y_i) & p(x_2y_l) & p(x_2y_m) \\
   & \vdots & & & & \\
  x_i & p(x_i) & p(x_iy_j) & p(x_iy_i) & p(x_iy_l) & p(x_iy_m) \\
   & \vdots & & & & \\
  x_n & p(x_n) & p(x_ny_j) & p(x_ny_i) & p(x_ny_l) & p(x_ny_m) \\
\end{array}
\]

\[
\begin{array}{cccccc}
  & y_j & y_i & y_l & \ldots & y_m \\
  & p(y_j) & p(y_i) & p(y_l) & \ldots & p(y_m) \\
\end{array}
\]

MARGINAL PROBABILITIES

2. CONDITIONAL PROB. MATRIX

\[
p(x_i/y_j) = \frac{p(x_i, y_j)}{p(y_j)}
\]

\[
p(y_j/x_i) = \frac{p(x_i, y_j)}{p(x_i)}
\]
TEST # 2 PLUG SHEET

MUTUAL INFORMATION: \( I(x, y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \)

\( I(x) \geq I(x; y) \)

TRANSINFORMATION: \( I(x; y) = H(x) - H(x/y) = H(y) - H(y|x) \)

CHANNEL CAPACITY: \( C = \max I(x; y) \)

REDUNDANCY: \( R = \frac{C}{I(x; y)} = 1 - \alpha \)

RELATIVE REDUNDANCY: \( \alpha_r = \frac{1 - I(x; y)}{H(x)} \)

CHANNEL EFFICIENCY: \( \eta = \frac{H_x}{H_y} \)

TRANSMISSION RATE: \( R_t = \frac{H(x)}{\sum_i t_i p(x_i)} \)

INFORMATION CHANNEL

BINARY SYMMETRIC CHANNEL: \( \begin{array}{c}
0 \quad 0
\end{array} \begin{array}{c}
1 \quad 1
\end{array} \begin{array}{c}
0 \quad 1
\end{array} \begin{array}{c}
1 \quad 0
\end{array} \)

UNIFORM CHANNELS: \( p[lb_i, a_i] \), THINGS ADD UP

BINARY ERASURE CHANNEL:

MUROGA'S TECHNIQUE

BLOCK CODE, NON-SINGULAR, UNIQUELY DECODABLE, INSTANTANEOUS

PREFIX PROPERTY

KRAFT'S INEQUALITY: \( \sum_{i=1}^{l_0} N_i f^{-l_i} \leq 1 \)

McMILLAN'S INEQUALITY

SHANNON'S FIRST THEOREM: \( H_p(x) \leq \sum_{i=0}^{L_0} \frac{L_i}{n} H_k \)

AVERAGE WORD LENGTH: \( \bar{l} = \sum_{i=1}^{l_0} t_i p(x_i) l_i \)

CODING PROCEDURES:

SHANNON, SHANNON-FANO

HUFFMAN \( q = \frac{1}{r^\alpha} \alpha \geq \alpha \in \mathbb{R} \)

ALSO MAY USE \( \log \frac{1}{p_i} \leq l_i \leq \log \frac{1}{p_i} + 1 \)
**Mutual Information**

\[ I(x_i; y_j) = \sum_i \sum_j p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \]

- **Self Information:** \[ I(x_i) = -\log p(x_i) \]
- **A Priori Knowledge \( x_i \) is being transmitted and \( y_j \) is being received:** \[ p(x_i) \]
- **Posterior Knowledge:** \[ p(x_i; y_j) \]
- **The difference of these two is the gain in information:** \[ I(x_i; y_j) = \log p(x_i; y_j) - \log p(x_i) \]

**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. \[ I(x_i) = I(x_i; x_i) = \sum_j p(x_i) I(x_i; y_j) \]

Also, \[ I(y_j) = I(y_j; y_j) \geq I(x_i; y_j) \]

**C. Average Information Gain**

\[ I(x; y) = \frac{I(x_i; y_j)}{2} \]

\[ = \sum_i \sum_j p(x_i, y_j) I(x_i; y_j) \]

\[ = H(x) + H(y) - H(x, y) \]

\[ = H(x) - H(x|y) = H(y) - H(y|x) \]

"On the average, observation of any \( y \) gives us \( I(x; y) \) bits of info."
Channel Parameters

A. Channel Capacity
\[ C = \max I(x; y) = \max [H(x) - H(x/y)] \]

B. Rate of Information Transmission
\[ C_t = \frac{c}{t} \text{ bits/sec} \]

\( t \) = time that it takes to xmit each symbol
\[ C_t = \frac{\log n}{t} \text{ for a noise free channel} \]

C. Redundancy

1. Absolute Redundancy: difference twixt maximum capacity and actual info;
\[ R = C - I(x; y) \leftarrow \text{general no!} \]
\[ = \log n - H(x) \leftarrow \text{noise free} \]

2. Relative Redundancy
\[ R_r = \frac{\log n [\log n - H(x)]}{I(x; y)} \leftarrow \text{noise free} \]
\[ = 1 - \frac{I(x; y)}{\log n} \leftarrow \text{general} \]

D. Channel Efficiency
\[ n = \frac{I(x; y)}{\log n} = 1 - R_r \]

E. Generalized Transmission Rate
\[ R_t = \frac{H(x)}{\sum p(x_i) t_i} \]
\[ = \frac{\sum p(x_i) \log p(x_i)}{\sum p(x_i) t_i} \]
THE INFORMATION CHANNEL

A, B; A \rightarrow B, is described by giving an input alphabet, A; \{a_i, i=1,...,r\} and an output alphabet, B; \{b_j, j=1,...,s\} and a set of conditional probabilities, p(b_j/a_i) for i,j.

We denote p(b_j/a_i) = P_{i,j}. Clearly, \sum_j P_{i,j} = 1.

B. BINARY SYMMETRIC CHANNEL

\bar{p} = p(0/0) = p(1/1) \quad p = p(0/1) = p(1/0)

\begin{align*}
0 & \quad \bar{p} \quad 1 \\
1 & \quad 0 \quad \bar{p}
\end{align*}

C. INFO. CHANNEL EXTENSION

CONSIDER INFO. CHANNEL WITH ALPHABETS

A \rightarrow B AND PROBABILITY (COND.) MATRIX P;

DEFN: LET A^n \rightarrow B^n BE THE nTH EXTENSIONS OF A \rightarrow B:

A^n = \{a_i^n, i=1,...,r^n\} \quad B^n = \{b_j^n, j=1,...,s^n\}

THE CHANNEL MATRIX, \Pi, IS THEN

\[ \Pi = \left[ \pi_{11} \pi_{12} \ldots \pi_{1s^n} \right] \left[ \pi_{21} \pi_{22} \ldots \pi_{2s^n} \right] \ldots \left[ \pi_{r^n 1} \pi_{r^n 2} \ldots \pi_{r^n s^n} \right] \]

WHERE \alpha_i^n = \{a_i, a_{i2}, \ldots a_{in}\} \quad \beta_j^n = \{b_{j1}, b_{j2}, \ldots b_{jn}\}

THEN \[ \Pi_{i,j} = p(\beta_j^n/\alpha_i^n) = p_{i,j_1} p_{j_2,j_2} \ldots p_{j_n,j_n} \]

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

\[ H(A) = -\sum_i p(a_i) \log p(a_i) =\text{A PRIORI ENTROPY} \]

\[ H(A/b_j) = -\sum_i p(a_i/b_j) \log p(a_i/b_j) =\text{A POSTERIORI ENT.} \]
D. UNIFORM CHANNELS

-DEFN: AN INFO CHANNEL IS UNIFORM IF ALL ROWS
  & COLUMNS OF \( P = [p_{ij}] = [p_i/q_i]^j \) ARE
  PERTURBATIONS OF EACH OTHER.

\[
\begin{align*}
\text{EX:} & & b_1 & b_2 & b_3 & b_4 \\
& a_1 & 1/2 & 1/2 & 1/2 & 1/2 \\
& a_2 & 1/2 & 1/2 & 1/2 & 1/2
\end{align*}
\]

-CHANNEL CAPACITY

\[
C = \max I(x; y) = \max \left[ H(y) - H(y/x) \right]
\]

FOR A UNIFORM CHANNEL: \( H(y/x_i) = h = \text{CONST.} \)

\[
\Rightarrow H(y/x) = \sum_i p(x_i) H(y/x_i) = h
\]

\( H(y) \) IS MAX FOR \( p(y_j) = \frac{1}{2} \) \( \Rightarrow H(y) = \frac{H}{2} \)

AND \( C = \frac{H}{2} = h \)

E. CHANNEL CAPACITY FOR BSC

\[
\begin{align*}
0 & \xrightarrow{p} 0 & \quad p(0) &= \alpha & \quad p(1) &= \bar{\alpha} \\
1 & \xrightarrow{\bar{p}} 1 & \quad p(0) &= \bar{\alpha} & \quad p(1) &= \alpha
\end{align*}
\]

\[
\begin{align*}
H(x) &= -\alpha \log_2 \alpha - \bar{\alpha} \log_2 \bar{\alpha} \\
H(y/x) &= -p \log_2 p - q \log_2 q \\
I(x; y) &= H(y) - H(y/x) \\
&= H(y) + p \log_2 p + q \log_2 q \\
C &= \max \left[ I(x; y) \right] = 1 + p \log_2 p + q \log_2 q \text{ BITS}
\end{align*}
\]
F. BEC (BINARY ERASURE CHANNEL)

\[
\begin{array}{ccc}
0 & \overset{p}{\rightarrow} & 0 \\
0 & \overset{q}{\rightarrow} & 1
\end{array}
\]

\[
P(Y/X) = \begin{cases} 
0 & p \quad q \\
1 & q \quad p
\end{cases}
\]

Let \( p(0) = \alpha \) and \( p(1) = \overline{\alpha} \)

Let's find the channel capacity:

\[H(X) = -\alpha \log \alpha - \overline{\alpha} \log \overline{\alpha}\]

\[P(X/Y) = \begin{cases} 
0 & \alpha/\alpha = 1 \\
0 & \alpha/\alpha = 0 \\
1 & \overline{\alpha}/\overline{\alpha} = \alpha/\alpha = 1
\end{cases}
\]

\[H(X/Y) = \alpha \log 2 - \alpha \log \alpha - \overline{\alpha} \log \overline{\alpha} - \alpha \log 2
\]

\[= -q (\alpha \log \alpha + \overline{\alpha} \log \overline{\alpha}) = q H(X)
\]

\[I(X;Y) = H(X) - H(X/Y) = (1-q) H(X) = p H(X)
\]

\[C = \min I(X;Y) = p \min H(X) = p \text{ bits}
\]

G. BINARY SYMMETRIC CHANNEL EXTENSIONS

KRONECKER MATRICES
Muroga's Technique

1. Consider

\[
\begin{bmatrix}
P_{11} & P_{12} \\
\vdots & \vdots \\
P_{n1} & P_{n2}
\end{bmatrix}
\begin{bmatrix}
Q_1 \\
\vdots \\
Q_n
\end{bmatrix}
= \begin{bmatrix}
-H(P_{11}) \\
\vdots \\
-H(P_{n2})
\end{bmatrix}
\]

Then it can be shown:

\( I(x; y) = - (p_1 \log_2 p_1 + p_2 \log_2 p_2) + p_1 Q_1 + p_2 Q_2 \)

where \( p_1' = \text{Pr}[Y_1] \) and \( p_2' = \text{Pr}[Y_2] \).

We wish to maximize \( I \) via Lagrange multipliers, \( \mu \).

Let

\[
U = - (p_1 \log_2 p_1 + p_2 \log_2 p_2) + p_1' Q_1 + p_2' Q_2 + \mu (p_1' - p_2')
\]

\[
\frac{dU}{dp_1} = - (\log_2 p_1 + \log_2 e) + Q_1 + \mu \quad (\text{2})
\]

\[
\frac{dU}{dp_2} = - (\log_2 p_2 + \log_2 e) + Q_2 + \mu \quad (\text{3})
\]

Setting \( \frac{dU}{dp_1} = \frac{dU}{dp_2} = 0 \) gives

\[
\mu = -Q_1 + (\log_2 e + \ln p_1')
\]

\[
\mu = -Q_2 + (\log_2 e + \ln p_2')
\]

Substituting into (1) gives

\[
C = \max I(x; y) = Q_1 - \log_2 p_1' = Q_2 - \log_2 p_2'
\]

\[
\Rightarrow p_1' = 2^{Q_1 - C} \quad p_2' = 2^{Q_2 - C} = 1 - p_1'
\]

Thus \( C = \log_2 (2^{Q_1} + 2^{Q_2}) \) bits.

2. Comments

(1) In general, we must solve:

\[
\begin{bmatrix}
P_{11} & \cdots & P_{1n} \\
\vdots & \ddots & \vdots \\
P_{n1} & \cdots & P_{nn}
\end{bmatrix}
\begin{bmatrix}
Q_1 \\
\vdots \\
Q_n
\end{bmatrix}
= \begin{bmatrix}
-H_1 \\
\vdots \\
-H_n
\end{bmatrix}
\]

\[
C = \log_2 \sum_{i=1}^{n} 2^{Q_i}
\]

(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.
0 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.
\( \sum_{i=1}^{q} r^{-l_i} \leq 1 \)

Proof: Let \( \beta_i \) be the number of code words of length \( l_i \), if \( \text{Max}[l_i] = l \), then

\( \sum_{i=1}^{l} \beta_i = q \)

There are \( q \) words of the code book.

Now

\( \sum_{i=1}^{q} r^{-l_i} = \sum_{i=1}^{l} \beta_i r^{-l} \)

We show sufficiency by writing

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

\( \Rightarrow \sum_{i=1}^{l} \beta_i r^{-l} \leq 1 \)

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

or \( 0 \leq n_{l} r^{-l} \leq 1 - n_1 r^{-1} - n_2 r^{-2} - \ldots - n_{l-1} r^{-(l-1)} \)

\( 0 \leq n_{l} \leq \frac{1}{r} - n_1 r^{-1} - \frac{1}{r^2} - n_2 r^{-2} - \ldots - \frac{1}{r^{l-1}} - n_{l-1} r^{-l} \)

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

\( 0 \leq n_{l-2} \leq \frac{r^2}{r^3} - n_1 r^{-1} \frac{r^2}{r^3} - n_2 r^{-2} - \ldots - \frac{1}{r^{l-3}} - n_{l-3} r^{-l+2} \)

\( 0 \leq n_{l-3} \leq \frac{r^3}{r^4} - n_1 r^{-1} \frac{r^3}{r^4} - n_2 r^{-2} - \ldots - \frac{1}{r^{l-4}} - n_{l-4} r^{-l+3} \)

\( 0 \leq n_{l-4} \leq \frac{r^4}{r^5} - n_1 r^{-1} \frac{r^4}{r^5} - n_2 r^{-2} - \ldots - \frac{1}{r^{l-5}} - n_{l-5} r^{-l+4} \)

\( \vdots \)

\( 0 \leq n_3 \leq \frac{r^3}{r^4} - n_1 r^{-1} \frac{r^3}{r^4} - n_2 r^{-2} \)

\( 0 \leq n_2 \leq \frac{r^2}{r^3} - n_1 r^{-1} \frac{r^2}{r^3} \)

\( 0 \leq n_1 \leq \frac{r}{r^2} - n_1 r^{-1} \frac{r}{r^2} \)
(CODE PROPERTIES)

Now

\[ D_1 \leq r \]

That is, we may, at most, build \( r \) words with length \( \ell_1 = 1 \). Thus, if we use \( n_1 \) words of length \( \ell_1 = 1 \), 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:

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

etc., one may prove the neg. 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 existence of a U.D. code, McMillan's Inequality (the same as Kraft's) says that Kraft's inequality is also necessary for the existence of a U.D. code.
PROOF: CONSIDER
\[
\left( \sum_{i=1}^{q} r^{-l_i} \right)^n = \left( r^{-l_1} + r^{-l_2} + \ldots + r^{-l_n} \right)^n = \sum_{k=1}^{n} r^{-k}.
\]

\[
k = l_1 + l_2 + \ldots + l_n.
\]

IF \( \max l_i = l \), THEN IT IS CLEAR THAT
\( n \leq k \leq n \).

NOW, IF \( N_K = \# \text{OF TERMS OF } r^{-l_k} \), THEN
\[
\sum_{k=1}^{q} r^{-l_k} = \sum_{k=n}^{n} r^{-k}.
\]

TO BE UNIQUELY DECODABLE, WE REQUIRE
\[
N_K \leq r^k \Rightarrow \sum_{k=n}^{\infty} N_K r^{-k} \leq \sum_{k=n}^{\infty} N_k r^{-k} \leq n n^2 \leq n^2 \Rightarrow \left( \sum_{i=1}^{q} r^{-l_i} \right)^n \leq n^2.
\]

NOW, IF \( x^n \leq n^2 \), THEN \( x \leq 1 \)
\[
\Rightarrow \sum_{i=1}^{q} r^{-l_i} \leq 1.
\]
0. SHANNON'S FIRST THEOREM

A. \( \bar{L} = \text{AVE LENGTH} = \sum_i p(x_i) \ell_i \)

B. THE AVE LENGTH OF A U.D. CODE IS
   LOWER BOUNDED BY \( H(x) / \log r \) THAT IS
   \[ \bar{L} \geq \frac{H(x)}{\log r} \]

PROOF:
   \( \sum_{i=1}^{n} q_i \log q_i \leq \sum_{i=1}^{n} q_i \log p_i \)

LET \( q_i = \frac{r^{-\ell_i}}{\sum_{i=1}^{n} r^{-\ell_i}} \)

\[ \Rightarrow H(x) = \sum_{i=1}^{n} p(x_i) \log \frac{r \ell_i}{r \sum_{i=1}^{n} r^{-\ell_i}} \]
\[ \leq - \sum_{i=1}^{n} p(x_i) \log r \sum_{i=1}^{n} r^{-\ell_i} \]
\[ \leq - \sum_{i=1}^{n} p(x_i) \log r^{-\ell_i} + p(x_i) \log \frac{r \sum_{i=1}^{n} r^{-\ell_i}}{\sum_{i=1}^{n} r^{-\ell_i}} \]
\[ \leq - \sum_{i=1}^{n} \ell_i p(x_i) \log r + \sum_{i=1}^{n} p(x_i) \log \frac{r \sum_{i=1}^{n} r^{-\ell_i}}{\sum_{i=1}^{n} r^{-\ell_i}} \]
\[ \leq - \sum_{i=1}^{n} \ell_i p(x_i) \log r + \sum_{i=1}^{n} p(x_i) \log \frac{r \ell_i}{\sum_{i=1}^{n} r^{-\ell_i}} \]
\[ \leq - \sum_{i=1}^{n} \ell_i p(x_i) \log r + \sum_{i=1}^{n} p(x_i) \log \frac{r \ell_i}{\sum_{i=1}^{n} r^{-\ell_i}} \]
\[ \Rightarrow \sum_{k=1}^{\infty} r^{-\ell_k} \leq 1 \]
\[ \Rightarrow \sum_{i=1}^{n} r^{-\ell_i} \leq 0 \]
\[ \therefore H(x) \leq \ell \log r \]

OR \( \bar{L} \geq \frac{H(x)}{\log r} \)

C. CRITERION FOR MEETING LOWER BOUND

(BASE 2)

\[ H(x) \leq \ell \log_2 2 = \ell \]
\[ \Rightarrow \ell \log_2 p(y_{k}) \leq \ell_k \]
\[ \Rightarrow \frac{1}{p(y_{k})} \leq 2^{\ell_k} \]
\[ p(y_{k}) \geq 2^{-\ell_k} \]

\therefore \text{EQUALITY IS MET ONLY IF } p(x_i) = 2^{-\ell_k}.

IN GENERAL, IF \( p(x_i) = r^{-\ell_k} \)
\[ c. \quad \lim_{n \to \infty} \frac{L_n}{n} = H_r(s) \]

**SHANNON'S 1ST THM**

* \( n \) refers to the \( n^{th} \) extension and \( y \) to the number of alphabet symbols

**PROOF:** \( H_r(s) \leq I \leq H_r(s) + 1 \)

*This is good always use the \( n^{th} \) ext. then:

\[ H_r(s) \leq I \leq H_r(s) + 1 \]

\[ n \cdot H_r(s) \leq I_n \leq n \cdot H_r(s) + 1 \]

\[ \Rightarrow H_r(s) \leq \frac{I_n}{n} \leq H_r(s) + \frac{1}{n} \]

Obviously, \( \lim_{n \to \infty} \frac{I_n}{n} = H_r(s) \)

Interpret \( \frac{I_n}{n} \) as the number of code symbols used from the original source per sample symbol (eh?)
CODING PROCEDURES (CLASSICAL)

A. SHANNON'S BINARY CODING PROCEDURE

1. FOR AN OPTIMAL CODE, WE CAN ACHIEVE $H(X) = \sum h_i p_i$

2. ARRANGE PROBABILITIES IN DECREASING ORDER
   \[ p_1 \geq p_2 \geq \ldots \geq p_q \]

3. COMPUTE $\alpha_i$'S:
   \[ \alpha_1 = 0 \]
   \[ \alpha_2 = p(x_1) \]
   \[ \alpha_3 = p(x_2) + \alpha_2 \]
   \[ \alpha_4 = p(x_3) + \alpha_3 \]
   \[ \vdots \]
   \[ \alpha_q = p(x_{q-1}) + \alpha_{q-1} \]
   \[ \alpha_{q+1} = p(x_q) + \alpha_q = 1 \]

4. FIND THE SET OF INTEGERS (SMALLEST) WHICH SATISFY:
   \[ 2^i p(x_i) \geq 1 \]

5. EXPAND EACH $\alpha_i$ TO $l_i$ PLACES (IN BINARY FORM) AND NO FURTHER.

EXAMPLE: $p_2 = \{0.4, 0.3, 0.2, 0.1\}$

\[ \alpha_1 = 0 \quad \alpha_3 = 0.7 \]
\[ \alpha_2 = 0.4 \quad \alpha_4 = 0.9 \]

\[ l_1 : 2^{l_1} \frac{4}{10} \geq 1 \Rightarrow l_1 = 2 \]
\[ l_2 : 2^{l_2} \frac{3}{10} \geq 1 \Rightarrow l_2 = 2 \]
\[ l_3 : 2^{l_3} \frac{2}{10} \geq 1 \Rightarrow l_3 = 3 \]
\[ l_4 : 2^{l_4} \frac{1}{10} \geq 1 \Rightarrow l_4 = 4 \]

\[ \alpha_1 = 0 \rightarrow x_1 = 00 \]
\[ \alpha_2 > 0.4 \geq (0.01)_2 \Rightarrow x_2 = 01 \]
\[ \alpha_3 > 0.7 \geq (1.01)_2 \Rightarrow x_3 = 101 \]
\[ \alpha_4 > 0.9 \geq (1.110)_2 \Rightarrow x_4 = 1110 \]
B. THE SHANNON-FANO CODING SCHEME

1. ARRANGE PROBABILITIES IN DECREASING ORDER.

2. DIVIDE INTO \( r \) SECTIONS OF "NEARLY EQUAL" PROBABILITIES. ASSIGN SAME SYMBOL TO EACH COMPONENT IN EACH SECTION.

3. SIMILARLY DIVIDE EACH SECTION AND REPEAT 2.

EXAMPLE: \( r = 2 \)

\[
\begin{array}{cccc}
x_1 & 0.25 & 0 & 0 \\
x_2 & 0.25 & 0 & #1 1 \\
x_3 & 0.125 & 1 & 0 & 0 \\
x_4 & 0.125 & 1 & 0 & #2 1 \\
x_5 & 0.0625 & 1 & 1 & 0 & 0 \\
x_6 & 0.0625 & 1 & 1 & 0 & #3 0 \\
x_7 & 0.0625 & 1 & 1 & 0 & #4 \\
x_8 & 0.0625 & 1 & 1 & 1 & #4 \\
\end{array}
\]
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 \( \frac{1}{2} \) FILL THINGS IN

B. EX. \( r = 2 \)

\[
\begin{array}{c|c|c|c|c|c|c|c|c|c}
X_1 & \frac{1}{3} & 001 & \frac{2}{9} & 01 & \frac{2}{9} & 000 & \frac{2}{9} & 11 & \frac{2}{9} & 11 & \frac{2}{9} \text{(OD)} & \frac{4}{9} (1) \\
X_2 & \frac{1}{3} & 110 & \frac{1}{9} & 001 & \frac{2}{9} & 01 & \frac{2}{9} & 000 & \frac{2}{9} & 10 & \frac{2}{9} (01) & \frac{3}{9} (0) \\
X_3 & \frac{1}{9} & 101 & \frac{1}{9} & 110 & \frac{1}{9} & 001 & \frac{2}{9} & 01 & \frac{2}{9} & 01 & \frac{2}{9} (10) & \frac{2}{9} (01) \\
X_4 & \frac{1}{9} & 110 & \frac{1}{9} & 101 & \frac{1}{9} & 110 & \frac{1}{9} & 001 & \frac{2}{9} & 000 & \frac{2}{9} (11) & \frac{2}{9} (11) \\
X_5 & \frac{1}{9} & 111 & \frac{1}{9} & 110 & \frac{1}{9} & 101 & \frac{1}{9} & 001 & \frac{2}{9} & 100 & \frac{2}{9} (10) & \frac{2}{9} (10) \\
X_6 & \frac{1}{9} & 111 & \frac{1}{9} & 111 & \frac{1}{9} & 111 & \frac{1}{9} & 111 & \frac{1}{9} & 111 & \frac{1}{9} (111) & \frac{1}{9} (111) \\
X_7 & \frac{1}{9} & 0000 & \frac{1}{9} & 111 & \frac{1}{9} & 111 & \frac{1}{9} & 111 & \frac{1}{9} & 111 & \frac{1}{9} (111) & \frac{1}{9} (111) \\
X_8 & \frac{1}{9} & 101 & \frac{1}{9} & 0001 & \frac{1}{9} & 0001 & \frac{1}{9} & 0001 & \frac{1}{9} & 0001 & \frac{1}{9} (0001) & \frac{1}{9} (0001) \\
X_9 & \frac{1}{9} & 011 & \frac{1}{9} & 0001 & \frac{1}{9} & 0001 & \frac{1}{9} & 0001 & \frac{1}{9} & 0001 & \frac{1}{9} (0001) & \frac{1}{9} (0001) \\
\end{array}
\]

C. COMMENTS

- HUFFMAN CODING, THO NOT UNIQUE, WILL
  ALWAYS GENERATE MINIMUM \( c \)

- WHEN \( r \geq 2 \), THEN THE \# OF WORDS TO
  BE CODED MUST BE \( q = r + (r-1) \alpha \)

\( \alpha \) IS ANY INTEGER, USE DUMMY

WORDS WITH PROB. ZERO TO COMPLETE

THE BOOK
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:

\[
p = p[x_1 + x_2 = 5] \quad \text{and} \quad p = 1 - p = p[x_1 + x_2 \neq 5]
\]

Clearly, \( x_1 \) and \( x_2 \) are independent.

Now, sum mutually exclusive events:

\[
P = p[x_1 = 2, x_2 = 3] + p[x_1 = 3, x_2 = 2]
\]

\[
= p[x_1 = 1, x_2 = 4] + p[x_1 = 4, x_2 = 1]
\]

\[
P(x_1, x_2) = P(x_1) \cdot P(x_2)
\]

and \( P(x_1) = \frac{1}{6} \)

\[
\Rightarrow p = \frac{4}{36} \left( \frac{1}{6} \right) \frac{1}{6}
\]

\[
= \frac{4}{216} = \frac{1}{54}
\]

Associated entropy is:

\[
H(5) = -p \log p - \left( 1-p \right) \log \left( 1-p \right)
\]

\[
= \frac{1}{6} \log \frac{5}{6} + \frac{5}{6} \log \frac{1}{6}
\]

\[
= 0.349 \quad \text{bits}
\]

Self-information

\[
\log_2 9 = 3.17 \quad \text{bits}
\]
\[ S = \left\{ \begin{array}{c} H \in T \\ \frac{1}{2} \leq \frac{H}{2p} \end{array} \right\} \quad \# \text{\(H\) = NUMBER OF HEADS} \]

We have here a binomial distribution

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

\[ p(H=h) = \sum_{k=0}^{h} \binom{10}{k} 2^{-10} = 2^{-10} \sum_{k=0}^{h} \binom{10}{k} \]

Thus

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

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

\[ = 2^{-10} \left[ 1 + 10 + \frac{10 \times 9}{2} + \frac{10 \times 9 \times 8}{3!} + \frac{10 \times 9 \times 8 \times 7}{4!} + \frac{10 \times 9 \times 8 \times 7 \times 6}{5!} \right] \]

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

\[ = \frac{638}{2^{10}} = 0.6230 \]

That's the way it is, do it!
2. \( X_1 = \text{First throw} = \{1, 2, \ldots, 6\} \)
\[ X_2 = \text{Second throw} \ldots = \{1, 2, \ldots, 6\} \]
\( X_i \neq X_j \) are ind.
\( \Rightarrow A \neq B \) are ind.

Thus (by inspection)
\[ \Rightarrow P(A) = \frac{1}{2} \]
\[ \Rightarrow P(B) = \frac{1}{2} \]
\[ \Rightarrow P(A/B) = P(C) = \frac{1}{2} \]
\[ \Rightarrow P(B/A) = P(B) = \frac{1}{2} \]

Now, consider \( C \):
\[ P[C] = P[1, 2] + P[1, 4] + P[1, 6] \]
\[ + P[6, 1] + P[6, 3] + P[6, 5] \]

since \( P[X_{i,j}] = P[X_{j,i}] = \frac{1}{6} \) \( \forall \ i \neq j \in \{1, \ldots, 6\} \)
\[ \Rightarrow P[X_{i,j}] P[X_{j,i}] = P[X_{i,j}, X_{j,i}] = \frac{1}{36} \]

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

(cont ->)
(2 cont)

Now

\[ P[C_9 A] = P[1, 2] + P(1, 4) + P[1, 6] \]
\[ + P[3, 2] + P(3, 4) + P(3, 6) \]
\[ + P[5, 2] + P(5, 4) + P(5, 6) \]
\[ P(C_9 A) = \frac{3 \times 3 \times 3}{4} = \frac{9}{4} = \frac{1}{4} \]

Now

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

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

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

Thus

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

and

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

(cont -- >)
(2 cont.)

Two events, \(E \neq C\), are statistically independent if

\[ P[E, C] = P[E]P[C] \]

Here, we associate the event

\[ E = A \text{ and } B = A, B \]

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

We have established that \(A \neq B\)

\[ \Rightarrow P(A)P(B) = P(A, B) = \left( \frac{1}{2}, \frac{1}{2} \right) = \frac{1}{4} \]

Now, obviously, if both \(A\) and \(B\) occur (i.e., both \(X_1\) and \(X_2\) are odd), then \(C\) may never be true. That is, the sum of 2 odd #'s is even.

Thus

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

Thus, the events \((AB)\) and \(C\)

are not statistically independent.

For \(\ge 2\) events, stat. indep. follows if all \(n\) events are pairwise, tri-wise, ... and \(n\)-wise independent.
\[ R \]

\[
\begin{pmatrix}
  y_1 & y_2 & y_3 & y_4 \\
  x_1 & y_1^{1/4} & 0 & 0 & 0 \\
  x_2 & y_2^{1/4} & 3/10 & 0 & 0 \\
  x_3 & 0 & 57/100 & 1/10 & 0 \\
  x_4 & 0 & 0 & 51/100 & 1/10 \\
  x_5 & 0 & 0 & 5/100 & 5/100 \\
  n & 1/10 & 35/100 & 20/100 & 1/10 \\
\end{pmatrix}
\]

\[
p(y_j) = \sum_{i} p(x_i, y_j) \\
p(y_1) = \frac{1/4}{4/10} = \frac{7}{20} \\
p(y_2) = \frac{35/100}{5/100} = \frac{7}{20} \\
p(y_3) = \frac{57/100}{51/100} = \frac{1}{5} \\
p(y_4) = \frac{1}{10} \\
p(y_5) = \frac{15/100}{5/100} = \frac{3}{25} \\
p(x_i) = \sum_{j} p(x_i, y_j) \\
p(x_1) = \frac{1/4}{4/10} = \frac{7}{20} \\
p(x_2) = \frac{1/10}{35/100} = \frac{3}{125} \\
p(x_3) = \frac{35/100}{5/100} = \frac{7}{20} \\
p(x_4) = \frac{51/100}{5/100} = \frac{3}{25} \\
p(x_5) = \frac{15/100}{5/100} = \frac{3}{25} \\
\]

**CONDITIONAL MATRIX**

\[
p(s/R) \Rightarrow p(x_i/y_j) = p(x_i, y_j) / p(y_j) \\
x_1 \quad \frac{1/4 \cdot 20/100}{7/20} = \frac{5}{7} \\
x_2 \quad \frac{1/10 \cdot 20/100}{7/20} = \frac{2}{7} \\
x_3 \quad 0 \quad \frac{5/100 \cdot 20/100}{7/20} = \frac{3}{7} \\
x_4 \quad 0 \quad \frac{5/100 \cdot 20/100}{7/20} = \frac{1}{10} \\
x_5 \quad 0 \quad \frac{5/100 \cdot 20/100}{7/20} = \frac{1}{10} \\
\]

\[
\sum \rightarrow 1 \quad 1 \quad 1 \quad 1 \quad 1 \\
\]
Conditional Matrix $P(R|S) \Rightarrow p(y_j|x_k)$

$$p(y_j|x_k) = \frac{P(x_k, y_j)}{P(x_k)}$$

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1$</td>
<td>1</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>0.4</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$Y_3$</td>
<td>0</td>
<td>0.6</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>$Y_4$</td>
<td>0</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>$Y_5$</td>
<td>0</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Next, find $H(S)$

$$H(S) = \sum_{i=1}^{5} \frac{1}{p(x_i)} \log_2 \frac{1}{p(x_i)}$$

$$= \frac{1}{4} \log_2 4 + \frac{2}{5} \log_2 5 + \frac{3}{25} \log_2 25 + \frac{1}{25} \log_2 25$$

$$= 1.851 \text{ NATS}$$

$$H(Y) = \sum_{j=1}^{4} \frac{1}{p(y_j)} \log_2 \frac{1}{p(y_j)}$$

$$= -2 \log_2 0.2 + \frac{3}{10} \log_2 5 + \frac{1}{10} \log_2 10$$

$$= 1.287 \text{ NATS}$$

$$H(X_i, Y) = \sum_{j} \frac{1}{p(x_i, y_j)} \log_2 \frac{1}{p(x_i, y_j)}$$

$$= \frac{1}{4} \log_2 4 + \frac{3}{10} \log_2 10 + \frac{3}{20} \log_2 20$$

$$= 1.848 \text{ NATS}$$

CONT
b. Verify \( H(x, y) \leq H(x) + H(y) \)

\[
1.848 \leq 1.287 + 1.287 = 2.634
\]

Continuing:

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

\[
= \frac{1}{4} \ln 1 + \frac{1}{10} \ln 4 + \frac{3}{10} \ln \frac{4}{5} + \frac{5}{100} \ln \frac{3}{2}
\]

\[
+ \frac{1}{10} \ln \frac{3}{2} + \frac{5}{100} \ln \frac{3}{2} + \frac{1}{10} \ln 3^{1/2}
\]

\[
= \frac{1}{10} \ln 4 + \frac{1}{10} \ln 3 + \frac{3}{10} \ln \sqrt{2} + \frac{7}{10} \ln 3^{1/2}
\]

\[
= 0.4159 \text{ NATS}
\]

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

\[
= \frac{1}{4} \ln \frac{7}{5} + \frac{1}{10} \ln \frac{7}{2}
\]

\[
+ \frac{3}{10} \ln \frac{7}{6} + \frac{1}{20} \ln 7 + \frac{1}{10} \ln 2 + \frac{5}{10} \ln 2
\]

\[
+ \frac{3}{10} \ln \frac{7}{6} + \frac{1}{20} \ln 4 + \frac{1}{10} \ln 17
\]

\[
= \frac{1}{4} \ln \frac{7}{5} + \frac{1}{10} \ln \frac{7}{2}
\]

\[
+ \frac{3}{10} \ln \frac{7}{6} + \frac{1}{20} \ln 7 + \frac{1}{10} \ln 2 + \frac{1}{10} \ln 4
\]

\[
= 0.5609 \text{ NATS}
\]

\[
H(S/R) + H(R) = 0.5609 + 1.387 \text{ NATS}
\]

\[
= 1.848 = H(R, S)
\]
4. \( S_1 = \left( s_1, s_2, s_3, \ldots, s_{q_1} \right) \) \( \Rightarrow H(S_1) = H_1 \)

\( S_2 = \left( k_1, k_2, k_3, \ldots, k_{q_2} \right) \) \( \Rightarrow H(S_2) = H_2 \)

\( S_\lambda = \left( s_1, s_2, \ldots, s_{q_1}, k_1, k_2, \ldots, k_{q_2} \right) \)

\( H(S_\lambda) = -\sum_{i=1}^{q_1} \lambda p_i \log \lambda p_i - \sum_{j=1}^{q_2} \lambda q_j \log \lambda q_j \)

\( = -\lambda \sum_i p_i \log \lambda - \lambda \sum_j q_j \log \lambda \\
- \lambda \sum_i p_i \log \lambda - \lambda \sum_j q_j \log q_j \)

\( \sum_i p_i = \sum_j q_j = 1 \)

\( \Rightarrow H(S_\lambda) = -\lambda \log \lambda - \lambda \log \lambda \\
+ \lambda H_1 + \lambda H_2 \)

The entropy, \( H(\lambda) \), associated with

\( S = \left( \lambda, \chi \right) \) is

\( H(\lambda) = -\lambda \log \lambda - \lambda \log \lambda \)

Thus,

\( H(S_\lambda) = \lambda H_1 + \lambda H_2 + H(\lambda) \) \( \Rightarrow \)
(cont.)

to maximize, take

\[ \frac{dH(\xi)}{d\lambda} = 0 \]

For simplicity, use natural logs

\[ \frac{dH(\xi)}{d\lambda} = H_1 - H_2 - \frac{d}{d\lambda} \left[ \lambda \log \lambda + (1 - \lambda) \log (1 - \lambda) \right] \]

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

\[ = H_1 - H_2 - \left[ (\log e \lambda + 1) - \log (1 - \lambda) - 1 \right] \]

Thus,

\[ H_2 - H_1 = \lambda e^{H_2 - H_1} \]

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

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

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

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

\[ \lambda_0 = \frac{e^{H_1 - H_2}}{1 + e^{H_1 - H_2}} = \frac{1}{e^{H_1 - H_2} + 1} \]
(4 cont) \[ \bar{\lambda}_0 = 1 - \frac{1}{e^{H_2-H_1} + 1} = \frac{e^{H_2-H_1} - 1}{e^{H_2-H_1} + 1} = \frac{e^{H_2-H_1}}{1 + e^{H_2-H_1}} \]

Now \[ H(\lambda_0) = -\lambda_0 \ln \lambda_0 - \bar{\lambda}_0 \ln \bar{\lambda}_0 \]
\[ = \frac{1}{e^{H_2-H_1} + 1} \ln \left( e^{H_2-H_1} + 1 \right) \]
\[ + \frac{1}{1 + e^{H_2-H_1} \ln \left( e^{H_2-H_1} + 1 \right)} \]

Thus

\[ H(S_{\lambda_0}) = \lambda_0 H_1 + \bar{\lambda}_0 H_2 + H(\lambda) \]
\[ = \frac{H_1}{e^{H_2-H_1} + 1} + \frac{H_2}{e^{H_2-H_1} + 1} \]
\[ + \frac{1}{e^{H_2-H_1} + 1} \ln \left( e^{H_2-H_1} + 1 \right) \]
\[ + \frac{1}{1 + e^{H_1-H_2} \ln \left( e^{H_1-H_2} + 1 \right)} \]
\[ = \frac{1}{e^{H_2-H_1} + 1} \left[ H_1 + \ln \left( e^{H_2-H_1} + 1 \right) \right] \]
\[ + \frac{1}{e^{H_2-H_1} + 1} \left[ H_2 + \ln \left( e^{H_1-H_2} + 1 \right) \right] \text{ Nats.} \]

where again, we have used base e. for base r
\[ \text{ Nats } \log_2 r = \frac{\log r}{\log 2} \]

\[ H(S_{\lambda_0}) = H(S_{\lambda_0}) \log_2 r; \text{ r-ary} \]

is the maximum value the entropy may achieve for this mixed source.
5. a. \( H(X,Y) = H(X) + H(Y) \) with equality iff \( X \perp Y \) are statistically independent.
   b. \( H(Y/X) \leq H(Y) \) with equality only if \( X \perp Y \) are statistically independent random vectors.
   c. In general, \( \log_b x = \log_a x / \log_a b \).

    # Bits = \( \log_2 \) \( x \).
    # Bits = \( \log_2 x = \frac{\log_{10} x}{\log_{10} 2} \).
    # Hartleys = \( \log_{10} x = \log_2 x \times \log_{10} 2 \).
    \( \# \) Bits = \( 7.2 \times 0.3010 = 2.17 \) Hartleys.

    Similarity, we may show
    \( 1.44 \) Bit = \( 1 \) nat.
    Where nats are from \( \log_e \).

    Thus
    \( 7.2 \) Bits = \( 7.2 \times 1.44 \) Bits = \( \frac{\text{NAT}}{1 \text{NAT}} \times 1.44 \text{ Bits} = 5 \text{ Nats} \).
(5 cont)

d. Additive property on Entropy

\[ S = \sum_{i=1}^{n} S_i \]

Divide event \( S_n \) into \( m \) disjoint events:

\[ S_n = \{ r_1, r_2, \ldots, r_m \} \]

And define

\[ S' = S_1, S_2, \ldots, S_{n-1}, r_1, r_2, \ldots, r_m \]

Then the entropies of \( S' \) are related by

\[ H'(p_1, p_2, \ldots, p_{n-1}, q_1, q_2, \ldots, q_m) \]

\[ = H(p_1, p_2, p_3, \ldots, p_n) \]

\[ + p_n H_{S_n}(q_1, q_2, \ldots, q_m/p_n) \]

Where \( H_{S_n} \) is the entropy associated with the source \( S_n \) in (1).
For the source shown above, we have

\[ S = \left\{ \frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{1}{10}, \frac{2}{10} \right\} \]

where \( P[E_5] = P[F_2] + P[F_1] \)

The corresponding entropy of \( S \) is

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

\[ = 1.505 \text{ Nats} \]

Now consider the augmented source \( S' \),

\[ S' = \left\{ \frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{1}{10}, \frac{2}{10}, \frac{3}{10} \right\} \]

The corresponding entropy of \( S' \) is

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

\[ = 1.666 \text{ Nats} \]

Now \( \frac{q_1}{p_n}, \ldots, \frac{q_m}{p_n} \) are

\[ q_1 = \frac{1}{10}, \quad q_2 = \frac{2}{10}, \quad q_3 = \frac{3}{10}, \quad q_4 = \frac{1}{10}, \quad q_5 = \frac{2}{10}, \quad q_6 = \frac{3}{10} \]

\[ H_{s_n} = \frac{1}{3} \ln 3 + \frac{2}{3} \ln \frac{2}{3} \]

\[ = 0.6365 \]

\[ 1.505 + \frac{3}{10} (0.6365) = 1.67 \text{ bits} \]
6. \( S = \{ x_1, x_2, p_1, p_2 \} \)

\[ S' = \{ x_1', x_2', p_1', p_2' \} \]

\[ p_1' = p_1 - \Delta p \quad \text{and} \quad p_2' = p_2 + \Delta p \]

Assume that \( p_1 - \Delta p \geq p_2 + \Delta p < 1 \)

(which also says \( p_1 \geq p_2 \) for \( p_1 > \Delta p > 0 \))

i.e., the probabilities in \( S' \) are numerically close.

Now:

\[ H(S) = H = -p_1 \log p_1 - p_2 \log p_2 \quad \checkmark \]

\[ H(S') = H' = -p_1' \log p_1' - p_2' \log p_2' \]

\[ = -(p_1 - \Delta p) \log (p_1 - \Delta p) - (p_2 + \Delta p) \log (p_2 + \Delta p) \quad \checkmark \]

Consider the difference:

\[ H' - H = p_1 \log (p_1 - \Delta p) + \Delta p \log (p_1 - \Delta p) \]

\[ + p_2 \log (p_2 + \Delta p) + \Delta p \log (p_2 + \Delta p) \]

\[ - p_1 \log p_1 - p_2 \log p_2 \]

\[ = p_1 \log \frac{p_1 - \Delta p}{p_1} + \Delta p \log \frac{p_2 + \Delta p}{p_1 - \Delta p} \]

\[ + p_2 \log \frac{p_2 + \Delta p}{p_2} \quad \checkmark \]

Use the inequality:

\[ \log x \leq 1 \times \sqrt{x} \]

\( (x-1) \text{ See } p16 \)
It follows that

\[ H - H' \leq p_1 \left[ 1 - \left( 1 - \frac{\Delta p}{p_1} \right) \right] + p_2 \left[ 1 - \left( 1 + \frac{\Delta p}{p_2} \right) \right] + \Delta p \log \frac{p_2 + \Delta p}{p_1 - \Delta p} - \Delta p \log \frac{p_2 + \Delta p}{p_1 - \Delta p} \]

Now, from 1, \( \frac{p_2 + \Delta p}{p_1 - \Delta p} \leq 1 \) implies \[ \log \frac{p_2 + \Delta p}{p_1 - \Delta p} \leq 0 \] Since \( \log(\cdot) \) is a monotonic increasing function.

Thus \[ H - H' \leq \Delta p \log \frac{p_2 + \Delta p}{p_1 - \Delta p} \leq 0 \]

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

\[ \text{QED} \]
7. FROM TEXT: PROB. 2-5 p. 41

a. \[ S = \{ s_1, s_2, \ldots, s_i, \ldots \} \]
\[ P(s_i) = a \alpha^i \]

**WE REQUIRE THAT**
\[ 1 = \sum_{i=1}^{\infty} P(s_i) = \sum_{i=1}^{\infty} a \alpha^i \]
\[ = a \frac{\alpha}{1-\alpha} \]
\[ \Rightarrow a = \frac{1-\alpha}{\alpha} \sqrt{\frac{1}{\alpha}} \]
\[ \therefore P(s_i) = \frac{1-\alpha}{\alpha} \alpha^i \]

b. \[ H(S) = \sum_{i=1}^{\infty} P(s_i) \log \frac{1}{P(s_i)} \]
\[ = -\sum_{i=1}^{\infty} P(s_i) \log \frac{1-\alpha}{\alpha} \alpha^i \]
\[ = -\sum_{i=1}^{\infty} \log \alpha \frac{1-\alpha}{\alpha} \log \alpha^i \]
\[ = -\left[ \sum_{i=1}^{\infty} \frac{1-\alpha}{\alpha} \alpha^i \right] \log \frac{1-\alpha}{\alpha} \log \alpha \sum_{i=1}^{\infty} \alpha^i \]

**BUT** \[ \sum_{i=1}^{\infty} \frac{1-\alpha}{\alpha} \alpha^i = \sum_{i=1}^{\infty} P(s_i) = 1 \]

**AND** \[ \sum_{i=1}^{\infty} \alpha^i = \frac{1}{1-\alpha} \]
\[ \Rightarrow H(S) = -\log \frac{1-\alpha}{\alpha} + \frac{1-\alpha}{\alpha} \frac{1}{1-\alpha} \log \alpha \]
\[ = \log \frac{1}{\alpha} - \frac{1}{1-\alpha} \log \alpha \]
\[ = \log \alpha - \log 1-\alpha - \log \frac{1}{\alpha} \log \alpha \]
\[ = \left[ 1 - \frac{1}{1-\alpha} \right] \log \alpha - \log 1-\alpha \]
\[ = - \frac{\alpha}{1-\alpha} \log \alpha - \log (1-\alpha) \]

**HENCEFOROHT, USE LOG BASE E. ie \[ \log \frac{A}{B} = \log A - \log B \]

(Note that for \( \alpha = \frac{1}{2} \), we get \( H(S) = 2 \) bits)
1. \[ \lim_{\alpha \to 0^+} H(s) = \lim_{\alpha \to 0^+} \frac{-\alpha \ln \alpha}{\ln \frac{1}{\alpha}} = \lim_{\alpha \to 0^+} \frac{-\alpha}{\ln \frac{1}{\alpha}} = \lim_{\alpha \to 0^+} \frac{-\alpha}{\ln \alpha} = 0^+ \]

**Using L'Hopital:**
\[ \lim_{\alpha \to 0^+} H(s) = \lim_{\alpha \to 0^+} \frac{-1}{\frac{1}{\ln \alpha}} = \lim_{\alpha \to 0^+} \frac{-\alpha}{\ln \alpha} = \lim_{\alpha \to 0^+} \frac{-\alpha}{\ln \alpha} = 0^+ \]

2. \[ \lim_{\alpha \to 1^-} H(s) = \lim_{\alpha \to 1^-} \frac{\ln \alpha}{\alpha^2 - 1} = \lim_{\alpha \to 1^-} \frac{\ln (\alpha - 1)}{\ln (1^2 - 1)} = \lim_{\alpha \to 1^-} \frac{\ln (\alpha - 1)}{\ln (1^2 - 1)} = -\infty \]

**Now:**
\[ \lim_{\alpha \to 1^-} \frac{\ln \alpha}{\alpha^2 - 1} = \lim_{\alpha \to 1^-} \frac{\ln ^2 \alpha}{\alpha - 1} = \lim_{\alpha \to 1^-} \frac{\ln ^2 \alpha}{\alpha - 1} = -\infty \]

\[ \therefore \lim_{\alpha \to 1^-} H(s) = \infty - (-\infty) = \infty \]

3. **Checking for Extreme:**
\[ \frac{d}{d\alpha} H(s) = \frac{\alpha}{\alpha - 1} \ln \alpha - \frac{\ln ^2 \alpha}{\alpha - 1} + \frac{\ln \alpha}{\alpha - 1} \]
\[ = \frac{\alpha}{\alpha - 1} \ln \alpha - \frac{\ln ^2 \alpha}{\alpha - 1} + \frac{\ln \alpha}{\alpha - 1} \]
\[ = \frac{\alpha}{\alpha - 1} \ln \alpha + \frac{\ln \alpha}{\alpha - 1} \]

\[ \therefore \text{No extreme for finite } \alpha, \text{ since } H(s) \bigg|_{\alpha = 1} = \infty \]

4. **On } \alpha \]

* In order for all } P(s_i) \text{ in (1) to be positive, we require } \alpha > 0

* In order for the (geometric) series in (2) to converge, } |\alpha| < 1

**Thus:**
\[ 0 < \alpha < 1 \]

**To plot } H(s) \text{ in (5), use HP-25 program:**

\[ \begin{array}{c}
\text{(a)} \\
\text{STO } 0 & - & 1 & \ln \\
\text{STO } 0 & + & \text{RCLO} & \div \\
\text{RCLO } & \text{RCLO} & - & \text{GTO } 00 \\
\text{STO } 1 & \times & \text{RCLO} & \div \\
\end{array} \]
The striking aspect of this problem is the infinite entropy for $\alpha$ near 1. That is, we can "tune" our source to as high an entropy as desired by letting $\alpha$ go correspondingly close to 1.

Intuitively, what is happening is as follows. For $\alpha \approx 0$, the probabilities are roughly:

$$(p_1, p_2, p_3, \ldots, p_k, \ldots) \approx (1, 0, 0, \ldots, 0, \ldots)$$

where we have interpreted $H(\alpha) \big|_{\alpha=0} = 1 \leq 0$.

For $\alpha$ near 1, we essentially have

$$(p_1, p_2, \ldots, p_k, \ldots) \approx (\epsilon, \epsilon, \epsilon, \ldots, \epsilon, \ldots)$$

where $\epsilon \ll 1$. That is, we approach a condition of having an infinite number of infinitesimally equally probable events. This, then, constitutes a corresponding approach to infinite entropy.
1. During class work it was shown that the efficiency of a coding scheme could be enhanced by source extension. The purpose of this exercise is to investigate the effect of no. of symbols 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.

\[ X = \{ x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10} \} \]

\[ P = \{ \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{64} \} \]

Let \( r = 2, 3, 4, 5 \) and 7.

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 posts are given in a certain form, the codes resulting from Shannon-Fano method are compact.
\[ \sqrt{2} \]

A source alphabet \( S = \{s_1, s_2\} \) has prob. distribution

\[ P : \left( \frac{3}{4}, \frac{1}{4} \right) \]

Derive compact codes for \( S \), its 2nd, and 4th extensions.

Calculate the three efficiencies and verify, approximately, the result

\[ \frac{1}{n} \xrightarrow[n \to \infty]{} 1 \] for "large" \( n \)

H(\( S \))

Note: This prob. distribution requires Huffman Coding for compactness.

\[ \sqrt{3} \]

A channel is described by the following source-receiver relationship. The numbers are conditional probabilities.

\[ a_1 \rightarrow \frac{y_3}{2/3} \rightarrow b_1 \]

\[ a_2 \rightarrow \frac{y_3}{1/3} \rightarrow b_2 \]

\[ a_3 \rightarrow \frac{1}{1} \rightarrow b_3 \]

Use Minogor's technique to determine the maximum of the information that can be associated with the arrival of the received message i.e. the channel capacity.
4. Derive the following codes for the ensemble

\[ A = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7\} \]

\[ P = \{\frac{4}{10}, \frac{2}{10}, \frac{12}{100}, \frac{8}{100}, \frac{8}{100}, \frac{5}{100}, \frac{1}{100}\} \]

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

Use \([0,1]\) as the alphabet and verify the ensemble arrived at in class that (c) will yield compact codes. (a) & (b) may not 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 Code
(b) Uniquely Decodable Code
(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:

\[
\text{No. of words of length } l = 0
\]
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) \( \{0, 1, 2\} \)

<table>
<thead>
<tr>
<th>Word Length, ( c_i )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of words of length ( c_i ) in each code</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code A</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Code B</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Code C</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Code D</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

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 Amp 17 or 18.

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

\[ n = \text{EFFICIENCY} = \frac{T}{\text{HR}(s)} \]

(a) \( n = 3 \)  (Huffman)

\[
\begin{array}{cccccccc}
X_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
X_2 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
X_3 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
X_4 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
X_5 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
X_6 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
X_7 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
X_8 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
X_9 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
X_{10} & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
\end{array}
\]

\[
\begin{array}{cccccccc}
\frac{1}{2} & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\frac{1}{8} & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\frac{1}{16} & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\frac{1}{32} & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\frac{1}{64} & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\frac{1}{128} & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\frac{1}{256} & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\end{array}
\]

\[
L = \frac{3}{4} + \frac{3}{8} + \frac{3}{16} + \frac{4}{32} + \frac{4}{64} + \frac{4}{128} + \frac{5}{256} + \frac{6}{512} + \frac{6}{1024}
\]

\[
= 1 + \frac{6}{2} + \frac{12}{16} + \frac{5}{32} + \frac{12}{64}
\]

\[
= \frac{3}{2} \left[ 24 + 24 + 5 + 6 \right]
\]

\[
= \frac{59}{32}
\]
(b) \( r = 3 \Rightarrow q = r + (r-1)\alpha = 3 + 2\alpha \Rightarrow \alpha = 4 \Rightarrow q = 11 \)

\[
\begin{align*}
x_1 & : \frac{1}{4} & 2 & 2 & \frac{1}{4} & 2 & \frac{1}{4} & 2 & \frac{1}{4} & 2 \\
x_2 & : \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\
x_3 & : \frac{1}{8} & 0 & 2 & \frac{1}{8} & 2 & \frac{1}{8} & 2 & \frac{1}{8} & 2 \\
x_4 & : \frac{1}{16} & 10 & 10 & \frac{1}{16} & 10 & \frac{1}{16} & 10 & \frac{1}{16} & 10 \\
x_5 & : \frac{1}{16} & 11 & 11 & \frac{1}{16} & 11 & \frac{1}{16} & 11 & \frac{1}{16} & 11 \\
x_6 & : \frac{1}{16} & 12 & 12 & \frac{1}{16} & 12 & \frac{1}{16} & 12 & \frac{1}{16} & 12 \\
x_7 & : \frac{1}{16} & 0 & 0 & \frac{1}{16} & 0 & \frac{1}{16} & 0 & \frac{1}{16} & 0 \\
x_8 & : \frac{1}{32} & 0 & 11 & \frac{1}{32} & 11 & \frac{1}{32} & 11 & \frac{1}{32} & 11 \\
x_9 & : \frac{1}{4} & 0 & 12 & \frac{1}{4} & 12 & \frac{1}{4} & 12 & \frac{1}{4} & 12 \\
x_{10} & : \frac{1}{4} & 010 & 010 & \frac{1}{4} & 010 & \frac{1}{4} & 010 & \frac{1}{4} & 010 \\
x_{11} & : 0 & 011 & 011 & 0 & 011 & 0 & 011 & 0 \\
\end{align*}
\]

\[
\begin{align*}
L_2 &= \frac{1}{4} + \frac{2}{4} + \frac{2}{8} + \frac{2}{16} + \frac{2}{16} + \frac{3}{32} + \frac{4}{64} + \frac{4}{64} \\
&= \frac{3}{4} + \frac{4}{6} + \frac{7}{16} + \frac{3}{32} + \frac{8}{64} \\
&= \frac{1}{32} \left[ 24 + 16 + 14 + 3 + 4 \right] \\
&= \frac{1}{32}
\end{align*}
\]
(c) $r = 4 \Rightarrow q = r + (r-1)\alpha = 4 + 3\alpha \Rightarrow \alpha = 2 \Rightarrow q = 10$

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$\frac{1}{4}$</th>
<th>1</th>
<th>$\frac{1}{4}$</th>
<th>1</th>
<th>$\frac{3}{4}$</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_2$</td>
<td>$\frac{1}{4}$</td>
<td>2</td>
<td>$\frac{1}{4}$</td>
<td>2</td>
<td>$\frac{1}{4}$</td>
<td>1</td>
</tr>
<tr>
<td>$X_3$</td>
<td>$\frac{1}{8}$</td>
<td>0.1</td>
<td>$\frac{1}{8}$</td>
<td>0.1</td>
<td>$\frac{1}{8}$</td>
<td>0</td>
</tr>
<tr>
<td>$X_4$</td>
<td>$\frac{1}{8}$</td>
<td>0.2</td>
<td>$\frac{1}{8}$</td>
<td>0.2</td>
<td>$\frac{1}{8}$</td>
<td>2</td>
</tr>
<tr>
<td>$X_5$</td>
<td>$\frac{1}{16}$</td>
<td>0.3</td>
<td>$\frac{1}{16}$</td>
<td>0.3</td>
<td>$\frac{1}{16}$</td>
<td>3</td>
</tr>
<tr>
<td>$X_6$</td>
<td>$\frac{1}{16}$</td>
<td>0.4</td>
<td>$\frac{1}{16}$</td>
<td>0.4</td>
<td>$\frac{1}{16}$</td>
<td>4</td>
</tr>
<tr>
<td>$X_7$</td>
<td>$\frac{1}{16}$</td>
<td>0.3</td>
<td>$\frac{1}{16}$</td>
<td>0.3</td>
<td>$\frac{1}{16}$</td>
<td>0.3</td>
</tr>
<tr>
<td>$X_8$</td>
<td>$\frac{1}{32}$</td>
<td>3.1</td>
<td>$\frac{1}{32}$</td>
<td>3.1</td>
<td>$\frac{1}{32}$</td>
<td>3.1</td>
</tr>
<tr>
<td>$X_9$</td>
<td>$\frac{1}{64}$</td>
<td>3.2</td>
<td>$\frac{1}{64}$</td>
<td>3.2</td>
<td>$\frac{1}{64}$</td>
<td>3.2</td>
</tr>
<tr>
<td>$X_{10}$</td>
<td>$\frac{1}{64}$</td>
<td>3.3</td>
<td>$\frac{1}{64}$</td>
<td>3.3</td>
<td>$\frac{1}{64}$</td>
<td>3.3</td>
</tr>
</tbody>
</table>

\[ L = \frac{3}{4} + \frac{4}{8} + \frac{6}{16} + \frac{2}{32} + \frac{4}{64} \]
\[ = \frac{3}{4} + \frac{4}{8} + \frac{6}{16} + \frac{4}{64} \]
\[ = \frac{3}{4} + \frac{4}{8} + \frac{4}{16} + \frac{1}{16} \]
\[ = \frac{3}{4} + \frac{1}{2} + \frac{1}{16} \]
\[ = 1 + \frac{3}{4} \]
\[ = \frac{7}{4} \]
(d) \( r = 5 \Rightarrow q = r + (r-1) = 5 + 4 \Rightarrow q = 2 \frac{1}{4} \Rightarrow q = \frac{9}{4} \)

\[
\begin{align*}
x_1 & \quad \frac{1}{4} \\
x_2 & \quad \frac{1}{4} \\
x_3 & \quad \frac{1}{8} \\
x_4 & \quad \frac{1}{16} \\
x_5 & \quad 00 \\
x_6 & \quad 01 \\
x_7 & \quad 02 \\
x_8 & \quad \frac{1}{32} \\
x_9 & \quad \frac{1}{64} \\
x_{10} & \quad 030 \\
x_{11} & \quad 031 \\
x_{12} & \quad 032 \\
x_{13} & \quad 033 \\
x_{14} & \quad 034 \\
\end{align*}
\]

\[
\sqrt{7} = \frac{3}{4} + \frac{\frac{3}{8}}{16} + \frac{\frac{6}{16}}{16} + \frac{\frac{1}{16}}{64} = \frac{3}{4} + \frac{7}{16} + \frac{8}{32} = \frac{1}{32} [2 \times 4 + 14 + 3] = \sqrt{\frac{2}{1} \frac{14}{3} \frac{1}{41}}
\]
\[ e, r = 7 \Rightarrow q = 7 + 6 \alpha \Rightarrow \alpha = 1 \Rightarrow q = 13 \]

\[
\begin{array}{cccccccccccc}
X_1 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\
X_2 & \frac{1}{2} & 1 & \frac{1}{4} & 1 \\
X_3 & \frac{1}{8} & 3 & \frac{1}{8} & 2 \\
X_4 & \frac{1}{16} & 4 & \frac{1}{16} & 3 \\
X_5 & \frac{1}{16} & 5 & \frac{1}{16} & 4 \\
X_6 & \frac{1}{16} & 6 & \frac{1}{16} & 5 \\
X_7 & \frac{1}{16} & 20 & \frac{1}{16} & 6 \\
X_8 & \frac{1}{32} & 21 & \frac{1}{16} & 7 \\
X_9 & \frac{1}{64} & 22 & \frac{1}{32} & 8 \\
X_{10} & \frac{1}{24} & 23 & \frac{1}{64} & 9 \\
X_{11} & 0 & 24 & \frac{1}{32} & 10 \\
X_{12} & 0 & 25 & \frac{1}{64} & 11 \\
X_{13} & 0 & 26 & \frac{1}{128} & 12 \\
\end{array}
\]

\[
\bar{I} = \frac{3}{4} + \frac{3}{8} + \frac{4}{16} + \frac{3}{32} + \frac{1}{64}
\]

\[
= \frac{3}{4} + \frac{1}{4} + \frac{2}{16}
\]

\[
= \frac{7}{8}
\]
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 elementary probabilities

\[ P(\text{Head}) = P(\text{Tail}) = \frac{1}{2} \]

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 events (A and B) and C. Are they statistically independent? Your answer should include 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

\[ P(S, R): \begin{bmatrix}
\frac{1}{4} & 0 & 0 & 0 \\
\frac{1}{10} & \frac{3}{10} & 0 & 0 \\
0 & \frac{5}{100} & 0.10 & 0 \\
0 & 0 & \frac{5}{100} & 0.10 \\
0 & 0 & 0 & 0 \\
\end{bmatrix} \]

Determine \( P(x_i | y_i) \) and \( P(y_j | x_j) \). Build up the conditional matrices \( P(S | R) \) and \( P(R | S) \) and hence determine \( H(S), H(R), H(R | S), H(S | R) \)

and \( H(R, S) \).

Verify that \( H(R, S) = H(R) + H(S | R) \)
4. Consider 2 zero memory sources

\[ S_1: \quad (s_1, s_2, \ldots \rightarrow S_1) \]
\[ S_2: \quad (k_1, k_2, \ldots \rightarrow S_2) \]

with \( S_1 \) having prob. structure \( P_1, P_2, \ldots \rightarrow P_{q_1} \)
and \( S_2 \) having prob. structure \( Q_1, Q_2, \ldots \rightarrow Q_{q_2} \).

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 prob. \( \lambda P_i \), \( i = 1, 2, \ldots \rightarrow q_1 \), and the last \( q_2 \)
symbols have prob. \( (1-\lambda) Q_i \), \( i = 1, 2, \ldots \rightarrow q_2 \).

Find the value of \( \lambda \) 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). \( H(X, Y) \leq H(X) + \ldots \ldots \) with equality iff \ldots \ldots \ldots

b). \( H(\ldots /\ldots) \leq H(Y) \) \quad with equality iff \ldots \ldots \ldots

(c). \( 7.2 \text{ bits} = \ldots \text{ Hz} \); \( \therefore \) \text{ Hz} = \ldots \text{ bits}
(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 \( P(E_1) = P(E_4) = 0.1 \)

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

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

Verify the additive bound 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 thru b41 due 7-26-76
1. a. \( x_1 = \text{Event on first die } \in (1, 2, 3, 4, 5, 6) \)
\( x_2 = \text{\"Second die\"} \)

We divide our sample according to:

\[ p = p[x_1 + x_2 = 5] \quad \text{and} \quad \overline{p} = 1 - p = p[x_1 + x_2 \neq 5] \]

Clearly, \( x_1 \) and \( x_2 \) are independent.

Now, sum mutually exclusive events:

\[ P = P[x_1 = 2, x_2 = 3] + P[x_1 = 3, x_2 = 2] + P[x_1 = 1, x_2 = 4] + P[x_1 = 4, x_2 = 1] \]

\[ P(x_1, x_2) = P(x_1)P(x_2) \]

and:

\[ P(x_i) = \frac{1}{6} \]

\[ \Rightarrow p = \frac{4}{36} \left( \frac{1}{2}\times\frac{1}{2} \right) \]

\[ = \frac{4}{36} = \frac{1}{9} \]

Associated entropy is:

\[ H(5) = -p \ln p - \overline{p} \ln \overline{p} \]

\[ = \frac{1}{4} \ln 9 + \frac{3}{4} \ln \frac{3}{8} \]

\[ = 0.349 \text{ bits} \]

Self-entropy:

\[ \log_2 9 = 3.17 \text{ bits} \]
\[ S = \{ H, T \} \quad \# H = \text{NUMBER OF HEADS} \]

We have here a binomial distribution

\[
\begin{align*}
p(\#H \leq h) &= \sum_{k=0}^{h} \binom{10}{k} p^k (1-p)^{10-k} \quad ; \ h = 0, 1, 2, \ldots, 10 \\\np(\#H = h) &= \binom{10}{h} 2^{-10} = 2^{-10} \sum_{k=0}^{h} \binom{10}{k}
\end{align*}
\]

Thus

\[
p(\#H \leq 5) = \sum_{k=0}^{5} \binom{10}{k} 2^{-10} = 2^{-10} \left[ \binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{3} + \binom{10}{4} + \binom{10}{5} \right]
\]

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

\[
= 2^{-10} \left[ 1 + 10 + 45 + 120 + 210 + 252 \right] = \frac{638}{2^{10}} = 0.6230
\]

That's The way I do it!
2. \(X_1, \text{ first throw } = \{1, 2, \ldots, 6\}\)
\(X_2, \text{ second throw } = \{1, 2, \ldots, 6\}\)
\(X_1, X_2\) are ind.
\(\Rightarrow A, B\) are ind.

Thus (by inspection)

\[
P(A) = \frac{1}{2} \quad P(B) = \frac{1}{2}
\]

\[
P(A/B) = P(A) = \frac{1}{2}
\]

\[
P(B/A) = P(B) = \frac{1}{2}
\]

Now consider \(C:\)

\[
\]

Since \(P[X_i] = P[X_j] = \frac{1}{6} \quad \forall X_i, X_j \in \{1, \ldots, 6\}\)
\(\frac{1}{6} P[X_i] P[X_j] = P[X_i, X_j] = 1/36\)

we have

\[
P(C) = 3 \times 6 \times \frac{1}{36} = \frac{18}{36} = \frac{1}{2}
\]

(cont. ->)
(2 cont)

Now
\[ P[C, A] = P[1, 2] + P(1, 4) + P[1, 6] \]
\[ + P[3, 2] + P(3, 4) + P(3, 6) \]
\[ + P[5, 2] + P(5, 4) + P(5, 6) \]
\[ P(C|A) = \frac{3 \times 3 \times \frac{1}{6}}{3 \times \frac{1}{6}} = \frac{9}{36} = \frac{1}{4} \]

Now
\[ \Rightarrow P[C|A] = \frac{P[A, C]}{P(A)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2} \]

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

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

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

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

(Cont...)
(2 cont.)

Two events, \( E \not\subset C \), are statistically independent if

\[ P[E, C] = P[E] P[C] \]

Here, we associate the event

\[ E = A \text{ AND } B = A, B \]

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

We have established that \( A \not\subset B \)

\[ \Rightarrow P(A) P(B) = P(A, B) = \left( \frac{1}{2} \times \frac{1}{2} \right) = \frac{1}{4} \]

Now, obviously, if both \( A \) and \( B \) occur (i.e., both \( X_1 \) and \( X_2 \) are odd), then \( C \) may never be true. That is, the sum of 2 odd \#s is even.

Thus,

\[ P[E, C] = P[A, B, C] = 0 \not= \frac{1}{4} \]

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

For 72 events, stat. indep. follows if all \( n \) events are pair-wise, tri-wise, ... and \( n \)-wise independent.
\[ \begin{align*}
R &= \begin{bmatrix}
Y_1 & Y_2 & Y_3 & Y_4 \\
x_1 & \frac{9}{40} & 0 & 0 & 0 \\
x_2 & \frac{1}{10} & \frac{3}{10} & 0 & 0 \\
x_3 & 0 & \frac{5}{100} & \frac{1}{10} & 0 \\
x_4 & 0 & 0 & \frac{5}{100} & \frac{1}{10} \\
x_5 & \frac{1}{40} & \frac{35}{100} & \frac{20}{100} & \frac{1}{10}
\end{bmatrix}
\end{align*} \]

\[ P(Y_i) = \sum_j P(x_j, y_i) \]

\[ P(Y_1) = \frac{9}{40} = \frac{9}{20} \]
\[ P(Y_2) = \frac{35}{100} = \frac{7}{20} \]
\[ P(Y_3) = \frac{20}{100} = \frac{1}{5} \]
\[ P(Y_4) = \frac{1}{10} \]

\[ \text{CONDITIONAL MATRIX} \]

\[ P(s|R) = P(x_s|y_i) = \frac{P(x_s, y_i)}{P(y_i)} \]

\[ \begin{bmatrix}
Y_1 & Y_2 & Y_3 & Y_4 \\
x_1 & \frac{\frac{9}{40}}{\frac{9}{20}} & 0 & 0 & 0 \\
x_2 & \frac{1}{10} & \frac{\frac{35}{100}}{\frac{7}{20}} & 0 & 0 \\
x_3 & 0 & \frac{\frac{20}{100}}{\frac{1}{5}} & \frac{1}{2} & 0 \\
x_4 & 0 & 0 & \frac{\frac{5}{100}}{\frac{1}{4}} & 1 \\
x_5 & 0 & 0 & \frac{1}{4} & 1 \\
\end{bmatrix} \]
**Conditional Matrix** \( P(R|S) \Rightarrow p(y_j|x_i) \)

\[ p(y_j|x_i) = \frac{P(x_i, y_j)}{P(x_i)} \]

\[
\begin{array}{cccccc}
X_1 & Y_1 & Y_2 & Y_3 & Y_4 & \text{Total} \\
1 & 1 & 0 & 0 & 0 & 1 \\
2 & \frac{3}{10} & \frac{1}{4} & \frac{3}{10} & \frac{1}{4} & \frac{1}{4} \\
3 & 0 & \frac{5}{100} & 0 & \frac{10}{100} & \frac{1}{4} \\
4 & 0 & 0 & \frac{5}{100} & \frac{10}{100} & \frac{1}{4} \\
5 & 0 & 0 & 0 & 1 & \frac{1}{4} \\
\end{array}
\]

Next, find \( H(R) \)

\[
H(R) = \sum_{i=1}^{5} p(x_i) \log_2 \left( \frac{1}{p(x_i)} \right)
\]

\[
= \frac{1}{4} \log_2 4 + \frac{1}{5} \log_2 2 + \frac{3}{25} \log_2 25 + \frac{1}{25} \log_2 25
\]

\[
= 1.35 \text{ Nats}
\]

\[
H(RS) = \sum_{i=1}^{4} \sum_{j=1}^{5} p(x_i, y_j) \log_2 \left( \frac{1}{p(x_i, y_j)} \right)
\]

\[
= \frac{1}{4} \log_2 4 + 3 \times \frac{3}{10} \log_2 10 + 3 \times \frac{1}{20} \log_2 20
\]

\[
= 1.848 \text{ Nats}
\]
b. verify,  \( H(x, y) \leq H(x) + H(y) \)
\[ 1.848 \leq 1.287 + 1.35 = 2.638 \]

cantaining:

\[ H(R/S) = - \sum_i \sum_j p(x_i, y_j) \ln p(y_j/x_i) \]
\[ = \frac{1}{4} \ln 1 + \frac{1}{10} \ln 4 + \frac{3}{10} \ln \frac{4}{3} + \frac{5}{100} \ln 3 \]
\[ + \frac{1}{10} \ln \frac{3}{2} + \frac{5}{100} \ln \frac{2}{3} + \frac{1}{10} \ln \frac{3}{2} \]
\[ = \frac{1}{10} \ln 4 + \frac{1}{10} \ln 3 + \frac{3}{10} \ln \frac{4}{3} + \frac{2}{10} \ln \frac{3}{2} \]
\[ = 0.4159 \text{ NATS} \]

\[ H(S/R) = - \sum_i \sum_j p(x_i, y_j) \ln p(x_i/y_j) \]
\[ = \frac{1}{4} \ln \frac{3}{2} + \frac{10}{10} \ln \frac{3}{2} \]
\[ + \frac{3}{10} \ln \frac{3}{2} + \frac{1}{20} \ln 7 + \frac{1}{10} \ln 2 + \frac{5}{100} \ln 7 \]
\[ + \frac{1}{20} \ln 4 + \frac{1}{20} \ln 4 + \frac{1}{10} \ln 7 \]
\[ = \frac{1}{4} \ln \frac{3}{2} + \frac{3}{10} \ln \frac{3}{2} \]
\[ + \frac{3}{10} \ln \frac{3}{2} + \frac{1}{20} \ln 7 + \frac{1}{10} \ln 2 + \frac{1}{10} \ln 4 \]
\[ = 0.5609 \text{ NATS} \]

\[ H(S/R) + H(R) = 0.5609 + 1.287 = 1.848 \text{ NATS} \]
\[ = 1.848 = H(R, S) \]
4. \( S_1 = (s_1, s_2, s_3, \ldots, s_q) \Rightarrow H(S_1) = H_1 \)
\( S_2 = (k_1, k_2, k_3, \ldots, k_q) \Rightarrow H(S_2) = H_2 \)
\( S_\lambda = (s_1, s_2, \ldots, s_q, k_1, k_2, \ldots, k_q) \)
\( H(S_\lambda) = \frac{-q}{q} \sum_{i=1}^{q} \lambda p_i \log \lambda p_i - \frac{q}{q} \sum_{j=1}^{q} \lambda q_j \log \lambda q_j \)
\[= -\lambda \sum_{i} p_i \log \lambda \lambda p_i - \lambda \sum_{j} q_j \log \lambda \lambda q_j \]
\[= -\sum_{i} \lambda p_i \log \frac{1}{\lambda} p_i - \sum_{j} \lambda q_j \log \frac{1}{\lambda} q_j \]
\[= \sum_{i} \lambda p_i = \sum_{j} \lambda q_j = 1 \]
\[\Rightarrow H(S_\lambda) = -\lambda \log \lambda - \lambda \log \frac{1}{\lambda} + \lambda H_1 + \lambda H_2 \]

The entropy, \( H(\lambda) \), associated with
\( S = (\lambda, \lambda) \) is
\[H(\lambda) = -\lambda \log \lambda - \lambda \log \frac{1}{\lambda} \]
Thus,
\[H(S_\lambda) = \lambda H_1 + \lambda H_2 + H(\lambda) \]
to maximize, take
\[
\frac{dH(s,y)}{d\lambda} = 0
\]
For simplicity, use natural logs
\[
\frac{dH(s,y)}{d\lambda} = H_1 - H_2 - \frac{d}{d\lambda} \left( \lambda \log \lambda + (1 - \lambda) \log (1 - \lambda) \right)
\]
\[
= H_1 - H_2 - \left( (\log \lambda + 1) + (\log (1 - \lambda) - (1 - \lambda)(1 - \frac{1}{\lambda}) \right)
\]
\[
= H_1 - H_2 - \left( \log \lambda + 1 - \log (1 - \lambda) - \lambda \right)
\]
\[
= H_1 - H_2 + \log \frac{1 - \lambda}{\lambda} = 0
\]
Thus
\[
\frac{H_2 - H_1}{\lambda} = e^{H_2 - H_1}
\]
\[
\lambda = \frac{1 - \lambda}{\lambda}
\]
\[
\lambda(1 - \lambda) = \frac{1}{2} e^{H_2 - H_1}
\]
\[
\lambda = (1 - \lambda)e^{H_1 - H_2}
\]
\[
\lambda(1 - \lambda) = e^{H_1 - H_2}
\]
\[
\lambda_0 = \frac{e^{H_1 - H_2}}{1 + e^{H_1 - H_2}} = \frac{1}{e^{H_2 - H_1} + 1}
\]
\( \lambda_0 = 1 - \frac{1}{e^{H_2-H_1} + 1} \)

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

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

\[
H(\lambda_0) = -\lambda_0 \ln \lambda_0 - \lambda_0 \ln \lambda_0
\]

\[
= \frac{1}{e^{H_2-H_1} + 1} \ln (e^{H_2-H_1} + 1)
\]

\[
+ \frac{1}{1 + e^{H_1-H_2}} \ln (e^{H_1-H_2} + 1)
\]

Thus

\[
H(S_{\lambda_0}) = \lambda_0 H_1 + \lambda_0 H_2 + H(\lambda)
\]

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

\[
+ \frac{1}{e^{H_2-H_1} + 1} \ln (e^{H_2-H_1} + 1)
\]

\[
+ \frac{1}{1 + e^{H_1-H_2}} \ln (e^{H_1-H_2} + 1)
\]

\[
= \frac{e^{H_2-H_1} + 1}{e^{H_2-H_1} + 1} \left[ H_1 + \ln \left( \frac{e^{H_2-H_1} + 1}{3} \right) \right]
\]

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

where again, we have used base e.

For base 2

\[
\# \text{nats } \log_e x = \log_2 e \times \log_2 x
\]

\[
H(S_{\lambda_0})_{r} = H(S_{\lambda_0})_{e} \log_e r
\] \( r \) nats.

is the maximum value the entropy may achieve for this mixed source.
5. a. \( H(X, Y) \leq H(X) + H(Y) \) with equality iff \( X \perp Y \) are statistically independent random vectors.

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

c. In general, \( \log_a x = \frac{\log_b x}{\log_b a} \) Bits is \( \log_2 \) \\
\[ \# \text{BITS} = \log_2 x = \frac{\log_{10} x}{\log_{10} 2} \]

\[ \# \text{Hartleys} = \log_{10} x = \log_2 x \log_{10} 2 \]
\[ = \left( \# \text{BITS} \right) (0.3010) \]
\[ = 7.2 \text{ BITS} \times 0.3010 \]
\[ = 2.17 \text{ Hartleys} \]

Similarly, we may show:

1.44 BIT = 1. nats.

Where nats are from \( \log_e \)

Thus

\[ 7.2 \text{ BITS} = 7.2 \text{ BITS} \times 1.44 \text{ BITS} \]
\[ = 5 \text{ NATS} \]
(5 cont.)

d. Additive property on Entropy

\[ S = \sum_{i=1}^{n} \left( p_1, p_2, \ldots, p_{n-1}, p_n \right) \]

Divide event \( S_n \) into \( m \) disjoint events:

\[ S_n = \{ r_1, r_2, \ldots, r_m \} \]

And define

\[ S' = S_1, S_2, \ldots, S_{n-1}, r_1, r_2, \ldots, r_m \]

Then the entropies of \( S \) and \( S' \) are related by

\[ H'(p_1, p_2, \ldots, p_{n-1}, q_1, q_2, \ldots, q_m) \]

\[ = H(p_1, p_2, p_3, \ldots, p_m) \]

\[ + p_n H_s(q_1, q_2, \ldots, q_m) \]

Where \( H_{s_n} \) is the entropy associated with the source \( S_n \) in (i).

\( \Box \)
c. For the source shown,
we have
\[ S = \left\{ E_1, E_2, E_3, E_4, E_5 \right\} \]
\[ \frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{1}{10}, \frac{3}{10} \]
where \( P[E_5] = P[F_2] + P[F_1] \)
The corresponding entropy is
\[ H(S) = 2 \times \frac{1}{10} \ln 10 + 2 \times \frac{3}{10} \ln \frac{10}{3} + \frac{2}{10} \ln \frac{2}{3} \]
\[ = 1.505 \text{ Nats} \]
Now consider the augmented source
\[ S' = \left\{ E_1, E_2, E_3, E_4, F_1, F_2 \right\} \]
where \( P[F_2] = \frac{2}{3} \text{ and } P[F_1] = \frac{1}{3} \)
The corresponding entropy is
\[ H(S') = 3 \times \frac{1}{10} \ln 10 + 2 \times \frac{2}{10} \ln 5 + \frac{3}{10} \ln \frac{10}{3} \]
\[ = 1.663 \text{ Nats} \]
Now \( H\left( \frac{q_1}{p_n}, \ldots, \frac{q_m}{p_n} \right) = \frac{q_1}{p_n} = \frac{\frac{1}{10}}{\frac{3}{110}} = \frac{1}{3} \)
\( \frac{q_2}{p_n} = \frac{\frac{2}{10}}{\frac{3}{110}} = \frac{2}{3} \)
\[ H_{5n} = \frac{1}{3} \ln 3 + \frac{2}{3} \ln \frac{2}{3} \]
\[ = 0.6365 \]
\[ 1.505 \times 2 + \frac{3}{10} \times (0.6365) = 1.67 \text{ Bits} \]
6. \( S = \{ x_1, x_2 \} \)

\[ S' = \{ x_1', x_2' \} \]

\( p_1' = p_1 - \Delta p \)
\( p_2' = p_2 + \Delta p \)

Assume that \( p_1 - \Delta p \geq p_2 + \Delta p \Rightarrow \) (which also says \( p_1 \geq p_2 \) for \( \Delta p > 0 \))

so the probabilities in \( S' \) are numerically closer.

Now:

\[
H(S) = H = -p_1 \log p_1 - p_2 \log p_2 \quad \checkmark
\]

\[
H(S') = H' = -p_1' \log p_1' - p_2' \log p_2' = -(p_1 - \Delta p) \log (p_1 - \Delta p) \quad \checkmark
\]

\[
= (p_2 + \Delta p) \log (p_2 + \Delta p)
\]

Consider the difference

\[
H - H' = p_1 \log (p_1 - \Delta p) + \Delta p \log (p_1 - \Delta p)
\]
\[
+ p_2 \log (p_2 + \Delta p) + \Delta p \log (p_2 + \Delta p)
\]
\[
- p_1 \log p_1 - p_2 \log p_2
\]

\[
= p_1 \log \frac{p_1 - \Delta p}{p_1} + \Delta p \log \frac{p_2 + \Delta p}{p_1 - \Delta p} \quad \checkmark
\]

\[
+ p_2 \log \frac{p_2 + \Delta p}{p_2} \quad \checkmark
\]

Use the inequality

\[
\lg x \leq \frac{1 - x}{x - 1}
\]

\[
(x-1)
\]
on the contrary of the usual results. 2. Finally, more specifically, we have:

\[ H - H' \geq \Delta P \frac{p_3 - \Delta P}{p_3 + \Delta P} \leq 0 \]

since \( H \geq H' \).

Furthermore, \( H \in \mathbb{R}, H' \in \mathbb{R}, \Delta P \in \mathbb{R} \). Chromatic functions of bars \( \gamma \).

Consider \( f \) a monotonic function. \( \frac{p_3 - \Delta P}{p_3 + \Delta P} \) is positive. \( \Delta P \) is a positive real number:

\[ \int_{-\Delta P}^{\Delta P} \frac{p_3 - \Delta P}{p_3 + \Delta P} \, dp \leq 1 \]

or equivalently:

\[ \int_{-\Delta P}^{\Delta P} \frac{p_3 - \Delta P}{p_3 + \Delta P} \, dp = 0 \]

Therefore, from \( \gamma \), \( \int_{-\Delta P}^{\Delta P} \frac{p_3 - \Delta P}{p_3 + \Delta P} \, dp = 0 \).
7. FROM TEXT: PROB. 2-5. P. 41

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

\[ P(s_i) = a \alpha^i \]

**WE REQUIRE THAT**

\[ 1 = \sum_{i=1}^{\infty} P(s_i) = \sum_{i=1}^{\infty} a \alpha^i \]

\[ = a \sum_{i=1}^{\infty} \alpha^i \]

\[ = a \frac{1}{1-\alpha} \]

\[ \Rightarrow a = \frac{1}{\frac{1}{1-\alpha}} \]

\[ P(s_i) = \frac{1}{\frac{1}{1-\alpha}} \alpha^i \]

b. \( H(S) \equiv \sum_{i} P(s_i) \log \frac{1}{P(s_i)} \)

\[ = -\sum_{i} P(s_i) \log \frac{1}{\frac{1}{\alpha^i}} \]

\[ = -\sum_{i} \frac{1}{\alpha} \alpha^i \left[ \log \frac{1}{\alpha} + \log \alpha^i \right] \]

\[ = -\sum_{i} \frac{1}{\alpha} \alpha^i \left[ \log \frac{1}{\alpha} + i \log \alpha \right] \]

\[ = -\left[ \sum_{i} \frac{1}{\alpha} \alpha^i \right] \log \frac{1}{\alpha} + \frac{\alpha-1}{\alpha} \log \alpha \sum_{i} i \alpha^i \]

**BUT**

\[ \sum_{i} \frac{1}{\alpha} \alpha^i = \sum_{i} P(s_i) = 1 \]

**AND**

\[ \sum_{i} i \alpha^i = \frac{\alpha^2}{(1-\alpha)^2} \]

\[ \Rightarrow H(S) = -\frac{1}{\alpha} \log \frac{1}{\alpha} + \frac{1}{\alpha} \frac{\alpha^2}{(1-\alpha)^2} \log \alpha \]

\[ = \log \frac{1}{\alpha} - \frac{1}{\alpha} \log \alpha \]

\[ = \log \alpha - \log \frac{1}{\alpha} - \frac{\alpha^2}{(1-\alpha)^2} \log \alpha \]

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

\[ = \frac{\alpha}{1-\alpha} \log \alpha - \log \frac{1}{\alpha} \]

**HENCEFORTH, USE LOG BASE E. i.e. \( \log e \)**

**NOTE THAT FOR \( \alpha = \frac{1}{2} \), WE GET \( H(S) = 2 \text{ BITS} \)**
1. \( \lim_{\alpha \to 0^+} H(s) = \lim_{\alpha \to 0^+} \frac{-\alpha \lg \alpha}{\sqrt{\alpha}} \)

   USING L'HOSPITAL:

   \( \lim_{\alpha \to 0^+} H(s) = \lim_{\alpha \to 0^+} \frac{\sqrt{\alpha}}{\alpha^{-1/2}} \)

   \( = \lim_{\alpha \to 0^+} \alpha = 0^+ \)

2. \( \lim_{\alpha \to 1^-} H(s) = \lim_{\alpha \to 1^-} \frac{\ln \alpha}{\alpha^2 - 1} - \ln (1-\alpha) \)

   NOW \( \lim_{\alpha \to 1^-} \frac{\ln \alpha}{\alpha^2 - 1} = \lim_{\alpha \to 1^-} \frac{1}{1-\alpha^2} = \lim_{\alpha \to 1^-} \frac{1}{2(1-\alpha)} = \infty \)

   \( \therefore \lim_{\alpha \to 1^-} H(s) = \infty - (-\infty) = \infty \)

3. CHECKING FOR EXTREMA:

   \( \frac{d}{d\alpha} H(s) = \frac{d}{d\alpha} \frac{\alpha}{\alpha^2 - 1} \ln \alpha + \frac{d}{d\alpha} \frac{1}{\alpha} \ln (1-\alpha) \)

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

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

   \( \therefore \) NO EXTREMA FOR FINITE \( \alpha \), SINCE \( H(s) \bigg|_{\alpha=1} = \infty \)

4. ON \( \alpha \)

   • IN ORDER FOR ALL \( P(s_2) \) IN (i) TO BE POSITIVE, WE REQUIRE \( \alpha > 0 \)

   • IN ORDER FOR THE (GEOMETRIC) SERIES IN (ii) TO CONVERGE, \( |\alpha| < 1 \)

   \( \therefore 0 < \alpha < 1 \)

   (6)

5. TO PLOT \( H(s) \) IN (5), USE HP-25 PROGRAM:

   (\( \alpha \))

   STO 0 - 1 IN

   IN \( \div \) RCL 0 +

   RCL 0 RCL 0 - gto 00

   \( 1 \ \div \ x \ \times \ GTO \ 00 \)
THE STRIKING ASPECT OF THIS PROBLEM IS THE INFINITE ENTROPY FOR $\alpha$ NEAR 1.
THAT IS, WE CAN "TUNE" OUR SOURCE TO AS HIGH AN ENTROPY AS DESIRED BY LETTING $\alpha$ GO CORRESPONDingly
CLOSE TO 1.

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

$$(p_1, p_2, p_3, \ldots, p_\infty) \approx (1, 0, 0, \ldots, 0, \ldots)$$
WHERE WE HAVE INTERPRETED $H(0)|_{\delta=0} = 1 \leq 0$.
FOR $\alpha$ NEAR 1, WE ESSENTIALLY HAVE

$$(p_1, p_2, \ldots, p_\infty) \approx (\epsilon, \epsilon, \epsilon, \ldots, \epsilon, \ldots)$$
WHERE $\epsilon \ll 1$. THAT IS, WE APPROACH A CONDITION OF HAVING AN INFINITE NUMBER OF INFINTESIMALY EQUALLY PROBABLE EVENTS. THIS, THEN, CONSTITUTES A CORRESPONDING APPROACH TO INFINITE ENTROPY.
1. During class work it was shown that the efficiency of a coding scheme could be enhanced by source extension. The purpose of this exercise is to investigate the effect of no. of symbols in the encoding alphabet on the efficiency of compact codes.

Consider an ensemble of 10 messages arranged, for your convenience, in the 5 increasingly ordered of probabilities:

\[ X = \{ x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10} \} \]

\[ P = \left\{ \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{64} \right\} \]

Let \( r = 2, 3, 4, 5 \) and 7.

These points on the horizontal axis should enable you to figure out the dependency 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
   $P = \left(\frac{3}{4}, \frac{1}{4}\right)$

   Derive compact code for $S$, its 2nd, and 4th extension.

   Calculate the three efficiencies and verify, approximately,
   
   
   
   
   
   
   

   Note: This prob. distribution requires Huffman Coding for compactness.

   A channel is described by the following source-receiver relationship. The numbers are conditional probabilities.

   Use Minoga's technique to determine the maximum information that can be associated with the arrival of the received message i.e. the channel capacity.

\[
\begin{align*}
\frac{1}{9} - \frac{4}{9} = \frac{3}{9} \\
\frac{2}{9} = \frac{1}{3}
\end{align*}
\]
4. Derive the following codes for the ensemble

\[ A = \{ a_1, a_2, a_3, a_4, a_5, a_6, a_7 \} \]

\[ P = \left\{ \frac{4}{10}, \frac{2}{10}, \frac{12}{100}, \frac{8}{100}, \frac{8}{100}, \frac{8}{100}, \frac{11}{100} \right\} \]

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

Use \( \{0,1\} \) as the alphabet and verify the conclusion:

Use (c,1) as the alphabet and verify the conclusion that (c) will yield compact

armored at in class that (c) will yield compact

Codes. (a) or (b) may not 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 Code,

(b) Uniquely Decodable Code,

(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:

<table>
<thead>
<tr>
<th>No. of Words</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
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 where

(a) The alphabet is (0, 1)
(b) The alphabet is (0, 1, 2)

<table>
<thead>
<tr>
<th>No. of words of length li in each code</th>
<th>Code A</th>
<th>Code B</th>
<th>Code C</th>
<th>Code D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 1 2 4 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2 2 3 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 4 6 0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2 2 2 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 Amp 17 or 18.
3. Reading of Shannon's paper is mandatory and will count as a HW* problem.
In general
\[ \eta = \text{EFFICIENCY} = \frac{\bar{I}}{Hr(s)} \]

(a) \( r = 2^{-r} \) (Halfsmm)

\[
\begin{align*}
X_1 & = \frac{1}{4} & 00 & \frac{1}{4} & 00 & \frac{1}{4} & 00 & \frac{1}{4} & 00 \\
X_2 & = \frac{1}{4} & 01 & \frac{1}{4} & 01 & \frac{1}{4} & 01 & \frac{1}{4} & 01 \\
X_3 & = \frac{1}{8} & 101 & \frac{1}{8} & 101 & \frac{1}{8} & 101 & \frac{1}{8} & 101 \\
X_4 & = \frac{1}{8} & 110 & \frac{1}{8} & 110 & \frac{1}{8} & 110 & \frac{1}{8} & 110 \\
X_5 & = \frac{1}{16} & 1001 & \frac{1}{16} & 1001 & \frac{1}{16} & 1001 & \frac{1}{16} & 1001 \\
X_6 & = \frac{1}{16} & 1110 & \frac{1}{16} & 1110 & \frac{1}{16} & 1110 & \frac{1}{16} & 1110 \\
X_7 & = \frac{1}{16} & 1111 & \frac{1}{16} & 1111 & \frac{1}{16} & 1111 & \frac{1}{16} & 1111 \\
X_8 & = \frac{1}{32} & 10001 & \frac{1}{32} & 10001 & \frac{1}{32} & 10001 & \frac{1}{32} & 10001 \\
X_9 & = \frac{1}{64} & 100000 & \frac{1}{64} & 100000 & \frac{1}{64} & 100000 & \frac{1}{64} & 100000 \\
X_{10} & = \frac{1}{64} & 100001 & \frac{1}{64} & 100001 & \frac{1}{64} & 100001 & \frac{1}{64} & 100001 \\
\end{align*}
\]

\[
\Gamma = \frac{\frac{3}{4} + \frac{3}{4} + \frac{3}{8} + \frac{3}{8} + \frac{3}{16} + \frac{3}{16} + \frac{3}{32} + \frac{6}{32} + \frac{6}{64} + \frac{5}{64}}{\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}}
\]

\[
= \frac{1 + \frac{6}{8} + \frac{12}{16} + \frac{5}{32} + \frac{2}{4}}{\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}}
\]

\[
= \frac{1 + \frac{6}{32} + \frac{12}{32} + \frac{5}{32} + \frac{2}{32}}{\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}}
\]

\[
= \frac{1}{5} \times \left( \frac{24 + 24 + 5 + 6}{32} \right)
\]

\[
= \frac{59}{32}
\]
(b) \( r = 3 \Rightarrow q = r + (r-1) \alpha = 3 + 2\alpha \Rightarrow \alpha = 4 \Rightarrow q = 11 \)

\[
\begin{array}{ccccccc}
X_1 & \frac{1}{4} & 2 & \frac{1}{4} & 2 & \frac{1}{4} & 2 \\
X_2 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\
X_3 & \frac{1}{8} & 02 & \frac{1}{8} & 02 & \frac{1}{8} & 02 \\
X_4 & \frac{1}{8} & 10 & \frac{1}{8} & 10 & \frac{1}{8} & 10 \\
X_5 & \frac{1}{16} & 11 & \frac{1}{16} & 11 & \frac{1}{16} & 11 \\
X_6 & \frac{1}{16} & 12 & \frac{1}{16} & 12 & \frac{1}{16} & 12 \\
X_7 & \frac{1}{16} & 010 & \frac{1}{16} & 010 & \frac{1}{16} & 010 \\
X_8 & \frac{3}{32} & 011 & \frac{3}{32} & 011 & \frac{3}{32} & 011 \\
X_9 & \frac{1}{4} & 0120 & \frac{1}{4} & 0120 & \frac{1}{4} & 0120 \\
X_{10} & \frac{1}{4} & 0121 & \frac{1}{4} & 0121 & \frac{1}{4} & 0121 \\
X_{11} & 0 & 0122 & 0 & 0122 & 0 & 0122 \\
\end{array}
\]

\[
\begin{array}{ccccccc}
& \frac{1}{4} & 1 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\
& \frac{1}{4} & 2 & \frac{1}{4} & 1 & \frac{1}{4} & 1 \\
& \frac{1}{4} & 00 & \frac{1}{4} & 01 & \frac{1}{4} & 02 \\
& \frac{1}{8} & 01 & \frac{1}{8} & 02 & \frac{1}{8} & 02 \\
& \frac{1}{8} & 02 & \frac{1}{8} & 02 & \frac{1}{8} & 02 \\
\end{array}
\]

\[
\bar{L}_3 = \frac{1}{4} + \frac{2}{4} + \frac{2}{8} + \frac{2}{8} + \frac{2}{16} + \frac{2}{16} + \frac{2}{16} + \frac{2}{32} + \frac{4}{64} + \frac{4}{64} = \frac{3}{4} + \frac{4}{8} + \frac{2}{16} + \frac{3}{32} + \frac{3}{64} = \frac{1}{32} [24 + 16 + 14 + 3 + 4] = \frac{61}{32}
\]
(c) \( r = 4 \Rightarrow q = r + (r - 1) \alpha = 4 + 3 \alpha \Rightarrow \alpha = 2 \Rightarrow q = 10 \)

<table>
<thead>
<tr>
<th>( x_i )</th>
<th>( \frac{1}{4} )</th>
<th>( \frac{1}{4} )</th>
<th>( \frac{1}{4} )</th>
<th>( \frac{1}{4} )</th>
<th>( \frac{1}{4} )</th>
<th>( \frac{1}{4} )</th>
<th>( \frac{1}{4} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>( 1 )</td>
<td>( 2 )</td>
<td>( 3 )</td>
<td>( 0 )</td>
<td>( 2 )</td>
<td>( 3 )</td>
<td>( 1 )</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
</tr>
<tr>
<td>( x_3 )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
</tr>
<tr>
<td>( x_4 )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
</tr>
<tr>
<td>( x_5 )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
</tr>
<tr>
<td>( x_6 )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
</tr>
<tr>
<td>( x_7 )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
</tr>
<tr>
<td>( x_8 )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
</tr>
<tr>
<td>( x_9 )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
</tr>
<tr>
<td>( x_{10} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{1}{8} )</td>
</tr>
</tbody>
</table>

\[
\bar{z} = \frac{3}{4} + \frac{4}{8} + \frac{6}{16} + \frac{3}{32} + \frac{4}{64}
\]

\[
= \frac{3}{4} + \frac{1}{2} + \frac{1}{2}
\]

\[
= \frac{7}{4}
\]
\( d \) \( r=5 \Rightarrow q = r + (r-1) \alpha = 5 + 4 \alpha \Rightarrow \alpha = 2 \frac{1}{9} q = 12 \\

\[
\begin{array}{cccccccc}
  x_1 & \frac{4}{4} & 1 & 1 & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\
  x_2 & \frac{4}{4} & 2 & 2 & \frac{1}{4} & \frac{1}{4} & 1 & 1 \\
  x_3 & \frac{4}{16} & 3 & 3 & \frac{1}{8} & \frac{1}{8} & 2 & 2 \\
  x_4 & \frac{4}{16} & 4 & 4 & \frac{1}{8} & \frac{1}{8} & 3 & 3 \\
  x_5 & \frac{4}{16} & 00 & 00 & \frac{1}{8} & \frac{1}{8} & 4 & 4 \\
  x_6 & \frac{4}{16} & 01 & 01 & \frac{1}{8} & \frac{1}{8} & 0 & 0 \\
  x_7 & \frac{4}{16} & 02 & 02 & \frac{1}{8} & \frac{1}{8} & 0 & 0 \\
  x_8 & \frac{4}{32} & 030 & 030 & \frac{1}{16} & \frac{1}{16} & 0 & 0 \\
  x_9 & \frac{4}{32} & 04 & 04 & \frac{1}{16} & \frac{1}{16} & 0 & 0 \\
  x_{10} & \frac{4}{4} & 031 & 031 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
  x_{11} & 0 & 032 & 032 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
  x_{12} & 0 & 033 & 033 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
  x_{13} & 0 & 034 & 034 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
\end{array}
\]

\[ \sqrt{\bar{E}} = \frac{2}{4} + \frac{2}{8} + \frac{6}{16} + \frac{1}{16} + \frac{6}{64} \]
\[ = \frac{3}{4} + \frac{7}{16} + \frac{3}{32} \]
\[ = \frac{1}{32} \left[ 24 + 14 + 3 \right] \]
\[ = \frac{21}{32} \]

\[ = \frac{1}{41} \]

\[ \frac{1}{41} \]
\( e, r = 7 \Rightarrow q = 7 + 6a \Rightarrow x = 1 \Rightarrow q = 13 \)

| \( x_1 \) | \( \frac{1}{4} \) | 0 | \( \frac{1}{4} \) | 0 |
| \( x_2 \) | \( \frac{1}{4} \) | 1 | \( \frac{1}{8} \) | 2 |
| \( x_3 \) | \( \frac{1}{8} \) | 3 | \( \frac{1}{8} \) | 3 |
| \( x_4 \) | \( \frac{1}{16} \) | 4 | \( \frac{1}{8} \) | 4 |
| \( x_5 \) | \( \frac{1}{16} \) | 5 | \( \frac{1}{16} \) | 5 |
| \( x_6 \) | \( \frac{1}{16} \) | 6 | \( \frac{1}{16} \) | 6 |

\( \overline{L} = \frac{3}{4} + \frac{3}{8} + \frac{4}{16} + \frac{3}{8} + \frac{4}{16} \)
\( = \frac{3}{4} + \frac{1}{4} + \frac{3}{8} \)
\( = \frac{7}{8} \)
(b) Our channel now is
\[ P(y_i/x_i) \Rightarrow \begin{bmatrix} \bar{p} - \epsilon & p - \epsilon & 2\epsilon & 0 \\ p - \epsilon & \bar{p} - \epsilon & 0 & 2\epsilon \end{bmatrix} \]

Again, let
\[ q = P(x_1), \quad \bar{q} = P(x_2) \]

Thus
\[ P(x_i, y_i) = \begin{bmatrix} q(\bar{p} - \epsilon) & q(p - \epsilon) & 2q\epsilon & 0 \\ \bar{q}(p - \epsilon) & \bar{q}(p - \epsilon) & 0 & \bar{q}2\epsilon \end{bmatrix} \tag{15} \]

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):
\[ H(y/x) = -(\bar{p} - \epsilon) \log (\bar{p} - \epsilon) - (p - \epsilon) \log (p - \epsilon) \]
\[ -2\epsilon \log 2\epsilon \tag{16} \]

From (15):
\[ P(y_1) = (q + p - 2p\epsilon) - \epsilon \quad \text{(same as 4)} \]
\[ P(y_2) = 1 - (p + q - 2p\epsilon) - \epsilon \quad \text{(same as 5)} \]
\[ P(y_3) = 2q\epsilon \]
\[ P(y_4) = 2\bar{q}\epsilon = 2\epsilon - 2\epsilon \]

Thus
\[ -H(y) = \begin{bmatrix} (q + p - 2p\epsilon) - \epsilon \\ 1 - (p + q - 2p\epsilon) - \epsilon \\ 2q\epsilon \end{bmatrix} \log \begin{bmatrix} 1 - (p + q - 2p\epsilon) - \epsilon \\ 2q\epsilon \log \frac{2}{2\epsilon - 2q\epsilon} \end{bmatrix} \tag{17} \]
NOW, USING \((16) \div (17)\):
\[
I(x; y) = H(y/x) - H(y)
\]  \(\text{(16)}\)

AND, AS BEFORE
\[
\frac{d}{dq} I(x; y) = -\frac{d}{dq} H(y)
\]

FROM \((17)\):
\[
\frac{d}{dq} I(x; y) = (1-2p) \log \left[ \frac{(p+q-2pq)-e}{1-(p+q-2pq)-e} \right] + \left(1-\frac{2}{p}\right) - (1-2p) \log \left[ 1-(p+q-2pq)-e \right] - \left(1-\frac{2}{p}\right)
\]
\[
+ 2e \log 2q e - 2e - 2e \log (2e-2qe) - 2e
\]
\[
= (1-2p) \log \frac{(p+q-2pq)-e}{1-(p+q-2pq)-e} + 2e \log \frac{2q e}{2e(1-q)}
\]

SETTING THIS TO ZERO YIELDS
\[
(1-2p) \log \frac{(p+q-2pq)-e}{1-(p+q-2pq)-e} = 2e \log \frac{2e(1-q)}{2q e}
\]

BOTH SIDES ARE ZERO FOR \(q = \frac{1}{2}\).
\[
\therefore \, q = \bar{q} = \frac{1}{2}
\]
SUBSTITUTING INTO (12);

\[ C = I(x; y) \bigg|_{q=\frac{1}{2}} \]

\[ = -(\bar{p} - e) \log(\bar{p} - e) - (p - e) \log(p - e) - 2e \log z e \]
\[ + (\frac{1}{2} - e) \log(\frac{1}{2} - e) + (\frac{1}{2} - e) \log(\frac{1}{2} - e) \]
\[ + e \log e + e \log e \]
\[ = -(\bar{p} - e) \log(\bar{p} - e) - (p - e) \log(p - e) \]
\[ + 2(\frac{1}{2} - e) \log(\frac{1}{2} - e) \]
\[ + 2e \log e - 2e \log 2 - 2e \log e \]
\[ = -(\bar{p} - e) \log(\bar{p} - e) - (p - e) \log(p - e) \]
\[ + 2(\frac{1}{2} - e) \log(\frac{1}{2} - e) - 2e \log 2 \]  \[ (19) \]
(c) Denote the capacity in part (a) \((13)\) by \(C_a:\)

\[
\begin{align*}
\frac{C_a}{- (p-e) \log (p-e) - (p-e) \log (p-e)}
+ 2 \left( \frac{1}{2} - e \right) \log \left( \frac{1}{2} - e \right)
\end{align*}
\]

And that in part (b) \((19)\) by \(C_b:\)

\[
\begin{align*}
\frac{C_b}{- (p-e) \log (p-e) - (p-e) \log (p-e)}
+ 2 \left( \frac{1}{2} - e \right) \log \left( \frac{1}{2} - e \right) - 2e \log 2
\end{align*}
\]

It's clear that

\[
C_a - C_b = 2e \log 2 > 0
\]

Thus, channel \(a\) is always better than \(b\). (Even for \(e\) near 0)

(Note that, if \(C_a\) and \(C_b\) were measured in bits, then \(C_a - C_b = 2e\) bits.)
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, p81 of TEXT.

4. For \( m = 5 \), find corresponding \( k \).

   Choose two of the possible 25 numbers, and perform a Hamming code (single error correction). Compute

   \[
   P_f[ \text{error with Hamming code}] \\
   P_f[ \text{'' without Hamming code}]
   \]

   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

$$(1947)_{10} = (1111 0011 0111)_2$$
$$(8)_{10} = (1000)_2$$
$$(45)_{10} = (1111)_2$$

(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

\[
\begin{array}{cccccccccccc}
12 & 12 & 2 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
15 & 14 & 13 & 12 & 11 & 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\
\end{array}
\]

\( p_0 \oplus 0 \oplus 0 \oplus 1 \oplus 0 \oplus 0 \oplus 0 \oplus 0 = 0 \Rightarrow p_0 = 0 \)

\( p_1 \oplus 0 \oplus 0 \oplus 0 \oplus 0 \oplus 0 \oplus 0 = 0 \Rightarrow p_1 = 1 \)

\( p_2 \oplus 0 \oplus 0 \oplus 0 \oplus 0 \oplus 0 \oplus 0 = 0 \Rightarrow p_2 = 1 \)

\( p_3 \oplus 0 \oplus 0 \oplus \ldots \oplus 0 = 0 \Rightarrow p_3 = 0 \)

**Coding (15)\(_2 = (1111)\(_2 \)**

\[
\begin{array}{cccccccccccc}
15 & 14 & 13 & 12 & 11 & 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
\end{array}
\]

\( p_0 \oplus 1 \oplus 1 \oplus 1 \oplus 0 \ldots \oplus 0 = 0 \Rightarrow p_0 = 1 \)

\( p_1 \oplus 1 \oplus 1 \oplus 0 \ldots \oplus 0 = 0 \Rightarrow p_1 = 1 \)

\( p_2 \oplus 1 \oplus 1 \oplus 0 \ldots \oplus 0 = 0 \Rightarrow p_2 = 1 \)

\( p_3 \oplus 0 \ldots \oplus 0 = 0 \Rightarrow p_3 = 0 \)
CODING 1947:

\[
\begin{array}{cccccccc}
15 & 14 & 13 & 12 & 11 & 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
\end{array}
\]

\[p_0 \oplus 1 \oplus 1 \oplus 1 \oplus 0 \oplus 0 \oplus 1 = 0 \Rightarrow p_0 = 0\]
\[p_1 \oplus 1 \oplus 0 \oplus 1 \oplus 0 \oplus 0 \oplus 0 \oplus 1 = 0 \Rightarrow p_1 = 0\]
\[p_2 \oplus 1 \oplus 0 \oplus 1 \oplus 0 \oplus 0 \oplus 1 \oplus 0 \oplus 0 = 0 \Rightarrow p_2 = 0\]
\[p_3 \oplus 1 \oplus 0 \oplus 0 \oplus 0 \oplus 1 \oplus 1 \oplus 0 = 0 \Rightarrow p_3 = 1\]

OUR FINAL CODE FOR 8-15-1947 IS

000000000001000000000000000000000000000000000000000000000000000000000000000

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), LET'S DEAL ONLY WITH THE CODE FOR 1947 \Rightarrow
ERROR TO BE SPOTTED

\[ 0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 1 = 0 \Rightarrow p_0 = 1 \]
\[ 0 \oplus 1 \oplus 0 \oplus 1 \oplus 0 \oplus 1 \oplus 1 = 0 \Rightarrow p_1 = 1 \]
\[ 0 \oplus 1 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 = 0 \Rightarrow p_2 = 0 \]
\[ 1 \oplus 1 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 = 0 \Rightarrow p_3 = 1 \]

ERROR IS \( p_3p_2p_1p_0 = (1011)_2 = 11^{\text{TH}} \text{PLACE} \).

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 (Thm. 3.7.3 p. 82). We here paraphrase it. We use Ash's notation.

We may write:

\[ I(X; Y) = H(X) - H(X|Y) \]  

Where

\[ X = (x_1, ..., x_n) \quad Y = (y_1, ..., y_n) \]  

(i.e., an \( n \)th extension)

If we randomly choose our input signals (equally likely), so that

\[ p(x_i) = \frac{1}{s} \]  

where there are \( s \) input words the \( i \)th of which is \( x_i \).

It follows that

\[ H(X) = \sum_{i=1}^{s} \frac{1}{s} \log s = \log s \]  

Thus, (1) becomes

\[ I(X; Y) = \log s - H(X|Y) \]  

We are given that

\[ I(X; Y) \leq \sum_{i=1}^{n} I(x_i; y_i) \]  

Thus, we may write (5) as

\[ \log s - H(X|Y) \leq \sum_{i=1}^{n} I(x_i; y_i) \]

* Ash's \( I(X; Y) \) is obviously our \( I(X; Y) \).
Now, the channel capacity is
\[ C = \max \ I(x_i; y_j) \]

Thus,
\[ \sum_{i=1}^{n} I(x_i; y_j) \leq \sum_{i=1}^{n} C = nc \]

AND (6) BECOMES
\[ \log s - H(X|I) \leq nc \quad (6) \]

At this point, we utilize Fano's bound* (Theorem 3.7.1 in Ash):
\[ H(X|I) \leq H_E + P(E) \log s - 1 \quad (9) \]

WHERE THE ERROR ENTROPY IS
\[ H_E = -P_E \log P_E - (1-P_E) \log (1-P_E) \quad (10) \]

Note that
\[ \max H_E = \log 2 = 1 \text{ bit} \quad (11) \]

Rearranging (8):
\[ H(X|I) \geq \log s - nc \]

Thus, using (9):
\[ H_E + P(E) \log s - 1 \geq \log s - nc \]

OR, USING (11)
\[ \log 2 + P(E) \log s - 1 \geq \log s - nc \]

OR, since
\[ \log s > \log s - 1 \]

WE HAVE
\[ \log 2 + P(E) \log s \geq \log s - nc \]

*We must use \( P[E] \) (expected value) as opposed to \( P[e] \), since the input coding, \( \frac{1}{n} \) thus the \( P_0[error] \), are random processes.
SOLVING FOR $\log s$:

$$\log s \left[ \frac{1}{P(E)} - 1 \right] \geq \log 2 - nc$$

$$\log s \geq \frac{\log 2 - nc}{P(E) - 1}$$

OR

$$\log s \geq \frac{nc - \log 2}{1 - P(E)}$$

Q.E.D.

(WE HERE ABANDON ASH)
(b) Show if $s \geq 2^{n(c+s)}$, $b > 0$, then

\[ P(E) = 1 - \frac{C + \sqrt{n}}{C + s} \]

AND, AS $n \to \infty$, $P(E) \to 0$

FROM PART (a):

\[ \ln s \leq \frac{nc + \ln 2}{1 - P(E)} \]

SOLVE FOR \( P(E) \)

\[ 1 - P(E) \leq \frac{nc + \ln 2}{\ln s} \]

\[ P(E) \geq 1 - \frac{nc + \ln 2}{\ln s} \]

NOW, IF $s \geq 2^{n(c+s)} = \ln \text{(LOG BASE)}$

\[ \ln s \geq n(c+s) \ln 2 \quad (\ln \text{IS MONOTONIC)} \]

\[ \ln s \leq \frac{1}{n(c+s) \ln 2} \]

\[ \ln s \geq \frac{-1}{n(c+s) \ln 2} \]

AND

\[ P(E) \geq 1 - \frac{nc + \ln 2}{n(c+s) \ln 2} \]

\[ \geq 1 - \frac{C \ln 2 + \sqrt{n}}{C + s} \]

LETS USE $\ln (\cdot) = \log_2 (\cdot)$. Thus

\[ P(E) = 1 - \frac{C + \sqrt{n}}{C + s} \quad \forall n \]

AS $n \to \infty$, $\frac{1}{n} \to 0$, AND

\[ P(E) \geq 1 - \frac{C}{C + s} > 0 \]

THIS FOLLOWS FROM $\delta > 0$.
(c) Solving again for $P(E)$ in part (b):

$$\frac{P(E)}{NC+1} \geq 1 - \frac{\log s}{C+R}$$

where, again, $\log \left( \cdot \right) = \log_2 \left( \cdot \right)$

Consider, then,

$$s = 2^{nR} \Rightarrow \log s = nR$$

Thus,

$$\frac{P(E)}{NC+1} \geq 1 - \frac{nR}{C+R} \geq 1 - \left( \frac{\log (C+1/n)}{C+R} \right)$$

Let $R = C + \Delta \geq \Delta$ is a fixed positive number. Then:

$$\frac{P(E)}{NC+1} \geq 1 - \left( \frac{\log (C+1/n)}{C+\Delta} \right)$$

Clearly, as $n \to \infty$, we have

$$P(E) \geq 1 - \frac{C}{C+\Delta}$$

which, for a fixed $\Delta$, removes the possibility for equality.

Thus:

$$P(E) \geq 1 - \frac{1}{1 + \Delta/C}$$

Clearly, the larger $\Delta$, the worst asymptotically $P(E)$. And, obliquely, $\lim_{n \to \infty} P(E) \neq 0$

Note: On pg. 83, Ash shows that is worse than this, that, in fact, for $R > C$, that $P_n(E) \to 1$,
3. We begin by writing down the conditional matrix \( P(b_j | a_i) \):

\[
\begin{array}{cccc}
    & \vdots & \vdots & \vdots \\
\hline
p(b_j | a_i) & b_j & \leq & \frac{1}{r} \\
\hline
a_i & \vdots & \vdots & \vdots \\
\hline
r & \frac{r}{s} & \ldots & \frac{r}{s} \\
\end{array}
\]

Since the channel is uniform, each of its columns must add to \( \frac{r}{s} \).

Now, we are given that

\[
p(a_i) = \frac{1}{r}
\]

Thus, the equivocation may be written:

\[
H(A/B) = -\sum_{j=1}^{s} \sum_{i=1}^{r} p(a_i, b_j) \cdot \frac{1}{r} \cdot \log p(a_i/b_j) = -\sum_{j=1}^{s} \sum_{i=1}^{r} p(b_j) \cdot p(a_i/b_j) \cdot \frac{1}{r} \cdot \log p(a_i/b_j)
\]

Let's next generalize the joint matrix by multiplying all elements in (1) by (2):

\[
\begin{array}{cccc}
    & \vdots & \vdots & \vdots \\
\hline
p(a_i, b_j) & b_j & \leq & \frac{1}{r} \\
\hline
a_i & \vdots & \vdots & \vdots \\
\hline
\frac{1}{s} & \frac{1}{3} & \ldots & \frac{1}{3} & \frac{1}{11}
\end{array}
\]
Thus
\[ p(b_j) = \frac{1}{3} \]
and (3) becomes
\[ H(A/B) = -\frac{1}{3} \sum_j \sum_i p(a_i/b_j) \log p(a_i/b_j) \]

Now, since all columns have equivalent elements, we may write
\[ H(A/B) = -\frac{1}{3} \left[ \sum_i \leq p(a_i/b_1) \log p(a_i/b_1) \right] \]
\[ = -\sum_i p(a_i/b_1) \log p(a_i/b_1) \]

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
\[ H_e \triangleq -P_e \log P_e + (1-P_e) \log (1-P_e) \]

By the additive rule of entropies, it is obvious that
\[ H_e \leq -\sum_{i=1}^{r} p(a_i/b_i) \log p(a_i/b_i) \]
Thus, substituting into (6), we have our desired bound
\[ H(A/B) \geq H_e \]

Note that this bound is met for the $p(a_i/b_i)$ matrix
\[
\begin{bmatrix}
\frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} & \frac{1}{2}
\end{bmatrix}
\]

Thus, by also utilizing the Fano bound, we can write (for the given uniform channel):
\[ H_e \leq H(A/B) \leq H_e + P_e \log 2 \]
4. For \( m=5 \) (\( \Rightarrow k=4 \))*, we have 32 permissible codes. We'll choose two, and Hamming code them, an easy one to code is 000000.

The code for which is clearly**

\[
\begin{array}{cccccc}
0 & 0 & 0 & 0 & 0 & 0 \\
\uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\
0 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
\]

For our second code, choose \((31)_2 = 11111\):

Coding

\[
\begin{array}{cccccc}
\hline
8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\
\hline
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\hline
P_3 & P_2 & P_1 & P_0
\end{array}
\]

\[
\begin{align*}
P_0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 &= 0 \Rightarrow P_0 = 0 \\
P_1 \oplus 1 \oplus 1 \oplus 1 \oplus 1 &= 0 \Rightarrow P_1 = 1 \\
P_2 \oplus 1 \oplus 1 \oplus 1 &= 0 \Rightarrow P_2 = 1 \\
P_3 & = 0 \Rightarrow P_3 = 1
\end{align*}
\]

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 interpreted 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

$$P_E = \binom{n}{1} \bar{p}^n - \binom{n}{2} \bar{p}^{n-1} + \cdots + \binom{n}{n} \bar{p}^0 = 1 - (1 - \bar{p})^n$$

WHERE $\bar{p}$ IS THE PROB. A 0 IS RECEIVED AS A 1 AND VISA VERSA. FOR $p = 1/100$,

$$P_E = 1 - \left(\frac{99}{100}\right)^n = 4.901 \times 10^{-2} = .004901$$

For the coded scheme

$$P_E^{(c)} = \binom{n}{2} \bar{p}^2 \bar{p}^{n-2} + \binom{n}{3} \bar{p}^3 \bar{p}^{n-3} + \cdots + \binom{n}{n} \bar{p}^n = \bar{p} - \sum_{i=0}^{n-3} \binom{n}{i} \bar{p}^i \bar{p}^{n-1-i} - \binom{n}{0} \bar{p}^0 \bar{p}^{n-1} - \binom{n}{1} \bar{p}^1 \bar{p}^{n-1}$$

$$= (\bar{p} + p)^n - \bar{p}^n - np \bar{p}^{n-1}$$

$$= 1 - (1 - \bar{p})^n - np(1 - \bar{p})^{n-1}$$

FOR $n = 9$, $p = \frac{1}{100}$

$$P_E^{(c)} = 1 - \left(\frac{99}{100}\right)^9 - \frac{9}{100} \left(\frac{99}{100}\right)^8$$

$$= 0.003436 > 0.004901 \text{ (in Manning's better)}$$

**Figure of Merit** = $0.003436 \div 1.4265$
HOMEWORK

PROB: WRITE OUT $P(A_j, B_k, C_l)$ IN TERMS OF
VARIOUS CONDITIONAL & MARGINAL PROBABILITIES
FOR TWO EVENTS:

$P(A, B) = P(A/B) P(B)$

IT FOLLOWS THAT

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

CLEARLY, WE MAY INTERCHANGE $A, B, C$
IN ANY DESIRED FASHION
10.

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, \(P(x_1, x_2) = \frac{\text{total relevant events}}{\text{total # of events}}\).

Defining: \(\binom{n}{m} = \frac{n!}{(n-m)!}\)

Total # relevant events: \(\binom{10}{2} = 45\) = 90

\[
\begin{align*}
\text{Total ways to get } R, R &= \binom{6}{2} = 15 \times 5 = 30 \\
\text{Total ways to get } R, B &= 6 \times 4 = 24 \\
\text{Total ways to get } B, R &= 6 \times 4 = 24 \\
\text{Total ways to get } B, B &= \binom{4}{2} = 12
\end{align*}
\]

\[
\begin{array}{c|c|c}
\hline
& R & B \\
\hline
R & \frac{1}{3} & \frac{1}{15} \\
B & \frac{2}{15} & \frac{2}{15} \\
\hline
\end{array}
\]

\[
P(x_2 = R) = P(x_2 = R, x_1 = B) + P(x_2 = R, x_1 = R)
\]

\[
= \frac{1}{3} + \frac{2}{15} = \frac{7}{15}
\]

\[
P(x_2 = B) = 1 - P(x_2 = R) = \frac{8}{15}
\]

\[
\begin{array}{c|c|c}
\hline
& R & B \\
\hline
R & \frac{5}{15} & \frac{4}{15} \\
B & \frac{4}{15} & \frac{6}{15} \\
\hline
\end{array}
\]

\[
P(x_2 = R) P(x_1 = R) = P(x_2 = R) P(x_1 = B) = P(x_1 = B)
\]

\[
\text{conditional matrix: } P(x_2/x_1) = \frac{P(x_1, x_2)}{P(x_1)}
\]

\[
\begin{array}{c|c|c}
\hline
& R & B \\
\hline
R & \frac{5}{9} & \frac{1}{9} \\
B & \frac{4}{6} & \frac{2}{6} \\
\hline
\end{array}
\]

Here, \(P(x_2/x_1) = P(x_1/x_2)\).
Consider the joint PDF:

\[ f(x, y) = P(x \leq x, y \leq y) \]

Sketch the corresponding PDF.
Demonstrate the fact that the average uncertainty of a system is not affected by the arrangement of events

\[ P(A) = \frac{1}{4}, \quad P(B) = \frac{1}{4}, \quad P(C) = P(D) = \frac{1}{8} \]

**Case 1:**

\[
H(S) = -P(A) \ln P(A) - P(B) \ln P(B) - P(C) \ln P(C) - P(D) \ln P(D)
\]
\[= \frac{1}{4} \ln 4 + \frac{1}{4} \ln 2 + \frac{1}{8} \ln 8 + \frac{1}{8} \ln 8\]
\[= \frac{1}{4} 2 + \frac{1}{2} 1 + \frac{1}{8} 3 + \frac{1}{8} 3\]
\[= \frac{1}{2} + \frac{1}{2} + \frac{3}{4} = 1 \frac{3}{4} \text{ BITS} \]

**Case 2:**

\[
H(S) = H(P(A, P(A^c)) + P(A^c)) \ln \left( \frac{P(B)}{P(C)} \right) + \frac{P(C)}{P(A)} \ln \left( \frac{P(A)}{P(A) + P(A^c)} \right)
\]
\[= \frac{1}{2} \ln 2 + \frac{1}{2} \ln 2\]
\[+ \frac{1}{2} \left[ \frac{1}{2} \ln \frac{1}{2}, \ln \frac{1}{4} + \frac{1}{2} \ln \frac{1}{2}, \ln \frac{1}{2}, \ln \frac{1}{4} \right]\]
\[= 1 + \frac{1}{2} \left[ \frac{1}{2} \ln 2 + \frac{1}{2} \ln 4 \right]\]
\[= 1 + \frac{1}{2} \left[ \frac{1}{2} + 1 \right] = 1 + \frac{3}{4} = 1 \frac{3}{4} \text{ BITS} \]

**Case 3:**

\[
H(S) = H \left( \frac{P(A + B)}{P(C + D)} \right) + \frac{P(A + B)}{P(A)} H \left( \frac{P(B)}{P(A + B)} \right) + \frac{P(C + D)}{P(C + D)} H \left( \frac{P(C)}{P(C + D)} \right)
\]
\[P(A + B) = \frac{3}{4} \quad P(C + D) = \frac{1}{4}\]
\[\Rightarrow H(S) = \frac{3}{4} \ln 3 + \frac{1}{4} \ln 4 + \frac{3}{4} \left[ \frac{1}{4} \ln 3 + \frac{1}{4} \ln 3 \right]
\[+ \frac{1}{4} \left[ \frac{1}{2} \ln 2, \ln 2 \right]
\[= \frac{3}{4} \cdot 2 - \frac{3}{4} \ln 3 + \frac{1}{2} + \frac{3}{4} \left[ \frac{1}{3} \ln 3 + \frac{2}{3} \ln 3 - \frac{2}{3} \right] + \frac{1}{4}\]
\[= \frac{3}{2} + \frac{1}{2} - \frac{1}{2} + \frac{3}{4}\]
\[= \frac{7}{4} = 1 \frac{3}{4} \text{ BITS} \]
From Text (2-3) p. 41

\[ S_1 \Rightarrow (P_1, P_2, P_3, \ldots, P_q) ; H_1 \]
\[ S_2 \Rightarrow (Q_1, Q_2, Q_3, \ldots, Q_q) ; H_2 \]
\[ S(\lambda) = (r_1, r_2, \ldots, r_q ; s_1, s_2, \ldots, s_q) \Rightarrow (\lambda P_1, \lambda P_2, \ldots, \lambda P_q ; \lambda Q_1, \lambda Q_2, \lambda Q_3, \ldots, \lambda Q_q) \]

a) Show that
\[ H[S(\lambda)] = \lambda H_1 + \lambda H_2 + H(\lambda) \]
\[ H(S(\lambda)) = -\sum_{i=1}^{q} \lambda P_i \ln \lambda P_i - \sum_{j=1}^{q} \lambda Q_j \ln \lambda Q_j \]
\[ = \lambda \sum_{i=1}^{q} P_i \ln P_i - \lambda \sum_{j=1}^{q} Q_j \ln Q_j \]
\[ = \lambda \ln \lambda H_1 - \lambda \ln \lambda - \lambda \ln \lambda = \lambda H_1 + \lambda H_2 - \lambda \ln \lambda \]

Now \[ H(\lambda) = -\lambda \ln \lambda - \lambda \ln \lambda \quad (\text{i.e., } \lambda + \lambda = 1) \]
\[ \Rightarrow H[S(\lambda)] = \lambda H_1 + \lambda H_2 + H(\lambda) \]

Interpretation: By taking two sources, \( S_1 \) and \( S_2 \) together, and letting only one generate a symbol. The expected proportion of time \( S_1 \) generates is \( \lambda \), and \( S_2 \) is \( \overline{\lambda} = 1 - \lambda \). This problem gives resulting entropy.

b) Find \( \lambda_0 \) that maximizes \( H[S(\lambda)] \)
\[ \frac{d}{d\lambda} H[S(\lambda)] = H_1 - H_2 + \frac{d}{d\lambda} H(\lambda) \]
\[ \frac{d}{d\lambda} H(\lambda) = -\frac{d}{d\lambda} \lambda \ln \lambda - \frac{d}{d\lambda} (1 - \lambda) \ln (1 - \lambda) \]
\[ = -\ln \lambda - 1 - \frac{d}{d\lambda} \ln (1 - \lambda) + \ln (1 - \lambda) \]
\[ = -\ln \lambda - 1 + \frac{1}{1 - \lambda} + \ln (1 - \lambda) + \frac{1}{1 - \lambda} \]
\[ = -\ln \lambda - 1 + \frac{1}{1 - \lambda} - 1 \]
\[ = \ln \frac{1}{1 - \lambda} \]
\[ \Rightarrow \frac{d}{d\lambda} H[S(\lambda)] = 0 = H_1 - H_2 + \ln \frac{2 - \lambda}{\lambda} \]
\[ \Rightarrow \ln \lambda_0 = H_1 - H_2 \Rightarrow \frac{1}{1 - \lambda_0} = e^{H_1 - H_2} \]
\[ \lambda_0 = (1 - \lambda_0) e^{H_1 - H_2} \Rightarrow \lambda_0 (1 + e^{H_1 - H_2}) = e^{H_1 - H_2} \]
\[ \text{or } \lambda_0 = \frac{e^{H_1 - H_2}}{1 - e^{H_1 - H_2}} = \frac{1}{e^{H_1 - H_2} - 1} \]
\[ H[S(\lambda_0)] = e^{H_1 - H_2} + [1 - e^{H_1 - H_2}] \cdot \ln H_1 \]
\[ = \lambda_0 \ln \lambda_0 - \lambda_0 \ln \lambda_0 \]

etc.
(2.14) \[ S = \{ s_1, s_2, \ldots, s_q \} \Rightarrow \{ p_1, p_2, \ldots, p_q \} \]
\[ S' = \{ r_1, r_2, \ldots, r_{q+1}, q+2, \ldots, q+2q \} \Rightarrow \{ p'_1, p'_2, \ldots, p'_{2q} \} \]

\[ p'_i = \begin{cases} 
(1 - \epsilon) p_i & ; i = 1, 2, \ldots, q \\
\epsilon p_{i-q} & ; i = q+1, q+2, \ldots, 2q 
\end{cases} \]

\[ H[S] = -\sum_{i=1}^{q} p_i \log p_i \]

\[ H[S'] = -\sum_{i=1}^{q} (1 - \epsilon) p_i \log (1 - \epsilon) p_i - \sum_{i=1}^{q} \epsilon p_i \log \epsilon p_i \]

\[ = - (1 - \epsilon) \sum p_i \log (1 - \epsilon) - (1 - \epsilon) \sum p_i \log p_i - \epsilon \sum p_i \log \epsilon p_i \]

\[ = - (1 - \epsilon) \sum p_i \log (1 - \epsilon) - \epsilon \log \epsilon 
+ (1 - \epsilon) H(S) + \epsilon H(S) \]

\[ = H(S) + H(S) \]
LET $X = (x_1, x_2)$ \hspace{1cm} $P = \left( \frac{1}{4}, \frac{3}{4} \right)$

\[ P(Y_1/x_1) = \frac{1}{4} \hspace{1cm} P(Y_1/x_2) = 0.10 \]
\[ P(Y_2/x_1) = 0.35 \hspace{1cm} P(Y_2/x_2) = 0.7 \]
\[ P(Y_3/x_1) = 0.40 \hspace{1cm} P(Y_3/x_2) = 0.2 \]

1. FIND $H(X)$

\[ H(X) = \frac{1}{4} \ln 4 + \frac{3}{4} \ln 3 = 0.532 \text{ NATS} = 0.811 \text{ BITS} \]

2. FIND $H(X,Y)$

\[ P(Y_1,x_1) = P(Y_1/x_1)P(x_1) = (0.25)(0.25) = 0.0625 \]
\[ P(Y_1,x_2) = P(Y_1/x_2)P(x_2) = (0.10)(0.75) = 0.075 \]
\[ P(Y_2,x_1) = (0.35)(0.25) = 0.0875 \]
\[ P(Y_2,x_2) = (0.7)(0.75) = 0.525 \]
\[ P(Y_3,x_1) = (0.4)(0.25) = 0.1 \]
\[ P(Y_3,x_2) = (0.2)(0.75) = 0.15 \]

\[ H(X,Y) = 1.43 \text{ NATS} = 2.07 \text{ BITS} \]

3. FIND $H(Y)$

\[ P(Y_1) = 0.0625 + 0.075 = 0.1375 \]
\[ P(Y_2) = 0.0875 + 0.525 = 0.6125 \]
\[ P(Y_3) = 0.25 \]

\[ H(Y) = 0.920 \text{ NATS} = 1.33 \text{ BITS} \]

4. $H(Y/X) = H(X,Y) - H(X)$

\[ = 2.07 - 0.811 = 1.60 \text{ BITS} \]

5. $H(X/Y) = H(X,Y) - H(Y)$

\[ = 2.07 - 1.33 = 0.74 \text{ BITS} \]
\[ S = \left( \begin{array}{c} P \\ \frac{q}{4} \\ \frac{1}{4} \end{array} \right) \]

\[ P \Rightarrow \text{PASS} \]

\[ S = \left( \begin{array}{c} C \\ \bar{C} \end{array} \right) \]

\[ C \Rightarrow \text{DOWN CAR} \]

\[ P(C/P) = 0.1 \]

\[ S = \left( \begin{array}{c} f \\ \bar{f} \end{array} \right) \]

\[ f \Rightarrow \text{FRAT MEMBER} \]

\[ P(C/\bar{P}) = 0.5 \]

\[ P(C/f) = 1.0 \]

\[ P(f/\bar{C},P) = 1.0 \]

\[ P(f/C,\bar{P}) = 0.4 \]

\[ p(c) = P(C,P) + P(C,\bar{P}) \]

\[ = P(C/P)P(P) + P(C/\bar{P})P(\bar{P}) \]

\[ = (0.1)(0.75) + (0.5)(0.25) = 0.2 \]

\[ H(S_c) = 0.2 \log_2 5 + 0.8 \log_2 1.25 \]

\[ = 0.5 \text{ Nats} = 0.722 \text{ Bits} \]

\[ p(f) = P(f,C,P) + P(f,\bar{C},P) + P(f,C,\bar{P}) + P(f,\bar{C},\bar{P}) \]

\[ \text{ETC} \]
prove that:

1. \( Y \in x \Rightarrow H(Y/Z) + H(Z/X) \)
2. \( x \in Z \Rightarrow H(x/YZ) - H(x) \)
3. \( H(Y/Z) = H(YZ) - H(Z) \)
4. \( H(Z/X) = H( XYZ) - H(x) \)

\[
H(XYZ) - H(X) \geq [H(xZ) + H(YZ)] - [H(x) + H(z)]
\]

\[
H(XYZ) \leq H(xZ) + H(YZ) - H(Z)
\]

\[
H(XYZ) \leq H(xZ) + H(YZ) - H(Z)
\]

\[
H(xZ) \leq H(x) + H(YZ) - H(Z)
\]

\[
H(xZ) \leq H(x) + H(YZ) - H(Z)
\]

\[
H(xZ) \leq H(x) + H(YZ) - H(Z)
\]

SINCE, BY LEMMA, \( H(a, b) \leq H(a) + H(b) \),

OUR PROOF IS COMPLETE.

c. \( H(Z/X, Y) \leq H(Z/X) \)

\[
H(Z/X, Y) = H(XYZ) - H(x, Y)
\]

\[
H(Z/X) = H(XZ) - H(x)
\]

\[
H(XYZ) - H(X, Y) \leq H(XZ) - H(x)
\]

\[
H(XYZ) \leq H(x) + H(xZ) - H(x)
\]

\[
H(xZ) \leq H(x) + H(Zx/X) \leq H(x) + H(Z)
\]

BY SAME ARGUMENT, WE'RE DONE.

\[
H(Y, Z/X) = H(Y/X) + H(Z/XY)
\]

\[
H(Y, Z/X) = H(XYZ) - H(x)
\]

\[
H(Y/X) = H(Y) - H(x)
\]

\[
H(Z/XY) = H(XYZ) - H(xy)
\]

\[
H(xYZ) \leq H(Y) + H(xYZ) - H(xy)
\]

\[
0 \leq H(xY) + H(xYZ/xy)
\]

\[
= H(Y) + H(z)
\]

No!

PROBLEM SHOULD HAVE READ

\[
H(Y, Z/X) \geq H(Y/X) + H(Z/XY)
\]
H.W. ESTABLISH $S^2$ WHERE

$$S = \begin{pmatrix} S_1 & S_2 & S_3 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

$S^2$ WILL HAVE $S^2 = 4$ SYMBOLS

$$S^2 = \begin{pmatrix} S_1 S_1 & S_1 S_2 & S_1 S_3 \\ S_2 S_1 & S_2 S_2 & S_2 S_3 \\ S_3 S_1 & S_3 S_2 & S_3 S_3 \end{pmatrix}$$

$H(S) = \frac{1}{2} \log_2 2 + \frac{1}{4} \log_2 4 + \frac{1}{4} \log_2 4$

$$= \frac{1}{2} (1) + 2 \frac{1}{4} \cdot 2$$

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

$H^2(S) = \frac{1}{4} \log_2 4 + 4 \frac{1}{8} \log_2 8 + 4 \frac{1}{16} \log_2 16$

$$= \frac{1}{4} \cdot 2 + 4 \times \frac{1}{8} \cdot 3 + \frac{1}{16} \cdot 4$$

$$= \frac{1}{2} \times \frac{15}{16} = \frac{9 + 30 + 7}{16} \times \frac{15}{16}$$

$$= \frac{1}{2} + \frac{3}{2} + 1 = 3$$

THAT IS $H^2(S) = \eta H(S)$ AS EXPECTED.
P. Work out conditional matrices in terms of joint matrix.

The joint matrix is:

\[
\begin{array}{cccccc}
X_1 & P(X_1, Y_1) & P(X_1, Y_2) & \cdots & P(X_1, Y_j) & \cdots & P(X_1, Y_m) \\
X_2 & P(X_2, Y_1) & P(X_2, Y_2) & \cdots & P(X_2, Y_j) & \cdots & P(X_2, Y_m) \\
& \vdots & & & & & \vdots \\
X_n & P(X_n, Y_1) & P(X_n, Y_2) & \cdots & P(X_n, Y_j) & \cdots & P(X_n, Y_m) \\
\end{array}
\]

\[
P(Y_1) \quad P(Y_2) \quad \cdots \quad P(Y_j) \quad \cdots \quad P(Y_m)
\]

For conditional matrix \( H(X/Y) \), replace \((i,j)\)th element by

\[
P(X_i/Y_j) = \frac{P(X_i, Y_j)}{P(Y_j)}
\]

\( H(Y/X) \) by

\[
P(Y_j/X_i) = \frac{P(X_i, Y_j)}{P(X_i)}
\]
(i) **Discrete Noise-Free Channel: \( x \perp y \) Independent**

Given \( P(x;y) \) Matrix

\[
H(x,y) = - \sum_i \sum_j p(x_i, y_j) \log_2 p(x_i, y_j)
\]
\[
= - \sum_i \sum_j p(x_i) p(y_j) \log_2 p(x_i) p(y_j)
\]
\[
= - \sum_i p(x_i) \log_2 p(x_i)
\]
\[
= H(x) + H(y)
\]

(ii) **Discrete Noisy Channel with Noise**

<table>
<thead>
<tr>
<th>( y_1 )</th>
<th>( y_2 )</th>
<th>\ldots</th>
<th>( y_k )</th>
<th>\ldots</th>
<th>( y_n )</th>
<th>( P(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>( p_1 )</td>
<td>( p_1 )</td>
<td>\ldots</td>
<td>( p_1 )</td>
<td>( p_1 )</td>
<td>\ldots</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>( p_2 )</td>
<td>( p_2 )</td>
<td>\ldots</td>
<td>( p_2 )</td>
<td>( p_2 )</td>
<td>\ldots</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>\ldots</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>\ldots</td>
</tr>
<tr>
<td>( x_i )</td>
<td>( p_i )</td>
<td>( p_i )</td>
<td>\ldots</td>
<td>( p_i )</td>
<td>( p_i )</td>
<td>\ldots</td>
</tr>
<tr>
<td>( x_n )</td>
<td>( p_n )</td>
<td>( p_n )</td>
<td>\ldots</td>
<td>( p_n )</td>
<td>( p_n )</td>
<td>\ldots</td>
</tr>
</tbody>
</table>

\[
P_y = \sum_{i=1}^{n} p_i = \frac{1}{m}
\]

\[
\Rightarrow H(y) = - \sum_{i=1}^{m} \log_2 p_i = - \sum_{i=1}^{m} \log_2 p_y
\]

\[
H(x) = n \sum_{i=1}^{n} p_i \log_2 n p_i
\]
\[
= - n \sum_{i=1}^{n} p_i \log_2 n - n \sum_{i=1}^{n} p_i \log_2 p_i
\]
\[
= - n \left( \frac{1}{n} \right) \log_2 n - n \sum_{i=1}^{n} p_i \log_2 p_i
\]
\[
= - \log_2 n \ldots
\]
PROVE THAT \( I(A^n; B^n) = n I(A; B) \)
\[
\begin{align*}
A &= \{ a_1, a_2, \ldots, a_r \} \\
B &= \{ b_1, b_2, \ldots, b_s \} \\
A^n &= \{ \alpha_1, \alpha_2, \ldots, \alpha_{r^n} \} \\
B^n &= \{ \beta_1, \beta_2, \ldots, \beta_{s^n} \}
\end{align*}
\]

NOTATION
\[
\begin{align*}
P[a_i] &= P_i \\
P[b_j] &= P_j \\
P[a_i, b_j] &= P_{i,j} \\
P[a_i/b_j] &= P_{i,j} \\
P(\alpha_i) &= P_i^n \\
P(\beta_j) &= P_j^n \\
P(\alpha_i, \beta_j) &= P_{i,j}^n \\
P(\alpha_i/\beta_j) &= P_{i,j}^n
\end{align*}
\]

Now
\[ I(A; B) = H(A) - H(A/B) \]  
\[ \text{IT FOLLOWS THAT} \]
\[ I(A^n; B^n) = H(A^n) - H(A^n/B^n) \]  

WE HAVE SHOWN (EQ. 2-18 ON PG. 21) THAT
\[ H(A^n) = n H(A) \]  

THUS, IT REMAINS TO SHOW THAT
\[ H(A^n/B^n) = n H(A/B) \]  

NOW
\[ H(A/B) = - \sum_{x=1}^{r} \sum_{j=1}^{s} P_{x,j}^n \log P_{x,j}^n \]  

THUS
\[ H(A^n/B^n) = - \sum_{x=1}^{r} \sum_{j=1}^{s} P_{x,j}^n \log P_{x,j}^n \]  

NOW
\[
\begin{align*}
P_{i,j}^n &= P[a_i, a_{i_2} \ldots a_{i_n}, b_j, b_{j_2} \ldots b_{j_n}] \\
&= P[a_i, b_j] P[a_{i_2}, b_{j_2}] \ldots P[a_{i_n}, b_{j_n}] \\
&= P_{i,j} P_{i_2,j_2} \ldots P_{i_n,j_n}
\end{align*}
\]
Also
\[ p_{zi}^* = p \left[ \frac{\alpha_i}{b_j} \right] = p \left[ a_{i1} a_{i2} \ldots a_{in} / b_{j1} b_{j2} \ldots b_{jn} \right] = p \left[ a_{i1} / b_{j1} \right] p \left[ a_{i2} / b_{j2} \right] \ldots p \left[ a_{in} / b_{jn} \right] = p_{\alpha_i/\beta_j} p_{\alpha_2/\beta_j} \ldots p_{\alpha_n/\beta_j} \]

Thus
\[ H(A^n/B^n) = -\sum_{j=1}^{n} \sum_{d=1}^{m} \left[ \log p_{\alpha_i/\beta_j} p_{\alpha_2/\beta_j} \ldots p_{\alpha_n/\beta_j} \right] \]

But
\[ \sum_{j=1}^{n} \sum_{d=1}^{m} \frac{\sum_{i=1}^{r} p_{\alpha_i/\beta_j} \log p_{\alpha_i/\beta_j} p_{\alpha_2/\beta_j} \ldots p_{\alpha_n/\beta_j}}{d_{j1} d_{j2} \ldots d_{jn}} = \sum_{x=1}^{r} \sum_{j=1}^{n} p_{\alpha_i/\beta_j} \log p_{\alpha_i/\beta_j} p_{\alpha_2/\beta_j} \ldots p_{\alpha_n/\beta_j} \]

\[ = -\sum_{j=1}^{n} \sum_{d=1}^{m} p_{\alpha_i/\beta_j} \log p_{\alpha_i/\beta_j} p_{\alpha_2/\beta_j} \ldots p_{\alpha_n/\beta_j} = \sum_{x=1}^{r} \sum_{j=1}^{n} p_{\alpha_i/\beta_j} \log p_{\alpha_i/\beta_j} p_{\alpha_2/\beta_j} \ldots p_{\alpha_n/\beta_j} \]

\[ = \sum_{x=1}^{r} \sum_{j=1}^{n} p_{\alpha_i/\beta_j} \log p_{\alpha_i/\beta_j} p_{\alpha_2/\beta_j} \ldots p_{\alpha_n/\beta_j} \]

\[ \Rightarrow H(A^n/B^n) = -\sum_{j=1}^{n} \sum_{d=1}^{m} p_{\alpha_i/\beta_j} \log p_{\alpha_i/\beta_j} p_{\alpha_2/\beta_j} \ldots p_{\alpha_n/\beta_j} \]

\[ - \ldots - \sum_{x=1}^{r} \sum_{d=1}^{m} p_{\alpha_i/\beta_j} \log p_{\alpha_i/\beta_j} p_{\alpha_2/\beta_j} \ldots p_{\alpha_n/\beta_j} \]
FROM (4):
\[ H(A/B) = - \sum_{i_k=1}^{r} \sum_{j_k=1}^{s} p_{i_k j_k} \log p_{i_k j_k} \]

THUS
\[ H(A^n/B^n) = n \cdot H(A/B) \]

SUBSTITUTING THIS AND (3) INTO (2) Completes THE PROOF:
\[ I(A^n;B^n) = n \cdot H(A) - n \cdot H(A/B) \]
\[ = n \cdot I(A;B) \]
PROB: USING THE DEF. OF MUTUAL INFORMATION,
SHOW THAT: \( I(X;Y) = H(X) + H(Y) - H(X,Y) \) (a)
\[ = H(X) - H(X/Y) \]
\[ = H(Y) - H(Y/X) \]

\( I(X;Y) = I(x_i; y_i) \)
\( = \sum_i \sum_j p(x_i, y_j) \log_2 \frac{p(x_i | y_j)}{p(x_i)} \)
\[ = \sum_i \sum_j p(x_i, y_j) \log_2 \frac{p(x_i | y_j)}{p(x_i)} \]
\[ = -H(X,Y) - \sum_i \sum_j p(x_i, y_j) \log_2 p(x_i, y_j) + \sum_i p(x_i) \log_2 p(x_i) \]
\[ = -H(X) + H(X) + H(Y) \]
\[ = H(Y) - H(Y/X) \]

**b. STARTING FROM** (2):
\( I(X;Y) = \sum_i \sum_j p(x_i, y_j) \log_2 \frac{p(y_j | x_i)}{p(y_j)} \)
\[ = \sum_i \sum_j p(x_i, y_j) \log_2 \frac{p(y_j | x_i)}{p(y_j)} \]
\[ = -H(Y | X) - \sum_i \sum_j p(y_j) \log_2 p(y_j) + \sum_i p(x_i) \log_2 p(x_i) \]
\[ = -H(Y | X) + H(Y) \]
\[ = H(Y) - H(Y/X) \]
USING MUROGA'S TECHNIQUE, FIND THE CHANNEL CAPACITY FOR

\[
p = \begin{bmatrix}
\frac{1}{2} & \frac{1}{2} \\
\frac{1}{4} & \frac{3}{4}
\end{bmatrix} \quad \Rightarrow p'_1 = \frac{3}{4}, \quad p'_2 = \frac{5}{4}
\]

WE WANNA SOLVE

\[
\begin{bmatrix}
p_1' \\
p_2'
\end{bmatrix} \begin{bmatrix}
q_1 \\
q_2
\end{bmatrix} = \begin{bmatrix}
-H(\frac{1}{2}) \\
-H(\frac{1}{4})
\end{bmatrix}
\]

\[-H(\frac{1}{2}) = 1\text{ bit}
\]

\[-H(\frac{1}{4}) = \frac{1}{4} \log_2 4 + \frac{3}{4} \log_2 \frac{1}{3}
\]

\[= \log_2 4 - \frac{3}{4} \log_2 3
\]

\[= 0.811278 \quad (0)
\]

\[\begin{align*}
p^{-1} &= C^T / \det p \\
\det p &= \frac{3}{4} - \frac{1}{4} = \frac{1}{4} \\
C &= \begin{bmatrix}
\frac{3}{4} & -\frac{1}{4} \\
-\frac{1}{2} & \frac{1}{2}
\end{bmatrix} \\
C^T &= \begin{bmatrix}
\frac{3}{4} & -\frac{1}{2} \\
-\frac{1}{4} & \frac{1}{2}
\end{bmatrix}
\end{align*}
\]

\[\Rightarrow p^{-1} = \begin{bmatrix}
3 & -2 \\
-1 & 2
\end{bmatrix}
\]

\[
\begin{bmatrix}
q_1 \\
q_2
\end{bmatrix} = \begin{bmatrix}
3 & -2 \\
-1 & 2
\end{bmatrix} \begin{bmatrix}
-1 \\
-0.811278
\end{bmatrix}
\]

\[= \begin{bmatrix}
-3 + 2(0.811278) \\
1 - 2(0.811278)
\end{bmatrix}
\]

\[= \begin{bmatrix}
-1.3774438 \\
-0.622556
\end{bmatrix} \quad (3)
\]
\[ C = \log_2 2^{q_1} + 2^{q_2} \]
\[ = \log_2 2^{1.0344192} \]
\[ = 0.0486210 \text{ BITS} \]

**Using the graph**

\[ p_{11} = \frac{1}{2}, \quad p_{22} = \frac{3}{4} \Rightarrow q_1 = -1.35 \]
\[ q_2 = -0.6 \]
\[ C = 0.0486210 \text{ BITS} \]
GIVEN
\[
\begin{bmatrix}
p_{11} & p_{12} \\
p_{21} & p_{22}
\end{bmatrix}
\begin{bmatrix}
Q_1 \\
Q_2
\end{bmatrix}
= \begin{bmatrix}
H(p_{11}) \\
H(p_{22})
\end{bmatrix}
= \begin{bmatrix}
p_{11} \log p_{11} + p_{12} \log p_{12} \\
p_{21} \log p_{21} + p_{22} \log p_{22}
\end{bmatrix}
\]

\[p'_1 = p_{11} + p_{12}\]
\[p'_2 = p_{22} + p_{21}\]

SHOW THAT
\[I(x; y) = -p_1 \log p'_1 - p'_2 \log p'_2 + p'_1 Q_1 + p'_2 Q_2\]

NOW \[H(y) = -p_1 \log p_1 - p_2 \log p_2\]

\[\Rightarrow I(x; y) = H(y) + p'_1 Q_1 + p'_2 Q_2\]
\[= H(y) + \left( p_{11} Q_1 + p_{12} Q_2 \right) + \left( p_{21} Q_1 + p_{22} Q_2 \right)\]
\[= H(y) + \begin{bmatrix}
p_{11} \log p_{11} + p_{12} \log p_{12} \\
p_{21} \log p_{21} + p_{22} \log p_{22}
\end{bmatrix}\]
\[= H(y) - \sum_i \sum_j p_{ij} \log p_{ij}\]
\[= H(y) - H(y/x)\]
START WITH $\Phi_0$ AND BUILD AN INSTANT CODE

FOR $S = \Phi, S_1, S_2, \ldots, S_8$

$S_1 = 10$
$S_2 = 11$
$S_3 = 00$
$S_4 = 010$
$S_5 = 011$

FOR $S: \Phi, S_1, \ldots, S_8$

$S_1 = 01$
$S_2 = 111$
$S_3 = 001$
$S_4 = 000$
$S_5 = 101$
$S_6 = 100$
$S_7 = 110.0$
$S_8 = 1101$
TEXT: PROB 3-2, p. 63

(a), (b) INSTANTANEOUS CODES ARE

\[ a, c \hat{f} e \]

ALL INSTANT CODES ARE DECODABLE.

THUS \( a, c, \hat{f} e \) ARE U.D. \( b \) IS ALSO

U.D. (BUT DOES NOT OBEY THE

PREFIX PROPERTY). \( f \) IS NOT U.D

SINCE \( s_1 s_2 = s_3 s_1 = 00100 \)

\( d \) IS NOT U.D, SINCE \( s_2 s_3 = s_1 s_5 = 10110 \)

TO FIND AVERAGE WORD LENGTHS, USE HP

\[ a: \bar{L} = 3 \]
\[ b: \bar{L} = 2.1250 \]
\[ c: \bar{L} = 2.1250 \]
\[ d: \bar{L} = 1.9375 \]
\[ e: \bar{L} = 1.9375 \]
\[ f: \bar{L} = 2 \]
2. MUST CHECK VIA KRAFT

\[ \sum_{i=1}^{9} r^{-i} \leq \sum_{i=1}^{2} N_{i} r^{-i} \leq 1 \]

HERE \( r = 3 \)

\[ ( ) \quad 3 \quad R/S \quad + \]

CHS \( X \leftrightarrow Y \)

ENT \( Y \times \)

\( RCL 0 \quad 0 \times 0 \)

A : 0.91

B : 1.004

C : 0.99999

D : 0.99999

<table>
<thead>
<tr>
<th>( s )</th>
<th>A</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>11</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>210</td>
<td>12</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>211</td>
<td>20</td>
<td>220</td>
<td>220</td>
</tr>
<tr>
<td>2200</td>
<td>220</td>
<td>221</td>
<td>221</td>
</tr>
<tr>
<td>2201</td>
<td>221</td>
<td>2220</td>
<td>2220</td>
</tr>
<tr>
<td>2210</td>
<td>222</td>
<td>2221</td>
<td>2221</td>
</tr>
<tr>
<td>2211</td>
<td>210</td>
<td>22220</td>
<td>22220</td>
</tr>
<tr>
<td>22111</td>
<td>211</td>
<td>22221</td>
<td>22221</td>
</tr>
<tr>
<td>22211</td>
<td>2222</td>
<td>22222</td>
<td>22222</td>
</tr>
</tbody>
</table>
DEVISE TWO INTELLIGENT CODES FOR

\( \mathcal{A} = \{0, 3, 5\} \) AND \( \mathcal{C} = \{0, 1, 2\} \)

<table>
<thead>
<tr>
<th>CODE #1</th>
<th>CODE #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>5_1</td>
<td>00</td>
</tr>
<tr>
<td>5_2</td>
<td>01</td>
</tr>
<tr>
<td>5_3</td>
<td>10</td>
</tr>
<tr>
<td>5_4</td>
<td>1100</td>
</tr>
<tr>
<td>5_5</td>
<td>1101</td>
</tr>
<tr>
<td>5_6</td>
<td>1110</td>
</tr>
<tr>
<td>5_7</td>
<td>1111</td>
</tr>
<tr>
<td>5_8</td>
<td>2100</td>
</tr>
</tbody>
</table>
Verify the following code lengths will work for \( r = 3 \). We must satisfy Kraft's inequality:

\[
\sum_{i=1}^{q} r^{-l_i} = \sum_{i=1}^{2} N_i \cdot r^{-i} \leq 1
\]

1. \( l_i = (1, 2, 3, 4, 5, 6, 7) \)
   \( N_i = (1, 1, 1, 0, 0, 1, 0) \)
   \( \sum_{i=1}^{2} N_i \cdot r^{-i} = 0.48 \)

2. \( l_i = 1, 2, 3, 4, 5, 6, 7 \)
   \( N_i = 1, 2, 0, 3, 2, 1, 0 \)
   \( \sum_{i=1}^{2} N_i \cdot r^{-i} = 0.60 \)

3. \( N_i = 1, 0, 2, 0, 1, 2, 3 \)
   \( l_i = 1, 2, 3, 4, 5, 6, 7 \)
   \( = 0.42 \)
Compute $\bar{I}$ and $H(x)$ for

\[
\begin{array}{ccc}
  x_1 & 0.4 & 00 \\
  x_2 & 0.3 & 01 \\
  x_3 & 0.2 & 101 \\
  x_4 & 0.1 & 1110 \\
\end{array}
\]

$$
\bar{I} = (0.4)(2) + (0.3)(2) + (0.2)3 + 0.1(4) \\
= 2.40
$$

$H(x) = 1.25$ bits

And, as expected $\bar{I} \geq H(x)$
TEXT: p. 91 # 4-7: REQUIRE \( q = 3 + 2\alpha \Rightarrow \text{LET} \ \alpha = 3 \Rightarrow q = 2 \\
3. s_1 0.4 0 0.4 0 0.4 0 0.4 0
6. s_2 0.2 -2 0.2 2 0.2 2 \rightarrow 0.4 1
5. s_3 0.1 11 0.1 11 \rightarrow 0.2 \rightarrow 0.2 \rightarrow 0.2 2
4. s_4 0.1 12 0.1 12 \rightarrow 0.1 \rightarrow 0.1 12
8. s_5 0.05 101 \rightarrow 0.1 \rightarrow 100 \rightarrow 0.1 12
7. s_6 0.05 102 \rightarrow 0.05 \rightarrow 102 \rightarrow 100
9. s_7 0.05 \rightarrow 1000 \rightarrow 0.05 \rightarrow 102
8. s_8 0.05 \rightarrow 1001
9. s_9 0.0 \rightarrow 1002

ANOTHER CODE WOULD BE
3. s_1 .4 1 .4 1 .4 1 \rightarrow .4 0
6. s_2 .2 00 .2 00 \rightarrow .2 00 \rightarrow .4 1
5. s_3 .1 02 \rightarrow .1 01 \rightarrow .2 00 \rightarrow .2 2
4. s_4 .1 20 \rightarrow .1 02 \rightarrow .1 01
8. s_5 .05 21 \rightarrow .1 20 \rightarrow .1 02
7. s_6 .05 22 \rightarrow .05 21
9. s_7 .05 \rightarrow 010 \rightarrow .05 22
8. s_8 .05 011
9. s_9 .0 \rightarrow 012

ETC.
For $q = 6$, we should get five trees:

1. 

2. 

3. 

4. 

5. 

Etc.
CODE (HUFFMAN)

<table>
<thead>
<tr>
<th>$s_i$</th>
<th>$\frac{1}{9}$</th>
<th>$001$</th>
<th>$\frac{2}{9}$</th>
<th>$01$</th>
<th>$\frac{4}{9}$</th>
<th>$00$</th>
<th>$\frac{1}{9}$</th>
<th>$001$</th>
<th>$\frac{2}{9}$</th>
<th>$00$</th>
<th>$\frac{1}{9}$</th>
<th>$001$</th>
<th>$\frac{1}{9}$</th>
<th>$001$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_2$</td>
<td>$\frac{1}{9}$</td>
<td>$000$</td>
<td>$\frac{1}{9}$</td>
<td>$001$</td>
<td>$\frac{2}{9}$</td>
<td>$10$</td>
<td>$\frac{2}{9}$</td>
<td>$01$</td>
<td>$\frac{2}{9}$</td>
<td>$10$</td>
<td>$\frac{2}{9}$</td>
<td>$01$</td>
<td>$\frac{2}{9}$</td>
<td>$10$</td>
</tr>
<tr>
<td>$s_3$</td>
<td>$\frac{1}{9}$</td>
<td>$001$</td>
<td>$\frac{1}{9}$</td>
<td>$000$</td>
<td>$\frac{1}{9}$</td>
<td>$001$</td>
<td>$\frac{1}{9}$</td>
<td>$000$</td>
<td>$\frac{1}{9}$</td>
<td>$001$</td>
<td>$\frac{1}{9}$</td>
<td>$000$</td>
<td>$\frac{1}{9}$</td>
<td>$000$</td>
</tr>
<tr>
<td>$s_4$</td>
<td>$\frac{1}{9}$</td>
<td>$110$</td>
<td>$\frac{1}{9}$</td>
<td>$000$</td>
<td>$\frac{1}{9}$</td>
<td>$000$</td>
<td>$\frac{1}{9}$</td>
<td>$000$</td>
<td>$\frac{1}{9}$</td>
<td>$000$</td>
<td>$\frac{1}{9}$</td>
<td>$000$</td>
<td>$\frac{1}{9}$</td>
<td>$000$</td>
</tr>
<tr>
<td>$s_5$</td>
<td>$\frac{1}{9}$</td>
<td>$111$</td>
<td>$\frac{1}{9}$</td>
<td>$110$</td>
<td>$\frac{1}{9}$</td>
<td>$110$</td>
<td>$\frac{1}{9}$</td>
<td>$110$</td>
<td>$\frac{1}{9}$</td>
<td>$110$</td>
<td>$\frac{1}{9}$</td>
<td>$110$</td>
<td>$\frac{1}{9}$</td>
<td>$110$</td>
</tr>
<tr>
<td>$s_6$</td>
<td>$\frac{1}{9}$</td>
<td>$100$</td>
<td>$\frac{1}{9}$</td>
<td>$111$</td>
<td>$\frac{1}{9}$</td>
<td>$111$</td>
<td>$\frac{1}{9}$</td>
<td>$111$</td>
<td>$\frac{1}{9}$</td>
<td>$111$</td>
<td>$\frac{1}{9}$</td>
<td>$111$</td>
<td>$\frac{1}{9}$</td>
<td>$111$</td>
</tr>
<tr>
<td>$s_7$</td>
<td>$\frac{1}{9}$</td>
<td>$101$</td>
<td>$\frac{1}{9}$</td>
<td>$100$</td>
<td>$\frac{1}{9}$</td>
<td>$100$</td>
<td>$\frac{1}{9}$</td>
<td>$100$</td>
<td>$\frac{1}{9}$</td>
<td>$100$</td>
<td>$\frac{1}{9}$</td>
<td>$100$</td>
<td>$\frac{1}{9}$</td>
<td>$100$</td>
</tr>
<tr>
<td>$s_8$</td>
<td>$\frac{1}{9}$</td>
<td>$010$</td>
<td>$\frac{1}{9}$</td>
<td>$101$</td>
<td>$\frac{1}{9}$</td>
<td>$101$</td>
<td>$\frac{1}{9}$</td>
<td>$101$</td>
<td>$\frac{1}{9}$</td>
<td>$101$</td>
<td>$\frac{1}{9}$</td>
<td>$101$</td>
<td>$\frac{1}{9}$</td>
<td>$101$</td>
</tr>
<tr>
<td>$s_9$</td>
<td>$\frac{1}{9}$</td>
<td>$011$</td>
<td>$\frac{1}{9}$</td>
<td>$010$</td>
<td>$\frac{1}{9}$</td>
<td>$010$</td>
<td>$\frac{1}{9}$</td>
<td>$010$</td>
<td>$\frac{1}{9}$</td>
<td>$010$</td>
<td>$\frac{1}{9}$</td>
<td>$010$</td>
<td>$\frac{1}{9}$</td>
<td>$010$</td>
</tr>
</tbody>
</table>

$H(s) = 9 \times \frac{1}{9} \log_2 \frac{1}{9} = 3.178175$

$E = 7 \times 3 \frac{1}{9} + 2 \times 4 \frac{1}{9} = \frac{21 + 8}{9} = \frac{29}{9} = 3.2222$

$\Rightarrow N = \frac{E}{H(s)} = 0.9884$
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)
✓ MAROQA (PG 64 OF NOTES) (7-28-76)
✓ H.O #3 (MATRICES ON PG 62 OF NOTES) (7-28-76)
✓ H.O #4 (BINARY MULTIPLICATIVE CHANNEL)
✓ 4-6 FROM TEXT (8-4-76)

Pg 88 OF NOTES, PROVE THAT, FOR A HOFFMANN CODE, $H(x) \leq H(x) + 1 - 2p_{\text{min}}$ (8-4-76)

Pg 97 OF NOTES, (8-5-76) DATE CODING
✓ H.O #6 \rightarrow COMPUTER PROGRAM
✓ READING SHANNON'S PAPER

TAKE HOME TEST PROBLEMS
HW (1)

Experiment A has $M$ mutually exclusive possible outcomes $A_m$, Experiment B has $N$ mutually exclusive possible outcomes $B_n$. Show that

$$P(B_n / A_m) = \frac{P(A_m / B_n) \cdot P(B_n)}{\sum_{i=1}^{N} P(A_m / B_i) \cdot P(B_i)}$$
A change that tends in the following sense to equalize a set of probabilities \( p_1, \ldots, p_M \) always results in an increase in uncertainty:

Suppose \( p_1 > p_2 \). Define

\[
\begin{align*}
p_i' &= p_1 - \Delta p \\
p_i'' &= p_2 + \Delta p
\end{align*}
\]

where \( \Delta p > 0 \) and \( p_1 - \Delta p \geq p_2 + \Delta p \). Show that \( H(p_1', \ldots, p_M') > H(p_1, \ldots, p_M) \).

1.7 (Feinstein 1958). Let \( A = [a_{ij}] \) be a doubly stochastic matrix, that is, \( a_{ij} \geq 0 \) for all \( i, j \); \( \sum_{i=1}^{M} a_{ij} = 1 \), \( i = 1, \ldots, M \); \( \sum_{j=1}^{M} a_{ij} = 1 \), \( j = 1, \ldots, M \). Given a set of probabilities \( p_1, \ldots, p_M \), define a new set of probabilities \( p_1', \ldots, p_M' \) by

\[
p_i' = \sum_{j=1}^{M} a_{ij} p_j, \quad i = 1, 2, \ldots, M.
\]

Show that \( H(p_1', \ldots, p_M') \geq H(p_1, \ldots, p_M) \) with equality if and only if \( (p_1', \ldots, p_M') \) is a rearrangement of \( (p_1, \ldots, p_M) \). Show also that Problem 1.6 is a special case of this result.

1.8 Given a discrete random variable \( X \) with values \( x_1, \ldots, x_M \), 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 \( x_1, \ldots, x_M \); \( y_1, \ldots, y_L \) respectively. Let \( Z = X + Y \). 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 \( H(X) \leq H(Z) \).

1.10 Give an example in which \( H(X) > H(Z), H(Y) > H(Z) \).

1.11 Prove the generalized grouping axiom.

\[
H(p_1, \ldots, p_{r_1}; p_{r_1+1}, \ldots, p_{r_2}; \ldots; p_{r_{L-1}+1}, \ldots, p_0)
\]

\[
= H(p_1 + \cdots + p_{r_1}, p_{r_1+1} + \cdots + p_{r_2}, \ldots, p_{r_{L-1}+1} + \cdots + p_0)
\]

\[
+ \sum_{i=1}^{L} (p_{r_{i-1}+1} + \cdots + p_{r_i}) H\left( \frac{p_{r_{i-1}+1}}{\sum_{j=r_{i-1}+1}^{i} p_j}, \ldots, \frac{p_i}{\sum_{j=r_{i-1}+1}^{i} p_i} \right)
\]

1.12 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 \( p_1, \ldots, p_M \) such that \( p_1 > 0, \sum_{i=1}^{M} p_i = 1 \), then \( h(p) = -C \log p \).

1.13 Given a function \( h(p) \), \( 0 < p \leq 1 \), satisfying

a. \( h(p_1 p_2) = h(p_1) + h(p_2) \), \( 0 < p_1, p_2 \leq 1 \).

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 \( h(p) = -C \log p \) where \( C > 0, b > 1 \).

Our first application of this result will be to the problem of "noiseless" channel, that is, input to output. Thus we do not care about the exact nature of the input, but only that the input and output are related. In this way, the term "channel" is to be understood as a device that accepts an input from some set of symbols and produces an output. In general, a channel will involve transmission errors. To summarize, the ingrediant is:

1. A random variable \( X \), with probabilities \( p_1, \ldots, p_M \) respectively, and a fixed length of \( X \) symbol sets \( \{x_1, \ldots, x_M\} \) such that

2. A set \( \{a_i, \ldots, a_M\} \) called alphabet; each symbol \( x_i \) is a character called the code. We will correspond to the code words \( x_1 x_2 \ldots x_M \) is called a code. The code words
The problem undertaken in this section is the evaluation of the maximum rate of transmission of information of binary channels. 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. Matry

Fig. 3-17. BC. 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

\[ MA_1 = p_{11}, \quad OA_2 = p_{22} \]

The ordinates of the entropy curve at \( A_1 \) and \( A_2 \) are

\[ B_1A_1 = H(p_{11}), \quad B_2A_2 = H(p_{22}) \]

Now, for any given channel output probabilities such as \( OA = p \) and

\[ MA = 1 - p, \] where \( p \) is the probability of receiving 1, the transinformation can be geometrically identified. In fact,

\[
\begin{align*}
I(X;Y) &= H(Y) - H(Y|X) \\
I(X;Y) &= H(p) - p_1H(p_{11}) - p_2H(p_{22}) \\
I(X;Y) &= BA - FA
\end{align*}
\]

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.
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 $P_{22}$. 

The chart illustrates the relationship between the probability of error and the capacity of the channel. The curves represent different values of $P_{22}$, and the capacity $C$ is shown on the $y$-axis. The $x$-axis represents $P_{11}$. The chart helps in visualizing how changes in $P_{11}$ and $P_{22}$ affect the capacity of the channel.
THIS PROGRAM COMPUTES THE CHANNEL CAPACITY OF AN NXN CHANNEL WHEN THE CHANNEL (CONDITIONAL PROBABlLITY) MATRIX P(B/A) IS KNOWN.

THE CHANNEL CAPACITY IS DETERMINED USING MURGA'S TECHNIQUE.

THE SYSTEM OF LINEAR EQUATIONS THAT IS GENERATED BY THIS TECHNIQUE IS SOLVED BY MEANS OF A GAUSS-JORDAN REDUCTION.

************************************************************************ WARNING ************************************************************************

ONE MUST ARRANGE THE ROWS OF THE CHANNEL (CONDITIONAL PROBABlLITY) MATRIX P(B/A) SO THAT THE MAIN DIAGONAL ELEMENTS ARE NON-ZERO. REARRANGING THE ROWS DOES NOT AFFECT THE VALUE OF THE CHANNEL CAPACITY AND HELPS TO INSURE THAT THE GAUSS-JORDAN REDUCTION WILL YIELD THE PROPER SOLUTION TO THE SYSTEM OF LINEAR EQUATIONS.

*************************************************************************************************************************************************

THE CHANNEL (CONDITIONAL PROBABlLITY) MATRIX P(B/A) IS READ IN BY ROWS 1.E..A(1,1)*A(1,2)*...*A(1,N)*A(2,1)*A(2,2)*...*A(N,N).

DEFINITION OF VARIABLES

N DIMENSION(N) OF THE NXN CHANNEL MATRIX P(B/A)
EPS PARAMETER FOR CHECKING THE SINGULARITY OF THE COEFFICIENT MATRIX
\[ A(i,j) \quad \text{ELEMENT IN THE ITH ROW AND JTH COLUMN OF THE CHANNEL MATRIX} \]

\[ \text{SUM} \quad \text{ELEMENT IN THE (J+1)TH COLUMN OF THE AUGMENTED CHANNEL MATRIX} \]

\[ \text{MAGIC} \quad \text{FACTOR FOR CONVERTING LOGARITHMS TO THE BASE 2 TO LOGARITHMS TO THE BASE E, (MAGIC = \text{ALOG}(2))} \]

\[ \text{CHANNEL CAPACITY (C = MAGIC*\text{ALOG}(x_{11}) + 2*x_{22} + \ldots + 2*x_{NN}) WHERE } x_{ij} \text{ IS THE ELEMENT IN THE ITH ROW JTH COLUMN OF THE SOLUTION MATRIX} \]

\[ \text{DIMENSION } A = (N+1) \times (N+1) \]

\[ \text{REAL MAGIC} = (1.0, 0.59315) \]

\[ \text{THE MATRIX WHICH IS READ IS THE CHANNEL MATRIX AND THE ELEMENTS ARE READ IN BY ROWS} \]

\[ \text{READ(15); EPS} \]

\[ 100 \quad \text{FORMAT(15)} \]

\[ 100 \quad \text{NP1 = N + 1} \]

\[ 509 \quad \text{FORMAT()} \]

\[ 405 \quad \text{WRITE(16,509)} \]

\[ 101 \quad \text{FORMAT(100,5)} \]

\[ 1 \quad \text{SUM = 0.0} \]

\[ 2 \quad \text{DO I = 1,N} \]

\[ 3 \quad \text{READ(I,J)} \]

\[ 4 \quad \text{SUM = SUM + A(J,J)} \]

\[ 5 \quad \text{WRITE(201,4)} \]

\[ 201 \quad \text{FORMAT(17); J} \]

\[ 2 \quad \text{CONTINUE} \]

\[ 209 \quad \text{FORMAT(17)} \]

\[ C \quad \text{BEGIN GAUSS-JORDAN REDUCTION} \]

\[ \text{DETER = 1.0} \]

\[ \text{DO 9 K = 1,N} \]

\[ \text{DETER K = DETER*A(K,K)} \]

\[ 2 \text{IF(ABS(A(K,K)) < EPS) GO TO S} \]

\[ \text{WRITE(15); EPS} \]

\[ 9 \text{GO TO 111} \]

\[ 202 \quad \text{FORMAT(15); K} \]

\[ 5 \text{IF(I.EQ.K)} \]

\[ 6 \text{DO J = K+1,N} \]

\[ 7 \text{A(I,J) = A(I,J) - A(I,K)*A(K,J)} \]

\[ 8 \text{A(I,K) = 0.0} \]

\[ 9 \text{CONTINUE} \]

\[ \text{WRITE(16,425)} \]

\[ \text{DO 10 I = 1,N} \]

\[ \text{WRITE(16,501)} \]

\[ \text{WRITE(16,419)} \]

\[ 419 \quad \text{FORMAT(23X}; \text{AUGMENTED CHANNEL MATRIX} \]

\[ \text{WRITE(16,209)} \]

\[ 203 \quad \text{FORMAT(200,10X,14,10X,14)} \]

\[ \text{CALCULATION OF CHANNEL CAPACITY} \]

\[ \text{SUMX = 0.0} \]

\[ \text{DO 12 I = 1,N} \]

\[ \text{SUMX = SUMX + 2.0*A(I,NP1)} \]

\[ \text{12 CONTINUE} \]

\[ \text{WRITE(16,413)} \]

\[ 413 \quad \text{FORMAT(8X,C = LOG2( (2**X(1) + 2**X(2) + \ldots + 2**X(N) ) = 1.263) \]

\[ \text{INPUT DATA FOR EXAMPLE PROBLEM} \]

\[ \text{INPUT OA} \]

\[ \text{TA FOR EXAMPLE PROBLEM} \]

\[ \text{IF(SUMX < 15.14) CONTINUE} \]

\[ \text{GO TO 16} \]

\[ \text{CONTINUE} \]

\[ \text{WRITE(16,209)} \]

\[ 209 \quad \text{FORMAT(17)} \]

\[ 4 \quad \text{SOLUTION MATRIX FOR X} \]

\[ \text{WRITE(16,209)} \]

\[ 209 \quad \text{FORMAT(17)} \]

\[ 4 \quad \text{SOLUTION MATRIX FOR X} \]

\[ \text{WRITE(16,209)} \]
EE 5325  2nd Summer, 1976

Sometimes 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. Weing Mote.
4. Froeblich
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!

[Signature]

C. Prabhskar
7/27/76
The following is a sketch for Binary Multiplicative Channel (BMC). This channel is modelled as a Zero-memory channel with 4 possible inputs making up the alphabet \( \mathcal{A} = \{00, 01, 10, 11\} \)

\[
\begin{align*}
ad - BMC - b &= ac
\end{align*}
\]

(a) Develop the output alphabet and the channel (conditional) matrix

(b) Let the primary symbols be 0, 1 and the corresponding probs be \( p \) and \( q \) respectively. Derive an expression for \( I(A; B) \), the transmitted information.

(c) Hence determine the channel capacity, if you can!

Due 8/13/76
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 is noiseless.

**THIS PROGRAM DECODES A HUFFMAN CODE USING A GIVEN TABLE**

**C AND THE TREE SEARCH METHOD**

*****************************************************************

<table>
<thead>
<tr>
<th>PROB.</th>
<th>LETTER</th>
<th>HUFF. CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0046</td>
<td>A</td>
<td>0100</td>
</tr>
<tr>
<td>0.0127</td>
<td>B</td>
<td>011111</td>
</tr>
<tr>
<td>0.0216</td>
<td>C</td>
<td>11111</td>
</tr>
<tr>
<td>0.0317</td>
<td>D</td>
<td>01011</td>
</tr>
<tr>
<td>0.0416</td>
<td>E</td>
<td>111</td>
</tr>
<tr>
<td>0.0608</td>
<td>F</td>
<td>0011001</td>
</tr>
<tr>
<td>0.0915</td>
<td>G</td>
<td>011101</td>
</tr>
<tr>
<td>0.1267</td>
<td>H</td>
<td>1000</td>
</tr>
<tr>
<td>0.1755</td>
<td>I</td>
<td>0011001</td>
</tr>
<tr>
<td>0.0915</td>
<td>J</td>
<td>011100101</td>
</tr>
<tr>
<td>0.0569</td>
<td>K</td>
<td>01110010</td>
</tr>
<tr>
<td>0.0321</td>
<td>L</td>
<td>01010</td>
</tr>
<tr>
<td>0.0305</td>
<td>M</td>
<td>001101</td>
</tr>
</tbody>
</table>

**AVERAGE LENGTH: 4.1195**

*****************************************************************

The above table was taken from the following reference:

**REFERENCE: REZA, 'AN INTRODUCTION TO INFORMATION THEORY'**

M(i) is the binary digit (0 or 1) being observed. If = 1, 2, ..., L

Y(j) is the alphabet letter (or space) being formed from the code.
The message to be decoded is:

<table>
<thead>
<tr>
<th>i</th>
<th>J</th>
<th>Y(i,j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>D</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>E</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>F</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>G</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>H</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>I</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>J</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>K</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>L</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>M</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>N</td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td>O</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>P</td>
</tr>
<tr>
<td>17</td>
<td>18</td>
<td>Q</td>
</tr>
<tr>
<td>18</td>
<td>19</td>
<td>R</td>
</tr>
<tr>
<td>19</td>
<td>20</td>
<td>S</td>
</tr>
<tr>
<td>20</td>
<td>21</td>
<td>T</td>
</tr>
<tr>
<td>21</td>
<td>22</td>
<td>U</td>
</tr>
<tr>
<td>22</td>
<td>23</td>
<td>V</td>
</tr>
<tr>
<td>23</td>
<td>24</td>
<td>W</td>
</tr>
<tr>
<td>24</td>
<td>25</td>
<td>X</td>
</tr>
<tr>
<td>25</td>
<td>26</td>
<td>Y</td>
</tr>
<tr>
<td>26</td>
<td>27</td>
<td>Z</td>
</tr>
<tr>
<td>27</td>
<td>28</td>
<td>A</td>
</tr>
<tr>
<td>28</td>
<td>29</td>
<td>B</td>
</tr>
<tr>
<td>29</td>
<td>30</td>
<td>C</td>
</tr>
<tr>
<td>30</td>
<td>31</td>
<td>D</td>
</tr>
<tr>
<td>31</td>
<td>32</td>
<td>E</td>
</tr>
<tr>
<td>32</td>
<td>33</td>
<td>F</td>
</tr>
<tr>
<td>33</td>
<td>34</td>
<td>G</td>
</tr>
<tr>
<td>34</td>
<td>35</td>
<td>H</td>
</tr>
<tr>
<td>35</td>
<td>36</td>
<td>I</td>
</tr>
<tr>
<td>36</td>
<td>37</td>
<td>J</td>
</tr>
<tr>
<td>37</td>
<td>38</td>
<td>K</td>
</tr>
<tr>
<td>38</td>
<td>39</td>
<td>L</td>
</tr>
<tr>
<td>39</td>
<td>40</td>
<td>M</td>
</tr>
<tr>
<td>40</td>
<td>41</td>
<td>N</td>
</tr>
<tr>
<td>41</td>
<td>42</td>
<td>O</td>
</tr>
<tr>
<td>42</td>
<td>43</td>
<td>P</td>
</tr>
<tr>
<td>43</td>
<td>44</td>
<td>Q</td>
</tr>
<tr>
<td>44</td>
<td>45</td>
<td>R</td>
</tr>
<tr>
<td>45</td>
<td>46</td>
<td>S</td>
</tr>
<tr>
<td>46</td>
<td>47</td>
<td>T</td>
</tr>
<tr>
<td>47</td>
<td>48</td>
<td>U</td>
</tr>
<tr>
<td>48</td>
<td>49</td>
<td>V</td>
</tr>
<tr>
<td>49</td>
<td>50</td>
<td>W</td>
</tr>
<tr>
<td>50</td>
<td>51</td>
<td>X</td>
</tr>
<tr>
<td>51</td>
<td>52</td>
<td>Y</td>
</tr>
<tr>
<td>52</td>
<td>53</td>
<td>Z</td>
</tr>
</tbody>
</table>
THE DECODED MESSAGE IS:

ABC DEF GHI JKL MNC PQR STU VWX YZ ZY XHV UTF UQP UNM LKJ
HW1: EXPERIMENT A HAS N MUTUALLY EXCLUSIVE POSSIBLE OUTCOMES, \( A_1, A_2, ..., A_n \). EXPERIMENT B HAS N POSSIBLE MUTUALLY EXCLUSIVE OUTCOMES \( B_n \). SHOW THAT

\[
P[B_n | A_m] = \frac{P[A_m | B_n] P[B_n]}{\sum_{i=1}^{N} P[A_m | B_i] P[B_i]} \tag{1}
\]

WE SHALL MAKE USE OF

\[
P[A_i | B_j] = \frac{P[A_i, B_j]}{P[B_j]} \iff \text{MULTIPLICATIVE LAW OF PROBABILITY MEASURE} \tag{2}
\]

\[
P[A_i] = \sum_{B_j} P[A_i, B_j] \iff \text{MARGINAL DENSITY DEFN.} \tag{3}
\]

OUR EVENTS ARE:

\( A_i : A_1, A_2, ..., A_n \)

\( B_j : B_1, B_2, ..., B_n \)

START BY REWRITING (2) AS

\[
P[A_m | B_n] = \frac{P[A_m, B_n]}{P[A_m]} \tag{4}
\]

(2) MAY ALSO BE WRITTEN AS

\[
P[A_m, B_n] = P[A_m | B_n] P[B_n] \quad (= P[B_n, A_m]) \tag{5}
\]

FROM (3) WE WRITE

\[
P[A_m] = \sum_{j=1}^{N} P[A_m, B_j] \tag{5}
\]

WHICH, FROM (2) MAY BE WRITTEN

\[
P[A_m] = \sum_{j=1}^{N} P[A_m | B_j] P[B_j] \tag{6}
\]

SUBSTITUTING THIS AND (5) INTO (4) GIVES THE DESIRED RESULT:

\[
P[B_n | A_m] = \frac{\sum_{i=1}^{N} P[A_m | B_i] P[B_i]}{\sum_{i=1}^{N} P[A_m | B_i] P[B_i]} \tag{7}
\]
FROM TEXT

Generalize part 3 of (2-3) to n sources. We now have n sources the k'th of which is capable of generating \( q_k \) symbols:

\[
S_1 = \{ s_{11}, s_{12}, s_{13}, \ldots, s_{1q_1} \} \equiv q_1 \text{ symbols}
\]

\[
S_2 = \{ s_{21}, s_{22}, s_{23}, \ldots, s_{2q_2} \} \equiv q_2 \text{ symbols}
\]

\[
\vdots
\]

\[
S_k = \{ s_{k1}, s_{k2}, s_{k3}, \ldots, s_{kq_k} \} \equiv q_k \text{ symbols}
\]

\[
\vdots
\]

\[
S_n = \{ s_{n1}, s_{n2}, s_{n3}, \ldots, s_{nq_n} \} \equiv q_n \text{ symbols}
\]

Associate a probability \( p_{kj} = P[S_{kj}] \) with the \( j \) th element of the \( k \)th source such that

\[
\sum_{j=1}^{q_k} p_{kj} = 1 \quad \forall \ k
\]

The entropy of the \( k \)th source is

\[
H_K = - \sum_{j=1}^{q_k} p_{kj} \log p_{kj}
\]

Now consider the probability set \( \{ \lambda_k \} \) such that

\[
\sum_{k=1}^{n} \lambda_k = 1
\]

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

\[
\{ \lambda_1 p_{11}, \lambda_1 p_{12}, \lambda_1 p_{13}, \ldots, \lambda_1 p_{1q_1} \} \equiv \text{from } S_1
\]

\[
\lambda_2 p_{21}, \lambda_2 p_{22}, \ldots, \lambda_2 p_{2q_2}, \ldots \equiv \text{from } S_2
\]

\[
\vdots
\]

\[
\lambda_k p_{k1}, \lambda_k p_{k2}, \ldots, \lambda_k p_{kq_k}, \ldots \equiv \text{from } S_k
\]

\[
\lambda_n p_{n1}, \lambda_n p_{n2}, \ldots, \lambda_n p_{nq_n} \equiv \text{from } S_n
\]

(cont →)
THE ENTROPY, \( H[S(\lambda)] \), ASSOCIATED WITH THE MIXED SOURCE FOLLOWS AS

\[
H[S(\lambda)] = - \sum_{j_1=1}^{q_1} \lambda_1 p_{1j_1} \log \lambda_1 p_{1j_1} - \sum_{j_2=1}^{q_2} \lambda_2 p_{2j_2} \log \lambda_2 p_{2j_2} - \ldots - \sum_{j_k=1}^{q_k} \lambda_k p_{kj_k} \log \lambda_k p_{kj_k} - \ldots - \sum_{j_n=1}^{q_n} \lambda_n p_{nj_n} \log \lambda_n p_{nj_n}
\]

\[
= \sum_{k=1}^{n} \sum_{j_k=1}^{q_k} \lambda_k p_{kj_k} \log \lambda_k p_{kj_k}
\]

\[
= \sum_{k=1}^{n} \lambda_k \sum_{j_k=1}^{q_k} p_{kj_k} \left[ \log \lambda_k + \log p_{kj_k} \right]
\]

\[
= \sum_{k=1}^{n} \lambda_k \left( \sum_{j_k=1}^{q_k} p_{kj_k} \right) \log \lambda_k
- \sum_{k=1}^{n} \lambda_k \sum_{j_k=1}^{q_k} p_{kj_k} \log p_{kj_k}
\]

FROM (1) AND (2), THIS BECOMES:

\[
H[S(\lambda)] = - \sum_{k=1}^{n} \lambda_k \log \lambda_k + \sum_{k=1}^{n} \lambda_k H_K
\]

(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

\[
H[S(\lambda)] = H(\lambda) + \sum_{k=1}^{n} \lambda_k H_K
\]

(4)

THIS IS THE ENTROPY OF THE MIXED SOURCE. THE MIXED SOURCE IS FORMED BY COMBINING THE INDIVIDUAL SOURCES, THE KTH OF WHICH GENERATES A SYMBOL 100\% OF THE TIME. THAT IS, THE PROBABILITY THAT A SYMBOL GENERATED FROM THE MIXED SOURCE ORIGINATED FROM THE KTH COMPONENT SOURCE IS \( \lambda_k \).
(A SH)

The initial entropy is

\[ H = H(p_1, p_2, \ldots, p_M) = - \sum_{k=1}^{M} p_k \log p_k \]

\[ = p_1 \log p_1 - p_2 \log p_2 - \sum_{k=3}^{M} p_k \log p_k \]  \hspace{1cm} (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

\[ p_1' = p_1 - \Delta p \]

\[ p_2' = p_2 + \Delta p \]  \hspace{1cm} (2)

such that

\[ p_1 - \Delta p \geq p_2 + \Delta p \]  \hspace{1cm} (3)

where \( \Delta p > 0 \) must satisfy \( \Delta p < p_1 \).

The entropy of the perturbed source is

\[ H' = H(p_1', p_2', p_3, p_4, \ldots, p_M) \]

\[ = -p_1' \log p_1' - p_2' \log p_2' - \sum_{k=3}^{M} p_k \log p_k \]  \hspace{1cm} (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):

\[ H - H' = (p_1 - \Delta p) \log (p_1 - \Delta p) + (p_2 + \Delta p) \log (p_2 + \Delta p) \]

\[ - p_1 \log p_1 - p_2 \log p_2 \]

\[ = p_1 \log (p_1 - \Delta p) - \Delta p \log (p_1 - \Delta p) + p_2 \log (p_2 + \Delta p) + \Delta p \log (p_1 + \Delta p) \]

\[ - p_1 \log p_1 - p_2 \log p_2 \]  \hspace{1cm} \text{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_2 - \Delta p} \] (5)

Now, from the lemma (Lemma 1):
\[ \log x \leq 1 - x \]

Thus, we may write (5) as
\[ H - H' \leq p_1 \left[ 1 - (1 - \frac{\Delta p}{p_1}) \right] + p_2 \left[ 1 - (1 + \frac{\Delta p}{p_2}) \right] \]
+ \Delta p \log \frac{p_2 + \Delta p}{p_2 - \Delta p} \] (6)
\[ \leq (\Delta p) + (-\Delta p) + \Delta p \log \frac{p_2 + \Delta p}{p_2 - \Delta p} \]

Now, from (3)
\[ \frac{p_2 + \Delta p}{p_2 - \Delta p} \leq 1 \]

Since \( \log(x) \) is a monotonically increasing function, it follows that
\[ \log \frac{p_2 + \Delta p}{p_2 - \Delta p} \leq 0 \]

Furthermore, since \( \Delta p > 0 \), we may rewrite (6) as
\[ H - H' \leq 0 \]

or \[ H' \geq H \] (7)

which was to be proved. From (5), equality is achieved for \( \Delta p = 0 \).
It follows that, since both \( H' \) and \( H \) are convex, that \( H' \) is strictly greater than \( H \) for \( \Delta p \) strictly greater than zero. That is

\[ H' > H \] for \( \Delta p > 0 \)
(1.12) Given function $h(p) ; 0 < p \leq 1$.

$\exists h(p_1p_2) = h(p_1) + h(p_2)$

$h(p)$ is monotonically decreasing.

Show the only function satisfying this condition is $h(p) = -C \ln b p$.

$C > 0$

$b > 1$

Proof:

Define $H(p) = b^h(p)$

Where, $b > 0$

Also, define $G(p) = \frac{d}{dp} H(p)$

Plug (5) into (1):

$H(p_1p_2) = b^{h(p_1p_2)}$

$= b^{h(p_1) + h(p_2)}$

$= H(p_1)H(p_2)$

From (7):

$\frac{d}{dp_1} H(p_1p_2) = G(p_1p_2) \frac{d}{dp_1} (p_1p_2)$

$= p_2G(p_1p_2)$

$= H(p_2) \frac{d}{dp_1} H(p_1)$

Or

$G(p_1p_2) = \frac{1}{p_2} H(p_2) \frac{d}{dp_1} H(p_1)$

Similarly, we can show

$G(p_1p_2) = \frac{1}{p_1} H(p_1) \frac{d}{dp_2} H(p_2)$
COMBINING (5) \# (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, EQUVALENTLY

\[ \frac{p_1}{H(p_1)} \frac{d}{dp_1} H(p_1) = \frac{p_2}{H(p_2)} \frac{d}{dp_2} H(p_2) \]  \hspace{1cm} (10)

NOW, THE NUMBER TO WHICH THESE RELATIONS ARE EQUAL MUST BE INDEPENDENT OF BOTH \( p_1 \) AND \( p_2 \). AS SUCH, LET

\[ \frac{p}{H(p)} \frac{d}{dp} H(p) = c = \text{CONSTANT} \]

OR

\[ \frac{d}{dp} H(p) = \frac{c}{p} H(p) \]  \hspace{1cm} (11)

THE SOLN' TO THIS DIFFERENTIAL EQUATION IS UNIQUE:

\[ H(p) = e^{d \cdot p \cdot c'} \]

WHERE \( d = \text{REAL CONSTANT} \). FROM (5)

\[ d \cdot p \cdot c' = b \cdot H(p) \]

\[ \Rightarrow H(p) = \frac{b \cdot d \cdot p \cdot c'}{c} = \frac{b \cdot d}{c'} + C \cdot \frac{b}{c} \cdot p \]  \hspace{1cm} (12)

THIS RELATIONSHIP MUST SATISFY (1)

\[ \Rightarrow d = 1 \]

AND

\[ H(p) = c' \cdot \frac{b}{c} \cdot p \]  \hspace{1cm} (13)

THIS RELATION IS THE ONLY 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)

\[ \frac{d}{dp} h(p) = c' \frac{d}{dp} \log_b p = \frac{c'}{\log_b p} \frac{d}{dp} \log_b p \]
\[ = c' \log_b p \times \frac{1}{p} < 0 \]

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 (0).

\[ \log_b x = \log_b x / \log_b b = -\log_b x \]

NOTE: IN APPENDIX 2 OF SHANNON'S
PAPER, ANOTHER PROOF OF THIS
IS GIVEN, BUT, WITH DIFFERENT
INITIAL ASSUMPTIONS.
WE WISH TO FIND THE CHANNEL CAPACITY FOR
THE CHANNEL DESCRIBED BY

\[
p = \begin{bmatrix}
  \frac{3}{4} & \frac{1}{8} & \frac{1}{8} & 0 \\
  \frac{1}{8} & \frac{3}{4} & 0 & \frac{1}{8} \\
  \frac{1}{8} & \frac{1}{8} & \frac{3}{4} & 0 \\
  0 & 0 & \frac{1}{4} & \frac{3}{4}
\end{bmatrix}
\]

BEGIN BY FINDING

\[p^{-1} = C^T / |p|\]

WHERE C IS THE MATRIX OF COFACTORS.

LET'S FIND THESE FIRST:

\[C_{11} = \begin{bmatrix}
  \frac{3}{4} & 0 & \frac{1}{8} \\
  \frac{1}{8} & \frac{3}{4} & 0 \\
  0 & \frac{1}{4} & \frac{3}{4}
\end{bmatrix}
\]

\[= \begin{bmatrix}
  \left(\frac{3}{4}\right)^2 + \frac{1}{4} \left(\frac{1}{8}\right)^2 \\
  \frac{1}{4} \left(\frac{3}{4}\right)^2 + \frac{1}{4} \left(\frac{1}{8}\right)^2
\end{bmatrix} - \begin{bmatrix}0\end{bmatrix}
\]

\[= \frac{27}{64} + \frac{1}{256} = \frac{108 + 1}{256} = \frac{109}{256}
\]

\[C_{12} = \begin{bmatrix}
  \frac{1}{8} & 0 & \frac{1}{8} \\
  \frac{1}{8} & \frac{3}{4} & 0 \\
  0 & \frac{1}{4} & \frac{3}{4}
\end{bmatrix}
\]

\[= \begin{bmatrix}
  \left(\frac{3}{4}\right)^2 \frac{1}{8} + \frac{1}{4} \left(\frac{1}{8}\right)^2 \\
  \frac{1}{4} \left(\frac{3}{4}\right)^2 \frac{1}{8} + \frac{1}{4} \left(\frac{1}{8}\right)^2
\end{bmatrix} = 0
\]

\[= \begin{bmatrix}
  \frac{9}{128} + \frac{1}{256}
\end{bmatrix} = -\frac{19}{256}
\]

\[C_{13} = \begin{bmatrix}
  \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\
  \frac{1}{8} & \frac{1}{8} & 0 \\
  0 & 0 & \frac{3}{4}
\end{bmatrix}
\]

\[= \frac{3}{4} \left(\frac{1}{8}\right)^2 - \frac{1}{8} \left(\frac{3}{4}\right)^2
\]

\[= \frac{3}{256} - \frac{9}{128}
\]

\[= -\frac{15}{256}
\]
\[ C_{14} = \begin{bmatrix} \frac{1}{8} & \frac{3}{4} & 0 \\ \frac{1}{8} & \frac{1}{2} & \frac{3}{4} \\ 0 & 0 & \frac{1}{4} \end{bmatrix} \]

\[ = - \left[ \left( \frac{1}{8} \right)^2 \frac{1}{4} - \frac{1}{8} \cdot \frac{3}{4} \cdot \frac{1}{4} \right] \]

\[ = \left[ \frac{1}{256} - \frac{6}{256} \right] \]

\[ = \frac{5}{256} \]

\[ C_{21} = \begin{bmatrix} \frac{1}{8} & \frac{1}{2} & 0 \\ \frac{1}{8} & \frac{3}{4} & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} \end{bmatrix} \]

\[ = - \left[ \left( \frac{3}{4} \right)^2 \frac{1}{8} - \left( \frac{1}{8} \right)^2 \frac{3}{4} \right] \]

\[ = - \left[ \frac{9}{256} - \frac{3}{256} \right] \]

\[ = -\frac{15}{256} \]

\[ C_{22} = \begin{bmatrix} \frac{3}{4} & \frac{1}{8} & 0 \\ \frac{1}{8} & \frac{3}{4} & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} \end{bmatrix} \]

\[ = \left[ \left( \frac{3}{4} \right)^3 - \frac{3}{4} \left( \frac{1}{8} \right)^2 \right] \]

\[ = \frac{27}{64} - \frac{3}{256} \]

\[ = \frac{105}{256} \]

\[ C_{23} = \begin{bmatrix} \frac{3}{4} & \frac{1}{8} & 0 \\ \frac{1}{8} & \frac{1}{4} & 0 \\ 0 & 0 & \frac{3}{4} \end{bmatrix} \]

\[ = \left[ \left( \frac{3}{4} \right)^2 \frac{1}{8} - \left( \frac{1}{8} \right)^2 \frac{3}{4} \right] \]

\[ = - \left[ \frac{9}{128} - \frac{3}{256} \right] = -\frac{15}{256} \]

\[ C_{24} = \begin{bmatrix} \frac{3}{4} & \frac{1}{8} & \frac{1}{2} \\ \frac{1}{8} & \frac{1}{8} & \frac{3}{4} \\ 0 & 0 & \frac{1}{4} \end{bmatrix} \]

\[ = \left[ \frac{1}{4} \left( \frac{3}{4} \right)^2 \frac{1}{8} - \left( \frac{1}{8} \right)^2 \frac{3}{4} \right] \]

\[ = \frac{5}{256} \]
\[ C_{31} = \begin{bmatrix} \frac{1}{8} & \frac{1}{8} & 0 \\ \frac{3}{4} & 0 & \frac{1}{8} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{bmatrix} \]

\[ = - \left( \frac{1}{8} \right)^2 \frac{1}{4} - \frac{1}{8} \left( \frac{3}{4} \right)^2 \]

\[ - \frac{1}{256} - \frac{9}{128} \]

\[ C_{32} = \begin{bmatrix} \frac{3}{4} & \frac{1}{8} & 0 \\ \frac{1}{8} & 0 & \frac{1}{8} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{bmatrix} \]

\[ = - \left[ \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{8} - \left( \frac{1}{8} \right)^2 \left( \frac{3}{4} \right) \right] \]

\[ = \frac{3}{128} + \frac{3}{256} \]

\[ C_{33} = \begin{bmatrix} \frac{3}{4} & \frac{1}{8} & 0 \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ 0 & 0 & \frac{3}{4} \end{bmatrix} \]

\[ = \left[ \left( \frac{3}{4} \right)^3 - \left( \frac{1}{8} \right)^2 \left( \frac{3}{4} \right) \right] \]

\[ = \frac{27}{64} - \frac{3}{256} \]

\[ = \frac{105}{256} \]

\[ C_{34} = \begin{bmatrix} \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \]

\[ = \left[ \left( \frac{3}{4} \right)^2 \frac{1}{4} - \left( \frac{1}{8} \right)^2 \frac{1}{4} \right] \]

\[ = - \left[ \frac{19}{64} - \frac{1}{256} \right] \]

\[ = - \left[ \frac{35}{256} \right] \]
\[ C_{y_1} = \begin{bmatrix} \frac{1}{8} & \frac{1}{8} & 0 \\ \frac{3}{4} & 0 & \frac{1}{2} \\ \frac{1}{8} & \frac{3}{4} & 0 \end{bmatrix} \]

\[ = \begin{bmatrix} \left(\frac{1}{8}\right)^3 - \left(\frac{1}{8}\right)^2 \frac{3}{4} \\ \frac{1}{512} - \frac{6}{512} \end{bmatrix} \]

\[ = \frac{5}{512} \]

\[ C_{y_2} = \begin{bmatrix} \frac{3}{4} & \frac{1}{8} & 0 \\ \frac{1}{8} & 0 & \frac{1}{2} \\ \frac{1}{8} & \frac{3}{4} & 0 \end{bmatrix} \]

\[ = \begin{bmatrix} \left(\frac{1}{8}\right)^3 - \left(\frac{3}{4}\right)^2 \frac{1}{8} \\ \frac{1}{512} - \frac{9}{128} \\ \frac{1}{512} - \frac{36}{512} \end{bmatrix} \]

\[ = \frac{-35}{512} \]

\[ C_{y_3} = \begin{bmatrix} \frac{3}{4} & \frac{1}{8} & 0 \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & 0 \end{bmatrix} \]

\[ = \begin{bmatrix} \left(\frac{1}{8}\right)^3 - \left(\frac{3}{4}\right)^2 \frac{3}{4} \\ \frac{1}{512} - \frac{6}{512} \end{bmatrix} \]

\[ = \frac{5}{512} \]

\[ C_{y_4} = \begin{bmatrix} \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{4} & 0 \\ \frac{1}{8} & \frac{1}{8} & \frac{3}{4} \end{bmatrix} \]

\[ = \begin{bmatrix} \left(\frac{3}{4}\right)^3 + \left(\frac{1}{8}\right)^3 - \left(\frac{1}{8}\right)^2 \left(\frac{3}{4}\right) - \left(\frac{1}{8}\right)^2 \left(\frac{3}{4}\right) \end{bmatrix} \]

\[ = \frac{27}{64} + \frac{1}{512} - \frac{6}{256} \]

\[ = \frac{205}{512} \]
THUS

\[
C = \begin{bmatrix}
\frac{109}{256} & \frac{-19}{256} & \frac{-15}{256} & \frac{5}{256} \\
\frac{-19}{256} & \frac{105}{256} & \frac{-15}{256} & \frac{5}{256} \\
\frac{-19}{256} & \frac{9}{256} & \frac{105}{256} & \frac{-35}{256} \\
\frac{5}{512} & \frac{-35}{512} & \frac{5}{512} & \frac{205}{512}
\end{bmatrix}
\]

\[
CT = \begin{bmatrix}
\frac{109}{256} & \frac{-19}{256} & \frac{-15}{256} & \frac{5}{256} \\
\frac{-19}{256} & \frac{105}{256} & \frac{-15}{256} & \frac{5}{256} \\
\frac{-19}{256} & \frac{9}{256} & \frac{105}{256} & \frac{-35}{256} \\
\frac{5}{512} & \frac{-35}{512} & \frac{5}{512} & \frac{205}{512}
\end{bmatrix}
\]

\[
= \frac{1}{512} \begin{bmatrix}
218 & -30 & -38 & 5 \\
-38 & 210 & 18 & -35 \\
-38 & -30 & 210 & 5 \\
10 & 10 & -70 & 205
\end{bmatrix}
\]

IT REMAINS TO FIND THE DETERMINANT OF \( p \):

\[
|p| = \frac{1}{8} C_{24} + \frac{3}{4} C_{44}
\]

\[
= \frac{1}{8} \left( \frac{5}{256} \right) + \frac{3}{4} \left( \frac{205}{512} \right)
\]

\[
= \frac{5 + 615}{2048} = \frac{620}{2048} = \frac{155}{512}
\]

\[\therefore p^{-1} = \frac{1}{155} \begin{bmatrix}
218 & -30 & -38 & 5 \\
-38 & 210 & 18 & -35 \\
-38 & -30 & 210 & 5 \\
10 & 10 & -70 & 205
\end{bmatrix}
\]

CHECK

\[
p p^{-1} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]
NOW TO FIND THE ENTROPIES OF EACH ROW OF $p$:

\[
H_1 = H\left(\frac{3}{4}, \frac{1}{8}, \frac{1}{8}\right) = 1.061278 \text{ BITS} \tag{10}
\]
\[
H_2 = H\left(\frac{1}{8}, \frac{3}{4}, \frac{1}{8}\right) = 1.061278 \text{ BITS}
\]
\[
H_3 = H\left(\frac{1}{8}, \frac{1}{8}, \frac{3}{4}\right) = 1.061278 \text{ BITS}
\]
\[
H_4 = H\left(\frac{3}{4}, \frac{3}{4}\right) = 0.8112781 \text{ BITS} \tag{11}
\]

\[
\begin{bmatrix}
Q_1 \\
Q_2 \\
Q_3 \\
Q_4
\end{bmatrix}
= [\rho^{-1}]
\begin{bmatrix}
-H_1 \\
-H_2 \\
-H_3 \\
-H_4
\end{bmatrix}
\]

\[
= \frac{-1}{155}
\begin{bmatrix}
218 & -30 & -38 & 5 \\
-38 & 210 & 18 & -35 \\
-30 & -30 & 210 & 5 \\
10 & 10 & -70 & 205
\end{bmatrix}
\begin{bmatrix}
1H_1 \\
+H_2 \\
+H_3 \\
+H_4
\end{bmatrix}
\]

\[
= \begin{bmatrix}
-1.0532 \\
-1.1178 \\
-1.0532 \\
-0.7306
\end{bmatrix} \tag{12}
\]
\[
\text{THUS}
\]
\[
C = \log_2 2^{\Phi_1} + 2^{\Phi_2} + 2^{\Phi_3} + 2^{\Phi_4}
\]
\[
= \log_2 2 \times 2^{\Phi_1} + 2^{\Phi_2} + 2^{\Phi_4}
\]
\[
= \log_2 2.02724
\]
\[
= 1.01952 \text{ bits}
\]
WE WISH TO FIND THE MAXIMUM TRANSFORMATION WITH THE FOLLOWING CONDITIONAL MATRICES:

(a) \[
\begin{bmatrix}
0.9 & 0.1 \\
0.1 & 0.9
\end{bmatrix}
\]

(b) \[
\begin{bmatrix}
0.9 & 0.1 \\
0.2 & 0.8
\end{bmatrix}
\]

THESE MATRICES CORRESPOND TO

WE SHALL FIRST CONSTRUCT A GRAPH OF THE ENTROPY FUNCTION

\[H(p) = p \log_2 \frac{1}{p} + \bar{p} \log_2 \frac{1}{\bar{p}}\]

WHERE \(\bar{p} = 1 - p\)

---

**HP-25 PROGRAMS**

\(\text{STO } 0 \quad \text{RCL } 0\)

\(1 \quad \frac{1}{x}\)

\(- \quad \ln\)

\(\text{CHS} \quad 2\)

\(\text{STO } 1 \quad \ln\)

\(\frac{1}{x} \quad \div\)

\(\ln \quad \text{RCL } 0\)

\(2 \quad x\)

\(\ln \quad +\)

\(\text{GTO } 00\)

\(\text{RCL } 1\)

\(x\)
a. To find the channel capacity, it remains to follow the instructions on handout #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:
\[
p[y=0] = 0.45 \quad p[y=1] = 0.55
\]
and \( C = 0.40 \) bits

(cont) \( \Rightarrow \)
FINDING THE SOURCE PROBS.

Let $q = P(X = 0)$. Then

\[
\begin{align*}
P(Y = 0) &= P(Y = 0 | X = 0)P(X = 0) + P(Y = 0 | X = 1)P(X = 1) \\
\Rightarrow P(Y = 0) &= P(Y = 0 | X = 0)P(X = 0) + P(Y = 0 | X = 1)P(X = 1) \\
\Rightarrow P(Y = 0) &= p_{11}q + p_{21}q \\
\Rightarrow P(Y = 0) &= qP_{11} + qP_{21}
\end{align*}
\]

\[
\begin{align*}
P(Y = 1) &= P(Y = 1 | X = 0)P(X = 0) + P(Y = 1 | X = 1)P(X = 1) \\
\Rightarrow P(Y = 1) &= P(Y = 1 | X = 0)P(X = 0) + P(Y = 1 | X = 1)P(X = 1) \\
\Rightarrow P(Y = 1) &= p_{11}q + p_{21}q \\
\Rightarrow P(Y = 1) &= qP_{11} + qP_{21}
\end{align*}
\]

But $p_{11} + p_{12} = 1 \Rightarrow p_{12} = 1 - p_{11}$

And $p_{21} + p_{22} = 1 \Rightarrow p_{21} = 1 - p_{22}$

\[
\Rightarrow \begin{cases}
P = p_{11}q + (1 - p_{22})q \quad \text{(1)} \\
P = (1 - p_{11})q + p_{22}q \quad \text{(2)}
\end{cases}
\]

Let's see if these equations are consistent. (1) becomes

\[
\begin{align*}
\overline{P} &= 1 - p_{11}q - (1 - p_{22})q \\
\overline{P} &= 1 - p_{11}q - (1 - p_{22})(1 - q) \\
\overline{P} &= 1 - p_{11}q - (1 - p_{22} - q + q p_{22}) \\
\overline{P} &= 1 - p_{11}q - 1 + p_{22} + q - q p_{22} \\
\overline{P} &= (1 - p_{11})q + p_{22}(1 - q) \\
\overline{P} &= (1 - p_{11})q + p_{22}q
\end{align*}
\]

Which is the same as (2). Thus, (1) and (2) are consistent (in equivalent).

Let's solve for $q$ in (1)

\[
\begin{align*}
P &= p_{11}q + (1 - p_{22})(1 - q) \\
P &= p_{11}q + 1 - p_{22} - q + q p_{22} \\
P &= q(p_{11} - 1 + p_{22}) + 1 - p_{22} \\
\Rightarrow q(p_{11} - 1 + p_{22}) &= P - 1 + P_{22}
\end{align*}
\]

\[
q = \frac{P - 1 + P_{22}}{P_{11} - 1 + P_{22}}
\]
FOR MATRIX (a), WE HAVE
\[ p = \frac{1}{2}, \quad p_{11} = p_{22} = 0.9 \]
\[ \therefore q = \frac{(0.5 - 1 + 0.9)}{(0.9 - 1 + 0.9)} \]
\[ = 0.500 \]
\[ \bar{q} = 0.500 \]

FOR MATRIX (b)
\[ p = 0.45, \quad p_{11} = 0.9, \quad p_{22} = 0.8 \]
\[ \therefore q = \frac{(0.45 - 1 + 0.8)}{(0.9 - 1 + 0.8)} \]
\[ = 0.357 \]
\[ \bar{q} = 0.643 \]
**Binary Multiplicative Channel**

\[
\begin{array}{c}
a \\
b = ac \\
c
\end{array}
\]

We assume the binary (multiplication) operations

\[
0 \times 0 = 0 \times 1 = 0 \times 0 , \quad 1 \times 1
\]

b. Since the BMC is assumed noiseless, the conditional probability matrix,

\[
P[b/a,c],
\]

is

\[
\begin{array}{c|cc}
& 0 & 1 \\
\hline
0 & 0 & 1 \\
1 & 1 & 0 \\
\end{array}
\]

We have here specified the binary output alphabet \( b = \{0, 1\} \).

b. We now wish to develop an expression for the transformation 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 \( a \) and \( c \) are furthermore assumed to be statistically independent so that:

\[
P(a=0, c=1) = P(a=1, c=0) = pq
\]

\[
P(a=0, c=0) = p \bar{q}
\]

\[
P(a=1, c=1) = q \bar{p}
\]

*Following notation in Prob. 5-9 of Text, A is here assumed to be comprised of the (not) events \( a \neq c \).
IT FOLLOWS THAT

\[-H(A) = 2pq \log pq + p^2 \log p^2 + q^2 \log q^2 \]  

FROM THE INPUT PROBABILITIES AND THE CONDITIONAL PROB. MATRIX, WE WRITE \( P(A;B) = P(B|A) P(A) \):

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>( p^2 )</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>( pq )</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>( pq )</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>( q^2 )</td>
</tr>
</tbody>
</table>

Thus

\[-H(A;B) = p^2 \log p^2 + q^2 \log q^2 + 2pq \log pq = -H(A) \]  

Since \( p(B) = \sum_A p(A;B) \), we have

\[
\begin{array}{c|c}
  b & p(b) \\
  \hline
  0 & p^2 + 2pq \\
  1 & q^2 \\
\end{array}
\]

And

\[-H(B) = (p^2 + 2pq) \log (p^2 + 2pq) + q^2 \log q^2 \]
NOW
\[ I(A; B) = H(A) + H(B) - H(A; B) \]

BUT, FROM (2), \( H(A; B) = H(A) \). THUS
\[ I(A; B) = H(B) \]
\[ = -(p^2 + 2pq) \log (p^2 + 2pq) \]
\[ - q^2 \log q^2 \]  \hspace{1cm} (4)

WHERE \( q = 1 - p \). SIMPLIFYING;
\[ I(A; B) = -[p^2 + 2p(1-p)] \log [p^2 + 2p(1-p)] \]
\[ - (1-p)^2 \log (1-p)^2 \]
\[ = -[p^2 + 2p - 2p^2] \log [p^2 + 2p - 2p^2] \]
\[ - 2(1-p)^2 \log (1-p) \]
\[ = (p^2 - 2p) \log (2p - p^2) - 2(1-p)^2 \log (1-p) \]
\[ = -p(2-p) \log p(2-p) - 2(1-p)^2 \log (1-p) \]  \hspace{1cm} (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
\[ p^2 + 2pq = q^2 \]  \hspace{1cm} \[\rightarrow\]
SOLVING FOR P:
\[ p^2 + 2p(1-p) - (1-p)^2 = 0 \]
\[ p^2 + 2p - 2p^2 - (1-2p+p^2) = 0 \]
\[ -p^2 + 2p - 1 + 2p - p^2 = 0 \]
\[ -2p^2 + 4p - 1 = 0 \]

\[ p = \frac{-4 \pm \sqrt{16-8}}{-4} \]
\[ p = 1 \pm \sqrt{\frac{1}{4}} \]
\[ p = 1 \pm \frac{1}{2} \]
\[ p = 1 \pm \frac{1}{2} \]

SINCE 0 < P < 1, WE HAVE
\[ p = 1 - \sqrt{\frac{1}{2}} \approx 0.293 \]
\[ q = 1 - p = \frac{1}{2} \approx 0.707 \]

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."
Our source and associated prob.'s are
\[ S = \left\{ s_1, s_2, s_3, s_4, s_5, s_6, s_7 \right\} \]
\[ \frac{1}{3}, \frac{1}{3}, \frac{1}{9}, \frac{1}{9}, \frac{1}{27}, \frac{1}{27}, \frac{1}{27} \]

2. Let's first compute \( H_3(s) \), since it's the easiest:
\[
H_3(s) = \frac{1}{3} \log_3 3 + \frac{1}{3} \log_3 3 + \frac{1}{4} \log_3 9 + \frac{1}{4} \log_3 9 \\
+ \frac{1}{27} \log_3 27 + \frac{1}{27} \log_3 27 + \frac{1}{27} \log_3 27 \\
= \frac{2}{3} \log_3 3 + \frac{2}{4} \log_3 3 + \frac{3}{27} \log_3 3 \\
= \frac{2}{3} (1) + \frac{2}{9} (2) + \frac{1}{9} (3) \\
= \frac{13}{9}
\]

Also, since all probabilities are of the form \( 3^{-l} \), we can expect a compact code here.

Now
\[ H_2(s) = \frac{H_3(s)}{\log_3 2} \]

But
\[ \log_3 2 = \frac{\ln 2}{\ln 3} \]
\[ \Rightarrow H_2(s) = \frac{13}{9} \cdot \frac{\ln 3}{\ln 2} = 2.289390279 \]
(bc) For $X = \{0, 1, \frac{2}{3}\}$ (Huffman Coding):

\[
\begin{align*}
S_1 & \quad 1/3 \quad 00 \quad 1/3 \quad 00 \quad 1/3 \quad 00 \quad 1/3 \quad 1 \quad 2/3 \quad 0 \\
S_2 & \quad 1/3 \quad 01 \quad 1/3 \quad 01 \quad 1/3 \quad 01 \quad 1/3 \quad 10 \quad 2/3 \quad 3 \\
S_3 & \quad 1/9 \quad 100 \quad 1/9 \quad 100 \quad 1/9 \quad 11 \quad 2/9 \quad 10 \quad 3/9 \quad 101 \\
S_4 & \quad 1/9 \quad 101 \quad 1/9 \quad 101 \quad 1/9 \quad 110 \quad 2/9 \quad 11 \\
S_5 & \quad 1/27 \quad 111 \quad 2/27 \quad 110 \quad 1/27 \quad 101 \\
S_6 & \quad 1/27 \quad 1100 \quad 1/27 \quad 111 \\
S_7 & \quad 1/27 \quad 1101 \\
\end{align*}
\]

\[
\bar{L} = 4 \cdot \frac{1}{3} + 6 \cdot \frac{1}{9} + 11 \cdot \frac{1}{27} \\
\text{\centering \\
\begin{align*}
&= \frac{1}{27} \left[ 36 + 18 + 11 \right] \\
&= \frac{65}{27} \approx 2.40740741 > H_2(S) \text{ as expected.}
\end{align*}
\]

For $X = \{0, 1, 2, 3\}$, we may write $p_2 = 2^{-\frac{2}{2}}$

\[
\Rightarrow \bar{L}_1 = \bar{L}_2 = 1 \quad \bar{L}_3 = 2 \quad \bar{L}_4 = \bar{L}_5 = \bar{L}_6 = 3
\]

A code (instant), obeying the Prefix Property, is

\[
\begin{align*}
S_1 & \quad 1/3 \quad 0 \\
S_2 & \quad 1/3 \quad 1 \\
S_3 & \quad 1/9 \quad 20 \\
S_4 & \quad 1/9 \quad 21 \\
S_5 & \quad 1/27 \quad 220 \\
S_6 & \quad 1/27 \quad 221 \\
S_7 & \quad 1/27 \quad 222
\end{align*}
\]

Due to Probability Structure, we are assured that this code is compact. Now,

\[
\bar{L} = 2(\frac{1}{3}) + 4(\frac{1}{9}) + 9(\frac{1}{27})
\]

\[
\text{\centering \\
\begin{align*}
&= \frac{2}{3} + \frac{4}{9} + \frac{9}{27} \\
&= \frac{1}{3} \left[ 6 + 4 + 3 \right] \\
&= 13/9 = H_3(S) \text{ as expected}
\end{align*}
\]
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. LET'S BEGIN BY ESTABLISHING NOMENCLATURE. WE ASSUME Q SOURCE SYMBOLS THE I^TH OF WHICH HAS PROBABILITY P\_i;

\[ S = \{ s_1, s_2, \ldots, s_i, \ldots, s_q \} \]

\[ \{ p_1, p_2, \ldots, p_i, \ldots, p_q \} \]

WITHOUT LOSS OF GENERALITY, LET

\[ p_1 \geq p_2 \geq p_3 \geq \ldots \geq p_i \geq \ldots \geq p_q \]

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 $C_j$:

$$
\bar{L}_j = \sum_{k=1}^{q-j} \ell_k p_k^{(j)}
$$

Without loss of generality, assume that

$$p_1^{(k)} \geq p_2^{(k)} \geq \ldots \geq p_i^{(k)} \geq \ldots \geq p_{q-k}^{(k)} \quad \forall \ k$$

Thus, following the Huffman coding procedure, we will take $p_j^{(j)}$ and $p_{q-j}^{(j)}$ (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:

$$
\bar{L}_{j-1} = \sum_{k=1}^{q-j+1} \ell_k p_k^{(j-1)} p_{q-j-1}^{(j-1)}
$$

$$
= \bar{L}_j + p_{q-j}^{(j)} \times (1. \text{Binit}) + p_{q-j-1}^{(j)} \times (1. \text{Binit})
$$

$$
= \bar{L}_j + p_{q-j}^{(j)} + p_{q-j-1}^{(j)}
$$

(3)
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 $L_j$ is, we will assume the contrary and show a contradiction. Assume $\exists$ a code $C_{j-1}$ with average length $\bar{L}_{j-1} < L_{j-1}$.

Assume that the code consists of word lengths $\bar{L}_1, \bar{L}_2, \ldots, \bar{L}_{q-j+1}$.

We now argue that one of the code words, say corresponding to $\bar{L}_i$, must differ from another (adjacent) code word, say corresponding to $\bar{L}_{i+1}$, only in the last digit. Otherwise, we could drop the last digits of each word, retain the prefix property, and reduce $\bar{L}_{j-1}$. (i.e., $C_{j-1}$ would not be compact). It follows, then, that we can construct a code $\tilde{C}_j$ by combining the code words corresponding to $\bar{L}_i$ and $\bar{L}_{i+1}$ (i.e., keeping their $\bar{L}_{i-1} = \bar{L}_{i+1} - 1$ first binit which are equivalent) and keeping all other code words the same. Then, if $\bar{L}_j$ is the average word length of
CODE $C_j$, THEN

$$\tilde{L}_j = \tilde{L}_{j-1} - \tilde{P}_j - \tilde{P}_{j-1}$$

WHERE $\tilde{P}_j$ AND $\tilde{P}_{j-1}$ CORRESPOND TO THE
SYMBOLS ASSOCIATED WITH WORD
LENGTHS $\tilde{L}_j$ AND $\tilde{L}_{j-1}$. WE MAY
REWRITE THIS AS

$$\tilde{L}_{j-1} = \tilde{L}_j + \tilde{P}_j + \tilde{P}_{j+1}$$

BUT WE HAVE ESTABLISHED, FROM
CONSIDERATIONS PREVIOUS TO
(2) THAT

$$\min (\tilde{P}_j + \tilde{P}_{j+1}) = \rho(j) + \rho(q-j)$$

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 CONTRADICTION TO
OUR ASSUMPTION THAT $\tilde{L}_{j-1} < L_{j-1}$
AND THE PROOF IS COMPLETE.
In writing these "trees", we shall use the idea of 
words... per level.
For example:

```
  +-- LEVEL 3 (2 words)
     +-- LEVEL 2 (1 word)
         +-- LEVEL 1 (3 words)
```

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

```
  +-- 2 words
     +-- 1 word
         +-- 3 words
```

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 its 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 a three points.

Also, note that 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 two words in the lowest level tree diagram. This is somewhat intuitive.
(2) \( r = 3 \quad q = 8 \)

**ONE WORD IN TOP LEVEL**

```
(1)
(5)
(2)
```

**TWO WORDS IN TOP LEVEL**

```
(2)  (2)
(1)  (5)
```

```
(2)  (2)
(2)  (2)
(2)  (2)
```

**ZERO WORDS AT TOP LEVEL**

```
(0)  (8)
```

Looks like, for \( r = 3 \), \( q = 8 \), there's a possible compact codes.
(b) \( r = 3 \quad q = 9 \)

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

\[ H(s) \leq \ell \leq H(s) - 2p_{\text{min}} + 1 \]  

We will have to assume binary coding. We have already established (in Sec. 4-1 of text) that

\[ H(s) \leq \ell \]  

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

<table>
<thead>
<tr>
<th>S</th>
<th>S'</th>
<th>( q=2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_1 )</td>
<td>( p_1' )</td>
<td>( p_1'' )</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( p_2 )</td>
<td>( p_2' )</td>
</tr>
<tr>
<td>( p_3 )</td>
<td>( p_3' )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( p_{q-1} )</td>
<td>( p_{q-1}' )</td>
</tr>
<tr>
<td>( p_q )</td>
<td>( p_q' )</td>
<td></td>
</tr>
</tbody>
</table>

The code associated with \( S \) is \( C_0 \), denote the symbols in \( S \| S' \) by \( \{s_1^1, s_2^1, \ldots, s_{q-1}^1, s_q^1, s_1^2, s_2^2, \ldots, s_{q-1}^2, s_q^2 \} \) and the corresponding word lengths by \( l_1, l_2, \ldots, l_q \).
Also, \( p_{i+1} \leq p_i \) and \( p_{i+1}' \leq p_i' \). 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 \( l_i' = l_q - 1 = l_{q-1} \). (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 \( q = 2 \) (i.e. \( S = \{S_1, S_2\} \)), we write:

\[
0 \leq H(S) \leq 1 \text{ (bits)}
\]

\[
\frac{1}{2} \leq p_{\text{min}} \leq 1 \text{ binit (5)}
\]

(1) is obviously satisfied for these values.

To complete the proof, we will assume (1) is true for some Huffman derived code, \( C' \), then as a consequence, it is true for code \( C \). Let us denote the symbols for \( C' \) by

\[
S' = \{s_{i_1}', s_{i_2}', \ldots, s_{i_q}'\}
\]

Again, their probs are arranged as

\[
p_{i_1}' \geq p_{i_2}' \geq \ldots \geq p_{i_{q-1}}' \geq p_{i_q}'
\]
AND C BY

\[ S = (s_1, s_2, \ldots, s_q) \quad (8) \]
\[ p_1 \geq p_2 \geq \ldots \geq p_{q-2} \geq p_{q-1} \geq p_q \quad (9) \]

NOW, IF THE ENTROPY OF CODE C IS \( H \) AND THE ENTROPY OF \( C' \) IS \( H' \), THEN, FOLLOWING HUFFMAN CODING PROCEDURE AND UTILIZING THE "ADDITIONAL PROPERTY OF ENTROPY" GIVES

\[ H = H' + (p_{q-1} + p_q) H_{q,q-1} \quad (10) \]

WHERE

\[ H_{q,q-1} = \frac{1}{H} \left( \frac{p_{q-1}}{p_q + p_{q-1}} \right) \]
\[ = \frac{p_{q-1}}{p_q + p_{q-1}} \log \frac{p_q + p_{q-1}}{p_{q-1}} + \frac{p_q}{p_q + p_{q-1}} \log \frac{p_q + p_{q-1}}{p_q} \quad (11) \]

WE MAY REWRITE (10) AS

\[ H = H' + (p_{q-1} + p_q) H_{q,q-1} \quad (12) \]
Following the spirit of inductive proof, we assume (1) is true for code \( C' \). That is

\[ H' \leq L' \leq H' + 1 - 2p'_{\text{min}} \tag{13} \]

where \( L' \) is the average word length associated with \( C' \) and

\[ p'_{\text{min}} = \min(p'_{q-1}, p'_{q-2}) = p'_{q-1} \tag{14} \]

If \( L \) is the average word length associated with \( C \), then (from Eq 4-25)

\[ L = L' + p_q + p_{q-1} \tag{15} \]

Substituting this into (13)

\[ H' \leq L - (p_q + p_{q-1}) \leq H' + 1 - 2p'_{\text{min}} \tag{16} \]

Substituting (10):

\[ H' \leq L - (p_q + p_{q-1}) \leq H - (p_{q-1} + p_q)H_{q-1} + 1 - 2p'_{\text{min}} \tag{17} \]

We know (from Sec 4-1) that

\[ L \geq H \tag{16} \]
SO THAT (17) MAY BE WRITTEN AS

\[ H \leq L \leq H \cdot (p_{q-1} + p_q) H_{q,q-1} + (p_q + p_{q-1}) + 1 - 2p'_{\min} \]

OR

\[ H \leq L \leq H + \left(1 - H_{q,q-1}\right) \left(p_{q-1} + p_q\right) + 1 - 2p'_{\min} \]  \hspace{1cm} (19)

DEFINE

\[ p_{\min} = \min \left[ p_{q-1}, p_q \right] = p_q \]  \hspace{1cm} (20)

OBVIOUSLY

\[ 2p_{\min} \leq p_{q-1} + p_q \]  \hspace{1cm} (21)

SINCE

\[ 1 - H_{q,q-1} \geq 0 \]  \hspace{1cm} (22)

WE CAN REWRITE (19) AS

\[ H \leq L \leq H + \left(1 - H_{q,q-1}\right) 2p_{\min} + 1 - 2p'_{\min} \]  \hspace{1cm} (22a)

NOW, IN ORDER TO PROVE (1), WE MUST SHOW THAT, WITH REFERENCE TO (22b)

\[ H + \left(1 - H_{q,q-1}\right) 2p_{\min} + 1 - 2p'_{\min} \geq H + 1 - 2p_{\min} \]

OR, EQUIVALENTLY

\[ 2 \left(2 - H_{q,q-1}\right) p_{\min} - 2p'_{\min} \geq 0 \]  \hspace{1cm} (23)
FROM HUFFMAN PROCEDURE, AND DUE TO THE PROBABILITY ORDERING, IT IS OBVIOUS THAT

\[ p_{\text{min}} \geq p_{\text{min}} \]  \hspace{1cm} (24)

NOW THE MAXIMUM VALUE THE L.H.S. OF (23) CAN ACHIEVE IS WHEN \( H_q, q-1 = 0 \) AT WHICH TIME IT IS EQUAL TO

\[ 2p_{\text{min}} - 2p_{\text{min}}' \]

OBVIOUSLY, \( p_{\text{min}}' \geq p_{\text{min}} \). Thus

\[ 2p_{\text{min}} - 2p_{\text{min}}' \leq 0 \]  \hspace{1cm} (25)

THIS COMPLETES THE PROOF.

IN SUMMARY, WE HAVE SHOWN THAT (23) AND (22) ARE TRUE. SINCE WE HAVE PROVEN \( \text{I} \) FOR \( C_2 \) \((\leq q = 2)\) ASSUMED IF FOR \( C' = C_{q-1} \) AND SHOWN AS A CONSEQUENCE, THAT \( \text{I} \) IS TRUE FOR \( C = C_q \). IT FOLLOWS THAT \( \text{I} \) 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):

\[ H + (1 - H_{q,q-1}) 2^{p_{\min}} + 1 - 2^{p_{\min}} \leq H + 1 - 2^{p_{\min}} \]

THEN, IT FollowS FROM (22b) THAT

\[ H \leq l \leq H + 1 - 2^{p_{\min}} \hspace{1cm} (26) \]

AND THE PROOF IS COMPLETE.

IN SUMMARY, WE HAVE SHOWN 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.
We wish to code 7-4-1776 ala Hamming using single error correcting code now.

\[(1776)_{10} = (11011110000)_{2} \Rightarrow 11 \text{ bits}\]

We will separately encode 7, 4, and 1776 (as instructed) using \(m=11\) for each. (It should be noted that this is not an optimal coding scheme for coding a date in terms of word length).

The expansion of 7 is

\[000000000111\]

For \(m=11\), \(k=4\). Thus, set up the table:

\[
\begin{array}{cccccccccccc}
15 & 14 & 13 & 12 & 11 & 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
\]

We'll use even 1 bit parity. Now,

\[
P_{0} + 1 + 0 + 0 + \cdots + 0 + 0 = 0 \Rightarrow P_{0} = 0
\]
\[
P_{1} + 1 + 0 + 0 + \cdots + 0 + 0 = 0 \Rightarrow P_{1} = 0
\]
\[
P_{2} + 1 + 0 + 0 + 0 + 0 + 0 + 0 = 0 \Rightarrow P_{2} = 0
\]
\[
P_{3} + 0 \cdots + 0 = 0 \Rightarrow P_{3} = 0
\]
\[ \begin{align*}
\text{FOR } (4)_0 &= 100 \\
\begin{array}{cccccccc}
15 & 14 & 13 & 12 & 11 & 10 & 9 & 8 \\
| & | & | & | & | & | & |
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}
\quad p_3 \quad p_2 \quad p_1 \quad p_0
\end{align*} \]

\[ \begin{align*}
p_0 \oplus 0 \oplus 0 \oplus 0 \oplus 0 &= 0 & \Rightarrow p_0 &= 0 \\
p_1 \oplus 0 \oplus 1 \oplus 0 \oplus 0 &= 0 & \Rightarrow p_1 &= 1 \\
p_2 \oplus 0 \oplus 1 \oplus 0 \oplus 0 &= 0 & \Rightarrow p_2 &= 1 \\
p_3 \oplus 0 &= 0
\end{align*} \]

\[ \begin{align*}
\text{FOR } (1776)_{12} &= 110111100000 \\
\begin{array}{cccccccc}
15 & 14 & 13 & 12 & 11 & 10 & 9 & 8 \\
| & | & | & | & | & | & |
\hline
1 & 1 & 0 & 1 & 1 & 1 & 0 & 0
\end{array}
\quad p_3 \quad p_2 \quad p_1 \quad p_0
\end{align*} \]

\[ \begin{align*}
p_0 \oplus 0 \oplus 0 \oplus 0 \oplus 0 \oplus 0 \oplus 1 \oplus 0 \oplus 0 \oplus 0 &= 0 & \Rightarrow p_0 &= 1 \\
p_1 \oplus 0 \oplus 0 \oplus 0 \oplus 0 \oplus 1 \oplus 1 \oplus 0 \oplus 0 \oplus 0 &= 0 & \Rightarrow p_1 &= 0 \\
p_2 \oplus 0 \oplus 0 \oplus 0 \oplus 0 \oplus 1 \oplus 0 \oplus 0 \oplus 1 \oplus 0 &= 0 & \Rightarrow p_2 &= 1 \\
p_3 \oplus 1 \oplus 1 \oplus 1 \oplus 0 \oplus 0 \oplus 1 \oplus 0 \oplus 0 &= 0 & \Rightarrow p_3 &= 0
\end{align*} \]

\[ \text{STRINGING THESE TOGETHER GIVES} \]

\[ \begin{align*}
000000000011010000000000000000001010101101111000001011
\end{align*} \]
I HAVE READ
"MATHEMATICAL THEORY
OF COMMUNICATION"
PARTS I THRU V
by CLAUDE SHANNON

[Signature]
COMPUTER PROGRAM
DONE WITH WES REDUS
Given the Fano Bound

\[ H(X/Y) \leq H[p(e), 1-p(e)] + p(e) \log (S-1) \]

and the fact

\[ I(X_i, X_n | Y_1, Y_2, \ldots, Y_n) \leq \sum_{i=1}^{n} I(X_i/Y_i) \]

with equality iff \( Y_i \)'s are independent.

prove that,

1. \( p(e) \), the prob. of error satisfies the

relation

\[ \log S \leq \frac{nC + \log 2}{1 - p(e)} \]

2. if \( S \geq 2^{n(C+\delta)} \), \( \delta > 0 \), then

\[ \frac{1}{p(e)} \geq 1 - \frac{C + \frac{1}{n}}{\delta} \rightarrow 1 - \frac{C}{e + \delta} > 0 \]
where \( C \) = channel capacity
\( R \) is a + number being the source uncertainty
\( \overline{p(x)} \) = 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 \( \left[ \left[ 2^{nR} \right], n \right] \) can have an average probability which \( \rightarrow 0 \) as \( n \rightarrow \infty \); hence, no sequence of codes \( \left[ \left[ 2^{nR} \right], n, \lambda_n \right] \) can exist

d with \( \lim\limits_{n \rightarrow \infty} \lambda_n = 0 \)

**Note:** The code designation \( [A, B, \xi] \) means that

- \( A \) is the no. of input \( n \)-sequences number
- \( B \) is the number \( n \), the extension of primary \((0,1)\) alphabet
- \( \xi \) is the max. prob. of error (like \( \epsilon \)
7-13-76 (TUES)

HOUSEKEEPING

J.C. PRABHUDAR, EE104 (AFTERNOON)

5375 - INFORMATION THEORY

TEXT: INFORMATION THEORY & CODING

by NORMAN ABRAMSON (U.O.C. HAWAII)

McGRAW HILL, 1963

OTHER REFERENCES

1. "AN INTRODUCTION TO INFO. THEORY" (RECOMMEND)

by F.M. REZER (McGRAW HILL, 1968)

2. "INFORMATION THEORY" (EXCELLENT)

by ROBERT ASH (INTERSCIENCE, WILEY)

3. "INFORMATION AND CODING THEORY" (FAIRLY GOOD)

by F.M. INCELS (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
   5. "Modern" Coding Schemes
     - Coding Theorems
     - Parity Checking Codes
     - Group Codes
     - Bose-Chaudhury Coding Schemes

(Reading Assignment)
- The Mathematics of Codes.

Grading:
- 3 Quizes (Class Time: 2 hrs) (60%)
- 10-15 Homework Starred Problems (40%)
- Possibly a Final

Also:
- Suggested Homework Problems.
7-19-76

READ THIS PAPER:

"THE MATHEMATICAL THEORY OF COMMUNICATION"

by CLAIRE SHANNON & W. WEAVER

IN BELL SYSTEM TECHNICAL JOURNAL

27 379-423, 623-656 (1948)

ALSO READ BEFORE 7-19-76 (NON) PGS. 1-10 TEXT

I. PROBABILITY THEORY: SOME STATEMENTS AND RELATIONS

(2) FREQUENCY OF EVENTS APPROACH:

FORM A BASIC EXPERIMENT.

E.G. PULLING OUT A SPADE FROM A PACK OF 52

IN GENERAL, LET \( n(x_k) = \text{NUMBER} \)

OF TIMES AN EVENT \( x_k \)

occurs \( \frac{n(x_k)}{N} = \text{TOTAL \# OF EVENTS} \).

THEN, ONE SAYS \( P[ x_k \text{ WILL OCCUR} ] = \frac{n(x_k)}{N} \)

AS \( N \) GETS BIG.

IT IS CLEAR THAT \( 0 \leq n(x_k) \leq N \)

\( 0 \leq \frac{n(x_k)}{N} \leq 1 \)

\( \Rightarrow 0 \leq P(x_k) \leq 1 \)

\( \therefore P(x_k) \text{ IS SINGLE VALUED \& REAL} \)

\( P(x_k) = 0 \Rightarrow \text{IMPOSSIBLE EVENT} \)

\( P(x_k) = 1 \Rightarrow \text{CERTAIN EVENT} \)
(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, $\Omega_k$, in a sample space. Then, $m(\Omega_k)$ (in for "measure") is a real half single valued function called the probability measure and the prob. measure on an event $\sum$ of prob. measure of possible outcomes $\{\Omega_k\}$ that make up that event.

Ex. for a pack of cards, $m(\Omega_k) = \frac{1}{52}$

$P[\text{choosing a space}] = \frac{13}{52} = \frac{1}{4}$

* Two events are disjoint if they contain no outcomes in common, i.e., "cannot happen simultaneously."

On this basis

(1) $0 \leq m(\Omega_k)$

(2) $m[A \cup B] = m(A) + m(B)$ if $A \cap B$ are disjoint

This is called the "additivity property" of the measure.
Also, \( m(x) = 0 \) only if \( x = \emptyset \) (null event)  
\( m(x) = 1 \) only if \( x = U \) (universal set)  
Also (i) \( m(A) \leq m(B) \) if \( A \subseteq B \)  
"\( C \subseteq \) is a subset of "  
(2) \( m(\overline{A}) = m(B) - m(\overline{B} - \overline{A}) \) if \( A \subseteq B \)  
(3) \( m(A^\prime) = m(\overline{A}) = m(U - A) \)  
\[ = m(U) - m(A) = 1 - m(A) \]  
(4) \( m(A \cup B) = m[(A - \overline{A}B) \cup B] \)  
\( \overline{AB} = A \) intersect \( B \)  
\( m(A \cup B) = m(A - \overline{A}B) - m(B) \)  
\[ = m(A) - m(\overline{A}B) + m(B) \]  
This is the general additive law of probability measure.  
(5) \( m[A \cup B \cup C] = m(A) + m(B) + m(C) \)  
\[ - m(\overline{A}B) - m(\overline{A}C) - m(\overline{B}C) \]  
\[ + m(\overline{A}B\overline{C}) \]  

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 \( \exists P(A) \geq 0 \) 
\( \forall \text{ event } A \in S \)

Axiom 2: \( P(S) = 1 \)

Axiom 3: If \( S_1, S_2, \ldots \) is a sequence of mutually exclusive (disjoint) events in \( S \), then \( S_i \cap S_j = \emptyset \) 
\( \forall i \neq j = 1, 2, \ldots \), 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 \( \Omega \) a probability (measure) function on \( S \), then prob that event \( A \) does not occur = \( 1 - P(A) \)
\( \Leftrightarrow \) \( P(\overline{A}) = P(A^c) = 1 - P(A) \)

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 \( P \) \( \forall S_0 \) if \( S_0 \) is a null set, then \( P(S_0) = 0 \).
Marginal, joint, and conditional probabilities. Let \( S \) be a sample space with \( n \) points with a probability (measure) \( \frac{1}{n} \). Partition \( S \) into \( r \) disjoint subsets \( A_1, A_2, A_3, \ldots, A_r \). Also partition \( S \) into \( s \) disjoint subsets \( B_1, B_2, \ldots, B_s \).

<table>
<thead>
<tr>
<th></th>
<th>( B_1 )</th>
<th>( B_2 )</th>
<th>( B_3 )</th>
<th>( \ldots )</th>
<th>( B_s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>( n_{11} )</td>
<td>( n_{12} )</td>
<td>( n_{13} )</td>
<td>( \ldots )</td>
<td>( n_{1s} )</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>( n_{21} )</td>
<td>( n_{22} )</td>
<td>( n_{23} )</td>
<td>( \ldots )</td>
<td>( n_{2s} )</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
</tr>
<tr>
<td>( A_r )</td>
<td>( n_{r1} )</td>
<td>( n_{r2} )</td>
<td>( n_{r3} )</td>
<td>( \ldots )</td>
<td>( n_{rs} )</td>
</tr>
</tbody>
</table>

\( n_{ij} = \# \) of outcomes with attribute \( A_i \) and \( B_j \).

\[
\sum_{i,j} n_{ij} = n
\]

Then the prob. of the event \( A_1 \) and \( B_3 \)

\[
P(A_1, B_3) = P(A_1 \cap B_3) = \frac{n_{13}}{n}
\]

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

\[
P[A_2] = \sum_{j=1}^{s} \frac{n_{2j}}{n} = \sum_{j=1}^{s} P(A_2, B_j)
\]
IN GENERAL,
\[ P[A_i] = \prod_j P[A_i, B_j] \]
AND
\[ P[B_j] = \prod_i P[A_i, B_j] \]
PROBS. SUCH AS \( P(A_i) \) OR \( P(B_j) \) ARE MARGINAL PROBS.

FROM THE POINT OF VIEW OF SET THEORY:
\[ A_i = (A_i \cap B_1) \cup (A_i \cap B_2) \cup (A_i \cap B_3) \cup \cdots \]
\[ \cdots \cup (A_i \cap B_s) \]
SINCE \((A_i \cap B_j) \cap (A_i \cap B_{j'}) = \emptyset \)
WHERE \( j \neq j' \).

THEN, FROM AXIOM 3:
\[ P(A_i) = P(A_i \cap B_1) + P(A_i \cap B_2) + \cdots + P(A_i \cap B_s) \]
\[ = \prod_j P(A_i, B_j) \]
\[ = \prod_{j=1}^s P(A_i, B_j) / n \]
ALSO
\[ P(A_i, C_k) = \prod_j P(A_i, B_j, C_k) \]
ALSO
\[ P(C_k) = \prod_j \prod_i P(A_i, B_j, C_k) \]
*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 $n_{23}$.

\[ P(A_2/B_3) = \frac{n_{23}}{\sum_{i=1}^{r} n_{i3}} \]

Or, in general

\[ P(A_{i}/B_{j}) = \frac{n_{ij}}{\sum_{i=1}^{r} n_{i}} \]

\[ = \frac{n_{ij}}{n} \]

\[ \Rightarrow P(A_{i}/B_{j}) = \frac{P[A_{i}, B_{j}]}{P[B_{j}]} \]

This relation is sometimes referred to as multiplicative law of probability measure:

\[ P(A_{i}, B_{j}) = P(B_{j}) P(A_{i}/B_{j}) \]
FOR THREE ATTRIBUTES:

\[ P[A_i, B_j | C_k] = \frac{P[A_i, B_j, C_k]}{P[C_k]} \]

\[ P[A_i | B_j, C_k] = \frac{P[A_i, B_j, C_k]}{P[B_j, C_k]} \]

HOMEWORK:

1. WRITE OUT \( P(A_i, B_j, C_k) \) IN TERMS OF VARIOUS CONDITIONAL AND MARGINAL PROBABILITIES.

2. 6 RED BALLS
4 BLACK BALLS

2 BALLS DRAWN W/O REPLACEMENT
FIND \( P[2^{nd} \text{ red} | 1^{st} \text{ is red}] \)
(VERIFY RIGOROUSLY)

\[ P(A, B) = \text{joint} \]
\[ = \frac{\text{relevant outcomes}}{\text{possible outcome}} \]
\[ = \frac{6C_2}{10C_2} = \frac{1}{3} \]

\[ P[A/B] = \frac{P[A, B]}{P[B]} = \frac{\frac{1}{3}}{\frac{6}{10}} = \frac{5}{9} \]
STATISTICAL INDEPENDENCE

\[ P(B|A) = \frac{P(A,B)}{P(A)} \]

If \( P(B|A) = P(B) \),
then \( P(B|A) = \frac{P(A,B)}{P(A)} = P(B) \)
thus \( P(A,B) = P(A)P(B) \)

then events \( A \neq B \) are said to
be statistically independent.

\[ : P(A,B) = P(A|B)P(B) = P(B|A)P(A) \]
\[ = P(A)P(B) = P(B)P(A) \]

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: \( B_1 \leftrightarrow \overline{B_1} = (A_1 \text{ or } A_2) \)
\( B_2 = (A_1 \text{ or } A_3) \)
\( B_3 = (A_1 \text{ or } A_4) \)

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"
\[ P(B_1, B_2) = P(A_1) = \frac{1}{4} \]
\[ = P(B_1)P(B_2) \]
\[ \therefore B_1 \text{ and } B_2 \text{ are stat. ind. } \Rightarrow \]
SAME HOLDS FOR $P(B_1, B_2) \neq P(B_2, B_3)$.

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 $1 \leq i < j < k, \ldots < N$, if the following relationships hold:

$$P(A_i, A_j) = P(A_i)P(A_j) \quad \text{< pairwise indep.}$$

$$P(A_i, A_j, A_k) = P(A_i)P(A_j)P(A_k)$$

$$\vdots$$

$$P(A_i, A_j, A_k, \ldots, A_N) = P(A_i)P(A_j)P(A_k)\ldots P(A_N)$$
**Random Variable (A Matter of Associating Real #’s with Outcome of Experiments)**

A rigorous def'n 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 infinite
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 random variable \( x \) is bounded by an
arbitrarily chosen real number \( x \).

\( P \left[ x \leq x \right] \triangleq \text{PDF of the random variable } x \)

\[ P \left[ x < -\infty \right] = 0 \]
\[ P \left[ x < \infty \right] = 1 \]

\[ P \left[ x \leq b \right] = P \left[ x < a \right] = P \left[ a < x < b \right] \text{ if } b > a \]

Thus, the PDF is a non-decreasing function of \( x \).

JOINT PDF (PROBABILITY DISTRIBUTION FUNCTION)

\( \triangleq P \left[ x \leq x, y \leq y \right] \)

"\( \leq \)" = "AND"

= probability that the sample point
is in the appropriate quadrant.

Obviously:

\[ P \left[ x < -\infty, y \leq y \right] = 0 \]
\[ P \left[ x \leq \infty, y \leq \infty \right] = 1 \]

(This encompasses the entire
\( xy \) plane or sample space)

MARGINALY:

\[ P \left[ x \leq x, y \leq \infty \right] = P \left[ x \leq x \right] \]
\[ P \left[ x \leq \infty, y \leq y \right] = P \left[ y \leq y \right] \]

These PDF's are called
"marginal" PDF's.
Consider a discrete r.v. \( X \) which can take on 6 values \( X_1, \ldots, X_6 \).

We have probabilities:

\[
\begin{align*}
P[X_1] &= 0.2 \\
P[X_2] &= 0.1 \\
P[X_3] &= 0.4 \\
P[X_4] &= 0.1 \\
P[X_5] &= 0.2 \\
P[X_6] &= 0
\end{align*}
\]

Clearly:

\[
\sum_{k=1}^{6} P[X_k] = P[X \leq X]
\]

And:

\[
P[X < \infty] = \sum_{k=1}^{6} P[X_k] = 1
\]
Consider joint discrete situation,
Let prob. plot be as follows:

\[
\begin{align*}
Y_2 & \uparrow \frac{3}{10} \\
Y_1 & \uparrow \frac{2}{10} \quad \uparrow \frac{1}{10} \\
x_1 & x_2 \quad x_3
\end{align*}
\]

Homework: Sketch the joint cumulative distribution function.

\[p(x_K) = \sum_m p(x_K, y_m)\]
\[p(y_m) = \sum_k p(x_K, y_m)\]

\[
\text{and } p(x_K, y_m) = p(y_m | x_K) p(x_K) = p(x_K | y_m) p(y_m)
\]

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 values \(-\infty < x < \infty\).)

Moreover, if the PDF is also differentiable with a continuous derivative, thus, its prob. density function (pdf)

\[
p(x) = \frac{d}{dx} P(x < x)
\]

\[
= \lim_{\Delta x \to 0} \frac{P(x \leq x) - P(x \leq x - \Delta x)}{\Delta x}
\]

thus, \( p(x) \Delta x = P(x - \Delta x < x < x] \)

i.e., 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), \( \frac{1}{2} \) (3) constitute the properties of a pdf of a continuous r.v.

Also: \( p(x = x_0) = 0 \) for continuous r.v.
Consider the distribution:

\[ \text{Rayleigh Distribution} \]

\[ \text{Gaussian or Normal pdf} \]

\[ (\text{7-16-76 Homework Handout}) \]
(1) Joint Probability Density Function.

In the two-dimensional sample space, the joint pdf \( P[x \leq X, y \leq Y] \) is everywhere continuous and possesses a mixed continuous mixed 2nd derivative everywhere, then

\[
\text{Joint pdf} \rho_{x,y} = \frac{\partial^2}{\partial x \partial y} P[x \leq X, y \leq Y]
\]

Then

\[
P[x \leq X, y \leq Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(x, y) \, dx \, dy
\]

Thus,

\[
P[x, y] = \int_{-\infty}^{x} \int_{-\infty}^{y} \rho(x, y) \, dx \, dy + \int_{-\infty}^{\infty} \int_{y}^{\infty} \rho(x, y) \, dx \, dy
\]

in the differential form:

\[
\rho(x, y) \, dx \, dy = P[X < x < x + dx, Y < y < y + dy]
\]

Recall the similar relation for single variable, \( \rho(x, x) \) is called the joint pdf.
Again, since the joint PDF is a non-negative, non-increasing function of its arguments: \( p(x, y) \geq 0 \)

and the probability a sample point falls in a region \( R \) of our space is:

\[ P \{ s \in R \} = \int_R p(x, y) \, dx, \, dy \]

and thus:

\[ \int_\infty^\infty p(x, y) \, dx \, dy = 1 \]

and, as before, marginally:

\[ \int_\infty^\infty p(x, y) \, dy = p(x) \]

\[ \int_\infty^\infty p(x, y) \, dx = p(y) \]

and, also:

\[ \int_\infty^\infty \int_\infty^\infty p(x, y) \, dx \, dy = P \{ x < x, y \} \]

\[ \int_\infty^\infty \int_\infty^\infty p(x, y) \, dx \, dy = P \{ y < y \} \]
**Conditional pdf**

Here, what is the pdf that will R.V. $Y$ follow in the hypothesis that a second R.V. $X - \Delta x < x < \Delta x$?

\[
P \left[ y \in \frac{x - \Delta x}{x} < x < \Delta x \right] \]

\[
= \frac{P \left[ x - \Delta x < x < \Delta x \mid y \right]}{P \left[ x - \Delta x < x < \Delta x \right]}
\]

\[
= \int_{y - \Delta y}^{y} \int_{x - \Delta x}^{x} p(x, y) dx dy
\]

\[
= \int_{x - \Delta x}^{x} p(y) dx
\]

**Note to find pdf, we must require that pdf is well defined.**

Now, we rewrite the conditional pdf as:

\[
P \left[ y \in \frac{x}{x} \right] = \frac{\int_{y - \Delta y}^{y} p(x, y) dy - \Delta x}{p(x) \Delta x}
\]

\[
= \int_{y - \Delta y}^{y} \frac{p(x, y)}{p(x)} dy / p(x)
\]

**Differentiating w.r.t. $y$:**

\[
p(y/x) = \frac{p(x, y)}{p(x)}
\]

Also, as before

\[
p(y/x) \geq 0
\]

\[
P [a < y < b / x] = \int_{a}^{b} p(y/x) dy
\]

Of course

\[
\int_{-\infty}^{\infty} p(y/x) dy = 1
\]
ON THE INFORMATION OF INTEREST TO COMMUNICATION. LINE.
COMMUNICATION PROCESS FLOW OF SOME
INFORMATION AVAILING COMMUNITY
A SIMPLE MODEL IS:

\[
\begin{array}{c}
\text{TLOSSER} \\
(\text{SOURCE})
\end{array} \xrightarrow{\text{CHANNEL}} \begin{array}{c}
\text{RECEIVER} \\
(\text{SINK})
\end{array}
\]

A MORE GENERAL MODEL IS:

\[
\begin{array}{c}
\text{TLOSSER} \\
(\text{SOURCE})
\end{array} \xrightarrow{\text{ENCODER}} \begin{array}{c}
\text{CHANNEL} \\
(\text{MEDIUM})
\end{array} \xrightarrow{\text{DECODER}} \begin{array}{c}
\text{RECEIVER} \\
(\text{SINK})
\end{array}
\]

ENCODER PERFORMS A 1 TO 1 MAPPING \( F(1) \).
DECODER, \( F^1 \), GIVES \( F^1(1) \). ON A PROBABALISTIC BASIS.
WE TIE TO ASK AND ANSWER THE QUESTIONS:
(1) THE SOURCE IS NON-BINARY, HOW DOES
ON MEASURE INFORMATION AND WHAT UNIT
SHALL WE MEASURE IT IN?
(2) HAVING DEFINED THIS UNIT, WHAT IS THE MEAN, AT WHICH THE
SOURCE SUPPLY INFORMATION?
(3) WHAT IS A CHANNEL? HOW DO WE CHARACTERIZE IT
EACH SOURCE-SHARE CHANNEL, WHAT IS COMBINED RATE OF INFORMATION?
(4) WHAT EFFECT DOES "NOISE" HAVE ON
THE PERFORMANCE OF THE CHANNEL?
(ANSWERED BY SHANNON'S THEOREM.)
A catalog with $n$ models designates \((x_1, y_1, x_2, \ldots, x_n)\) in \(m\) colors \((c_1, c_2, \ldots, c_m)\). The total \(\gamma\) of articles is \(\alpha > 1\).

A very key statement: The desired amount of the \(I(x_\alpha)\) associated with the collection of a particular model \(x_\alpha\) should be some function of the prob. with which that model \(x_\alpha\) "occurs" in the catalog. [RECOMMEND BY HARTLEY (1922)]

\[ I(x_\alpha) = \int [p(x_\alpha)] \]

If all models are equally likely, then \(p(x_\alpha) = \frac{1}{n}\) \(\forall\) \(1 < I(x_\alpha) = \frac{1}{n} \int (V_n) \)

Next, exclude a choice part color \(I_2(c_\beta) = \int (V_n) \)

If the choice of model \(x_\alpha\) color is independent, \(\forall\) \(1 < I(x_\alpha) = \frac{1}{n} \int (V_n) \)

Associated with a specific color is \((x_\alpha, y_1)\) \(I_1(c_\beta) = \int (V_n) \)

But what is the info associated with extraction \(X_\alpha\) and \(c_\beta = \int (V_n)\), clearly

\[ \int (\frac{1}{n \cdot m}) = \int (\frac{1}{n}) + \int (\frac{1}{m}) \]
There are many functions \( f(n) \)
which will satisfy
\[
f(n^m) = f(n^i) + f(n^j)
\]

a. Let \( f = \frac{\log P[X_n]}{\log n} \)
b. \( f = \# \) of factors in the factorization
of \( n \) in the product of primes\( (n^2 = 1) \)

Example: \( n = 16 = 2 \cdot 2 \cdot 2 \Rightarrow f(n^2) = 3 \)

\( n = 8 = 2 \cdot 2 \cdot 2 \Rightarrow f(n^2) = 3 \)

\( f(n^4) = 6 \)
OF,

7/19/76

- \log p_i =

SELF INFORMATION OF THE SYMBOL X, WHOSE
PROB. OF OCCURRENCE WAS p_i

= UNCERTAINTY ASSOCIATED WITH X_i

= THE AMOUNT OF INF. ASSOCIATED WITH
THE SELECTION OF X_i

IF, FOR A GIVEN SOURCE, THERE ARE
ONLY 2 SYMBOLS X_1, X_2 EACH
EQUALLY LIKELY:

\[
\text{TRANSMITTER} \quad \xrightarrow{\quad} \quad \begin{array}{c}
X_1, X_2 \\
\end{array}
\]

\[ p(x_1) = \frac{1}{2} \Rightarrow -\log \frac{1}{2} = \log 2 \]

\[ = \log_2 2 = 1 \text{ BIT OF INFO.} \]

\[ p(x_2) = \frac{1}{2} \text{ BIT} \]

THE AVERAGE UNCERTAINTY ASSOCIATED
WITH THE ENTIRE SOURCE = AVE \[ -\log p_i \]
NOTING THAT EACH SYMBOL X_i
OCURR. WITH PROB. p_i

\[
\text{ASIDE: EXPECTED VALUES (AVERAGE)}
\]

CONSIDER X_1, X_2, \ldots, X_n

\[ p(x_1), p(x_2), \ldots, p(x_n) \]

THEN \[ \frac{1}{n} \sum_{i=1}^{n} x_i p(x_i) \]

\[ = \text{STATISTICAL AVERAGE} \]
\[ E[- \log p_i] = - \sum_{i=1}^{n} P_i \log P_i \]

ENTROPY = AVERAGE INFO

\[ H(p_1, p_2, \ldots, p_n) > 0 \]

USING DEF: \( 0 < p_i < 1 \) \& \( i \in (0,1) \)

CONSIDER

\[
\begin{align*}
\begin{cases}
x_1, x_2, \ldots, x_{10} \\
0, 1, 2, 3, \ldots
\end{cases}
\end{align*}
\]

\( \rho(x_2) = \frac{1}{10} \)

\[ I = - \log \frac{1}{10} \]

\[ = \log_{10} 10 \]

\[ = 1 \text{ BARTLE} \]

\[
\begin{align*}
\text{BASE 2} & \Rightarrow \text{BIT} \\
\text{BASE 10} & \Rightarrow \text{HARTLE} \\
\text{BASE e} & \Rightarrow \text{NAT}
\end{align*}
\]

1 HARTLEY = 3.32 BITS

1 NAT = 1.44 BITS ETC

\[ \text{THIS IS FROM:} \]

\[ \log_a x = \log_b x / \log_b a \]

\[ \text{TEST} \]

\[
\begin{align*}
S_1 & : [A_1, A_2] \Rightarrow \left( \frac{1}{266}, \frac{255}{266} \right) \\
S_2 & : [B_1, B_2] \Rightarrow \left( \frac{1}{2}, \frac{1}{2} \right) \\
S_3 & : [C_1, C_2] \Rightarrow \left( \frac{7}{16}, \frac{9}{16} \right)
\end{align*}
\]

DEFN: IF \( \frac{1}{T} p_T = 1 \)

1 # OF SYMBOLS IS FINITE

THEN SOURCE IS CALLED

"A DISCRETE COMPLETE SOURCE"
\( H_1(\cdot) = \frac{1}{256} \log 256 + \frac{255}{256} \log \frac{256}{255} = \frac{1}{256} \log 2^8 + \frac{255}{256} \left[ \log 2^8 - \log 255 \right] = 0.0369 < < 1 \)

\( H_2(\cdot) = \frac{1}{2} \text{ bit} \)

\( H_3(\cdot) = \frac{7}{16} \log \frac{16}{7} + \frac{9}{16} \log \frac{16}{9} = 0.989 \text{ bits} \)

**The Mathematics of H**

**The Requirements on H**

1. **Continuity with regard to \( p_i \):**

2. **Symmetry:** \( H(p_1, p_2, \ldots, p_n) = p(p_{\text{s}}, p_2, \ldots, p_n, \ldots, p_{\text{a}}, p_{\text{a}-1}) \)

3. **Extremal Property:**

\( H(p_1, p_2, \ldots, p_n) \) is maximum when \( p_i = \frac{1}{n-1} \)

4. **If one of the events is divided 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) \[ H(p_1, p_2, \ldots, p_n) = -\sum_{k=1}^{n} p_k \ln p_k \]

(2) SYMMETRY (OBVIOUS)

(3) MAXIMALITY (OR EXTREMALITY) OF \( H \)
\[
\frac{dH}{dp_k} = \sum_{i=1}^{n} \frac{\delta H}{\delta p_i} \cdot \frac{dp_i}{dp_k}
= -\frac{1}{p_k} \ln p_k \cdot \frac{dp_k}{dp_k}
- \frac{1}{p_n} (p_n \ln p_n) \cdot \frac{dp_k}{dp_n} + 0
\]

ALSO, DUE TO COMPLETENESS,

\[ p_n = 1 - (p_1 + p_2 + \ldots + p_{n-1}) \]

\[
\frac{dH}{dp_k} = -(\ln e + \ln p_k) + (\ln e + \ln p_n)
\]

IF THE ORIGINAL LOG IS BASE \( \ln \)
\[ \frac{dH}{dp_k} = \ln p_n - \ln p_k = 0 \text{ FOR MAX OR MIN} \]

\[ \Rightarrow p_n = p_k \]

FOR MAXIMALITY, CONSIDER \( H[1,0,0,\ldots] = 0 \)

\[ \therefore \text{HERE} \ H\left[\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right] \text{ IS MAX} \]

(4) \[ H[q_1, q_2, \ldots, q_m] = - \left[ q_1 \ln q_1 + \ldots + q_m \ln q_m \right] \]
\[ = - \sum_{k=1}^{m} q_k \ln q_k \]
\[ = \sum_{i=1}^{n} p_i \ln p_i + p_n \ln p_n - \sum_{k=1}^{m} q_k \ln q_k \]
\[ = H(p_1, p_2, \ldots, p_n) + p_n \sum_{k=1}^{m} \frac{q_k}{p_n} \ln p_n - p_n \sum_{k=1}^{m} \frac{q_k}{p_n} \ln q_k \]
\[ = H(p_1, p_2, \ldots, p_n) + p_n \sum_{k=1}^{m} \frac{q_k}{p_n} \ln \left( \frac{p_n}{q_k} \right) \]
\[ = H(p_1, p_2, \ldots, p_n) + p_n H\left( \frac{q_1}{p_n}, \frac{q_2}{p_n}, \ldots, \frac{q_m}{p_n} \right) \]

= ADDITIVITY PROPERTY
9-20-76 (TUES)

**Additivity Property**

1. \( X: \{x_1, x_2, x_3\} \quad P: \left(\frac{1}{3}, \frac{4}{15}, \frac{8}{15}\right) \) \( \Rightarrow P = 1 \)

A DISCRETE & COMPLETE SCHEME

We will look at

1. \( H \left( \frac{1}{3}, \frac{4}{15}, \frac{8}{15} \right) \)

2. \( X: \{x_1, x_2, x_2 \cup x_3\} \quad \frac{1}{3} \quad P \left(\frac{1}{3}, \frac{12}{15}\right) \)

3. \( X: \{x_1, x_2 \cup x_3, x_2 \cup x_3\} \quad \frac{1}{3} \quad P \left(\frac{1}{3}, \frac{12}{15}, \frac{8}{12/15}, \frac{3}{12/15} \right) \Rightarrow \left(\frac{1}{3}, \frac{2}{3}\right) \)

**Question**

Relate \( H \left(\frac{1}{3}, \frac{4}{15}, \frac{8}{15}\right) \) to

\[
H \left(\frac{1}{3}, \frac{12}{15}\right) + \frac{1}{3} \quad H \left(\frac{1}{3}, \frac{2}{3}\right)
\]

\[
H \left(\frac{1}{3}, \frac{4}{15}, \frac{8}{15}\right) = \frac{1}{3} \ln 5 + \frac{4}{15} \ln \frac{15}{4} + \frac{8}{15} \ln \frac{15}{8}
\]

\[
H \left(\frac{1}{3}, \frac{2}{3}\right) = \frac{1}{3} \ln 5 + \frac{2}{3} \ln 5/4
\]

**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**

\[
P(A) = \frac{1}{2} \quad P(B) = \frac{1}{4} \quad P(C) = P(D) = \frac{1}{2} \quad \Rightarrow \text{prob} = 1
\]
For Case 1:

\[ H_1 = \frac{1}{2} \ln 2 + \frac{3}{4} \ln 4 + \frac{1}{8} \ln 8 + \frac{1}{8} \ln 8 \]
\[ = \left[ \frac{1}{2} + \frac{3}{4} + \frac{1}{8} + \frac{1}{8} \right] \text{bits} \]
\[ = 1 \frac{3}{4} \text{ bits} \]

For Case 2:

\[ H_2 = -P(A) \ln P(A) - [1-P(A)] \ln [1-P(A)] \]
\[ = \left[ 1-P(A) \right] \left[ \frac{P(b)}{1-P(a)} \ln \frac{P(b)}{1-P(a)} + \frac{P(c)}{1-P(a)} \ln \frac{P(c)}{1-P(a)} + \frac{P(d)}{1-P(a)} \ln \frac{P(d)}{1-P(a)} \right] \]
\[ = \frac{1}{2} \ln 2 + \frac{1}{2} \ln 2 \]
\[ + \frac{1}{2} \left[ \frac{1}{4} \ln 4 + \frac{1}{8} \ln 8 + \frac{1}{8} \ln 8 \right] \]
\[ = \frac{1}{2} \left[ \frac{1}{2} + \frac{3}{2} \right] \]
\[ = 1 \frac{3}{4} \text{ bits} \]

For Case 3:

\[ H_3 = -(P_A + P_B) \ln (P_A + P_B) - (P_c + P_d) \ln (P_c + P_d) \]
\[ + (P_A + P_B) \left[ -\frac{P_A}{P_A + P_B} \ln \frac{P_A}{P_A + P_B} \right] \]
\[ + \frac{P_a}{P_a + P_b} \ln \frac{P_a + P_b}{P_A + P_B} \]
\[ + (P_c + P_d) \left[ -\frac{P_c}{P_c + P_d} \ln \frac{P_c}{P_c + P_d} \right] \]
\[ + \frac{P_c}{P_c + P_d} \ln \frac{P_c + P_d}{P_c + P_d} \]
\[ = 1 \frac{3}{4} \text{ bits} \]
**LEMMAS PERTAINING TO H FUNCTION**

**THE LOG FUNCTION IS A CONVEX (UPWARD) FUNCTION.**

A necessary and sufficient condition for a convex function is that 
\[ \frac{d^2y}{dx^2} \leq 0 \]

Now 
\[ \frac{d}{dx} \ln(x) = \frac{1}{x}, \]
\[ \frac{d^2}{dx^2} \ln(x) = -\frac{1}{x^2}, \quad 0 < x < \infty \]

Also, for any convex function:
\[ \frac{1}{2} \left[ f(x_1) + f(x_2) \right] \leq f\left( \frac{x_1 + x_2}{2} \right) \]

Assume \( x_1, x_2 > 0 \) so that
\[ \frac{1}{2} \left[ \ln(x_1) + \ln(x_2) \right] = \ln\left( \frac{x_1 + x_2}{2} \right) \]
\[ \ln\sqrt{x_1 x_2} \leq \frac{x_1 + x_2}{2} \leq x_1, x_2 > 0 \]

geometric mean arithmetic mean
Lemma 2: \( \ln x \leq x - 1 \)

\[
\begin{align*}
&\text{OR} \\
&\ln x \ln_2 e \\
&= \ln_2 x \leq x - 1 \ln_2 e
\end{align*}
\]

We shall use this lemma to prove extremal property of \( H \).

Let \( X = x_1, x_2, \ldots, x_m \in \mathbb{R} \)

\( P(x_1) \neq P(x_2) \neq \ldots \neq P(x_m) \).

We wish to prove

\[
H(x) \leq -m \left( \frac{1}{m} \ln m \right)
\]

\[
\leq \ln \frac{1}{m}
\]

Be definition,

\[
H(x) - \ln m = \sum_{i=1}^{m} p_i \ln p_i + \ln \frac{1}{m}
\]

\[
= \sum_{i=1}^{m} p_i \ln p_i + \left( \sum_{i=1}^{m} p_i \right) \ln \frac{1}{m}
\]

\[
= \sum_{i=1}^{m} p_i \ln m p_i
\]

\[
\leq \sum_{i=1}^{m} p_i \left( \frac{1}{mp_i} - 1 \right) \ln_2 e \quad \text{(by Lemma 2)}
\]

\[
\leq \sum_{i=1}^{m} \left( \frac{1}{m} - p_i \right) \leq 0
\]

Hence

\[
H(x) \leq \ln m \Rightarrow \text{EXTREMA PROPERTY 3}
\]

\[
H(s) = \ln \left( \frac{s(x_1,x_2)}{p(s,w)} \right)
\]

\[
H(w) = - (w \ln w + \bar{w} \ln \bar{w}) \equiv \text{FIXED FOR} \quad \text{FIXED} \quad w
\]

\[
H(w) = \text{ENTROPY FUNCTION}
\]
**COR:** The entropy function is convex

\[
H(w) = -(w \log w) = -\sum x \log x
\]

Consider 3 sources each with 2 symbols:

- \(S_1(A,B)\) \(P_1 = \left(\frac{1}{3}, \frac{2}{3}\right)\)
- \(S_2(C,D)\) \(P_2 = \left(\frac{1}{4}, \frac{3}{4}\right)\)
- \(S_3(E,F)\) \(P_3 = \left(\frac{2}{24}, \frac{17}{24}\right)\) \(\Leftarrow\) any of \(S_1 \neq S_2\)

Show that

\[
\frac{1}{2}[H(x_1) + H(x_2)] \leq H\left(\frac{x_1 + x_2}{2}\right)
\]

- \(H(x_1) = \frac{1}{3} \log 3 + \frac{2}{3} \left[\log 3 - \log 2\right] = 0.85\)
- \(H(x_2) = \frac{1}{4} \log 4 + \frac{3}{4} \left(\log 4 - \log 3\right) = 0.81\)
- \(H(x_3) = \frac{7}{24} \log \frac{21}{17} + \frac{17}{24} \log \frac{17}{24} = \)

**Lemma 3:** Let \(x_1, x_2, \ldots, x_q \neq y_1, y_2, \ldots, y_q\) be two sets of complete schemes.

\[\frac{1}{q} \sum_{i=1}^{q} x_i = \frac{1}{q} \sum_{i=1}^{q} y_i = 1\]

Then

\[\sum_{i=1}^{q} x_i \log_{10} \frac{1}{x_i} \leq \sum_{i=1}^{q} x_i \log_{10} \frac{1}{y_i}\]

Notice that

\[\log_{10} x = \log_{10} x / \log_{10} 2 = \frac{1}{\log_{10} 2} x_i \ln_{10} x_i / x_i\]

Also

\[\sum_{i=1}^{q} x_i \ln_{10} \frac{y_i}{x_i} \leq \sum_{i=1}^{q} x_i \ln_{10} \left(\frac{y_i}{x_i} - 1\right)\] (from Lemma)

\[\Rightarrow \sum_{i=1}^{q} x_i \ln_{10} \frac{y_i}{x_i} \leq \frac{1}{\log_{10} 2} \sum_{i=1}^{q} x_i \left(\frac{y_i}{x_i} - 1\right) = \]

\[\sum_{i=1}^{q} x_i \ln_{10} \frac{y_i}{x_i} \leq 0\]

\[\Rightarrow \sum_{i=1}^{q} x_i \ln_{10} \frac{1}{x_i} \leq \sum_{i=1}^{q} x_i \ln_{10} \frac{1}{y_i}\]
**Lemma 4:** Related to joint entropy functions

**Sample Space**

\[ P : \{ P(E_k) \} \]

\[ P : \{ P(F_k) \} \]

\[ \text{Joint or "Product" Space} \]

\[ E \times F \]

\[ P \begin{pmatrix} E_1 F_1 & E_1 F_2 & \ldots & E_1 F_m \\ E_2 F_1 & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ E_n F_1 & \ldots & \ldots & E_n F_m \end{pmatrix} \]

\[ \text{Joint Probability Matrix} \]

**Source**

\[ E \]

\[ n \text{ symbols} \]

**Channel**

**Receiver**

\[ F \]

\[ m \text{ symbols} \]

It is clear that

\[ P[E_i] = P[E_1 F_1 U E_2 U \ldots U E_m F_m] \]

\[ = \prod_{i=1}^{m} P[E_i F_i] \]

\[ \text{same as} \]

\[ P[F_2] = \prod_{i=1}^{n} F[E_i, F_2] \]

\[ \text{MARGINAL PROB} \]

On this basis, it is reasonable to define the joint entropy of \( \{ EF \} \) space as

\[ H(x, y) = \sum_{i,j} P(x_i, y_j) \ln \frac{P(x_i, y_j)}{P(x_i) P(y_j)} \]

over the entire plane.
\[
H(x, y) = -\sum_{x_i, y_j} \frac{1}{M} p(x_i, y_j) \ln p(x_i, y_j)
\]
\[
x \sim x_i, y \sim y_j, \quad i = 1, 2, \ldots, M
\]
\[
y \sim y_j, \quad j = 1, 2, \ldots, L
\]

Thus, there are all outcomes of interest.

Lemma 4: (Will connect the individual entropies \(H(x), H(y)\) with joint \(H(x, y)\) as defined above):

\(H(x, y) \leq H(x) + H(y)\)

Note that

\[
p(x_i) = \sum_j \frac{1}{M} p(x_i, y_j)
\]
\[
p(y_j) = \sum_i \frac{1}{M} p(x_i, y_j)
\]

\[H(x) = -\sum_{i=1}^M \frac{1}{M} p(x_i) \ln p(x_i)\]
\[H(y) = -\sum_{j=1}^L \frac{1}{M} p(y_j) \ln p(y_j)\]

\[H(x) + H(y) = -\sum_{i=1}^M \frac{1}{M} \sum_{j=1}^L p(x_i, y_j) \ln p(x_i) p(y_j)\]
\[\geq \sum_{i=1}^M \frac{1}{M} \sum_{j=1}^L p(x_i, y_j) \ln q_{ij} \geq q_{ij} = p(x_i) p(y_j)\]
where

\[
H(X, Y) = -\sum_{y \in Y} \sum_{x \in X} p_{xy} \ln p_{xy}
\]

\[
P_{ij} = p(Y_j | X_i)
\]

\[
H(X, Y) = -\sum_{i=1}^{M} \sum_{j=1}^{L} p_{ij} \ln q_{ij}
\]

\[
\leq -\sum_{i=1}^{M} \sum_{j=1}^{L} p_{ij} \left[ \ln p(x_i) + \ln p(y_j) \right]
\]

\[
\leq -\frac{1}{2} \sum_{i=1}^{M} p(x_i) \ln p(x_i) - \frac{1}{2} \sum_{j=1}^{L} p(y_j) \ln p(y_j)
\]

\[
\leq H(X) + H(Y)
\]

which was to be shown. Equality will occur when \(X \perp Y\) are statistically independent.

**Cor. 1:**

\[
H(X_1, X_2, \ldots, X_n)
\]

\[
\leq H(X_1) + H(X_2) + \ldots + H(X_n)
\]

Equality prevailing when \(X_i\)'s are statistically independent.

**Cor. 2:**

\[
H(X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m)
\]

\[
\leq H(X_1, X_2, \ldots, X_n) + H(Y_1, Y_2, \ldots, Y_m)
\]

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

$$H[Y / X] = H[Y / X = x_i]$$

$$= p(x_1) H(Y / X = x_1) + p(x_2) H(Y / X = x_2) + \ldots + p(x_M) H(Y / X = x_M)$$

$$= \sum_{i=1}^{M} p(x_i) H(Y / X = x_i)$$

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

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

As an extension, make sure that the following is true:

(1) $H(Y, Z / X) = -\sum_{i=1}^{M} p(x_i, y_i, z_i) \ln p(y_i, z_i / x_i)$

(2) $H(Z / X, Y) = -\sum_{i=1}^{M} p(x_i, y_i, z_i) \ln p(z_i / x_i, y_i)$

(3) $H(Y_1, Y_m / X_1, X_2, \ldots, X_n) = \text{Homework}$
LEMMA 9.5: (CHANGES H(x) \rightarrow H(x,y)) WE SHALL SHOW

\[ H(y/x, H(x,y)) \]

\[ H(x,y) = H(x) + H(y/x) \]

\[ H(x,y) = H(x) + H(y/x) \]

\[ H(x,y) = H(x) + H(y/x) \]

\[ H(x,y) = H(x) + H(y/x) \]

As an example, consider

\[ H(x,y) = H(x) + H(y/x) \]

\[ H(x,y) = H(x) + H(y/x) \]

\[ H(x,y) = H(x) + H(y/x) \]

\[ H(x,y) = H(x) + H(y/x) \]

\[ H(x,y) = H(x) + H(y/x) \]

\[ H(x,y) = H(x) + H(y/x) \]
Lemma 6: (Relates the conditional $H(Y/X)$ to $H(Y)$)

$H(Y/X) \leq H(Y)$ with equality when $X \not\perp Y$ are independent.

From before:

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

---

**Homework**

Text 2.3 p. 41/2.4 p. 42/2.14 p. 44

1. From Uncle's Book

Let $X = (X_1, X_2)$, $Y = (\frac{1}{4}, \frac{3}{4})$

Let $Y = (Y_1, Y_2, Y_3)$

$p(Y_1/X_1) = \frac{1}{4} \quad p(Y_1/X_2) = \frac{1}{10}$

$p(Y_2/X_1) = 0.35 \quad p(Y_2/X_2) = \frac{7}{10}$

$p(Y_3/X_1) = \frac{1}{10} \quad p(Y_3/X_2) = \frac{7}{10}$

Find: $H(Y), H(Y), H(X,Y), H(Y/X), H(X/Y)$

And verify the relations from Lemma's 4, 5, 6.

**Assignment**

18. Suppose that in a certain city, $\frac{2}{3}$ of the high school pass. $\frac{4}{9}$ 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 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, and frat status are transmitted by 3 successive binary digits, how much info is conveyed by each digit.

Ask 1-4. Establish the following:

a) $H(y,z|x) \leq H(y|x) + H(z|x)$

With equality iff

$p(x_i, z_k|x_i) = p(y_i|x_i) p(z_k|x_i)$ 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

$p(y_j, z_k|x_i) = p(y_j|x_i) p(z_k|x_i)$ for all $i, j, k$
**AN EXTENDED SOURCE: SOURCE EXTENSION**

1. If we can only send a digit.

   For info, we only have 2 messages.

   For 2 at a time (word length 2),

   we get $2^2 = 4$ 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 $9$ symbols.

\( S = \{ s_1, s_2, \ldots, s_9 \} \)

\[ p = [ p(s_1), p(s_2), \ldots, p(s_9) ] \]

\[ \sigma^n = [ \sigma_1, \sigma_2, \ldots, \sigma_9^n ] \]

\[ p = [ p(\sigma_1), p(\sigma_2), \ldots, p(\sigma_9^n) ] \]

We will show:

\[ H[\sigma^n] = nH(S) \] if source has no memory.

If the successive symbols emitted from $\sigma^n$ 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)

5: \([s_1, s_2, \ldots, s_q]\)

5^* : \([o_1, o_2, \ldots, s_q]\)

\(H(s) = \sum_{i=1}^{q} p(s_i) \log p(s_i)\)

\(H^*(s) = H(s) + \sum_{i=1}^{q} p(o_i) \log \frac{1}{p(o_i)}\)

\(= \sum_{i=1}^{q} p(s_i) \log \frac{1}{p(s_i)}\)

\(= \sum_{i=1}^{q} p(o_i) \log \frac{1}{p(o_i)} + \sum_{i=1}^{q} \frac{1}{p(s_i)} = 1\)

(1) IT NEEDS TO BE SHOWN THAT 5^* IS COMPLETE

\(\sum_{i=1}^{q} p(o_i) = \sum_{i=1}^{q} p_i \cdot p_{i_1} \cdot p_{i_2} \ldots p_{i_n}\)

(2) \(H^*(s) = \sum_{i=1}^{q} p(o_i) \log \frac{1}{p(o_i)}\)

\(= \sum_{i=1}^{q} p(o_i) \log \frac{1}{p_i} \cdot p_{i_2} \ldots p_{i_n}\)

\(= \sum_{i=1}^{q} p(o_i) \log \frac{1}{p_i} + \sum_{i=1}^{q} \frac{1}{p_i} = 1\)

CONSIDER THE FIRST TERM

\(\sum_{i=1}^{q} p(o_i) \log \frac{1}{p_i} = \sum_{i=1}^{q} p_i \cdot p_{i_2} \ldots p_{i_n} \log \frac{1}{p_i}\)

\(= \sum_{i=1}^{q} p_i \log \frac{1}{p_i} = H(s)\)

CLEARLY: \(H^*(s) = n \cdot H(s)\)
EXAMPLE (FROM BOOK)

\[ S : (s_1, s_2, s_3) \Rightarrow P = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}) \]

\[ S^2 : (s_1s_1, s_1s_2, s_2s_1, s_2s_2, s_2s_3, s_3s_1, s_3s_2, s_3s_3) \]

\[ P : (\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}) \]

\[ H^2 = \frac{1}{4} \log_2 4 + 4 \cdot \frac{1}{8} \log_2 8 + 4 \cdot \frac{1}{16} \log_2 16 \]

\[ = \frac{1}{4} \times 2 + \frac{1}{2} \times 3 + \frac{1}{4} \times 4 = 3 \text{ bits} \]

\[ H(S) = 1\frac{1}{2} \text{ bits} \]

⇒ HOMWORK: TRY \( n = 3 \).

CHANNELS

\( H(X) = \text{AVERAGE INFORMATION PER CHARACTER (SYMBOL) AT THE SOURCE} \)

\[ \begin{array}{ccc}
\text{SOURCE} & \rightarrow & \text{CHANNEL} \\
\hline
X & \rightarrow & Y \\
\hline
\text{RECEIVER} & \rightarrow & \end{array} \]

\( H(Y) = \text{AVERAGE INFORMATION PER CHARACTER AT RECEIVER} \)

\( H(X,Y) = \text{AVERAGE INFORMATION PER PAIR OF CHARACTERS} \)

\( \omega = \text{AVERAGE UNCERTAINTY OF THE COMMUNICATION SYSTEM AS A WHOLE} \)

\( H(Y/X) = H(Y/X = x_i) = \text{A MEASURE OF INFORMATION ABOUT THE RECEIVING PORT KNOWING WHAT WAS TRANSMITTED, IT IS MEASURE OF "NOISE" OR "ERROR" IN THE CHANNEL} \)

\( H(X/Y) = H(X/Y = y_j) = \text{EQUIVOCATION OF THE CHANNEL} \quad \frac{1}{4} \text{ IS A MEASURE OF THE RECOVERY OF THE INPUT KNOWING THE OUTPUT} \)
CHANNEL CHARACTERIZATION

(1) A SOURCE WITH A PDF

(2) A JOINT PROB. MATRIX, WHICH DESCRIBES THE INPUT-OUTPUT RELATIONS ON A PROBABALISTIC MEASURE

\[ y_1, y_2, \ldots, y_n \text{ received} \]

\[
\begin{cases}
  (x_1, p(x_1, y_1), p(x_1 y_2), \ldots, p(x_1 y_n)) \\
  (x_2, p(x_2, y_1), p(x_2 y_2), \ldots, p(x_2 y_n)) \\
  \vdots \\
  (x_n, p(x_n, y_1), p(x_n y_2), \ldots, p(x_n y_n))
\end{cases}
\]

This, the joint probability matrix, completely describes the channel. From it, we can find \( p(x_k) \overset{\dagger}{=} p(y_k) \)

\[
\frac{1}{n} \text{ thus } H(x) = H(y)
\]

Can also get conditional probabilities \( \frac{1}{n} \) matrices.

CONDITIONAL PROB. MATRIX

\[ y_1, y_2, \ldots, y_n \]

\[
\begin{cases}
  x_1, p(x_1 | y_1) \\
  x_2, p(x_2 | y_2) \\
  \vdots \\
  x_n
\end{cases}
\]

\[ \Rightarrow \text{Homework: Work out conditional matrices in terms of joint matrix.} \]
Two Simple Channels

(1) Discrete Noise-Free Channel

\[ P(x, y) = x_1, x_2, \ldots, x_n \]

\[ y_1, P(x, y_1) = 0 \]

\[ y_2, 0, P(x_2, y_2) = 0 \quad \text{INXDIAGONAL} \]

\[ y_n, 0, 0, P(x_n, y_n) \]

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

\[ P(x, y) = y_1, y_2, \ldots, y_n \]

\[ x_1, p_1, p_1, p_1 \]

\[ x_2, p_2, p_2, p_2 \]

\[ \vdots \]

\[ x_n, p_n, p_n, p_n \]

\[ \sum_{i=1}^{n} p(x_i) = 1/m \]

\[ P(x_i) = mp_i \]

Joint:

\[ P(x, y) = \frac{p(x, y)}{P(x)P(y)} = P(x) \]

Conditional:

\[ P(x_i|y) = P(x_i) = mp_i \]

\[ P(x_i|y_j) = P(y_j|X_i) = \frac{1}{m} \]

Entropies:

\[ H(x, y) = -\sum_{i=1}^{n} mp_i \ln(mp_i) \]

\[ H(x) = -\sum_{i=1}^{n} mp_i \ln m = \ln m \]

\[ H(y) = -m^2 p_i \ln m = \ln m \]

\[ H(x|y) = -\sum \rho_i \ln \rho_i = -m^2 \rho_i \ln \rho_i = H(x) \]

\[ H(y|x) = H(y) \]
**Example:** The joint prob. matrix with

\[ s(x_1, x_2, \ldots, x_5) \quad r(y_1, y_2, y_3, y_4) \]

\[
\begin{array}{cccc|c}
  & y_1 & y_2 & y_3 & y_4 & p(x_i) \\
 x_1 & y_1 & y_2 & y_3 & y_4 & 0.25 \\
x_2 & 0 & 0.05 & 0.10 & 0 & 0.40 \\
x_3 & 0 & 0 & 0.05 & 0.15 \\
x_4 & 0 & 0 & 0 & 0.15 \\
x_5 & 0 & 0 & 0 & 0.05 \\
\end{array}
\]

\[ 0.35 \quad 0.35 \quad 0.2 \quad 0 \]

**Conditional** \( P(x_i/y) \)

\[
\begin{array}{cccc|c}
  & y_1 & y_2 & y_3 & y_4 \\
x_1 & \frac{25}{35} & 0 & 0 & 0 \\
x_2 & \frac{10}{35} & \frac{20}{35} & 0 & 0 \\
x_3 & 0 & \frac{5}{35} & \frac{15}{35} & 0 \\
x_4 & 0 & 0 & 5/20 & 1 \\
x_5 & 0 & 0 & \frac{1}{4} & 0 \\
\end{array}
\]

**Conditional** \( P(y_i/x) \)

\[
\begin{array}{ccccccc}
  x_1 & x_2 & x_3 & x_4 & x_5 \\
 y_1 & 1 & & & & \\
 y_2 & & & & & \\
 y_3 & & & & & \\
 y_4 & & & & & \\
\end{array}
\]

---

**Homework:** Figure out

\[ H(x), H(y), H(x,y), H(\frac{y}{x}), H(\frac{x}{y}), H(\frac{y}{x}) \]

\[ \frac{1}{2} \text{ verify } H(x,y) \leq H(x) + H(y) \]

\[ 2.665 \quad 2.066 \quad 1.856 \]
7/23/76   FIRST TEST
7/26/76   (MON)

HOMWORK

1.  PROB 5-4  p.142

2. USING THE DEFINITION 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
(1) Let us define a new element, namely mutual information:
\[ I(x_i; y_k) = \log_2 \frac{p(x_i | y_k)}{p(x_i)} \]
\[ = \log_2 \frac{p(x_1, y_k) / p(y_k)}{p(x_i) / p(y_k)} \]

Compare with self information:
\[ I(x_i) = -\log_2 p(x_i) \]

(2) The a priori knowledge that \( x_i \) is being transmitted
\[ = \sum \text{Prob} (x_i \text{ is being transmitted and is being received as any } y_k) = p(x_i) \]

The posterior knowledge of the observer is based on the conditional prob of \( x_i \) being transmitted given that a particular \( y_k \) is received = \( p(x_i | y_k) \)

The difference (information theoreticalwise) is the gain in information
\[ = \log \text{ ratio of two probs} \]
\[ = \log_2 \frac{p(x_i | y_k)}{p(x_i)} \]

It can be shown:
(1) \( I(x_i; y_k) \) is continuous
(2) \( I(x_i; y_k) = I(y_k; x_i) \) (symmetry)
(3) Self info: \( I(x_i; x_i) = -\log_2 p(x_i) \)
\[
\begin{align*}
I(x) &= I(x_i; x) \geq I(x_i; y)\\
I(y) &= I(y_j; y) \geq I(x_i; y_j)
\end{align*}
\]

(4) Let us average this "gain" into over the entire set of \((x_i, y_j)\) pairs:

\[
I(x; y) = \frac{I(x_i; y_j)}{\sum \sum p(x_i; y_j) \log \frac{p(x_i; y_j)}{p(x_i)} p(y)}
\]

If we go thru the summations, one would see that

\[
I(x; y) = H(x) + H(y) - H(x, y) \geq H(x) - H(x/y) \geq H(y) - H(y/x)
\]

Thus, on an average, the observance of any \(y\) provides us with \(I(x; y)\) bits of information.

Consider:

To max PWR when \(R\) is matched

\[ R \]

compare with

\[ C_N \]

We wanna maximize \(I(x, y)\).

The max of \(I(x, y)\) is channel capacity.
SHANNON'S DEFINITION OF CHANNEL CAPACITY

\[ C = \max I(x, y) \]

\[ = \max \left[ H(x) - H(x|y) \right] \]

**Example (Independent Input/Output)**

\[
\begin{array}{c|ccc}
  y_1 & y_2 & y_3 & y_4 \\
  x_1 & 1/4 & 1/4 & 1/2 & H(x) = 1 \text{ bit} \\
  x_2 & 1/4 & 1/4 & 1/2 & H(y) = 1 \text{ bit} \\
  y_2 & 1/2 & 1/2 & H(x,y) = 2 \text{ bits} \\
\end{array}
\]

\[ I(x,y) = H(x) + H(y) - H(x,y) \]

\[ = 0 \quad \text{(Intuitive! Input \& output are independent.)} \]

*If each symbol takes \( t \) sec to propagate, then the rate of information transmission of info, can be written as*

\[ C_t = \frac{C}{t} \text{ bits/sec} \]

*For a noise free system:*

\[ C_t = \frac{C}{t} = \frac{\log_2 n}{t} \quad \text{bits/sec} \]

*If the channel is noisy, the difference twixt max possible value of \( I(x;y) \) (channel capacity) \( \frac{1}{t} \)*

*The actual rate is called the absolute redundancy:

\[ C - I(x,y) \rightarrow \text{GENERAL} \]

\[ = \log_2 n - H(x) \rightarrow \text{NOISE FREE} \]

*And the relative redundancy:

\[ = \frac{\left[ \log_2 n - H(x) \right]}{\log_2 n} \rightarrow \text{GENERAL} \]

\[ = \left[ 1 - \frac{H(x)}{\log_2 n} \right] \rightarrow \text{NOISE FREE} \]
THE EFFICIENCY OF THE CHANNEL:
\[ R = \frac{I(x; y)}{\log n} = \frac{H(x)}{\log n} \]

= 1 - RELATIVE REDUNDANCY

IF A PARTICULAR SYMBOL \( x_i \) TAKES \( t_i \) SEC TO PROPAGATE THRU A
NOISELESS CHANNEL:

\[ R_t = \frac{H(x)}{\sum p(x_i) t_i} = \frac{\sum_i p(x_i) \log p(x_i)}{\sum_i p(x_i) t_i} \]
INFORMATION CHANNEL

OPEN: AN INF. CHANNEL IS DESCRIBED BY GIVING AN ALPHABET

\[ A : \{ a_1, a_2, \ldots, a_n \} \subseteq \text{INPUT ALPHABET} \]

\[ B : \{ b_1, b_2, \ldots, b_s \} \subseteq \text{OUTPUT ALPHABET} \]

A SET OF CONDITIONAL PROBABILITIES

\[ p(b_j/a_i) \] WHEN \[ p(b_j/a_i) \]

COND. PROB. WITH WHICH \[ b_j \] WAS RECEIVED IF

INPUT SYMBOL \[ a_i \] WAS SENT

\[ b_1, b_2, \ldots, b_s \]

\[ a_1, p(b_1/a_1), p(b_2/a_1), \ldots, p(a_n/b_1) \]

\[ a_2, p(b_1/a_2), p(b_2/a_2), \ldots, p(a_n/b_2) \]

\[ \vdots \]

\[ a_r, p(b_1/a_r), p(b_2/a_r), \ldots, p(a_n/b_r) \]

DEF.: \[ p(b_j/a_i) = p_{ij} \]

\[ b_1, b_2, \ldots, b_s \]

\[ a_1, p_{11}, p_{12}, \ldots, p_{1s} \]

\[ a_2, p_{21}, p_{22}, \ldots, p_{2s} \]

\[ \vdots \]

\[ a_r, p_{r1}, p_{r2}, \ldots, p_{rs} \]
Clearly:
\[ \sum_{i} p_{ij} = 1 \quad \forall \ i = 1, 2, \ldots, K \]

Since, if \( a_i \) is sent it must be received as some \( b_j \).

**DEF:** A binary symmetric channel (BSC) is "described" by

\[
\begin{pmatrix}
1 - \bar{p} & \bar{p} \\
\bar{p} & 1 - \bar{p}
\end{pmatrix}
\]

Where \( \bar{p} = p(0/0) = p(1/1) = p_{\text{correct transmission}} \)
\( p = p(0/1) = p(1/0) = p_{\text{error}} \)

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 alphabet \( A = \{ a_i, i = 1, 2, \ldots, K \} \)
and output alphabet \( B = \{ b_j, j = 1, 2, \ldots, S \} \).

And a primary channel described by the cond. prob. matrix:

\[
P = \begin{bmatrix}
p_{11} & \cdots & p_{1S} \\
\vdots & \ddots & \vdots \\
p_{K1} & \cdots & p_{KS}
\end{bmatrix}
\]
Next, consider the source $A^n$, the $n^{th}$ extension of $A$ (memoryless) where $A^n = \{x_i\}_{i=1}^n$, $x_i = 1, \ldots, r^n$. Also, the receiver alphabet $B^n = \{\beta_j\}_{j=1}^n$, $j = 1, \ldots, s^n$. The channel matrix will then be

$$\Pi = \begin{bmatrix}
\Pi_1 & \Pi_2 & \ldots & \Pi_{s^n}
\end{bmatrix}$$

where each input $x_i$, 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 $\Pi_{ji} = P(\beta_j | x_i) = \text{product corresponding elementary probs (due to lack of memory)}$.

**Ex.** 2$^{nd}$ extension of $P = \begin{bmatrix} 0 & 0 \end{bmatrix}$

$$P = \begin{bmatrix}
0 & 0,1 & 10 & 11
\end{bmatrix}$$

$$\begin{bmatrix}
\begin{array}{cccc}
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}
\end{bmatrix} = \begin{bmatrix}
\bar{p}p & p \bar{p} & \bar{p}p & p \bar{p}
\end{bmatrix} = \text{"Kronecker square" for } n = 2.$
IF ONE KNOWS $P = \begin{bmatrix} p_{11} & \ldots & p_{15} \\ \vdots & \ddots & \vdots \\ p_{51} & \ldots & p_{5s} \end{bmatrix}$ AND $P(q_i | v_i)$ (INPUT ALPS), THEN IT HAS BEEN SEEN THAT $P(b_0 | v_i)$ CAN BE COMPUTED $\forall i$ (ON TEST 1). THIS LEADS TO $\{P(q_i | b_j)\}_i$. THUS, ONE CAN DETERMINE THE "BACKWARD" COND. PROB. MATRIX FROM THE "FORWARD" COND. PROB. MATRIX.

**FORWARD** $\Rightarrow P(b_0^i | q_i)$

**BACKWARD** $\Rightarrow P(q_i | b_j^i)$

NOTICE THAT $\{P(q_i)\}_i$ YIELDS

$H(A) = -\sum_i P(q_i) \log P(q_i)$

= A PRIORI ENTROPY

Also $H(A | b_j^i) = -\sum_{(a)} P(q_i | b_j^i) \log P(q_i | b_j^i)$

= A POSTERIORI ENTROPY

**EX. CONSIDER A NOISY BINARY CHANNEL**

\begin{align*}
0 &\xrightarrow{2/3} 0 & 1 &\xrightarrow{2/3} 1 \\
0 &\xrightarrow{1/10} 1 & 1 &\xrightarrow{9/10} 0
\end{align*}

\[ P(\text{COND}) = \begin{bmatrix} 2/3 & 1/3 \\ 1/10 & 9/10 \end{bmatrix} \]

$P(q=0) = \frac{3}{4}$

$P(q=1) = \frac{1}{4}$

THIS IS ALL THE INFO WE NEED TO CHARACTERIZE THE CHANNEL.
\( P[0 \text{ IS RECEIVED } \& \text{ OUTPUT}] \)

\[
\begin{align*}
\frac{3}{4} \times \frac{3}{2} + \frac{1}{4} \times \frac{1}{10} &= \frac{21}{40} \\
\frac{3}{4} \times \frac{1}{3} + \frac{1}{4} \times \frac{4}{10} &= \frac{19}{40}
\end{align*}
\]

\( P[1 \text{ IS RECEIVED } \& \text{ OUTPUT}] \)

\[
\begin{align*}
P[a=0 \mid b=0] &= \frac{P[a=0, b=0]}{P(b=0)} \\
&= \frac{P[a=0, b=0]}{P(b=0)} \\
&= \frac{21}{40} \\ 
&= \frac{21}{40} \\
&= \frac{20}{21}
\end{align*}
\]

\( P[b=1, a=1] = P[b=1 \mid a=1] P[a=1] \)

\[
\begin{align*}
P[b=1, a=1] &= P[b=1 \mid a=1] P[a=1] \\
&= \frac{9}{10} \times \frac{1}{14} \\
&= \frac{9}{140}
\end{align*}
\]

\( P[a=1 \mid b=0] = \frac{P[b=0, a=1]}{P[b=0]} = \frac{1}{21} \)

\( P[a=0 \mid b=1] = \frac{10}{9} \)

**Finding Apriori and A posteriori Probabilities**

\[
\begin{align*}
H(A) &= \frac{3}{4} \log_2 \frac{1}{3} + \frac{1}{4} \log_2 4 \\
&= \frac{3}{4} \log_2 \frac{1}{3} + \frac{1}{4} \log_2 4 - \frac{3}{4} \log_2 3 \\
&= 2 - 1.585 \left( \frac{3}{4} \right) = 0.811 \text{ bits}
\end{align*}
\]

\[
\begin{align*}
H(A \mid b=0) &= -p(a=0 \mid b=0) \log_2 P(a=0 \mid b=0) \\
&\quad -p(1 \mid 0) \log_2 P(1 \mid 0) \\
&= \frac{20}{40} \log_2 \frac{21}{20} + \frac{1}{20} \log_2 21 \\
&= \log_2 21 - \frac{20}{21} \log_2 20 \\
&= (4.392) - \frac{20}{21} (4.322) = 0.276 \text{ bits}
\end{align*}
\]

\[
\begin{align*}
H(A \mid b=1) &= -p(0 \mid 1) \log_2 P(0 \mid 1) \\
&\quad -p(1 \mid 1) \log_2 P(1 \mid 1) \\
&= \frac{10}{19} \log_2 \frac{19}{10} + \frac{9}{19} \log_2 \frac{9}{19} \\
&= 0.998 \text{ bits}
\end{align*}
\]
A uniform channel: A channel described by its forward cond. matrix is said to be uniform if the elements in every row and every column of its matrix consist of an arbitrary permutation of the terms in the first row.

\[ \begin{bmatrix} b_1 & b_2 & b_3 & b_4 \\ a_1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ a_2 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \]

C = Channel Capacity = Max \( I (x; y) \)

= Trans. Info = Mutual Info

= Max [ \( H(x) - H(x/y) \) ]

= Max [ \( H(y) - H(y/x) \) ]

Let \( p(y_i/x_i) = \alpha_{i,j} \)

\( p(x_i, y_i) = \alpha_{i,j} \alpha_{j,i} \implies p(x_i) = \alpha_{i,i} \)

\[ P = \begin{bmatrix} p_{11} & \ldots & p_{1s} \\ \vdots & \ddots & \vdots \\ p_{n1} & \ldots & p_{ns} \end{bmatrix} \]

\( H(y/x_i) = -\sum_{j=1}^{m} p(y_j/x_i) \log p(y_j/x_i) \)
Thus, for a uniform channel, 
\[ H(Y/X;^2) = h = \text{const}. \]

And
\[ H(Y/X) = \frac{1}{n} \sum_{i=1}^{n} a_i h = h \]

Consider, then
\[ C = \max \left[ H(Y) - h \right] \]
\[ = \log s - h \]

Consider
\[ a_i \]
\[ \begin{array}{ccc}
0 & 1 & 0 \\
0 & 1 & 0 \\
3 & 0 & 0 \\
\end{array} \]
\[ a_i' \]
\[ \begin{array}{ccc}
1 & 0 & 1 \\
0 & 1 & 0 \\
1 & 2 & 0 \\
\end{array} \]
\[ b_i \]
\[ \begin{array}{ccc}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
\end{array} \]
\[ b_i' \]
\[ \begin{array}{ccc}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
\end{array} \]

\[ C = \log r - h' = \log s - h' \]
7/28/76 (WEO)

BSC (BINARY SYMMETRIC CHANNEL)

\[
\begin{array}{c}
\begin{array}{ccc}
& 0 & p \\
0 & & q \\
p & & 1 \\
q & & 0 \\
1 & & p
\end{array}
\end{array}
\]

- \[ P(0) = \alpha, \quad P(1) = 1 - \alpha \]
- \[ P(0|0) = P(1|1) = p \]
- \[ P(0|1) = P(1|0) = q \]

\[ H(X) = H(\alpha, 1-\alpha) \]

\[ = \alpha \log \frac{1}{\alpha} + (1-\alpha) \log 1 - \alpha \]

\[ H(Y|X) = -p \log p - q \log q \]

\[ I(X;Y) = H(Y) - H(Y|X) \]

\[ = H(Y) + pl \log p + q \log q \]

\[ \text{Max} [I(X;Y)] = 1 + pl \log p + q \log q \]

BIITS

BEC (BINARY ERASURE CHANNEL)

\[
\begin{array}{c}
\begin{array}{ccc}
& 0 & p \\
0 & & q \\
p & & 1 \\
q & & 0 \\
1 & & q
\end{array}
\end{array}
\]

- \[ p + q = 1 \]

\[ P(Y|X) = \begin{pmatrix} 0 & p & 0 \\ p & 0 & q \\ 0 & p & 0 \end{pmatrix} \]

- \[ P(0) = \alpha, \quad P(1) = 1 - \alpha = \bar{\alpha} \]

\[ H(X) = \alpha \log \frac{1}{\alpha} + \bar{\alpha} \log \frac{1}{\bar{\alpha}} \]

\[ H(X|Y) \] (we first need \( P(Y|X) \))
\[
\rho(x, y) = \begin{array}{ccc}
0 & 1 & \alpha p \\
\alpha q & \alpha q & 0 \\
\alpha p & \alpha p & \alpha p
\end{array} = \rho(y = y_x)
\]

\[
\rho(x/y) = \begin{array}{ccc}
\frac{\alpha p}{\alpha p} & \frac{\alpha q}{\alpha p} & 0 \\
\frac{0}{\alpha p} & \frac{(1-\alpha)q}{\alpha p} & \frac{(1-\alpha)p}{\alpha p} \\
\frac{0}{\alpha p} & \frac{0}{\alpha p} & \frac{0}{\alpha p}
\end{array}
\]

Now, find

\[
H(x/y) = \begin{array}{c}
\alpha p \log_2 1 + \alpha q \log_2 \alpha \\
\alpha q \log_2 0 + 0 - \alpha \log_2 \alpha + \alpha p \log_2 1
\end{array}
\]

\[
= \frac{q}{\rho} H(x)
\]

\[
\therefore I(x; y) = H(x) - (1-p)H(x)
\]

\[
= \rho H(x)
\]

\[
\text{maximize} \quad \{H(x)\} = 1 \text{ bit}
\]

\[
\therefore C = \text{max} I(x; y) = \rho \text{ bits}
\]
SECOND EXTENSION OF BSC

CLEARLY, IF 0, 1 COMPRISCE THE INPUT ALPHABET, FOR THE PRIMARY CHANNEL, THEN 00, 01, 10, 11 COMPRISCE THE SYMBOLS FOR INPUT (\( \frac{1}{2} \) OUTPUT) OF THE SECOND EXTENSION. ASSUME THE CHANNEL HAS ZERO MEMORY:

\[
U = \begin{bmatrix} x_1, x_2, x_1, x_2, x_1, x_2 \end{bmatrix}
\]
\[
V = \begin{bmatrix} y_1, y_2, y_1, y_2, y_1, y_2 \end{bmatrix}
\]

\[
P(x_1, x_2) = \frac{1}{8} P(x_1)P(x_2) = P(U)
\]
\[
P(y_1, y_2) = \frac{1}{8} P(y_1)P(y_2) = P(V)
\]
\[
P(v/u) = P(y_1, y_2 | x_1, x_2) = P(y_1 | x_1) P(y_2 | x_2)
\]
\[
P(u/v) = P(x_1 | y_1) P(x_2 | y_2)
\]
\[
P(u, v) = P(u)P(v/u).
\]

SOURCE ENTROPY:

\[
H^2(x) = H(U) = 2H(X)
\]
\[
= -2[\alpha \log_2 \alpha + \beta \log_2 \beta]
\]
\[
= H(x_1) + H(x_2)
\]
\[
H(U, V) = H(X_1, Y_1) + H(X_2, Y_2)
\]
\[
H(U/V) = H(Y_1/X_1) + H(Y_2/X_2)
\]
\[
H(U/V) = H(X_1/Y_1) + H(X_2/Y_2)
\]
\[
I(U; V) = H(U) - H(U/V) = H(X_1) + H(X_2) - H(x_2/y_2)
\]
\[
= 2I(X; Y)
\]
*⇒ HOMEWORK: USE (THIS) GRAPHICAL TECHNIQUE FOR THE FOLLOWING CHANNELS:

(a) \[
\begin{bmatrix}
0.9 & 0.1 \\
0.1 & 0.9
\end{bmatrix}
\]  (b) \[
\begin{bmatrix}
9/10 & 1/10 \\
1/10 & 9/10
\end{bmatrix}
\]

IF YOU NEED TO, ASSUME APPROPRIATE NUMBERS FOR INPUT S/OUTPUT PROBABILITIES.

ANALYTICAL TECHNIQUE FOR DETERMINING CAPACITY OF BINARY CHANNEL

[S. Muroga, "ON THE CAPACITY OF A DISCRETE CHANNEL" J. PHYS. SOC. JAP. 21, 1952]

Consider

0. \( p_{11} q_1 + p_{12} q_2 = p_{11} \log p_{11} + p_{12} \log p_{12} = -H(p_{11}) \)

\[
\begin{bmatrix}
p_{11} & p_{12} \\
p_{21} & p_{22}
\end{bmatrix}
\]

ALSO, LET

(2) \( p_{21} q_1 + p_{22} = p_{21} \log p_{21} + p_{22} \log p_{22} \)

\( I(x;y) = H(y) - H(y|x) \)

\( = -p_1 \log p_1 + p_2 \log p_2 + p_1 (p_{11} \log p_{11} + p_{12} \log p_{12}) + p_2 (p_{21} \log p_{21} + p_{22} \log p_{22}) \)
\[ \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = - \begin{bmatrix} H(p_{11}) \\ H(p_{22}) \end{bmatrix} \]

Use this solution to write:

\[ I(x, y) = -(p_1' \log p_1' + p_2' \log p_2') + p_1' q_1 + p_2' q_2 \]

We want to maximize with respect to \( p_1' \) and \( p_2' \) and resort to the Lagrange functional:

\[ U = -(p_1' \log p_1' + p_2' \log p_2') + p_1' q_1 + p_2' q_2 + \mu (p_1' + p_2') \]

\[ \frac{dU}{dp_1} = - (\log p_1' + \log 2^e) + q_1 + \mu = 0 \]
\[ \frac{dU}{dp_2} = - (\log 2^e + \log p_2') + q_2 + \mu = 0 \]

Solve these for \( \mu \):

\[ \mu_1 = -q_1 + (\log 2^e + \log p_1') \]
\[ \mu_2 = -q_2 + (\log 2^e + \log p_2') \]

Use these values of \( \mu \) in \( 5 \), one by one, thus:

Channel capacity \( C = \max I(x, y) = q_1 = \log p_1' = q_2 = \log p_2' \)

Which says that

\[ p_i' = 2^{q_i - C}, \quad i = 1, 2 \]

Or
\[ C = \log \left[ 2^{q_1} + 2^{q_2} \right] \] bits
WE MAY FIND $q_1 \neq q_2$

1. BY SOLVING MATRIX *A* TO GET $-63$.
2. CURVES ON HANDOUT #4 7/22/76

Ex. Let $p_{11} = p_{12} = \frac{1}{2}$, $p_{21} = \frac{1}{4}$, $p_{22} = \frac{3}{4}$

$$\begin{pmatrix}
\frac{1}{2} & \frac{1}{2} \\
\frac{1}{4} & \frac{3}{4}
\end{pmatrix}$$

FROM CHART, $q_1 = -1.4$ \[q_1 = 7.378\]
$q_2 = -0.6$ \[q_2 = 0.622\]

AS HOMEWORK, WORK BY MATRIX INVERSION

NOW, FROM CHART, $C = 0.06$ BITES

AS HOMEWORK, VERIFY THAT $C = 0.0425$

Use Muroga's technique to find

$$\begin{pmatrix}
\frac{3}{4} & \frac{1}{2} & \frac{1}{6} & 0 \\
\frac{1}{6} & \frac{3}{4} & 0 & \frac{1}{8} \\
\frac{1}{6} & \frac{1}{6} & \frac{3}{4} & 0 \\
0 & 0 & \frac{1}{4} & \frac{3}{4}
\end{pmatrix}$$
7/29/76 (Wed)

Comments on Muroga's Technique

(1) For a \( m \times m \) matrix, one needs to set up \( m \) simultaneous equations:

\[
p_{11} Q_1 + \ldots + p_{1m} Q_m = \sum_{j=1}^{m} p_{1j} \log p_{1j}
\]

\[
p_{m1} Q_1 + \ldots + p_{mm} Q_m = \sum_{j=1}^{m} p_{mj} \log p_{mj}
\]

Then

\[
I(x, y) = -\sum_{i=1}^{m} p_{i} \log p_{i} + \sum_{i=1}^{m} p_{i} \log Q_{i}
\]

\( M = \# \) of input and output symbols

\[
C = \log \sum_{i=1}^{m} 2^Q_{i}
\]

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 technique may not meet the requirement

\[
0 \leq p_i \leq 1 \quad \sum_i p_i = 1
\]

so watch out

(4) Silverman, Chang \& Lobb solved this in mid 50's
Codes and Their "Elementary" Properties
(Material from text, p. 46 on).

Read p. 51

Defn: Let the set of symbols comprising a given alphabet be called \( S = \{s_1, s_2, \ldots, s_9\} \). Then a code is a mapping (transformation) of all possible sequences of the symbols of \( S \) into other sequences of some other alphabet \( X = \{x_1, x_2, \ldots, x_3\} \).

\( S \) is called source alphabet and \( X \) is called code alphabet.

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 alphabet \( S \) into a fixed sequence of symbols of \( X \) (code alphabet). Such fixed sequences is called code words. A collection of code words is a code book.

Ex:  
\[ \begin{array}{c}
S_1 & 0 \\
S_2 & 11 \\
S_3 & 00 \\
S_4 & 11 \\
\end{array} \]
A block code is non-singular if all the words of the code are distinct. Otherwise, the block code is singular.

Ex: $s_1 = 0$
    $s_2 = 1$
    $s_3 = 00$
    $s_4 = 01$

But, suppose we receive 0011, could be $s_1 s_2$ 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 fixed sequences 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 $A_i$ into words $X_i$ is the block code which maps sequences of source symbols $(S_{1_1}, S_{2_1}, \ldots, S_{n_1})$ into sequences of code words $(X_{i_1}, X_{i_2}, \ldots, X_{i_n})$.

Ex: $S_1 = 0$, $S_1 S_2 = 00$
$S_2 = 11$
$S_1 S_2 = 011$

Ex: $S_3 = 00$
$S_3 S_4 = 001$
$S_4 = 01$

$S_3 S_4 S_5 = 001$
DEF: A BLOCK CODE IS DECODABLE IF THE NTH EXTENSION OF THE CODE IS NON-SINGULAR FOR EACH FINITE N.

EX.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>S_1</td>
<td>00</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>S_2</td>
<td>01</td>
<td>10</td>
<td>01</td>
</tr>
<tr>
<td>S_3</td>
<td>10</td>
<td>110</td>
<td>011</td>
</tr>
<tr>
<td>S_4</td>
<td>11</td>
<td>1110</td>
<td>0111</td>
</tr>
</tbody>
</table>

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, you always gotta look at next symbol.
DEF: A UNIQUELY DECODABLE CODE IS INSTANTANEOUS 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 $x_i = x_{i_1}x_{i_2}...x_{i_n}$ be a word. The sequence $(x_{i_1}, x_{i_2}, ..., x_{i_j})$ $0 \leq j \leq n$ is called a prefix of $x_i$.

Ex. $x_i = 0111$
prefixes: 0, 01, 011, 0111
prefix property
a neccessary & 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 alphabet be \((0, 1)\), say the codebook has 5 words arising from:

\[
\begin{align*}
S_1 &= 0 \quad (1) \\
S_2 &= 10 \quad (2) \\
S_3 &= 110 \quad (3) \\
S_4 &= 1110 \quad (4) \\
S_5 &= 1111 \quad (4)
\end{align*}
\]

Another is:

\[
\begin{align*}
S_1 &= 00 \quad (2) \\
S_2 &= 01 \quad (2) \\
S_3 &= 10 \quad (2) \\
S_4 &= 110 \quad (3) \\
S_5 &= 111 \quad (3)
\end{align*}
\]

Homework: Start with \(10 \frac{1}{2}\) build...
7/30/76 (FRI)

Code $\rightarrow$ Block $\rightarrow$ Non-Singular $\rightarrow$ Uniquely Decodable $\rightarrow$ Instant.

A necessary and sufficient condition for instant code $\rightarrow$ meet the prefix property.

**Kraft (1949)**

Let us seek to build an instant code from $S: A_1, A_2, \ldots, A_q$ be a code alphabet.

$x: x_1, x_2, \ldots, x_p$.

We will consider $x: 0, 1$.

Let $x_1, x_2, \ldots, x_q$ are words out of the alphabet $x$ with the corresponding lengths $l_1, l_2, \ldots, l_q$. It is conceivable that some sequences of $x_i$ have the same $l_i$.

The length of code word $= \# \text{ digits in } x_i$.

Kraft's Inequality

A necessary and sufficient condition for the existence of instant codes with word length $l_1, l_2, \ldots, l_q$ is

$$\sum_{i=1}^{q} 2^{-l_i} \leq 1 \quad \exists r = \# \text{ of digits in the encoding alphabet. For } r = 2$$

$$\sum_{i=1}^{q} 2^{-l_i} \leq 1$$

Note: This does **not** assure that the code is instantaneous.
Look at some codes \((r=2)\)

<table>
<thead>
<tr>
<th>Source Symbol</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S_1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(S_2)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(S_3)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(S_4)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
A: & \quad 2^{-2} = 1 : A \text{ is also instant} \\
B: & \quad 2^{-1} + 2^{-3} + 2^{-3} + 2^{-3} = \frac{7}{8} : B \text{ is also instant} \\
C: & \quad 2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} \leq 1 : C \text{ is \underline{not} instant} \\
D: & \quad 2^{-1} + 2^{-3} + 2^{-3} + 2^{-2} < 1 \text{ but } D \text{ is \underline{not} instant} \\
E: & \quad 2^{-1} + 2^{-2} + 2^{-3} + 2^{-2} > 1 \\
\end{align*}
\]

This code \underline{cannot} be instantaneous.

Example using Kraft inequality:

Code the decimal alphabet \(0, 1, \ldots, 9\) by using the coding alphabet \((0, 1)\) \((r=2)\)

\[
\begin{align*}
0 & \rightarrow 0 \\
1 & \rightarrow 10 \\
2 & \\
3 & \\
4 & \\
5 & \\
6 & \\
7 & \\
8 & \\
9 & \\
\end{align*}
\]

Require \(L\) for the rest have equal length. We seek an instantaneous code.
Using Kraft's Inequality, we require

\[ 2^{-1} + 2^{-2} + 8 \times 2^{-2} \leq 1 \]
\[ 2^{-2} \leq \frac{1}{32} \]
\[ \Rightarrow \frac{1}{2^2} \leq \frac{1}{32} \Rightarrow L \geq 5, \text{ so let } L = 5 \]

Let's build it

<table>
<thead>
<tr>
<th>( j )</th>
<th>( \theta_j )</th>
<th>( \theta_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>1001</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>10010</td>
<td>9</td>
</tr>
</tbody>
</table>

Proof of Kraft's Inequality

Sufficiency: \( \sum_{i=1}^{q} \theta_i \leq 1 \) and we need to show that the selection of \( L \) based on this equality will generate instantaneous codes. Let \( \theta_i = \# \) of code words with length \( i \). \( \Rightarrow \max i = L \).

The codes \( X_1, X_2, \ldots, X_q \) have length \( L_1, L_2, \ldots, L_q \).

Then \( q = \sum_{i=1}^{L} \theta_i \Rightarrow L \) is longest length code in the book.

\( \theta_L \) can then be written as

\[ \sum_{i=1}^{L} \theta_i r^{-L} \leq 1 \]

(Convince yourself by opening summation on left.)

Count terms like \( \frac{1}{r}, \frac{1}{r^2}, \ldots, \frac{1}{r^L} \)
EXPANDING THIS:

\[ \sum_{i=1}^{d} n_i r^{-i} = n_1 r^{-1} + n_2 r^{-2} + \ldots + n_d r^{-d} \]

\[ \leq 1 \]

\[ \implies n_d r^{-d} + n_{d-1} r^{-d+1} + \ldots + n_2 r^{-2} = n_{d-1} r^{-d} \]

or

\[ n_d \leq r^{-d} - n_1 r^{-1} - n_2 r^{-2} - \ldots - n_{d-1} r^{-d} \]

\[ \implies 0 \leq r^{-d} - n_1 r^{-d-1} - n_2 r^{-d-2} - \ldots - n_{d-1} r^{-1} \]

\[ n_{d-1} r^{-d} \leq r^{-d} - n_1 r^{-d-1} - n_2 r^{-d-2} - \ldots - n_{d-2} r^{-2} \]

\[ \implies n_{d-1} \leq r^{-d-1} - n_1 r^{-d-2} - n_2 r^{-d-3} - \ldots - n_{d-2} r^{-2} \]

\[ \ldots \]

\[ n_2 \leq r^{-2} - n_1 r^{-1} \]

\[ n_1 \leq r \]

THE CLAIM IS THAT THESE \( n_i \) ARE ADEQUATE TO BUILD AN INSTANT CODE.

\[ n_1 \leq r \implies \text{for } r \text{ elements, you can only} \]

\[ \text{build, at most, } n_1 \text{ element instant code} \]

\[ \implies r - n \text{ are left to build other words of the book.} \]

\[ n_2 \leq r(r-n) = \# \text{ of ways we can build instant codes of length two.} \]

FIRST TWO LOCATIONS HAVE BEEN TAKEN UP. \( (r^2 - n_1 r - n_2) r = \# \text{ of ways instantaneously codes of length} \]

\[ n_3 = 3. \]

HOMEWORK: 3.2, 3.3

READ "NOTES" p. 61-62.
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 $X_1 \neq X_k$ 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
$n_i \leq r \cdot 2^r = \#$ OF DIGITS IN ENCODING
ALPHABET $T$. THEN THE NUMBER
OF ENCODED MESSAGES OF LENGTH $2$,
$n_2$ CANNOT BE $> (r - n_1) r$,
ALONG THE SAME LINES
$n_3 < \left[ (r - n_1) r - n_2 \right] r$
$< r^3 - n_1 r^2 - n_2 r$

PROCEEDING LIKEWISE:
$n_m = \#$ OF ENCODED WORDS WITH
LENGTH $m$
$< r^m - n_1 r^{m-1} - n_2 r^{m-2} - \ldots - n_{m-1} r$
$0 \leq r^m - n_1 r^{m-1} - n_2 r^{m-2} - \ldots - n_{m-1} r - n_m r^{m-1}$
$0 \leq 1 - n_1 r^{-1} - n_2 r^{-2} - \ldots - n_{m-1} r^{-m+1} - n_m r^{-m}$
$\sum_{i=1}^{m} n_i r^{-i} \leq 1$
IF $m$ IS THE # OF DIGITS IN THE LARGEST
WORD, THE $m=2$.

$$\sum_{i=1}^{2} n_i r^{-i} \leq 1.$$  

BUT, AS WE SAID ON FRIDAY:

$$\sum_{i=1}^{q} n_i r^{-i} = \sum_{i=1}^{q} r^{-i} \quad (which \ is \ K)$$

* WILL NOW SHOW THAT $\sum_{i=1}^{2} n_i r^{-i} = \sum_{i=1}^{q} r^{-i}$.

LET $X = x_1, x_2, \ldots x_7$

LET $l_1 = 2 \quad l_2 = 2 \quad l_3 = 3 \quad l_4 = 3 \quad l_5 = 3 \quad l_6 = 4 \quad l_7 = 5$

$L_i$ IS THE LENGTH OF $X$.

$n_1 = 0 \quad n_2 = 2 \quad n_3 = 3 \quad n_4 = 1 \quad n_5 = 1 \quad n_6 = 1 \quad n_7 = 1$

$n_q$ IS THE NUMBER OF WORDS WITH LENGTH $L_i$.

$L_i = \text{LENGTH OF LONGEST WORD} = 5$

COMPUTING ($q = 7$)

$$\sum_{i=1}^{2} n_i r^{-i} = 2 \frac{1}{r^2} + 3 \frac{1}{r^3} + 1 \frac{1}{r^4} + 1 \frac{1}{r^5}$$

$$= \frac{1}{r^2} + \frac{1}{r^2} + \frac{1}{r^3} + \frac{1}{r^3} + \frac{1}{r^3} + \frac{1}{r^4} + \frac{1}{r^5}$$

$$\sum_{i=1}^{7} r^{-i} = \frac{1}{r^2} + \frac{1}{r^3} + \frac{1}{r^3} + \frac{1}{r^3} + \frac{1}{r^4} + \frac{1}{r^5}$$

IN OTHERWORDS, 1) IS A # OF STRINGS

$N_k$ EACH OF THE TYPE $\frac{1}{r^k}$ HAVING

$K$ TERMS IN EACH STRING.
**McMillan's Inequality**

(for uniquely decodable codes

and 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 YES. In, given

a uniquely decodable code, is

satisfies Kraft's inequality.

(McMillan's inequality (1956))

Consider \((\sum_{i=1}^{q} r^{-l_i})^{n} \geq n\)

is real, \(n\) \(l_i\) is an integer,

\((\sum_{i=1}^{q} r^{-l_i})^{n} \geq (r^{-l_1} + r^{-l_2} + \ldots + r^{-l_q})^{n}

of the type \(r^{-k} = \sum_{i=1}^{q} r^{-l_i} = r^{-k}\)

\(k = l_1 + l_2 + l_3 + \ldots + l_q\)

If \(k\) is length of longest word,

\(n \leq k \leq n l\)
Let, as before, \( N_k = \# \) of terms of the form \( r^{-k} \). Then from before

\[
\left( \sum_{i=1}^{q} r^{-\ell_i} \right)^n = \sum_{k=0}^{n} N_k r^{-k} \quad \text{just as reordering}
\]

where \( N_k = \# \) of strings of \( n \) code words that can be found so that each strip is of length \( k \).

Obviously, \( N_k \leq r^k \)

\[
\left( \sum_{i=1}^{q} r^{-\ell_i} \right)^n = \sum_{k=0}^{n} N_k r^{-k} \leq \sum_{k=0}^{n} r^k r^{-k} = \sum_{k=0}^{n} 1 = n + 1 \leq n!
\]

This gives

\[
\left( \sum_{i=1}^{q} r^{-\ell_i} \right)^n \leq n!
\]

Recall, \( x^n \leq n! \Rightarrow x < 1 \)

Then this gives McMillan's inequality:

\[
\sum_{i=1}^{q} r^{-\ell_i} < 1
\]

This is Krafts inequality as applied to uniquely decodable codes.
EX: LET THE # OF WORDS WITH
EQUAL LENGTHS BE
\[ [n_1, n_2, n_3, n_4] = [0, 3, 0, 5] \]
# WORDS LENGTH 1 = 0
# WORDS LENGTH 2 = 3
# WORDS LENGTH 3 = 0
# WORDS LENGTH 4 = 5

\[ \Rightarrow \text{THE CODE BOOK HAS 5 WORDS.} \]

QUESTION: CAN ONE BUILD AN
INTEGRAL CODE WITH BINARY
ALPHABET \( (0, 1) \). THEN
\[ \sum_{i=0}^{4} n_i r^{-i} \leq 1 \]
\[ \sum_{i=0}^{4} n_i r^{-i} = \sum_{i=1}^{8} r^{-2i} \leq 1 \]
\[ \sum_{i=0}^{4} n_i r^{-i} = \frac{3}{r^2} + \frac{5}{r^4} \leq 1 \]
NOT SATISFIED. FOR \( r = 2 \), GOTTA
LOOK FOR ANOTHER, ALPHABET
\[ \frac{3}{r^2} + \frac{5}{r^4} < 1 \]
\[ r^2 + 5 \leq r^4 \]
\[ r^4 - 3r^2 - 5 \geq 0 \]
FOR EQUALITY,
\[ r^2 = \frac{3 \pm \sqrt{9 + 4 \cdot 5}}{2} = 4.2, \frac{3 - 5.4}{-2} \]
\[ r = 2.1 \]
\[ \Rightarrow r = 3 \]
LET, THEN, THE ALPHABET BE \((0, 1, 2)\)
A scheme for \(n = 5\) is
\[
\begin{array}{ccc}
00 & 1000 & 2000 \\
01 & 1001 & 2222 \\
02 & 1002 &
\end{array}
\]

As homework, devise two instant encoding schemes with \(r = (0, 1, 2)\) for this same example.

\((\text{ex})\): 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. \(\& n; \leq 3\).

Now,\[
\sum_{i=1}^{3} n_i r^{-i} \leq 1
\]
\(l_1 2^{-1} + l_2 2^{-2} + l_3 2^{-3} \leq 1\)
(Also, \(l_1 + l_2 + l_3 = q = 3\)
\(\iff l_1 + 2l_2 + l_3 \leq 8\)
\(l_1 + l_2 + l_3 = 0\)
\[ l_1, l_2, l_3 \Rightarrow \text{we have to satisfy} \]
\[ 4l_3 + 2l_2 + l_3 \leq 8 \quad \text{and} \quad l_1 + l_2 + l_3 = 3 \]

All these will work

\[ \rightarrow \text{check them} \]

\[ \text{Shannon's First Theorem} \]

Defn: Average length of a coding scheme:

\[ \overline{L} = \sum_i p(x_i)l_i \]

\[ \exists X = \{x_1, x_2, \ldots, x_n\} \]
\[ l_1, l_2, \ldots, l_n \]

We seek to keep \( \overline{L} \) as low as possible

It turns out

\[ \overline{L} \geq H(x)/\log \]
8/31/76

SHANNON'S FIRST THEOREM (NOISELESS CASE)

(a) The average length of a coding scheme cannot be reduced below \( \frac{H(x)}{\log r} \),

\( = H(x) \) for binary alphabet.

\[ L \leq \sum_{i=1}^{q} \frac{p(x_i)l_i}{\log r} \]

\( q \) = total # of messages

\( l_i \) = length of \( i \)th message

\( x_i = \) prob of occurrence of \( x_i \) message.

Proof: Consider \( p_i \equiv \frac{q_i}{q} \)

\[ \Rightarrow \sum_{i=1}^{q} p_i = \sum_{i=1}^{q} q_i = 1 \]

We have proved that

\[ -\sum_{i=1}^{q} q_i \log q_i \leq -\sum_{i=1}^{q} p(x_i) \log p(x_i) = H(x) \]

\[ \Rightarrow \sum_{i=1}^{q} p(x_i) \log p(x_i) \leq \sum_{i=1}^{q} p(x_i) \frac{q}{\sum_{i=1}^{q} q_i} \log \frac{q}{\sum_{i=1}^{q} q_i} \log r \]

\[ \leq \sum_{i=1}^{q} p(x_i) \log r^{-l_i} + \sum_{i=1}^{q} p(x_i) \log \frac{q}{\sum_{i=1}^{q} q_i} \log r^{-l_i} \]

\[ \leq \sum_{i=1}^{q} p(x_i) \log r + \sum_{i=1}^{q} p(x_i) \log \frac{q}{\sum_{i=1}^{q} q_i} \log r^{-l_i} \]

\[ \leq \sum_{i=1}^{q} p(x_i) \log r + \sum_{i=1}^{q} p(x_i) \log \frac{q}{\sum_{i=1}^{q} q_i} \log r^{-l_i} \]

But \( \sum_{k=1}^{q} r^{-l_k} < 1 \) for u.d.c.

or \( H(x) \leq \sum \frac{H(x)}{\log r} \)

\[ \Rightarrow \sum \geq \frac{H(x)}{\log r} \]

For uniquely decidable codes, the average length cannot be valued below \( \frac{H(x)}{\log r} \).
IF THE RESTRICTION ON THE CODE BEING U.D. IS RELAXED, A LOWER \( \bar{I} \) MAY BE ACHIEVED.

IN CASE THE LOWEST \( \bar{I} = 0 \) IS NOT ATTAINABLE, THE ATTAINABLE LOW AQUIRED AS UNDER:

CONSIDER A WORD (MESSAGE) OF LENGTH \( l_K \) =

0 \[ - \frac{1}{\log r} \frac{\log r}{\log r} \leq l_K \leq \frac{1}{\log r} + 1 \]

AVERAGE LHS OF INEQUALITY OVER \( P(x_i) \forall i \in \mathbb{N} \)

\[ H(x) \frac{1}{\log r} \leq l \leq 1 + \frac{H(x)}{\log r} \]

NOW, LET \( r = 2 \), RAISE THE LHS OF (1): \[ \left[ p(x_{k}) \right]^{-1} \leq 2 \cdot \frac{l}{l} \]

ON \( p(x_{k}) \geq 2^{-l} \)

WHICH SATISFIES KRAFT. THUS, UNIQUELY DECODABLE CODES EXIST WHICH FOLLOW THE PATTERN OF PROBABILITIES GIVEN BY \( p(x_1) = \frac{1}{2}, p(x_2) = \frac{1}{2}, p(x_3) = \frac{1}{16}, p(x_4) = \frac{1}{6} \)

IF THE RESTRICTION OF U.D. IS ELIMINATED, A LOWER \( \bar{I} \) IS POSSIBLY

CONSIDER:

\( x_1 = 1 \), \( p_1 = 0.4 \), \( H(x) = 0.46 \)

\( x_2 = 0 \), \( p_2 = 0.4 \), \( I = 1.4 \)

\( x_3 = 100 \), \( p_3 = 0.3 \), \( H(x)/\log_{10} 2 = 1.53 \)

\( \bar{I} > 1.4 \)
Example:

\[ X = x_1, x_2, x_3, x_4, x_5, x_6 \]

\[ P = 2^{-1} 2^{-2} 2^{-4} 2^{-4} 2^{-4} 2^{-4} \]

\[ \exists \mathbf{\Sigma}; P(x_1) = 1 \]

<table>
<thead>
<tr>
<th></th>
<th>c_1</th>
<th>c_2</th>
<th>c_3</th>
<th>c_4</th>
<th>c_5</th>
<th>c_6</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>111</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>x_2</td>
<td>10</td>
<td>011</td>
<td>10</td>
<td>110</td>
<td>01</td>
<td>01</td>
</tr>
<tr>
<td>x_3</td>
<td>110</td>
<td>010</td>
<td>110</td>
<td>101</td>
<td>0011</td>
<td>011</td>
</tr>
<tr>
<td>x_4</td>
<td>110</td>
<td>001</td>
<td>110</td>
<td>100</td>
<td>0010</td>
<td>0111</td>
</tr>
<tr>
<td>x_5</td>
<td>101</td>
<td>000</td>
<td>1110</td>
<td>011</td>
<td>0001</td>
<td>0111</td>
</tr>
<tr>
<td>x_6</td>
<td>110</td>
<td>110</td>
<td>1110</td>
<td>010</td>
<td>0000</td>
<td>0111</td>
</tr>
</tbody>
</table>

(1) Which one of these schemes is U.D.?

- C_1: Blows Kraft Ineq. (not instant or U.D.)
  \[ C_1: \frac{1}{2} + \frac{5}{8} = \frac{7}{8} \Rightarrow \text{Blows Kraft ("111111"}) \]

- C_2: \[ \frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64} = \frac{63}{64} \]
  \[ \Rightarrow \text{This code is U.D.,} \]

- C_3: OK
  \[ \frac{1}{2} + \frac{1}{4} + \frac{1}{16} = 1 \Rightarrow 0101 \]

- C_5: \[ \frac{1}{2} + \frac{1}{4} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64} = \frac{63}{64} \Rightarrow \text{OK} \]

(2) Which ones of U.D. codes observe the prefix property?

- Only C_3, C_4, C_5

(3) Find L for each UDC:

- \[ L_3 = 1 + \frac{1}{2} + \frac{2}{4} + \frac{1}{4} + \frac{1}{16} + \frac{3}{32} + \frac{1}{64} = \frac{21}{8} \]
- \[ L_4 = 3 \]
- \[ L_5 = 2 \]
- \[ L_6 = 2 \frac{1}{8} \]
(4) Does any of these codes reach the lowest $L = L_0$?

Well, $L \geq H(X)$ by Shannon's First

$$H(X) = \frac{1}{2} \log 2 + \frac{1}{2} \log 4 + \frac{1}{2} \log 8$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{3}{4} = 1 \frac{3}{4} \text{ bits}$$

Best code is $C_5$ (i.e., most efficient)

** SPECIFIC CODING PROCEDURES **

(1) Shannon's Binary Coding Procedure

We have seen that if $p(x_k) \geq 2^{-2k}$

Then it is possible to achieve a realistic low $L$.

Thus $H(X) \leq L \leq H(X) + 1$

Steps:

(a) Arrange the ensemble in decreasing order of probs.

$p_1 \leq p_2 \leq p_3 \leq ... \leq p_q$

(b) Compute $\alpha_k$'s.

$\alpha_1 = 0$ always

$\alpha_2 = p(x_1)$

$\alpha_3 = p(x_2) + p(x_1) = p(x_2) + \alpha_2$

$\alpha_4 = p(x_3) + p(x_2) + p(x_1) = p(x_3) + \alpha_3$

$\vdots$

$\alpha_{q+1} = p(x_q) + p(x_{q-1}) + ... + p(x_2) + p(x_1) = 1$
(c) Determine the set of integers (the smallest) which satisfies
\[ 2^{-2^i} p(x_i) \geq 4 \quad \text{(as per the theorem)} \]
(d) Expand each \( \alpha_i \) (which is in decimal form) into binary notation up to \( l_i \) places

\[ \frac{1}{10} \quad \frac{2}{10} \quad \frac{5}{10} \quad \frac{1}{10} \]

\begin{align*}
\alpha_1 &= 0.10 \quad \frac{1}{5} \geq 10 \quad \Rightarrow l_1 = 2 \\
\alpha_2 &= 0.101 \quad \frac{3}{10} \geq 10 \quad \Rightarrow l_2 = 2 \\
\alpha_3 &= 0.101 \quad \frac{2}{10} \geq 10 \quad \Rightarrow l_3 = 3 \\
\alpha_4 &= 0.110 \quad \frac{1}{10} \Rightarrow l_4 = 4 \\
\end{align*}

\[ x_1 = 00 \quad x_2 = 01 \quad x_3 = 101 \quad x_4 = 1110 \]

\[ \frac{p_1}{7} = 4 \quad \frac{p_2}{7} = 3 \quad \frac{p_3}{7} = 0.2 \quad \frac{p_4}{7} = 0.1 \]

\( \Rightarrow \) Compute \( E \left( \frac{1}{7} \right) H(x) \) and compare.
1. Construct a Huffman code for the following symbols, compare its average length, $\bar{L}$, with the average uncertainty $H(x)$.

\[ x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13} \]
\[ \{0.2, 0.18, 0.1, 0.10, 0.10, 0.061, 0.059, 0.01, 0.04, 0.04, 0.03, 0.01 \} \]

Huffman codes called minimum redundancy code.

2. Derive a Shannon code for

\[ x: \begin{cases} x_1, x_2, x_3 \end{cases} \text{ and find the codes, } \bar{L}, \text{ and corresponding efficiencies for 2nd and 3rd extension} \]

\[ \Rightarrow \star \text{ 3. Prove, that for a Huffman code:} \]

\[ H(x) \leq \bar{L} \leq H(x) + 1 - 2p_{\min} \]

\[ \Rightarrow p_{\min} = P[\text{least probable message in the ensemble}] \]
SHANNON-FANO CODING SCHEME
SEPARABLE BINARY CODES,
DEALS WITH BINARY CODE TREES,

IF \( X = (x_1, x_2, \ldots, x_n) \),
\( \{p_1, p_2, \ldots, p_n\} \),
THEN IT IS DESIRABLE TO
ASSOCIATE A SEQUENCE \( C_K \)
of \((0,1)\) OF UNSPECIFIED LENGTH
\( 2^K \) FOR \( X \), SUCH THAT

1) THE PREFIX PROPERTY IS OBSERVED.
2) THE TRANSMISSION OF THE ENCODED
MESSAGES IS "REASONABLY" EFFICIENT,
ie, IT APPEARS INDEPENDENTLY
AND WITH (ALMOST) EQUAL PROBS.
### Let \( X \):  

| \( x_1 \) | 0.25 | 00 | (2) |
| \( x_2 \) | 0.25 | 01 | (2) |
| \( x_3 \) | 0.125 | 100 | (3) | cut 1 |
| \( x_4 \) | 0.125 | 101 | (3) | #3 |
| \( x_5 \) | 0.0625 | 1100 | (4) | cut 2 |
| \( x_6 \) | 0.0625 | 1101 | (4) | #4 |
| \( x_7 \) | 0.0625 | 1110 | (4) | #3 |
| \( x_8 \) | 0.0625 | 1111 | (4) | #4 |

**Note:** Prob. s given in this example are the type \( P(x) = 2^{-Q(x)} \), in such a case, \( H(x) = \sum P(x) \log P(x) \) 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 |  |  
| \( x_2 \) | 0.11 | 100 | #1 |
| \( x_3 \) | 0.14 | 101 | #2 |
| \( x_4 \) | 0.07 | 1100 |  
| \( x_5 \) | 0.07 | 1101 | #3 |
| \( x_6 \) | 0.04 | 1110 |  
| \( x_7 \) | 0.02 | 11110 |  
| \( x_8 \) | 0.02 | 111110 |  
| \( x_9 \) | 0.01 | 111111 |  


Consider alphabet be 0, 1, 2.

<table>
<thead>
<tr>
<th>X</th>
<th>P</th>
<th>(P(X))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(\frac{3}{16})</td>
<td>0.1875</td>
</tr>
<tr>
<td>1</td>
<td>(\frac{1}{6})</td>
<td>0.1667</td>
</tr>
<tr>
<td>2</td>
<td>(\frac{1}{6})</td>
<td>0.1667</td>
</tr>
<tr>
<td>3</td>
<td>(\frac{1}{6})</td>
<td>0.1667</td>
</tr>
<tr>
<td>4</td>
<td>(\frac{1}{6})</td>
<td>0.1667</td>
</tr>
<tr>
<td>5</td>
<td>(\frac{1}{6})</td>
<td>0.1667</td>
</tr>
<tr>
<td>6</td>
<td>(\frac{1}{6})</td>
<td>0.1667</td>
</tr>
</tbody>
</table>

\[ P(0) = \frac{16}{3} \quad P(1) = \frac{13}{3} \quad P(2) = \frac{10}{39} \]

Then find \(H(X), \bar{E}, \frac{1}{2} N\)

\[ N = \frac{H(X)}{\bar{E}} \]

**Shannon's First Theorem, Part (b)**

(Or Shannon's Noiseless Coding Theorem)

We have seen that, if \(L_i\) is chosen \(\exists \frac{1}{\log r} \leq L_i \leq \frac{1}{\log r} + 1\)

Then:

\[ \frac{1}{P_i} \leq r \]

And also \(H_r(s) \leq I \leq H_r(s) + 1\)

If the source were extended (w/o memory), then

\[ H_r(s^n) \leq \bar{L} \leq H_r(s^n) + 1 \]

\[ n H_r(s) \leq \bar{L} \leq n H_r(s) + 1 \]

\[ \bar{E} = \sum \frac{1}{P_i} \cdot P(O_i) \cdot \lambda_i \]

\[ \geq P(O_i) = P(X_1) \ldots P(X_n) \text{ etc.} \]
\[ H_f(s) \leq \frac{\ln n}{n} \leq H_r(s) + \frac{1}{n} \]

As \( n \to \infty \), we get
\[ \lim_{n \to \infty} \frac{\ln n}{n} = H_r(s) \]

\[
\frac{\ln n}{n} \text{ Average # of code symbols used per sample symbols from the original source.}
\]

**Notes:**

1. **The price one pays for large extensions is the complexity of the coding.**
2. **Even though, asymptotically, one approaches the above limit, the procedure does not necessarily yield a monotonically increasing improvement (or efficiency)**
   \[ \ln n \geq \ln n - 1 \]
   **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 \( L_r/P_i \leq L_i \leq L_r/P_i + 1 \))
HUFFMAN'S MINIMUM REDUNDANCY (MAX EFFICIENCY) SEPARABLE (INST) CODES.


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:

\[ P(x_1) \geq P(x_2) \geq \ldots \geq P(x_q) \]

\[ \Rightarrow L(x_1) \leq L(x_2) \leq \ldots \leq L(x_q) \]

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 \( q - n_0 / r - 1 = \text{INTEGER} \).
8-5-76 (Thurs.)

HUFFMAN CODES (Cont.)

(1) ---

(2) ---

This statement suggests that the total # of messages must
where \( \alpha \) is an integer.

(3) SOURCE REDUCTION (Auxiliary Ensemble)

of order 1, 2, 3, ..., 

\[
\begin{align*}
S_1 &= 0.4 \\
S_2 &= 0.3 \\
S_3 &= 0.1 \\
S_4 &= 0.06 \\
S_5 &= 0.04 \\
S_6 &= 0.4 \\
S_7 &= 0.4 \\
S_8 &= 0.3 \\
S_9 &= 0.2 \\
S_{10} &= 0.1 \\
S_{11} &= 0.1 \\
S_{12} &= 0.1 \\
S_{13} &= 0.1 \\
S_{14} &= 0.1 \\
S_{15} &= 0.1 \\
S_{16} &= 0.1 \\
S_{17} &= 0.1 \\
S_{18} &= 0.1 \\
S_{19} &= 0.1 \\
S_{20} &= 0.1 \\
S_{21} &= 0.1 \\
S_{22} &= 0.1 \\
S_{23} &= 0.1 \\
S_{24} &= 0.1 \\
S_{25} &= 0.1 \\
S_{26} &= 0.1 \\
S_{27} &= 0.1 \\
S_{28} &= 0.1 \\
S_{29} &= 0.1 \\
S_{30} &= 0.1 \\
S_{31} &= 0.1 \\
S_{32} &= 0.1 \\
S_{33} &= 0.1 \\
S_{34} &= 0.1 \\
S_{35} &= 0.1 \\
S_{36} &= 0.1 \\
S_{37} &= 0.1 \\
S_{38} &= 0.1 \\
S_{39} &= 0.1 \\
S_{40} &= 0.1 \\
S_{41} &= 0.1 \\
S_{42} &= 0.1 \\
S_{43} &= 0.1 \\
S_{44} &= 0.1 \\
S_{45} &= 0.1 \\
S_{46} &= 0.1 \\
S_{47} &= 0.1 \\
S_{48} &= 0.1 \\
S_{49} &= 0.1 \\
S_{50} &= 0.1 \\
S_{51} &= 0.1 \\
S_{52} &= 0.1 \\
S_{53} &= 0.1 \\
S_{54} &= 0.1 \\
S_{55} &= 0.1 \\
S_{56} &= 0.1 \\
S_{57} &= 0.1 \\
S_{58} &= 0.1 \\
S_{59} &= 0.1 \\
S_{60} &= 0.1 \\
S_{61} &= 0.1 \\
S_{62} &= 0.1 \\
S_{63} &= 0.1 \\
S_{64} &= 0.1 \\
S_{65} &= 0.1 \\
S_{66} &= 0.1 \\
S_{67} &= 0.1 \\
S_{68} &= 0.1 \\
S_{69} &= 0.1 \\
S_{70} &= 0.1 \\
S_{71} &= 0.1 \\
S_{72} &= 0.1 \\
S_{73} &= 0.1 \\
S_{74} &= 0.1 \\
S_{75} &= 0.1 \\
S_{76} &= 0.1 \\
S_{77} &= 0.1 \\
S_{78} &= 0.1 \\
S_{79} &= 0.1 \\
S_{80} &= 0.1 \\
S_{81} &= 0.1 \\
S_{82} &= 0.1 \\
S_{83} &= 0.1 \\
S_{84} &= 0.1 \\
S_{85} &= 0.1 \\
S_{86} &= 0.1 \\
S_{87} &= 0.1 \\
S_{88} &= 0.1 \\
S_{89} &= 0.1 \\
S_{90} &= 0.1 \\
S_{91} &= 0.1 \\
S_{92} &= 0.1 \\
S_{93} &= 0.1 \\
S_{94} &= 0.1 \\
S_{95} &= 0.1 \\
S_{96} &= 0.1 \\
S_{97} &= 0.1 \\
S_{98} &= 0.1 \\
S_{99} &= 0.1 \\
S_{100} &= 0.1 \\
\end{align*}
\]

If \( r = \# \) symbols in encoding alphabet = 2.

Work backwards.

(4) Consider the \( j^{th} \) reduction \( S_{j0} \).

One of its messages has originated by the merger of
2 messages from \( S_{j-1} \). Let
that message be \( S_{j0} \). Let
the merging messages of \( S_{j-1} \) be
\( S_{j0} \) \( \leq S_{j1} \). Other messages of
\( S_{j} \) came directly from
\( S_{j-1} \) \( \text{w/o merging} \).

Now, assign to each message...
$S_{j-1}$ (except $S_0$ and $S_1$ which are merging) the code words used by corresponding messages of $S_j$. The code words $S_{j+1} = 0 \frac{1}{2} S_j$, are formed by adding $0 \frac{1}{2} 1$ respectively to the code word used for $S_j$.

HW's code reduction sample on previous page.

Ex: $X = x_1 \ldots x_9$, $\frac{1}{2} P(x_9) = \frac{1}{4} v_2$

\begin{align*}
X_1 & : 011, \\
X_2 & : 111, \\
X_3 & : 010, \\
X_4 & : 01, \\
X_5 & : 010, \\
X_6 & : 000, \\
X_7 & : 001, \\
X_8 & : 100, \\
X_9 & : 101.
\end{align*}

Also find $H(x) \neq$ redundancy

\[ \begin{array}{l}
\frac{4}{9} (0) \\
\frac{3}{9} (10) \\
\frac{2}{9} (11) \\
\frac{3}{7} (00) \\
\frac{2}{7} (01) \\
\frac{3}{7} (1) \\
\frac{5}{7} (0) \\
\end{array} \]
NOTES:

(1) REPLACEMENT OF EQUAL PROBABILITIES
    IN ANY $\delta_i$ IS ARBITRARY $\frac{1}{2}$ WILL
    GENERATE CODES OF THE SAME T
    (JUST A REORDERING OF 0'S $\frac{1}{2}$ 1'S)

(2) WHEN $r \neq 2$, $r > 2$, THEN THE # OF
    MESSAGES, $q$, TO BE ENCODED MUST
    BE $\geq q = r^*(r-1)x \geq x$ IS
    AN INTEGER. FOR $r > 0$, ANY
    $q$ WILL DO: $q = x + x$
    FOR, SAY, $r = 2$
    $q = r^*(r-1)x = \frac{q-4}{3} $
    $\Rightarrow q = 7, 10, 13, 16$ ETC.

EXAMPLE: THE CODE FOR THE
    ENSEMBLE $P(84)$, $q = 11$. THUS
    $q = 13$. Thus, $S = p(x_{12}) = p(x_{13}) = 0$

(3) A COMPACT CODE (DEFINED ON P. 66 OF TEXT)
    MAY BE MANY U.D. CODES WITH
    MINIMUM $L$. E.

EX: THE CODE ON p.80, FIG. 4.3 CAN
    GENERATE $L_i : 1, 2, 4, 4, 4, 4$
    OR $L_i : 1, 2, 3, 4, 5$
    BOTH OF THESE COMPACT CODES
    HAVE $L = 2.2$ BINITS.
HW 4: 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)  (9:55)

Noisy Channels

- Error Correcting Code

- \[ P[\text{making an error}] = p(e) \text{ or } P[E] \]

- Fano Bound [Relate \( H(A|B) \) bounded by Fano Bound]

- Shannon's (Celebrated) Second Theorem

- Hamming's Single \footnote{Double Error Correcting Code.}

\begin{align*}
\text{these dates:} \\
7-4-1776 \\
6-15-1947 \text{ (India's Indus Test) \footnote{With Simple Error Correcting Capability}}
\end{align*}

One may decide the nature of the symbol sent by the observation of the received signal. Clearly, there is an element of error involved in the associated prob. (Making that error.) We evolve for decision rules to test their validity for a min \( p(e) \).
RULE 1: \[ P[a^* / b_j] \]

\[
= \begin{cases} 
P(a_1 / b_1) & \\
P(a_2 / b_1) & \\

P(a_n / b_1) 
\end{cases}
\]

Pick \( P(a^* / b_j) \) such that \( P(a^* / b_j) \geq P(a_i / b_j) \) for all \( i \).

A JUDGEMENT BASED ON THIS RULE IS

THE "IDEAL OBSERVER DECISION RULE"

EX: LET AN AM PULSE MOD SYSTEM BE

\[
\begin{array}{c}
A \\
B \\
C \\
\end{array}
\]

THE CHANNEL ERROR (AN AVERAGE OVER THE ENTIRE RECEIVED OUTPUT)

\[ = \text{average } P[E/b_j] \text{ over all } b_j \]

\[ = \sum_j P(E/b_j) P(b_j) \]

\[ \geq P[E/b_j] = P[\text{MAKING AN ERROR WHEN } b_j \text{ IS RECEIVED}] \]

ALSO, \( P(a^* / b_j) + \sum_{a_i} P(a_i / b_j) = 1 \)

\[ = P(E / b_j) \]

\[ = \sum_{i \text{ remaining}} P(a_i / b_j) \]

\[ = 1 - P(a^* / b_j) \]
ERGO, THE AVERAGE ERROR ASSOCIATED WITH THE JUDGEMENT THAT $a^*$ WAS RECEIVED (BY RULE 1)

$$= P(E/b_i)_{a^*}$$

$$= 1 - P(a^*/b_i) \leq 1 - P(a_i/b_i)$$

FROM (2) ON PREVIOUS PAGE:

$$1 - P(a_i/b_i) = P(E/b_i)_{a_i} \quad \mathbb{②}$$

BY BAYES RULE:

$$P(a^*/b_i) = \frac{P(a^*) P(b_i/a^*)}{P(b_i)} \geq \frac{P(a_i) P(b_i/a_i)}{P(b_i)} \quad \mathbb{③}$$

(FROM ①)

IF ALL INPUTS ARE EQUALLY LIKELY:

$$P(a_i) = \frac{1}{q}$$

WHERE $q = \#$ OF INPUT SYMBOLS

THEN

$$P(a^*/b_i) = q P(b_i) \quad \mathbb{③}$$

③ BECOMES

$$P(E/b_i)_{a^*} = 1 - \frac{P(b_i/a^*)}{q P(b_i)} \quad \mathbb{④}$$
AVERAGING (4)

\[
P(E) = \sum_j P(b_j) \left[ 1 - \frac{p(b_j/A*)}{p(b_j)} \right]
\]

\[= 1 - \frac{1}{q} \sum_j p(b_j/A*)
\]

\[= P[\text{CHANNEL ERROR BASED ON IDEAL OBSERVER SCHEME}]
\]

**EX**

Consider the above four symbols being injected into the noisy channel \( \frac{1}{2} \) let them all be equally likely.

\[P(A) = P(B) = P(C) = P(D) = \frac{1}{4}\]

Also, let the noise be

\[P(\text{NOISE } < 0) = 0.
\]

\[P[0 \leq \text{NOISE } \leq 2.5] = \frac{3}{4}
\]

\[P[2.5 < \text{NOISE } \leq 7.5] = \frac{3}{16}
\]

\[P[7.5 < \text{NOISE } \leq 12.5] = \frac{1}{16}
\]

\[P[\text{NOISE } > 12.5] = 0.
\]
**Rule 2 (The Maximum Likelihood Rule)**

\[ p(b_d / a_*) \geq p(b_j / a_i) \]

| \( \frac{1}{q} a_1 \) | 3/4 | 3/16 | 1/16 | 0 |
| \( \frac{1}{q} a_2 \) | 0   | 3/4  | 3/16 | 1/6 |
| \( \frac{1}{q} a_3 \) | 0   | 0    | 3/4  | 1/4 |
| \( \frac{1}{q} a_4 \) | 0   | 0    | 0    | 1  |

Let us call this \( p(b/A) \).

Now, \( p(E) \) for entire channel is

\[
p(E) = 1 - \frac{1}{q} \sum_{j=1}^{4} p(b_j / a_*) \sum_{i=1}^{q} p(b_i / a_i) = 3/16
\]

\[
p(b_j / a_i) = p(b_j / a_*) = \frac{3}{4}
\]

\[
\Rightarrow p(E) = 1 - \frac{1}{4} \left[ (\frac{3}{4} + \frac{3}{4} + \frac{3}{4} + 1) \right] = 3/16
\]

When the input symbols to the channel are equally likely, then \( p(E) \) is the same based on Rule 1 or Rule 2.
EXAMPLE:

\[
\begin{array}{cccc|c}
  \ b_1 & b_2 & b_3 & b_4 & a_j \\
\hline
  a_1 & \frac{1}{4} \ & \frac{3}{16} & \frac{1}{16} & 0 \\
  a_2 & 0 & \frac{1}{4} & \frac{3}{4} & 0 \\
  a_3 & 0 & \frac{3}{4} & 0 & \frac{1}{4} \\
  a_4 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \\
\end{array}
\]

\[P(E) = 1 - \frac{1}{4} \left[ \frac{3}{4} + \frac{3}{4} + \frac{3}{4} + \frac{3}{4} \right] \]
\[= \frac{1}{16} \quad \text{(A BIT LARGER)} \]

**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 across the channel. This "rate" is characterized by \(H(A/B)\) (the equivocation).
The Fano bound explores that connection,

\[ P[E/b_j] = 1 - P(a_i/b_j) \]

\[ = \sum_{i \neq j} P(a_i/b_j) \]  

The conditional entropy,

\[ H(A/B) = \sum_{b_j} H(A/b_j) P(b_j) \]  

\[ H(A/b_j) = -\sum_{a_i} p(a_i/b_j) \log p(a_i/b_j) \]

\[ = -p(a_j/b_j) \log p(a_j/b_j) \]

\[ - \sum_{i \neq j} p(a_i/b_j) \log p(a_i/b_j) \]

\[ \text{and subtract } p(E/b_j) \log p(E/b_j) \text{ to get} \]

\[ -[1 - P(E/b_j)] \log [1 - P(E/b_j)] \]

\[ + P(E/b_j) \log p(E/b_j) \]

\[ - P(E/b_j) \sum_{i \neq j} p(a_i/b_j) \log p(a_i/b_j) \]

\[ (m-1) \text{ symbols maximum of } \log (m-1) \]

\[ H(A/b_j) \leq -[1 - P(E/b_j)] \log [1 - P(E/b_j)] \]

\[ - P(E/b_j) \log P(E/b_j) \]

\[ - P(E/b_j) \log (m-1) \]

\[ \text{turns out:} \]

\[ H(A/B) = H[P(E/b)] + P(E) \log (m-1) \]

This is the "Fano bound".
8-10-76

GRADIENT ON TEST #2
- 7, -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
\[ H(A/B) \leq 4 \left[ P(E) \right] + P(E) \frac{1}{\log_2(M-1)} \]
\[ M = \# \text{OF OUTPUT SYMBOLS} \]

- \( H(A/B) \), THE EQUIVOCATION, IS
A MEASURE OF THE INFORMATION LOST THRU THE CHANNEL
- AVERAGE ADDITIONAL INFO NEEDED TO DETERMINE WHICH Q; WAS SENT,
- \( H(P(E)) \) IS THE AVERAGE AMOUNT OF INFO REQUIRED TO FIGURE OUT IF AN ERROR HAS OCCURRED.
\[ H[P(E)] = H[P(E), P(E)] \]
\[ = P(E) \frac{1}{\log_2(M-1)} P(E) - P(E) \frac{1}{\log_2(M-1)} P(E) \]
- \( P(E) \frac{1}{\log_2(M-1)} \) MAY BE INTERPRETED AS THE ADDITIONAL INFO NEEDED TO FIGURE OUT WHICH OF THE REMAINING \((M-1)\) SYMBOLS HAS BEEN ERRONEOUSLY 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 THEOREM

(GENERAL STATEMENT)

IT IS POSSIBLE TO XMIT INFO WITH AS SMALL A $P(E)$ AS WE DESIRE, IF WE TRANSMIT INFORMATION AT A RATE $< C_{CH} \text{CAP}$ (RATE IN THE SENSE OF $\frac{\text{bits/symbol}}{\text{symbol}}$, OR $\frac{\text{bits/sec}}{\text{sec}}$, OR $\frac{I}{d}$)

ie, keep $H(\text{SOURCE}) < \text{CHANNEL CAPACITY}$

SOME MATH PRELIMINARIES

- STERLING'S FORMULA

\[ n! \approx 2\pi e^{-n} n^{n+\frac{1}{2}} \quad \text{FOR N LARGE} \]
Shannon's Second Theorem for Noisy B.S.C.
(Using Ingel's Simplified Proof)

Let the source have $m$ messages: $(s_1, s_2, \ldots, s_m)$ assumed transmitted independently and equally likely. Each message $s_t$ may be a sequence of $q_k$ symbols $s_t = a_1 a_2 \ldots a_p$ of length $q_t$ of the message $s_t$ may be very large. Then the same entropy $H(s) = \log m$.

Also, let us use Shannon's random coding technique.

If each sequence is $n$ digits long, $2^n$ sequences, assign any of the above $s_t$ to any one of the $2^n$ possible sequence each of length $n$. Clearly

$$2^n \geq m \Rightarrow n \geq \log_2 m$$

and the probability that a particular $n$ digit code word (message) out of a possible $2^n$ code words will be chosen which $a_i$ was sent to represent a specific $s_i$ out of $m$ possible messages $= \frac{m}{2^n}$
Let the channel be BSC \( E \)

\[ P(y_x / x_0) \rightarrow \begin{bmatrix} p & p \\ \bar{p} & p \end{bmatrix} \]

\( \bar{p} = p \) [Correct Xmission]

\( \bar{p} = 1 - p = p \) [Incorrect Xmission]

The criterion of decision:

Find the same message which differs from the received message in the least number 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

\[ \text{Prob Error} = \binom{n}{r} p^{n-r} q^r, r = 0 \ldots n \]

\[ \Rightarrow \binom{n}{n-r} = \binom{n}{r} = \frac{n!}{(n-r)!(r)!} \]

The expected value of \( r \)

(in a sequence of \( n \) symbols)

\[ \bar{r} = \sum_{r=0}^{n} r \cdot \binom{n}{r} p^{n-r} q^r \]

\[ \bar{r} = \sum_{r=1}^{n} r \cdot \binom{n}{r} p^{n-r} q^r \]

\[ = nq = n(p) \]
Due to random assignment, the # of code words that differ from a received system is exactly 1 digit is \( \binom{n}{1} \) and:

- Total # of code words that differ from a received sequence by

\[
\sum_{i=0}^{n_q} \binom{n}{i} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n_q}.
\]

- Utilize Sterling's formula to \( n! \approx \sqrt{2\pi n} \left( \frac{n}{e} \right)^n \) for \( n \) big.

For \( n > \frac{1}{2} \) (which is true for a worthwhile channel \( q = p = 1 - q \)), then:

\[
\binom{n}{n_q-1} < \binom{n}{n_q}, \quad \binom{n}{n_q-2} < \binom{n}{n_q-1}, \quad \binom{n}{n_q} < \binom{n}{n_q+1}.
\]

\[
M \leq (n_q+1) \binom{n}{n_q} = \frac{(n_q+1)(n!)}{(pq)!(n-n_q)!}.
\]

Invoking Sterling's:

\[
M \leq \sqrt{\frac{(n_q+1)^2}{2\pi npq}} q^{-n} p^{-np}.
\]
Now of the $M$ sequences that on the average can be considered 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, so the

$P$ of an $n$-digit sequence selected at random corresponds to one of the $M$ messages

$= \frac{M}{2^n}$

Since we got $M$ total possible received sequences each of which has a probability of correspondence to any of the randomly picked $n$-digit sequence.

The expected # of messages that could be changed by transmission error (and be confused) as correct

$N = \frac{Mn}{2^n} \leq m 2^{-n} \sqrt{\frac{(nq+1)^2}{2nnpq}} 4^{-np} \cdot 2^{-np}$

$\leq m \sqrt{\frac{(nq+1)^2}{2nnpq}} 2^{-n} (1 + p^2 p - 9 pq)$

Boils down to $H(x) = \frac{C}{n}$
(1) When this happens

\[ N \leq m \sqrt{\frac{(n^2+1)^2}{2 \pi n \rho q}} 2^{-nC} \]

let \( m = 2^{nc}/n \)

then as \( n \to \infty \)

\[ N \to 0 \]

1. It is a clear indication that \( P(e) \to 0 \) as \( n \to \infty \).

2. Also, \( H(x) \to C \) as \( n \to \infty \).

For a memoryless channel with a capacity \( C \) and a discrete source with entropy \( H \) and a \( \# E > 0 \) iff \( \rho C < C \), it is possible to encode sequences of \( M \) source symbols in codes of length \( n \) digits for transmission over the channel \( S \).

\[ P(\text{such a sequence will be incorrectly received}) < \epsilon \]

if \( n \) is chosen large enough

\[ \exists \ m = 2^{nH} \] under such conditions, \( M \) transmitted sequences, \( (u_1, u_2, \ldots, u_m) \) will be received as \( (v_1, v_2, \ldots, v_m) \) \( \exists \)

\[ P(u^*_i/v^*_j) \geq 1 - \epsilon \] \( \exists \) "*" denotes assumed sent symbol.
THE CONVERSE

If we try to send info thru 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 \( \geq \epsilon \).

Notes: Like Kraft \& McMillan, Shannon's Second Theorem does not give specific coding techniques, but merely suggests selection of \( m \) relative to \( C \geq H \).
8-15-76 (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. (127)

HANDOUT

ABOVE PROBLEMS DUE IN CLASS ON 8/16/76

CORRECTIONS IN SHANNON'S SECOND THEM.

\[ M \geq 2^n + 1 \] FOR THE PROOF FOR GENERAL CHANNEL

\[ M = \# \text{ OF MESSAGES} \]
ERROR DETECTION

(1) HAMMING DISTANCE BETWEEN TWO MESSAGES $A \neq B$.

$\Delta$ THE # OF DIGITS IN WHICH $A \neq B$ DIFFER.

EX: (1) $A : 100$

$B : 1011 \Rightarrow d(A, B) = 1$

(2) $A : 10101 \Rightarrow d(A, B) = 2$

$B : 100$

PROPERTIES OF $d$:

$d(A, B) = d(B, C)$

$d(A, B) = 0$ IFF $A = B$

$d(A, B) + d(A, C) \geq d(A, C)$

$d(A, B) = d(B, A)$

(II) PARITY CHECK

TWO CASES: (EVEN 1, EVEN 0)

(ODD 1, ODD 0)

01100111110011

# OF 1'S = 9 $\Rightarrow$ THIS MESSAGE DOES NOT SATISFY EVEN 1 PARITY.

# OF 0'S = 5 $\Rightarrow$ EVEN 0'S NOT SAT.

ODD 1'S NOT SATISFIED

ODD 0'S NOT SATISFIED

WE WILL WORK ALL OUR PROBLEMS IN EVEN 1 PARITY.
This check can be used to detect errors.

Ex: Let \( m = \# \text{ of message bits} = 7 \)

And let the message be

\[
0101110
\]

Using even \( l \) parity \( \Rightarrow 0 \) since 4 1's make up even parity check. Thus, we send

\[
01011100, \text{ an 8 bit word}
\]

If the received message has parity, we say that no odd # single error has occurred (1, 3, 5... errors have occurred \( \frac{1}{2} \) are detected).

III. Let the code book consist of messages be

\[
A = 000
\]

\[
B = 111
\]

\[
d(A; B)
\]

Let the following be received:

\[
\begin{align*}
001 & \quad \text{assign to} \quad 000 \\
010 & \quad \rightarrow 000 \\
100 & \quad \rightarrow 001 \\
011 & \quad \rightarrow 111 \\
101 & \quad \rightarrow 111 \\
110 & \quad \rightarrow 111
\end{align*}
\]
Thus, we can say that the distance \( d \) in \( \mathbb{F} \), such error can be detecting.

<table>
<thead>
<tr>
<th>Distance between words</th>
<th>Coding type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No detection or codification</td>
</tr>
<tr>
<td>2</td>
<td>1 error, 1 described</td>
</tr>
<tr>
<td>3</td>
<td>1 error, 1 corrective</td>
</tr>
<tr>
<td>4</td>
<td>1 error, 1 corrective 1/2 detect 2 errors</td>
</tr>
<tr>
<td>5</td>
<td>2 errors, corrected (and, of course, detected)</td>
</tr>
<tr>
<td>6</td>
<td>- detected</td>
</tr>
<tr>
<td>7</td>
<td>- detected</td>
</tr>
</tbody>
</table>

In general, \( d(A, B) \geq 2q + 1 \). In order to correct 9 errors in a string of \( n \) symbols, it is clear from the simple example that a finite \( (>1) \) of messages (each \( n \) symbols long) gotta be reserved for each message (in digit long) sent.
IV. THE MAX # OF WORDS (MESSAGES) IN A CODE BOOK = N = LENGTH OF EACH METHOD \( \frac{1}{2} \) 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 BE ASSIGNED TO ONE CODE WORD OF THE CODE BOOK.

IF \( \exists \) \( V \) WORDS IN THE CODE BOOK, WE MUST HAVE \( VM \) POSSIBLE WORDS AT THE OUTPUT.

\( VM \leq 2^n \)

OR, \( M \leq \frac{2^n}{\sum_{i=1}^{q} \binom{n}{i}} \)

THIS IS A NECESSARY CONDITION ON \( M \); IS IT SUFFICIENT? NO!
EX: \( n = 4 \), \( k = 1 \),
\[
\sum_{i=1}^{2} (\frac{4}{\lambda}) \leq 3.2
\]
TRY \( r = 3 \)
\( r = 3 \), \( \lambda \), IF THIS IS A SUFFICIENT CONDITION, WE SHOULD BE ABLE TO PICK OUT 3 WORDS OUT OF A POSSIBLE 16 (EACH 4 DIGITS LONG).
\[
d(\text{ANY PAIR} = 0) = (\text{AT LEAST} 3)
\]
BUT IT DON'T WORK, THUS, THIS IS NOT A SUFFICIENT CRITERIA.

HAMMING'S PAPER: ERROR CORRECTION
(R.W. HAMMING BSTJ VOL 29, 1950, P 1475)
(1) HAMMING INTRODUCED REDUNDANT DIGITS (K IN #) TO A MESSAGE (M DIGITS LONG) > THE # OF DIGITS IN A XMTED MESSAGE = \( n = m + k \) ? CALLS K BITS
THE PARITY CHECK BIT
\[
\text{e.g. } (1854)_{2} = \text{1111 0111110}, \text{ m = 11}
\]
(2) THE K BITS ARE APPROPRIATELY POSITIONED \( \Rightarrow \) THEY SATISFY CERTAIN (EVEN) PARITY Checks AND ALSO GIVE THE DECIMAL LOCATION OF THE SINGLE ERROR. THE K BIT CHECKING MUST HAVE ENOUGH POSSIBLE "STATES" TO
1. Identify if an error has occurred in any location or no error at all. K bits 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**

<table>
<thead>
<tr>
<th>$m$</th>
<th>$k$</th>
<th>$2^k \leq m + k + 1$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. 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 ...

$P_2$ ... for 4, 5, 6, 7, 12, 13, 14, 15 ...

$P_3$ ... for 8, 9, 10, 11, 12, 13, 14, 15 ...

<table>
<thead>
<tr>
<th>BIT POSITION</th>
<th>BINARY EQUIV</th>
<th>$P_0$ CHECKS</th>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$P_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0001</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0010</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0011</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0100</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0110</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0110</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0111</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>8</td>
<td>1000</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1001</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1010</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>11</td>
<td>1011</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>12</td>
<td>1100</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>13</td>
<td>1101</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>14</td>
<td>1110</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>15</td>
<td>1111</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Thus, since each \( P_a \) occurs once in a parity check, that \( P_a \) can be determined easily.

After coding (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.

Example:
It is clear that for \( m=11, k=4 \), the code \( 1254_2 \)

\[
\begin{array}{cccccccccccc}
15 & 14 & 13 & 12 & 11 & 10 & 9 & 8 & 7 & 6 & 5 & 4 \\
1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1
\end{array}
\]

\[
(1254)_2 = 1110011111.0
\]

Run parity (even 1) on

\[
P_0 = P_0 \oplus 0 \oplus 1 \oplus 1 \oplus 0 \oplus 0 \oplus 1 \oplus 1 \oplus 1 \Rightarrow P_0 = 1
\]

\[
P_1 = P_1 \oplus 0 \oplus 1 \oplus 0 \oplus 0 \oplus 0 \oplus 1 \oplus 1 \Rightarrow P_1 = 1
\]

\[
P_2 = P_2 \oplus 1 \oplus 0 \oplus 1 \oplus 1 \oplus 0 \oplus 0 \oplus 0 \oplus 1 \Rightarrow P_2 = 0
\]

\[
P_3 = P_3 \oplus 1 \oplus 1 \oplus 0 \oplus 0 \oplus 0 \oplus 1 \oplus 1 \oplus 0 \Rightarrow P_3 = 1
\]
NOTE 1: IN SOLVING 7-4 1776
WE GOT $M=11$, $K=4$ FOR 1776
FOR 7 ($M=3$, $K=3$ FOR 7)
NO GOOD. USE $M=4$

CHOOSE $M=11$ UNIFORML
FOR ALL PARTS OF THE
DATE

NOTE 2: HAMMING CODE ALSO
CORRECTS AN ERROR IN
A PARITY BIT.
FOR THE ARISE OF UNIFORMITY

\[ m = 11, 21, 11 \]

\[ n = 15, 15, 15 \]

2. THE HAMMING CODE WILL CORRECT THE ERRORS IN LOCATIONS OCCUPIED BY PARITY DIGITS

3. THE "FIGURE OF MERIT" FOR A B.S.G AND A HAMMING CODE

\[
P(\text{ERROR}) = P \quad \text{FOR A SINGLE DIGIT}
\]

\[
P(\text{NO ERROR}) = \bar{P}
\]

IF \( n \) IS THE # OF DIGITS IN A CODE WORD \( 2 \cdot np << 1 \)

\( (n = 10, \frac{1}{100} = \rho) \), THEN

\[
P[\text{RECEIVING AN INCORRECT WORD} \quad \text{TEN DIGITS LONG W/O HAMMING CODE}]
\]

\[
= \left( \begin{array}{c} n \\ 1 \end{array} \right) \rho (\bar{\rho})^{n-1} + \left( \begin{array}{c} n \\ 2 \end{array} \right) \rho^2 (\bar{\rho})^{n-2} + \cdots
\]

\[
= 1 - (1-\rho)^n
\]

\( \sim np \)

THE \( P[\text{RECEIVING AN INCORRECT WORD AFTER APPLYING HAMMINGS SINGLE ERROR CORRECTING CODE}] \)

\[
= \left( \begin{array}{c} n \\ 2 \end{array} \right) \rho^2 (1-\rho)^{n-2} + \left( \begin{array}{c} n \\ n \end{array} \right) \rho^n \bar{\rho}^0
\]

\[
= 1 - (1-\rho)^n - np (1-\rho)^{n-1}
\]

\[
= \frac{1}{2} (n-1) \rho^2 + \cdots
\]

\[ \leq np \]
In, the Hamming coding has reduced the prob (reception of an incorrect word. On this basis,
\[
\text{figure of merit} = \frac{1-(1-p)^n}{1-(1-p)^n - np(1-p)^n - 1}
\]
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 \( \log_2 N \). Note that \( 2^m \leq 2^{n+1} \). Thus, a tabulation can be drawn:

<table>
<thead>
<tr>
<th>( N )</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>512</th>
<th>1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m )</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>( n )</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>parity</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
**GROUP CONCEPTS.**

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)
\[a, b \in G \Rightarrow a \circ b = c \in G\]

2) ASSOCIATIVE LAW

IF \( a, b, c \in G \), THEN
\[(a \circ b) \circ c = a \circ (b \circ c)\]

3) IDENTITY ELEMENTS \( U \)

\[\exists \ U \ni U \circ a = a\]

4) EACH ELEMENT MUST HAVE
AN INVERSE \( \exists a^{-1} \ni a \circ a^{-1} = U \)

EX: +, AND -#'S FORM A GROUP
UNDER THE OPERATION OF +

DO THEY FORM A GROUP UNDER
MULT? NO. THE GROUP OF
FRACTIONAL #’S # 0 FORM
A GROUP UNDER MULTIPLICATION
IT IS POSSIBLE FOR TWO NUMBERS TO FORM A GROUP \((0, 1)\), THE OPERATION BEING HALF ADDING. THAT IS

\[
\begin{align*}
0 \oplus 0 &= 0 \\
0 \oplus 1 &= 1 \\
1 \oplus 0 &= 1 \\
1 \oplus 1 &= 0
\end{align*}
\]

IDENTITY: \(0 = 0\)
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 ELEMENTS FROM \(G\) WHICH, BY THEMSELVES, OBSERVE ALL OF THE PROPERTIES OF THE GROUP, THE IDENTITY MUST ALSO BELONG TO THE SUBGROUP.
\[ \text{eg. } G(\ldots, -10, -3, \ldots, 0, 1, 2, \ldots, 40) \]

A subgroup is
\[ \langle -14, -7, 0, 7, 14, 21, \ldots \rangle \text{ etc} \]
(under, of course, addition)

A coset: let a group \( G \)
(defined under mult) be
\[ \langle q_1, q_2, \ldots, q_m, s_1, \ldots, s_n \rangle \text{ and} \]
A subgroup \( \langle s_1, \ldots, s_n \rangle \)

\[
\begin{align*}
S & \rightarrow s_1, s_2, s_3, s_4, s_5, s_n, \\
& \quad q_1s_1, q_1s_2, q_1s_3, q_1s_4, q_1s_n \\
& \quad q_2s_1, q_2s_2, q_2s_3, q_2s_4, q_2s_n \\
& \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\
& \quad q_ms_1, q_ms_2, q_ms_3, q_ms_4, q_ms_n
\end{align*}
\]

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 \( s_i \). 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, \( q \) and \( q' \) of a group \( G \) are in the same left coset of a subgroup \( S \) iff \( q^{-1}q' \) is an element of \( S \).

2. Every element of \( G \) is in one and only one coset of a subgroup \( S \).

Eq. Let \( m=2 \) for a Hamming (Single Error Correcting Code)

\[
2^k \geq n+1+k \Rightarrow k = 2
\]

\[
\therefore n = n + k = 5 \quad \text{and we should have} \quad 2^5 \quad \text{possible words. The} \quad \# \quad \text{of messages cannot be more than} \quad 2^5 = 2^2 = 41
\]

\[
\begin{array}{cccccc}
1 & 1 & 1 & 0 & 1 & 0
\end{array}
\]

The four words \( x \)mitte \( \bar{b} \) are

\[
\begin{align*}
S_1 &= 000000 \\
S_2 &= 001111 \\
S_3 &= 110001 \\
S_4 &= 111110
\end{align*}
\]
Let these form the "subset" and let us generate the left cosets:

\[
\begin{align*}
S_1 &= 00000 \\
S_2 &= 00111 \\
S_3 &= 11001 \\
S_4 &= 11110
\end{align*}
\]

\[
\begin{array}{cccc}
10000 & 10111 & 01001 & 01110 \\
01000 & 01111 & 10001 & 10110 \\
00100 & 00011 & 11101 & 11010 \\
00010 & 00101 & 11011 & 11100 \\
00001 & 00110 & 11000 & 11111 \\
01100 & 01011 & 10101 & 10010 \\
01010 & 01101 & 10011 & 10100
\end{array}
\]

Words with least \# of 1's: $S_1 \oplus S_2 = S_3$

An array formed with coset leaders with a minimum Hamming rate is called a standard array. Note that none of the 25 words are repeated.

6-1 p 184

Use Hamming's single error correcting code for $n$ bits/word, message $5 = m$, $k = ?$ [only 2 cases]. For both codes, compute $P(\text{incorrectly encoding a word sent w/o error correction})$. Assume REC with delivering an error $p = \frac{1}{100}$.
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 embarrassment 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...
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 \( \alpha \) due to distance (assumed perfectly linear). The received signal power is denoted \( P \).
2. Additive white Gaussian noise \( n(t) \) arising in the receiver front end, with single-sided spectral density \( N_0 \).

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 \( \omega_0 \) 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 \( N_0 \), with some acceptable quality, we say that the system is operating at a signal-to-noise ratio per information bit \( E_s/N_0 = P/N_0R \). 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 \( E_s/N_0 \) 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 \( \tau \) seconds apart, and during any \( \tau \)-second period, sometimes called a baud, it can take on one of \( N \) discrete values, usually equally spaced. We let \( a_{i1} \) be the value in the \( i \)-th interval. If \( N \) is a power of two, say \( 2^n \), then the signaling rate is \( 1/\tau \) symbols (bauds) per second, and the transmitted rate \( n/\tau \) bits per second. Ideally, the bandwidth occupied is \( W = 1/\tau \) 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 \( \theta \) 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 \( i \)-th baud such a receiver forms the integral

\[
z_i = \int_{\tau}^{\tau+\tau} y(t) \cos \left[ \omega_0 t + \theta + \psi(t) \right] dt
\]

It is easily shown that \( z_i = a_t + n_i \), where \( a_t \) is the modulation amplitude (scaled) in the \( i \)-th 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 is based on the correlator outputs \( z_i \), and 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 \( \tau \) seconds, a level \( a_t \) (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. 2). 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 deci-

![FIGURE 1. System block diagram.](image-url)
sions, 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

\[
C = \frac{1}{2} \log_2 (1 + P/N_0 W) \quad \text{bits/baud}
\]

or

\[
W \log_2 (1 + P/N_0 W) \quad \text{bits/second}
\]

where \(P\) is the received signal power, \(N_0\) the single-sided noise spectral density, and \(W\) the nominal bandwidth.

Shannon showed that whenever the information rate \(R\) is less than \(C\), 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 \(E_b/N_0 = P/N_0 C\). 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:

\[
C/W = \log_2 [1 + (P/N_0 C)(C/W)]
\]

This relation is plotted in Fig. 4. We see that, for a fixed power-to-noise ratio \(P/N_0\) more and more efficient communication is possible as the bandwidth is increased, and that with no bandwidth limitations, \(E_b/N_0\) approaches a limit of \(\log 2 (\approx 0.69, \text{ or } -1.6 \text{ dB})\), 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 \(E_b/N_0\) limit is reached only for infinite bandwidth, at \(1/2\) bit per cycle of bandwidth (or a code rate of about \(1/4\) 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 \(E_b/N_0\) 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 \(E_b/N_0\) of 6.8 dB is required to obtain a bit error probability of \(10^{-5}\) and 9.6 dB to obtain \(10^{-4}\). On the other hand, the capacity theorem promises essentially zero error probability whenever \(E_b/N_0\) exceeds \(-1.6 \text{ dB}\). 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 \( P/N_c W \gg 1 \), with fixed \( N_c \) 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 \( N = 2^k \) increases). Thus, if \( R_{\text{AM}} \) 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

\[
P \to 4^k P
\]

\[
R_{\text{AM}} \to R_{\text{AM}} + k
\]

\[
C \to C + k
\]

\[
\frac{R_{\text{AM}}}{C} \to \frac{R_{\text{AM}} + k}{C + k} \to 1 \text{ as } k \to \infty
\]

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 \( \Pr(\varepsilon) = 10^{-4} \) to show how rapidly \( R_{\text{AM}} \) 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 pseudo-random 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 in an easy entry 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 where the 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 \( b_n \), so that the input \( b_n \) is expressed mathematically as

\[
b_n = b_1 + b_4 \pmod{2}
\]  

(1a)

or, using the notation \( \oplus \) for modulo-2 addition,

\[
b_n = b_1 \oplus b_4
\]  

(1b)

When we say “shift register,” we imply that whenever the circuit is pulsed by a clock pulse (not shown), \( b_n \) enters the left end, all other bits shift one place to the right, and the rightmost bit 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 \( 1 \cdot 1 = 0 \). 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 = * 

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.

---

**I. Modulo-2 arithmetic**

<table>
<thead>
<tr>
<th>Addition</th>
<th>Multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>\times</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

---

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.
a ⊕ c = a ⊕ 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., \( a + b + c = c + b + a \), \( ab + c = ab + ac \), 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 output, 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 \( M \)-bit sequences generated by these registers.

In a feedback circuit, the probability of seeing a "1" is \((M/2)(M - 1)\), or just

---

**FIGURE 8. State diagram of feedback circuit in Fig. 7.**

---

<table>
<thead>
<tr>
<th>II. Connections for MLSR generators</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m )</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>17</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>III. Code words in a ((15, 4)) code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Bits</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>0000</td>
</tr>
<tr>
<td>0001</td>
</tr>
<tr>
<td>1000</td>
</tr>
<tr>
<td>0100</td>
</tr>
<tr>
<td>0010</td>
</tr>
<tr>
<td>1001</td>
</tr>
<tr>
<td>1101</td>
</tr>
<tr>
<td>0110</td>
</tr>
<tr>
<td>0111</td>
</tr>
<tr>
<td>1011</td>
</tr>
<tr>
<td>1110</td>
</tr>
<tr>
<td>0011</td>
</tr>
<tr>
<td>0011</td>
</tr>
</tbody>
</table>

---

IEEE spectrum JUNE 1970
Furthermore, since all \( m \)-bit sequences except the all-zero sequence occur somewhere in the maximum-length sequence, the probability of seeing a "1" given any \( 1 \) or fewer preceding bits is still nearly one half. Some other statistical properties make a maximum-length sequence difficult to distinguish from a sequence generated truly randomly, as by flipping a coin, yet these sequences are easy to generate and repeatable. Thus they are commonly used to generate pseudorandom bits.

The class of maximum-length shift-register codes is representative of the major classes of algebraic block codes, 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 \( n \)th bit is some modulo-2 sum of previous bits, which are themselves each modulo-2 sums of information bits, so the \( n \)th 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 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 itself is 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-
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 \( \pm 1 \) 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—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.

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
special, 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. 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.)

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\(^6\) and generalized minimum-distance decoding\(^7\) 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\(^8\) and Berlekamp\(^9\) 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-\(1/2\) convolutional code: rate \(1/2\) 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 se-
quences 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 2^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 (working on a self-orthogonal code). Threshold
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\cite{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 is received against what would have been transmitted, even 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.\cite{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\cite{16} works with 16-ary frequency-hopping modulation and incoherent list detection; the NASA Ames decoder for the Pioneer satellites\cite{17} and the JPL general-purpose decoder\cite{18} work with PSK and soft (eight-level) decisions; and the Codex decoder, built for the U.S. Army Satellite Communication Agency,\cite{19} works with PSK (or DPSK or QPSK) and hard decisions, the choice in every case being based on system consider-
tions. 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 $E_b/N_0 = 1.4$ 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-$\frac{1}{2}$ codes, PSK, and hard decisions has a theoretical limit of $E_b/N_0 = 4.5$ dB; 2 dB of this loss is due to hard decisions, 1 dB to the choice of rate $\frac{1}{2}$ 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, which has an internal decoding 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 (SEC 25), 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, 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 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


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 AAS.
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 chiffres 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 verallgemeinert, entwickelt der Autor Kodes die es ermöglichen bei Übertragung binärer bits \( k \) Fehler zu korrigieren.

Обобщая работу Хэмминга, 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 \( n \) de chiffres binaires : 
\[
a_1 \ a_2 \ldots \ldots \ a_n
\]

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 déteeteur — ou correcteur — de \( k \) erreurs.

L'exemple le plus simple est le code déteeteur d'une erreur. Dans ce cas \( m = n - 1 \), en on choisis le chiffre de test de façon que le nombre total de chiffres \( I \) 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

\[
\log_2 (1 + n) \leq N < 1 + \log_2 (1 + n)
\]

Dans le cas général d'un code correcteur de \( k \) erreurs, le nombre de configurations d'erreurs possibles est :

\[
H = 1 + C_n^1 + C_n^2 + \ldots + C_n^n
\]

Par suite le code le plus économique utiliserait un nombre de chiffres de test égal à l'entier immédiatement supérieur à \( \log_2 H \). 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
\[
kN - \log_2 H
\]
est de l'ordre de \( \log_2 (k!) \), 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 \( \mathcal{C} \).

Les éléments de l'anneau \( \mathcal{C} \) sont les nombres entiers écrits en numération binaire.

A chaque élément de l'anneau \( \mathcal{C} \) nous faisons correspondre un polynôme ayant comme coefficients les chiffres de l'élément.

Toute opération sur les éléments de \( \mathcal{C} \) sera faite sur les polynômes correspondants — au cours de ces opérations tout coefficient pair sera remplacé par 0, tout coefficient impair par 1.

Le résultat sera un polynôme auquel correspondra un élément de l'anneau \( \mathcal{C} \).

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 \( (\mathcal{C}) \).

Exemples :

Addition : \( 101 + 111 = 10 \) \( (\mathcal{C}) \)

Multiplication : \( 101 \times 111 = 11.011 \) \( (\mathcal{C}) \)

Puissance : \( 101^5 = 10.001 \) \( (\mathcal{C}) \)

Division : \( 1.101 = 1111 + 10 \) \( (\mathcal{C}) \)

En particulier : \( p + p = 0, \ (p + q)^2 = p^2 + q^2 \) \( (\mathcal{C}) \)

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 \( \mathcal{C} \), d'autre diviseur que lui-même et l'unité).

On peut classer évidemment les nombres dans l'anneau \( \mathcal{C} \) 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.

Étant donné un mot écrit en binaire
\[
a_1 \ a_2 \ldots \ldots \ a_n
\]

nous attacherons à chaque indice \( i \) un nombre \( p_i \) de l'anneau \( \mathcal{C} \) et au mot lui-même nous attacherons le nombre
\[
T = p_1 + p_2 + \ldots + p_n \quad (\mathcal{C})
\]

C'est la considération du nombre \( T \) qui, grâce à un choix convenable des nombres \( p_i \), permettra de corriger les erreurs éventuelles.

3. Code de Hamming.

Nous retrouvons le code de Hamming en faisant
\[
p_i = 1
\]

Les chiffres de test sont les chiffres du mot d'indices
\[
1, 2, 2^2, \ldots, 2^{n-1} \quad (N\text{ défini par les inégalités} \ 11)\]

L'information sera portée par les chiffres
\[
a_2 \ a_4 \ a_8 \ a_{16} \ldots \ a_n
\]

On détermine les chiffres de test par la condition
\[
T = \sum p_i a_i = 0
\]

condition qui s'écrit ici
\[
a_2 + 2a_4 + 4a_8 + \ldots + 2^{n-1} a_{2^{n-1}} = 3a_2 + 5a_4 + 6a_8 + \ldots + na_n \quad (\mathcal{C})
\]

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 \( T = 0 \).

S'il y a une erreur, portant par exemple sur le chiffre \( a_2 \) remplacé par \( (1-a_2) \), le nombre \( T \) prendra la valeur :

T = a_1 + 2a_2 + \ldots + n(1 - a_2) + \ldots + ma_m = a \tag{36}

À la valeur de T sera l'indice du chiffre erronné.

S'il y a deux erreurs, portant sur les chiffres d'indice a et \( b \), T prendra la valeur :

T = a + b \neq 0 \tag{37}

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

\[
\begin{bmatrix}
0 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1
\end{bmatrix}
\]

Aux nombres 0010110 et 0010010 correspondent les matrices

\[
\begin{bmatrix}
0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

Le nombre T s'obtient en faisant suivre chaque ligne de la matrice de son chiffre de parité (§ 1). On obtient ici :

\[
\begin{array}{ccc}
0 & 0 & 1 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{array}
\]

Le premier nombre est correct, le 5\textsuperscript{e} 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) :  

\[ T = \sum a_i p_i = 0 \tag{38} \]

Si après transmission, les chiffres de rang

\[ a_1, a_2, \ldots, a_j \quad (j \leq k) \]

sont erronés, le nombre \( T \) calculé sur le mot déformé prendra la valeur :

\[ T = p_{a_1} + p_{a_2} + \ldots + p_{a_j} \tag{39} \]

Il faut que le nombre ainsi trouvé soit caractéristique des rangs

\[ a_0, a_0, \ldots, a_0 \quad (j \text{-ère ligne} \quad \text{et de rangs} \quad a_1, a_2, \ldots, a_j \quad (j \leq k) \]

lorsque

\[ f \neq k \]

et les deux ensembles

\( \{a_1, a_2, \ldots, a_j\} \quad \text{et} \quad \{a'_1, a'_2, \ldots, a'_j\} \]

sont identiques.

Cette condition peut encore s'écrire :

\[ p_{a_1} + p_{a_2} + \ldots + p_{a_j} \neq 0 \tag{40} \]

lorsque \( f \leq 2k \)

et les \( \lambda_1, \ldots, \lambda_k \) étant tous différents.

On devra donc choisir les nombres \( p \) tels que l'addition, dans l'anneau \( \mathbb{Z} \), 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 :

\[ 2, 2^2, 2^3, \ldots, 2^{2k-1} \]

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 réprènté par une colonne

\[ P_i = \begin{bmatrix} a_i^1 \\ a_i^2 \\ \vdots \\ a_i^K \end{bmatrix} \]

Si le rang de cette matrice (dans l'anneau \( \mathbb{Z} \)) 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 \( A \) de \( K' \) lignes dont le déterminant calculé dans l'anneau \( \mathbb{Z} \) ne sera pas nul. On aura donc

\[ \det A = 1 \]

puis que les seules valeurs possibles sont 0 ou 1. En multipliant la matrice \( M' \) par \( A^{-1} \), on obtiendra la matrice \( M'' \) formée de nombres \( p'' \) tels que

\[ p_{i+1} = \begin{bmatrix} a_i^{1-K} \\ a_i^{2-K} \\ \vdots \\ a_i^{K-K'} \end{bmatrix} \]

et par suite les nombres \( p'' \), vérifieront encore la condition (42). De plus, la matrice \( M'' \) contiendra à ce moment la matrice \( A^{-1} \times \delta \), c'est-à-dire la matrice unité, donc l'ensemble des \( p'' \) contiendra les puissances successives de 2 :

\[ 1, 2, 2^2, 2^3, \ldots, 2^{2k-1} \]

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 :

\[ p_1 = 1, \quad p_2 = 2, \quad p_3 = 2^2, \ldots, \quad p_{2k-1} = 2^{2k-1} \]

pui

\[ p_{2k+1} = 2^{2k} - 1 \]

\[ p_{2k+2} = 2^{2k} \]

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 \( i \) chiffres. Considérons l'ensemble des nombres \( p_i \), et de leurs sommes dans l'anneau \( \mathbb{Z} \), par groupes de 2, 3, \ldots (2k - 1). Tous les nombres obtenus ont au plus \( i \) chiffres.

S'il existe un nombre non contenu dans l'ensemble ainsi formé et compris entre \( p_i \) et \( p_{i+1} \), 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 \( p_{i+1} = 2^i \).

On peut ainsi continuer pas à pas jusqu'à l'obtention des \( n \) nombres \( p \). Si \( p_i \) à \( K \) chiffres, les nombres

\[ 1, 2, 2^2, \ldots, 2^{2k-1} \]

seront inclus dans la suite des \( p \). La suite sera donc directement utilisable pour形成 un code. Il restera à établir le tableau de correspondance entre les \( H \) valeurs de la somme

\[ p_{a_1} + p_{a_2} + \ldots + p_{a_j} \tag{41} \]

et la valeur des indices \( a_1, a_2, \ldots, a_j \).

Le procédé ainsi défini est assez long à expérimenter. Cependant, pour des valeurs raisonnables de \( n \) et \( K \), il ne dépose pas les possibilités d'une calculatrice de moyenne puissance.

La détermination à priori du nombre \( K \) de chiffres de test parait 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 préuves des opérations arithmétiques, va nous fournir un mode de calcul des nombres \( p \). Désignons par \( \theta \) un nombre irréductible de \( N - 1 \) chiffres et par \( q \) le reste de la division dans l'anneau \( \mathbb{Z} \), d'un nombre \( q \) par \( \theta \). Le nombre \( \theta \) aura au maximum \( N \) chiffres.

Nous poserons alors :

\[ p_i = l + 2^n(q') + 2^{n+1}(q') + \ldots + 2^{n+(n-1)}(q' - 1) \tag{42} \]

\( (i = 1, 2, \ldots, n) \)
7. Exemple.

Nous avons formé un code de 15 chiffres correcteur pour 2 erreurs. On voit qu'il y aura 8 chiffres de test.

En prenant \( q = 19 = 10,011 \), on calcule aisément les nombres \( p \) et la matrice M :

\[\begin{align*}
\text{M} &= \begin{bmatrix}
0 & 1 & 1 & 1 & 1 \ 0 & 0 & 1 & 1 & 1 \ 0 & 0 & 1 & 1 & 1 \ 0 & 0 & 0 & 0 & 0 \ 0 & 0 & 0 & 1 & 1 \ 0 & 0 & 0 & 1 & 1 \ 0 & 0 & 0 & 1 & 1 \ 0 & 0 & 1 & 1 & 1 \ 1 & 1 & 0 & 1 & 1 \ 1 & 1 & 0 & 1 & 1 \ 1 & 1 & 0 & 1 & 1 \ 1 & 1 & 0 & 1 & 1 \ 1 & 1 & 0 & 1 & 1 \ 1 & 1 & 0 & 1 & 1 \ 1 & 1 & 0 & 1 & 1 \ 1 & 1 & 0 & 1 & 1 \ 1 & 1 & 0 & 1 & 1
\end{bmatrix}
\end{align*}\]

Cette matrice M est de rang 8, car le déterminant formé avec les colonnes 1, 2, 4, 6, 12, 7, 14 (choisies parce qu'elles présentent le plus de zéros) vaut 1.

La matrice \( \Delta \) (§ 4) sera formée avec ces colonnes. En l'inverse on trouve la matrice :

\[\begin{align*}
\Delta^{-1} &= \begin{bmatrix}
0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \ 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}
\end{align*}\]

Le produit \( \Delta^{-1}M \) (dans l'anneau \( \mathbb{C} \)) donne la matrice définitive :

\[\begin{align*}
\begin{bmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\end{align*}\]

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 ferme la somme

\[T = \Sigma p_n \]

Si elle est nulle, il n'y aura pas 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.

<table>
<thead>
<tr>
<th>( T )</th>
<th>( 1 )</th>
<th>( 2 )</th>
<th>( 3 )</th>
<th>( 4 )</th>
<th>( 5 )</th>
<th>( 6 )</th>
<th>( 7 )</th>
<th>( 8 )</th>
<th>( 9 )</th>
<th>( 10 )</th>
<th>( 11 )</th>
<th>( 12 )</th>
<th>( 13 )</th>
<th>( 14 )</th>
<th>( 15 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

On remarque que le nombre \( T \) prend 121 valeurs possibles

\[\left(1 + c_0^2 + c_1^2\right)
\]

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'entrée par excès de \( \log_2\) H.
I want to introduce a system of transmission of a word concatenated by a number of \( n \) binary numbers \( a_1, a_2, \ldots, a_n \) such that the \( n \) received can differ from the initial word by a certain number of errors by certain \( m \) being altered \( (1-a_i) \), to try to detect and correct these errors one uses only \( m \) number of the word as a support of information.

In any remaining called \( m \) of tests, I must serve the verification of the word after the transmission.

To give a love to determination of these test number in future I must number it in such a way as to be able to detect and correct \( m \) a maximum number of errors.
that is to form a code detector - a corrector - of \( K \) errors.

The most simple \( K \) is in the code detector of one error.

In this case \( M = N - 1 \),

If one chooses the 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. E. T. Bell & Technical J. 1930 gave the law of formation of code correct of 1 error.

The number of test digits in the entire is determined by the \( 2^M \) equals \( \log \).
In the general case of the code consisting of $K$ errors, the number of configurations possible is

$$H = 1$$

Thus, the most common code would use $\log_K N$ and test digit $= \lim$ the whole immediate superior to $\log_2 H$.

Besides, the code of Hamming they have not been able to construct such a code. Those we propose use a number of test digits $= to$

$$n - m = KN$$

$+$ the difference

$$KN - \log_2 H$$

is to the order of $\log_2(K!)$, and it is weak through these codes would be satisfied.
After having defined a code in which we can make use of a Hamming code, we will suppose Hamming code under the optic of the principle of formation of codes which will lead us to a determination of partially experimental determination of a systematic of these codes.

We will illustrate by an example of a complete code of 2 errors.

2. Definition of Ring a

In the elements of the set \( \mathbb{Z} \) are the whole numbers written in binary numeration.

To each element of the set \( \mathbb{Z} \) we correspond a polynomial having \( a \) as a coefficient the digit.
of the element is the polynomial defined on the body of characteristic 2.

All operations on the elements of a will be made into corresponding polynomials during the course of these operations, with even coefficients will be replaced by 0, leaving uneven coefficients. The result will be a polynomial in which well correspond a element of the 0 a.

in all habitual operations on the whole numbers in order to avoid all possible ambiguities, all the calculated (figured) or suppressions followed according to the stated rule.
will be followed by the
indication (a).

Examples
add \[101 + 101 = 10\]

When the polynomial will be
unreducible on the body of the
\[\mathbb{C}\]
we will say that the
corresponding number is unreducible. (There
is not admitted in the \(\mathbb{O}_a\) another
divisor than itself + unity.)

You can naturally classify the \(\mathbb{O}_a\)
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 unreducible numbers.

To given a written word in binary
\((a, a_2, \ldots, a_n)\),
we will attach to each indication:

A number \(p_i\) of the circle \(a_i\)
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 by Hamming

We find again the Code by Hamming in doing \( p_i = i \).

The test digits are the digits of the word of the redundancies

\[ 1, 2, 2^2, \ldots, 2^{n-1} \] (\( N \) defined by the inequalities \( 1 \))

The information will be carried by the digits \( a_3 \) as \( a_6 \) at \( a_9 \), \( \ldots \) and

one determines the test digits by the conditions

\[ T = \sum p_i a_i = 0 \]

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, for example, the digit $a_2$ replaced by $(1 - a_2)$ the number $T$ will take the value

$$T = a_1 + 2a_2 + ...$$  (a)

The value of $T$ will be the indication of the digit which is wrong.

If there are two errors on the digits $a$ and $b$, $T$ will take the value

$$T = a + b \neq 0$$

If there are more than two errors, $T$ 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

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

to the numbers

0010110 and 0010010

will correspond the matrices

$$\begin{bmatrix} 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 \end{bmatrix}$$

The number $T$ is obtained in making it follow each line of the matrix of its digits to parity (31). One obtains here

$$\begin{bmatrix} 0 \end{bmatrix}$$

and $||b|| = 5$.

The first $\#$ is correct, the fifth digit of the second $\#$ is false.

Let's see how to which condition the numbers \( p \) must satisfy in order that the calculation \( \beta \) would permit to correct \( K \) errors. We will suppose that the first digits are insufficient so that knowing the digits \( q \) fingerprint one could realize the condition \( T = \sum a_i p_i = 0 \).

If after transmission the digits \( q \) the line \( a_1, a_2, \ldots, a_j \) (\( j \leq K \)) are erroneous the \( T \) calculated on the deformed word will take the value \( T = p_{a_1} + p_{a_2} + \ldots + p_{a_j} \). (a)

It is necessary that the number that is found be \( \delta \) of the rows \( a_1, a_2, \ldots, a_j \) that is to say

\[
p_{a_1} + p_{a_2} + \ldots + p_{a_j} \neq p_{a_1} + p_{a_2} + \ldots + p_{a_j},
\]

when \( j \leq K \) and \( j' \leq K \) and the two together

\( (a_1 a_2 \ldots a_j) (a_1' a_2' \ldots a_j') \) are not identical.

This condition can be written

\[
(4.2)
\]

when \( \delta \leq 2K \).
and the $A_1 A_2 \ldots A_7$ being all different. One ought to choose the $#5$ $p$ in such a way that the addition, in the ring $\mathbb{Z}$, at the most $2k$ of these numbers gives a result not null.

If one time determined a group of $n$ numbers $p$, it will be necessary to choose the test digits. It is helpful then to replace the group obtained by another group of $n$ numbers, but containing the successive powers of $2$:

$$2^0, 2^1, 2^2, \ldots, 2^{k-1}$$

$k$ designating the number of digits of the greatest number of $p$ obtained.

Place for this the numbers $p$ in a matrix $M$ of $n$ columns and $k$ lines ($k < n$) each $# p$ then being represented by a column.

$$A_i = \begin{pmatrix} \\
\end{pmatrix}$$

If the row of this matrix is the $0$-a $k^{\prime}$ line, there is $k - k^{\prime}$ lines of the matrix are $1$-linear combo $k^{\prime}$ lines remaining.

If one gets rid of $k - k^{\prime}$ lines you will obtain a matrix $m'$ of members $p$, which will verify again the condition (42).
This leaf is from the matrix $A$ with a column not being determined will have defined values $A = [0 \times 1]$. This is the only possibility determined value $A = [0 \times 1]$. Then, the matrix $A$ will have defined values $A = [0 \times 1]$. The matrix $A$ will contain only one column. This will explain why the matrix $A$ will contain only one column. This will explain why the matrix $A$ will contain only one column.
This is to say the matrix unity is for whole of the group of $p^n$ will contain the successive powers:

$$1, \ 2, \ 2^2, \ \ldots \ \ldots \ 2^{k-1}$$

The correct indication will be taken as test digits. The condition 41 will determine these digits in function of digits of

$$\Pi = 2^k \Pi$$

denying #'s of $k^n$ digits. All the problem comes back to constructing the group of #’s satisfying condition 42.

Formation & nearest near to the following (what comes after the $4^k$):

Take 1st $P_1 = 1 \ \& \ P_2 = 2^2 \ \ldots \ \Pi = 2^{2k-1}$
Then $P^{2K} + 1 = 2^{2K} - 1$

$P^{2K} + 2 = 2^{2K}$

These $H$'s satisfy already the conditions (42).

To extend this following in the order of $P$ increasing, suppose we arrive at the next $P$ of $l$ digits.

Let us consider the group of $\#'$s $P$ to $P_{i}$ of other sums in the 0 as by groups of 2, 3, ... ($2K+1$)

All the numbers obtained are at the most $l$ digits.

If a $\#$ digit not contained in the group then formed it included between $P_{i}$ and this number will be taken $2^{l}$.
the value of \( P_i + 1 \) (if there are
several \( P_i \)'s you will chose
naturally the smallest.)

If not you will take
\( P_i + 1 = 2 \).

One can thus continue step by step until you obtain
2 members \( P_i \).

All \( P_i \) to the base \( P \) has
18 digits, the \( 12 \)

\( 1, 2, 2^2, \ldots, 2^{k-1} \)

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 \( Z \) corresponding
to the \( P_i \) values of the sum

\( P_2 = P + \ldots, P_i = \ldots \)
the value of indication
\[ \begin{align*}
& \theta_1, \theta_2, \ldots, \theta_n \\
\end{align*} \]

the process thus defined is

- long to figure out
- enough to utilize

However for reasonable values of \( n + k \), it doesn't surpass the possibilities of the calculator or medium power.

The determination of the existence of effective systems is difficult enough.

therefore we are going to expose a more systematic process of research for \#\'s P.
The theory of Consequences, so much used in proofs of arithmetic operations, is going to furnish us a Method of Calculating numbers.

Let us designate by \( \xi \) a number unreduceable \( \xi \), \( N+1 \) digits \( \gamma \) by \( \xi' \) the rest of the division in the \( \gamma \) of a number \( \xi \) by \( \gamma \). The number \( \xi' \) will have the maximum \( N \) digits.

We shall place then:

\[
\xi' = \gamma + 2^N (3) + 
\]

that is to say that the number \( \xi' \) is formed by juxtaposition of the successive remainders of the division by \( \gamma \) by unequal powers \( \gamma \) of \( \gamma' \). The aim is going to show that these numbers \( \xi' \) satisfy the condition 42.

In effect suppose:

\[\text{(61)}\]

And that will bring about:

\[\text{(62)}\]

in placing: \( \xi = \ldots \)

and \( \xi' = \ldots \)

but if we can considered the product

\[
\Pi (\lambda_1 + 1) \quad i = 1, 2, \ldots
\]

\[y = 1, 2, \ldots \]

This product can be written in the form of a determinant of Van der Monde, of which the square will contain the
Then one of the factors for example \( A_i + A_j \), will be divisible by \( q \).
As the sum in \( \Omega \) of \( A_i + A_j \) has less digits than the number \( q \), it would result that
\[
A_i + A_j = 0 \quad i = j.
\]

Thus the hypotheses \( \Omega \) 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 on \( k \) errors. Naturally we will transform them as indicated in paragraph 4 to form a following containing the powers of 2. In the general case, \( p \) including \( k \) digits, there will be \( (you \ would \ use) \) \( k \) test digits.
\[ e = 19 = 10.8 \]

\[ M = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix} \]

1. Taking the numbers \( p \) and the matrix \( M \), the determinant \( \det(M) \) is 8, a square, \( 2 \times 2 \) form.

2. The determinant \( \det(M) \) is 8, forming the columns because they have the most zeroes, \( \det(M) = 8 \).

3. The matrix \( A \) (see Appendix) will be formed with these columns. Converting line-by-line.

4. The product \( A^{-1}M \) in the (\( A \)) gives a definitive matrix.

5. The characteristic of the range \( \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} \) gives \( d = 1 \).

6. The digits of the range \( \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} \) will serve as test digits, the seven other digits will be the supports for the info.
To verify and correct a word, you will make the sum
\[ T = \Sigma \varphi_i a_i \]
If it is null, there will be no error (no 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) by using the following table which gives the possible values of T followed in parentheses by the range of the false digits.

One will remark that the # T takes a 121 possible values \((1 + C_1^2 + C_2^2)\) 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 8 log 2T.