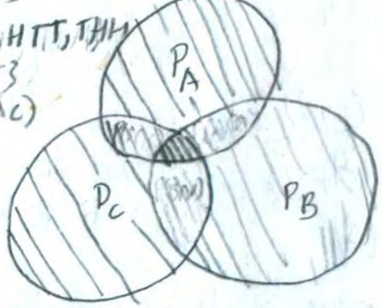
Chapter 1: 1a. Sample Space: {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

b. 1) {HHH, HHT, THH, HTH} 2) HHH, HHT 3) HHT, THT, TTT

c. A^c = "complement": the elements in the space which are not A. {HTT, TTT, TTH, THT}

$A \cap B$ = "intersection": the event both A and B occur.: {HHT, HHH}

$A \cup B$ = "Union": events of A and B, and A or B. {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}



2. a) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

$= P(A \cup B) \cup P(C) = \underbrace{[P(A) + P(B) - P(A \cap B)]}_{\text{Addition Law}} \cup P(C)$

$= P(A \cup C) + P(B \cup C) - P(A \cap B) \cup P(C)$

$= P(A) + P(C) - P(A \cap C) + P(B) + \cancel{P(C)} - P(B \cap C) - \cancel{P(C)} - P(A \cap B) + P(A \cap B \cap C)$

$\boxed{= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)}$

3. Semi

R R R G G W ; 3 draws ; RRR RRG RRW RGG GGW

RGR RWR GRG GWG WGR

GRR WRR GGR WGG WRG

$n = 6$
$k = 3$

$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{6!}{3!(3!)} = \frac{6 \cdot 5 \cdot 4}{6} = \frac{20}{6}$; $\dfrac{\frac{20}{6}}{\frac{6!}{1!(5!)} = 6} = \frac{20}{6 \cdot 6} = \frac{5}{9}$

Event A: 1 Draw

$\frac{P(R) + P(G) + P(W)}{P(G \cap R \cap W)} = \frac{\binom{3}{1} + \binom{2}{1} + \binom{1}{1}}{\binom{6}{1}} = \dfrac{\frac{3!}{1!(2!)} + \frac{2!}{1!\,1!} + \frac{1!}{1!\,0!}}{\frac{6!}{1!\,5!}} = \left(\frac{3}{6}\right) + \left(\frac{2}{6}\right) + \left(\frac{1}{6}\right)$

Event B: 2 Draw

$\frac{P(R) + P(G) + P(W)}{P(G \cap R \cap W)} = \ldots$ Should write Unions and intersection instead.

4. Prove $P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$; $P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1) + P(A_2) + \cdots + P(A_n) - P(A_1 \cap A_2) - \cdots - P(A_1 \cap A_n) - P(A_2 \cap A_3) - P(A_2 \cap A_n)$

$\sum_{i=1}^{n} P(A_i) = P(A_1) + P(A_2) + \cdots + P(A_n)$

$\boxed{P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)}$;

5. Let (A, not B) or (B, not A) be C

$\boxed{C = \neg A \cap B = (A \cap \neg B) \vee (\neg A \cap B) = A + B - A \cup B = A^c \cap B \vee A \cap B^c}$

Two six-sided dice are thrown: A) Sample space:

Dice 1:	Dice 2:
1	1
1	2
1	3
1	4
1	5
1	6
2	1
2	2
2	3
2	4
2	5
2	6
3	1
3	2
3	3
3	4
3	5
3	6
4	1
4	2
4	3
4	4
4	5
4	6
5	1
5	2
5	3
5	4
5	5
5	6
6	1
6	2
6	3
6	4
6	5
6	6

B)(1) A = sum of the two values is at least 5.

(1,5) (5,1) (1,6) (6,1) (4,4)
(2,4) (4,2) (2,6) (6,2) (4,5)
(3,3) (3,3) (3,6) (6,3) (5,4)
(2,5) (5,2) (4,6) (6,4)
(3,4) (4,3) (5,6) (6,5)
(3,5) (5,3) (6,6) (6,6)

(2) B = the value of the first dice greater than the second.

(2,1) (3,2) (4,3) (5,4) (6,5)
(3,1) (4,2) (5,3) (6,4)
(4,1) (5,2) (6,3)
(5,1) (6,2)
(6,1)

(3) C = the first value is 4

(4,1) (4,4)
(4,2) (4,5)
(4,3) (4,6)

c) $A \cap C = (4,2),(4,3),(4,4),(4,5),(4,6)$

$B \cup C = (2,1),(3,1),(5,1),(6,1),(3,2),(5,2),(6,2)$
$(5,3),(6,3),(5,4),(6,4),(6,5),(4,2),(4,3)$
$(4,4),(4,5),(4,6)$

$A \cap (B \cup C) = (4,2),(4,3),(4,4),(4,5),(4,6).$

7. Bonferroni's equality: $P(A \cap B) \geq P(A) + P(B) - 1$

Addition Law: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$P(A \cap B) = P(A) + P(B) - P(A \cup B)$

Therefore, $P(A \cap B) = P(A) + P(B) - P(A \cup B) \geq P(A) + P(B) - 1$

$\boxed{P(A \cup B) \leq 1}$

De Morgan's Law: $(A \cup B)^c = A^c \cap B^c$ 
$(A \cap B)^c = A^c \cup B^c$ 

9. Probability of rain on Saturday (25%)
Probability of rain on Sunday (25%)
The probability of consecutive events would be the multiplication of the probability of the events $\left(\frac{1}{4} \cdot \frac{1}{4}\right) = \frac{1}{16} = 12.5$
and not 50% proposed.

Information Theory and Reference: David MacKay.

Example 1.1: (Prob f) is Heads of a coin. (N) tosses. What is the prob-dist of heads (r)?

Binomial: $P(r|f,N) = \underbrace{\binom{N}{r}}_{\text{Binomial coefficient}} \underbrace{f^r}_{\text{Prob heads}} \underbrace{(1-f)^{N-r}}_{\text{Prob tails}}$

Mean: $\mathcal{E}[r] \equiv \sum_r P(r|f,N) \cdot r$

$var \equiv \mathcal{E}[(r-\mathcal{E}[r])^2]$

$= \mathcal{E}[r^2] - (\mathcal{E}[r])^2 = \sum_{r=0}^{N} P(r|f,N) r^2 - (\mathcal{E}[r])^2$

Exercise 1.2: Prove error probability is reduced by using R_3 by computing the error probability for a binary symmetric channel with noise level f?

R_3 is defined as a bit sequence of XXX where $X \in 0,1$.

With probability of a bit flipped being f,

with the probability of two bits being flipped $3f^2(1-f)$

and the bits flipped having probability f^3.

the probability distributions are:

$r=1 \quad P(r=1|f,N=3) = \binom{3}{1} f (1-f)^{3-1}$

$r=2 \quad P(r=2|f,N=3) = \binom{3}{2} f^2 (1-f)^1$

$r=3 \quad P(r=3|f,N=3) = \binom{3}{3} f^3$

Exercise 1.3: a) Show probability of error R_n, over n-repetitions is

$P_b = \sum_{n=(N+1)/2}^{N} \binom{N}{n} f^n (1-f)^{N-n}$ For odd n.

even : $n = 2N$

odd : $n = 2N+1$

b) $\simeq \sum_{n=(N+1)/2}^{N} \binom{N}{(N+1)/2} f^{(N+1)/2} \cdot (1-f)^{N-\frac{N}{2}-\frac{1}{2}} = \sum_{(N+1)/2}^{N} \binom{N}{\lceil N/2 \rceil} f^{(N+1)/2} \cdot (1-f)^{(N-1)/2}$

The Binary Entropy Function $H_2(x) \equiv x \log \frac{1}{x} + (1-x) \log \frac{1}{(1-x)}$

C) Which relates to Sterling Aprox: $x \ln x - x + \frac{1}{2} \ln 2x = x!$

$\ln \binom{N}{r} = \ln \frac{N!}{(N-r)!\,r!} = \frac{N \ln N - N + \frac{1}{2} \ln 2N}{((N-R)\ln(N-r) - N - r + \frac{1}{2}\ln 2(N-r))(r \ln r - r + \frac{1}{2}\ln(2r))}$

$\simeq (N-r) \ln \left(\frac{N}{N-r}\right) + r \ln \frac{N}{r}$

If rewritten, $\log \binom{N}{r} \simeq N H_2(r/N)$; $\binom{N}{r} \simeq 2^{N H_2(r/N)}$

$\simeq N H_2(r/N) \cdot \frac{1}{2} \log \left[2\pi N \frac{N-r}{N} \frac{r}{N}\right]$

Back to the exercise,

$\binom{N}{K} = \frac{1}{N+1} 2^{N H_2(K/N)} \leq \binom{N}{K} \leq 2^{N H_2(K/N)} \Rightarrow \binom{N}{K} \simeq 2^{N H_2(K/N)}$

$P_b \simeq 2^{N H_2(K/N)} \cdot f^{N/2} (1-f)^{N/2} = 4f(1-f)^{N/2}$

d) A prob 10^{-15} requires $N \simeq 2 \frac{\log -15}{\log 4f(1-f)}$ Largest

Exercise 1.4: Prove HG^T

$H = [P \; I_3] = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$ "Parity Check"

$G^T = \begin{bmatrix} I_4 \\ P \end{bmatrix}$ "Generator"

Exercise 1.7: Permutation of XXXX ; X ∈ 0, 1 :

0001	1001
0010	1010
0011	1011
0100	1100
0101	1101
0110	1110
0111	1111
1000	

Exercise 1.8: Block Decoding Error

(7,4) Hamming Code

$$P_B = \sum_{n}^{\infty} \binom{7}{r} f^r (1-f)^{7-r}$$

$$= 7f(1-f)^6 + 7 \cdot 6 f^2 (1-f)^5 + \dots$$

Exercise 1.9: Prepare a bipartite graph. The (7,4) Hamming Code ie 7 circles and 3 squares as parity check.

- Parity-check is a row of H [N]
- Bit-node is a column of H [M]

$\boxed{(7,4), (30,11), (N;M)}$

Exercise 1.10:
- The amount of weight two patterns generated is $\binom{N}{2} + \binom{N}{1} + \binom{N}{0} = \frac{N!}{(N-2)!\,2!} + \frac{N!}{(N-1)!\,1!} + \frac{N!}{(N-0)!\,0!}$
- The amount of syndromes would be 2^{N-M} or $2^6 = 64$ syndromes.

$$= \frac{14!}{12!\,2!} + \frac{14!}{(13)!\,1!} + \frac{14!}{(14)!}$$

$$= 91 + 14 + 1 = \boxed{106 \text{ patterns}}$$

- the total amount of pattern would not be solved by the amount of syndromes.

Exercise 1.11: $2^{N-M} > \left[\binom{N}{2} + \binom{N}{1} + \binom{N}{0}\right] \Rightarrow \boxed{(30,11)}$

Exercise 1.12: Probability is represented as Binomial: $P = \sum_{m=0}^{n} \binom{n}{m} f^m (1-f)^{n-m}$

$$p(R_3^2) = \sum_{m=2}^{3} \binom{3}{2} f^2 (1-f)^{3-m} = 3[p(R_3)]^2 = 3\left[\sum_{m=1}^{3} \binom{3}{1} f^t (1-f)^{3-m}\right]^2 = 3(3f^2)^2 + \dots$$

$$P(R_9) = \sum_{m=1}^{9} \binom{9}{m} f^m (1-f)^{N-m} = \binom{9}{5} f^5 (1-f)^4 + \dots = 126 f^5$$

An advantage of the small R_3 encoder is ability to process smaller pieces of 3-bit code.

2.2 The datapoints of figure 2.2 are not independent because a joint probability $[p(X,y) = p(x|y) \cdot p(y) = p(y|x) \cdot p(x)]$ is not seperable in this instance.

2.3.

	Has Disease	No Disease	
Positive	0.95	0.05	1.00
Negative	0.05	0.95	1.00
	1.00	1.00	2.00

$$p(\text{Has Disease} \mid \text{Positive}) = \frac{p(\text{Positive} \mid \text{Has Disease}) \cdot P(\text{Has Disease})}{P(\text{Positive} \mid \text{Has Dis.})\, p(\text{Has}) + p(\text{Negative} \mid \text{No}) \cdot p(\text{No})}$$

$$= \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.95 \cdot 0.99} = \boxed{1\%}$$

Note: p(Has Disease) = 0.01 for Joe's family

2.4. a. Urn [K balls, B black, $W = K - B$; N draws with replacement]

Fraction of Black $= \frac{B}{K}$; The distribution of drawing with replacement $\boxed{P(n \mid f, N) = \binom{N}{n} f^{n} (1-f)^{N-n}}$

b. $E(p(n \mid f, N)) = \sum_{0}^{\infty} p(n \mid f, N) \cdot n = N \cdot f$; $Var(p(n \mid f, N)) = E\left([n - E(n)]^2\right) = \boxed{E[n^2] - E[n]^2}$

$= Nf(1-f)$

2.5 K = Total Balls; B = Black, $W = K - B$ White;

$f_B = B/K$; N = Draws without replacement.

Standard Deviation $(p(n \mid f, N)) = \sqrt{Nf(1-f)}$

$E[Z] = E\left[\frac{(n_B - f_B N)^2}{N f_B (1-f_B)}\right] = \frac{\sum n_B (n_B - f_B N)}{N f_B (1 - f_B)}$

$= 1$

$N = 5$; $\sigma = \sqrt{5 \frac{2}{10} \left(1 - \frac{2}{10}\right)} = \sqrt{8/10}$

$N = 400$; $\sigma = \sqrt{400 \cdot \frac{2}{10} \left(1 - \frac{2}{10}\right)} = 8$.

$\boxed{\text{Probability Distribution: } N = 5, f_B = 1/5;\ Z = \frac{5}{4}(n_B - 1)^2 \text{ where } n_B = 1,2,3,4,5 \text{ is } P(n_B) = \binom{N}{n_B} f^{n} (1-f)^{N-n}. \text{ The values of the probability distribution less than 1 are } n_B = 1,\ P(n_B = 1) = 0.4096.}$

Example 2.6 $u \in \{0,1,2 \ldots 10\}$ each containing 10 balls. u has u black balls, $10-u$ white balls.

$$P(u, n_B \mid N) = P(n_B \mid u, N)\, P(u); \quad P(u \mid n_B, N) = \frac{P(u, n_B \mid N)}{P(n_B \mid N)} = \frac{P(n_B \mid u, N)\, P(u)}{P(n_B \mid N)} = \frac{\binom{N}{n_B} u^{n_B} (1-u)^{N-n_B} \cdot \left(\frac{1}{11}\right)}{P(n_B \mid N)}$$

$$P(n_B \mid N) = \sum_{u=0}^{10} P(u) \cdot P(n_B \mid u, N) = \frac{1}{10}\left[\sum P(n_B \mid u, N)\right] = \frac{1}{10}\left[\binom{10}{3}\left(\frac{0}{10}\right)^3 \left(1 - \frac{0}{10}\right)^{10-3} + \binom{10}{3}\left(\frac{1}{10}\right)^3\left(1 - \frac{1}{10}\right)^{10-3} + \cdots + \binom{10}{3}\left(\frac{10}{10}\right)^3\left(1 - \frac{10}{10}\right)^{10-3}\right]$$

$$= \frac{1}{10}[0.8287] = \boxed{0.0829}$$

u	0	1	2	3	4	5	6	7	8	9	10
$P(u \mid n_B, N)$	0	0.063	0.22	0.29	0.24	0.13	0.047	0.01	0.00	0.000	0

$$P(\text{Next Ball} \mid n_B, N) = \sum P(\text{Next Ball} \mid u, n_B, N)\, P(u \mid n_B, N) = \sum_{u=1}^{10} \frac{u}{10} P(u \mid n_B, N)$$

$$P(\text{Next Ball} \mid n_B = 3, N = 10) = \boxed{0.33}$$

Example 2.7. N tosses; $P(\text{Heads}) = f_H$; n_H = Number of Heads

What is the PDF of f_H? $P(n_B \mid f_H, N) = \binom{N}{n_B} f_H^{n_B} (1 - f_H)^{N - n_B}$

$\boxed{P(\text{Next Ball} \mid n_B, N) = P(\text{Next Ball} \mid f_H, n_B, N) \cdot \cancel{P(n \mid n_B, N)}}$

Exercise 2.8 . Prior $= P(n_B \mid f_n, N)$; Marginalization $= P(f_n)$

a. $$P(f_n, n_H = 0 \mid N = 3) = \frac{\sum_{u=0} P(f_n) P(n_H = 0 \mid f_n, N = 3)}{\sum P(n_H = 0 \mid N = 3)} = \frac{\binom{3}{0} f_n^{0} (1 - f_n)^{3-0}}{\binom{3}{0} \int_0^1 f_n^{0} \cdot (1 - f_n)^{3-0}\, df_n} = \boxed{\frac{(1 - f_n)}{\frac{\Gamma(1)\Gamma(3)}{\Gamma(5)}}}$$

b. $$P(f_n, n_H = 2 \mid N = 3) = \frac{\binom{3}{2} f_n^2 (1 - f_n)^{3-2}}{\binom{3}{2} \int_0^1 f_n^2 (1 - f_n)^{3-2}\, df_n} = \frac{f_n^2 (1 - f_n)}{\frac{\Gamma(3)\Gamma(2)}{\Gamma(5)}} = \frac{4!}{2!\,1!} f_n^2 (1 - f_n) = \boxed{2 f_n^2 (1 - f_n)}$$

c. $$P(f_n, n_H = 3 \mid N = 10) = \frac{\binom{10}{3} f_n^3 (1 - f_n)^{10-3}}{\binom{10}{3} \int_0^1 f_n^3 (1 - f_n)^{10-3}\, df_n} = \frac{f_n^3 (1 - f_n)^7}{\frac{\Gamma(4)\Gamma(8)}{\Gamma(12)}} = \frac{11!}{3!\,7!} f_n^3 (1 - f_n)^7 = \boxed{1320 f_n^3 (1 - f_n)^7}$$

d. $$P(f_n, n_H = 29 \mid N = 300) = \frac{\binom{300}{29} f_n^{29} (1 - f_n)^{271}}{\binom{300}{29} \int_0^1 f_n^{29} (1 - f_n)^{271}\, df_n} = \boxed{\frac{301!}{29!\,271!} f_n^{29} (1 - f_n)^{271}}$$

s, of length K,

inction of the

.

ttabits

when each circle is even.

$= G^T s$

1111111

$\ldots 0, 0)$

ise level f

$\simeq 21 p^2 f^5$

Example 2.9

```
int[] threebitCompression(String bits){
    int size = bits.size()%3;
    int[size] compression;
    int [8] model = {000,001,010,011,100,101,110,111}
    for(int i=0; i<size; i++){
      for(int j=0; j<8; j++){
        if(bits.substr(3*i, 3*(i+1)) == model[j]){
          compression[i] += 1;
        }
      }
    }
    return compression;
}
```

Example 2.10. $P(\text{Urn A} \mid \text{Black Ball}) = \dfrac{P(\text{Black Ball} \mid \text{Urn A})\, P(\text{Urn A})}{P(\text{Black Ball} \mid \text{Urn A}) \cdot P(\text{Urn A}) + P(\text{Black Ball} \mid \text{Urn B}) \cdot P(\text{Urn B})}$

$$= \frac{(\frac{1}{3})(\frac{1}{2})}{(\frac{1}{3})(\frac{1}{2}) + (\frac{2}{3})(\frac{1}{2})} = \frac{\frac{1}{6}}{\frac{1}{6} + \frac{2}{6}} = \boxed{\frac{1}{3}}$$

Example 2.11 $P(\text{Urn A} \mid \text{Black Ball}) = \dfrac{P(\text{Black Ball} \mid \text{Urn A})\, P(\text{Urn A})}{P(\text{Black Ball} \mid \text{Urn A})\, P(\text{Urn A}) + P(\text{Black Ball} \mid \text{Urn B})\, P(\text{Urn B})}$

$$= \frac{(\frac{1}{5})(\frac{1}{2})}{(\frac{1}{5})(\frac{1}{2}) + (\frac{2}{5})(\frac{1}{2})} = \frac{\frac{1}{10}}{\frac{1}{10} + \frac{2}{10}} = \boxed{\frac{1}{3}}$$

Example 2.12 Using Table 2.9: $H(x) = \sum_{i=1}^{27} p(x) \cdot \log \frac{1}{p(x)} = \boxed{4.1}$

Example 2.13 $H(x) = 1 \cdot \log \frac{1}{1/3} + \frac{1}{3} \log \frac{1}{1/10} + \frac{1}{3} \log \frac{1}{1/5} + \frac{1}{3} \log \frac{1}{1/21} = \boxed{1.48}$

Exercise 2.14 Proof of $E[f(x)] \geq f(E[x])$; $E[f(\lambda X_1 + (1-\lambda) X_2)] \geq \lambda f(E[X_1]) + (1-\lambda) f(E[X_2])$

if $\lambda = 1$; then $E[F(X_1)] \geq F(E[X_1])$ and $f(X_1) \geq \frac{1}{p(X_1)} f(E[X_1])$

if $\lambda = 0$; then $E[f(X_2)] \geq F(E[X_2])$ and $F(X_2) \geq \frac{1}{p(X_2)} F(E[X_2])$

if $0 < \lambda < 1$; then $f(X_1) \leq f(\lambda X_1 + (1-\lambda) X_2) \leq f(x_2)$

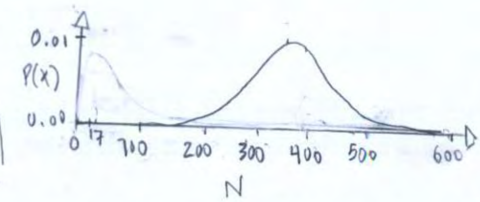
Example 2.15. Jensen's inequality: $E[f(x)] \geq f(E[x])$; $E[f(x)] = A = 100 m^2 \geq \bar{\ell}^2 = \boxed{10^2 m^2}$

Exercise 2.16 a) $P(X,y) = \text{bin}(n,p) = \binom{N}{n} p^n (1-p)^{N-n} = \binom{N}{2} (\frac{1}{6})^2 (1-\frac{1}{6})^{N-2}$

$P(x,y) = \text{bin}(n,p) = \boxed{\binom{N}{2} \left(\frac{1}{6}\right)^2 \left(1-\frac{1}{6}\right)^{N-2}}$

b) $P(X,y) = \text{bin}(n,p) = \boxed{\binom{N}{100} \left(\frac{1}{6}\right)^{100} \left(1-\frac{1}{6}\right)^{N-100}}$

$\boxed{E[x] = \sum_{100}^{600} n_i \cdot P(x_i)}$; $\boxed{SD[x] = \sqrt{\sum_{100}^{600} n_i p(x_i)(1-p(x_i))}}$



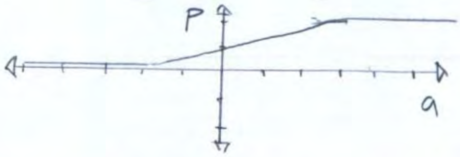
c) p_i, r_i = probabilities of Dice #1, #2 for a sum of $i = 1, 2, \ldots, 11, 12$ and options of 0-6 possible side labels.

$$\frac{1}{11}(x + x^2 + x^3 + \cdots + x^{12}) = (p_0 x^0 + p_1 x^1 + \cdots + p_6 x^6)(r_0 x^0 + r_1 x^1 + \cdots + r_6 x^6)$$

$= \boxed{P(S=1) = p_0 r_0 = P(S=12) = p_6 r_6}$ "50% 0's and 6's"

d) Yes, by crafting 100 Dice from wood, then labeling them $\{0,1,2,3,4,5\} \times 6^{100}$

Exercise 2.17. $q = 1-p$; $a = \ln p/q$; $e^{-a} = \frac{1-p}{p}$; $p(1+e^{-a}) = 1$; $p = \frac{1}{1+e^{-a}}$;



$$p = \frac{1}{1+e^{-a}} = \frac{2}{2}\left(\frac{1}{1+e^{-a}}\right) = \frac{1}{2}\left(\frac{2e^{-a}+2}{1+e^{-a}}\right) = \frac{1}{2}\left(\frac{1-e^{-a}}{1+e^{-a}}+1\right)$$

$$= \frac{1}{2}\left(\frac{e^{a/2}-e^{-a/2}}{e^{a/2}-e^{-a/2}}+1\right) = \frac{1}{2}(\tanh(a/2)+1) \; ; \text{ if } b = \log_2 q/p \; ; \boxed{p = \frac{q}{2^b}}$$

Exercise 2.18. $A_x = \{0,1\}$; Bayes Theorem: $\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{evidence}}$; $P(x|y) = \frac{P(y|x)P(x)}{P(y)}$

$$\log \frac{P(x=1|y)}{P(x=0|y)} = \boxed{\log \frac{P(y|x=1)P(x=1)}{P(y|x=0)P(x=0)}}$$

Exercise 2.19. Bayes Theorem: $\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$; $P(x|y) = \frac{P(y|x)P(x)}{P(y)}$

$$\frac{P(x=1|\{d_i\})}{P(x=0|\{d_i\})} = \frac{P(\{d_i\}|x=1)P(x=1)}{P(\{d_i\}|x=0)P(x=0)} = \boxed{\frac{P(d_1|x=1)P(d_2|x=1)P(x=1)}{P(d_1|x=0)P(d_2|x=0)P(x=0)}}$$

Exercise 2.20 Volume of an n-dimensional ball: $V_n(R) = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} R^n$

$$F = \frac{\text{Part of Volume}}{\text{Total Volume}} = \frac{\text{volume}(R)}{\text{Total Volume}} - \frac{\text{volume}(R-\epsilon)}{\text{Total Volume}} = 1 - \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} \frac{\Gamma(\frac{n}{2}+1)}{\pi^{n/2}} \left(\frac{R-\epsilon}{R}\right)^N = 1 - \left(1-\frac{\epsilon}{R}\right)^N$$

$N=2$; $\frac{\epsilon}{r} = 0.01$; $f = 1-(1-0.01)^2 = \boxed{0.0199}$

$\frac{\epsilon}{r} = 0.5$; $f = 1-(1-1/2)^2 = \boxed{0.75}$

$N=10$; $\frac{\epsilon}{r} = 0.01$; $f = 1-(1-0.01)^{10} = \boxed{0.096}$

$\frac{\epsilon}{r} = 0.5$; $f = 1-(0.5)^{10} = \boxed{0.999}$

$N=1000$; $\frac{\epsilon}{r} = 0.01$; $f = 1-(1-0.01)^{1000} = \boxed{0.99995}$

$\frac{\epsilon}{r} = 0.5$; $f = 1-(1-0.5)^{1000} = \boxed{1.0000}$

Conclusion: Higher dimensional Fractional sphere relationships approach singularity.

Exercise 2.21: $p_a = 0.1$; $p_b = 0.2$; $p_c = 0.7$; Let $f(a) = 10$, $f(b) = 5$, $f(c) = \boxed{10/7}$

$E[f(x)] = \sum P(x) \cdot f(x) = 0.1 \times 10 + 0.2 \times 5 + 0.7 \times \frac{10}{7} = \boxed{3.0}$

$E[1/P(x)] = \sum P(x) \cdot \left(\frac{1}{P(x)}\right) = \boxed{3.0}$

Exercise 2.22: $E[1/P(x)] = \boxed{\sum 1}$

Exercise 2.23: $p_a = 0.1$; $p_b = 0.2$; $p_c = 0.7$; $g(a) = 0$; $g(b) = 1$; $g(c) = 0$; $E[g(x)] = 0.2 \cdot 1.0 = \boxed{0.2}$

Exercise 2.24: $p_a = 0.1$; $p_b = 0.2$; $p_c = 0.7$; For a discrete value, $\boxed{p_b = 0.2}$

$P\left(\left|\log \frac{P(x)}{0.2}\right| > 0.05\right) = P(|\log(1)| > 0.05) = \boxed{0\%}$

Exercise 2.25: $H(x) \leq \log(|A_x|)$ with equality $p_i = 1/|A_x|$; Jensen's Equality: $E[f(x)] \geq f[E(x)]$

$H(x) = \sum P(x) \log \frac{1}{P(x)} \leq \log\left(\frac{1}{|A_x|}\right)$; Applying Jensen's Equality:

$E\left[|\log| \frac{1}{P(x)}\right] < \log \sum P(x) \frac{1}{P(x)}$

$E[1/P(x)] = H(x) \geq \log\left(\sum P(x) \cdot \frac{1}{|A_x|}\right) = \log\left(\sum P(x) \frac{1}{P(x)}\right) = 0$

$\boxed{H(x) \geq 0}$

Exercise 2.26: Kullbeck-Leibler Divergence: $D_{KL}(P\|Q) = \sum P(x)\log\frac{P(x)}{Q(x)}$.

Gibbs Inequality: $D_{KL}(P\|Q) \geq 0$

If $P=Q$; $D_{KL}(P\|Q) = \sum P(x)\log(1) = \boxed{0}$; Domain & Range of Log.

Exercise 2.27: Equation (2.43) $H(\vec{p}) = H(p_1, 1-p) + (1-p_1)H\left(\frac{p_2}{1-p_1}, \frac{p_3}{1-p_1}, \ldots, \frac{p_I}{1-p_I}\right)$

Equation (2.4.4) $H(\vec{p}) = H[(p_1+p_2+\cdots+p_m), (p_{m+1}+p_{m+2}+\cdots+p_I)]$

$$+(p_1+\cdots+p_m)H\left(\frac{p_1}{p_1+\cdots+p_m}, \ldots; \frac{p_m}{p_1+\cdots+p_m}\right)$$

$$+(p_{m+1}+\cdots+p_I)H\left(\frac{p_{m+1}}{p_{m+1}+\cdots+p_I}, \ldots, \frac{p_I}{p_{m+1}+\cdots+p_I}\right)$$

$H(p)$: Entropy Part #1 + Entropy Part #2

$$\boxed{\begin{aligned} &= H(p) + (1-p)H\left(\frac{p_2}{1-p_1}, \frac{p_3}{1-p_1}, \ldots, \frac{p_1}{1-p_1}\right) \\ &= H([p_1+p_2+\cdots+p_m], [p_{m+1}+p_{m+2}+\cdots+p_1]) \\ &\quad + (p_1+\cdots+p_m)H\left[\frac{p_1}{p_1+\cdots+p_m}, \ldots, \frac{p_m}{p_1+\cdots+p_m}\right] \\ &\quad + (p_{m+1}+\cdots+p_\ell)H\left[\frac{p_{m+1}}{p_{m+1}+\cdots+p_\ell}, \ldots, \frac{p_\ell}{(p_{m+1}+\cdots+p_\ell)}\right] \end{aligned}}$$

Exercise 2.28 $X \in \{0,1,2,3\}$; $P_A(\{0,1\}) = f$; $P_A(\{2,3\}) = 1-F$

$P_B(\{0\}) = g$; $P_B(\{1\}) = 1-g$ $\quad P(\{0,1,2,3\}|f,N) = \binom{N}{n} f^n (1-f)^{N-n}$

$P_C(\{2\}) = h$; $P_C(\{3\}) = 1-h$ $\quad =$

$H(x) = H(f) + f\cdot H(g) + (1-f)H(h)$

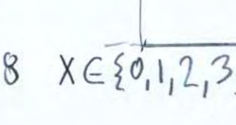
$= P(f)\cdot\log P(f) + f\cdot p(g)\log P(g) + (1-f)P(h)\cdot\log P(h)$

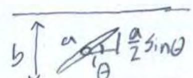
$\frac{dH(x)}{df} = P(g)\log P(g) - P(h)\cdot\log P(h) + \log\frac{1-x}{x} = \boxed{\log\frac{1-x}{x} + H(g) - H(h)}$

Exercise 2.29 $H(x) = \sum_{x=1}^{n} P(x)\log\frac{1}{P(x)} = \sum_{x=1}^{n}\binom{N}{x}(1/2)^x(1-1/2)^{N-x}\cdot[x\cdot\log(1/2) - (N-x)\log(1/2) + \log\binom{N}{x}]$

If $N=x$ because flips till heads, $= \boxed{\sum\left(\frac{1}{2}\right)^x[x\log 2]}$

Exercise 2.30 Urn $=\{w, \ldots, b\ldots\}$ $P(\text{Draw\#2}|\text{White}) = P(\text{Draw\#1}, P(\text{Draw\#2}|\text{White})) = \boxed{P(\text{Draw\#2}|\text{White})P(\text{Draw\#1})}$

Exercise 2.31  $a<b$ Fraction the coin will land in an Area $= \left[\frac{\text{Length}(b) - \text{Length}(a)}{\text{Length of a side}}\right]^2 = \boxed{\left(1-\frac{a}{b}\right)^2}$

Exercise 2.32 $P(a<b) = \int_0^{\pi/2}\int_0^{\frac{a}{2}\sin\theta}\frac{4}{\pi b}\,da\,d\theta = \int_0^{\pi/2}\frac{2a\sin\theta}{\pi\cdot b}\,d\theta = \frac{2a}{\pi b}$ as derived from the photo: 

Example 2.33.  ⊗ = Random Point.

Law of Cosines: $a^2 = b^2 + c^2 - 2bc\cos A$ $\boxed{\text{Eqn 2}}$

Requirements: $a = \frac{1}{2}$; $c = \frac{1}{2}(1-2b)$

$\boxed{\text{Eqn 1}}$ $a+b+c = 1$

$a = 1-b-c$

$(1-b-c)^2 = 1 + 2(bc-b-c) + c^2 + b^2$

$1 + 2(bc-b-c) = -2bc\cos A$;

$P(a,b,c) = P(a)\cdot P(c|b)$; $E[P(a=1/2)] = \sum P(a=1/2) = \frac{1}{100}$

$E[P(c=\frac{1}{2}(1-2b)|b)] = \sum P(c=\frac{1}{2}(1-2b)) = \frac{1}{50}$

$$\boxed{= \frac{1}{100}\cdot\frac{1}{50} = \frac{1}{5\times10^3}}$$

$\int a^x dx = a^x / \ln a$

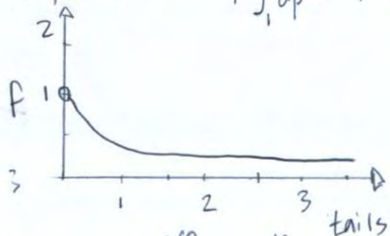
Exercise 2.34. $P(k = \text{tails}) = (1-p)^{k-1} \cdot p$ where $k = 1, 2, 3, \ldots n$

$E[P(k)] = \sum_1^\infty k(1-p)^{k-1} \cdot p = \int_1^\infty k(1-p)^{k-1} p \, dk = p \int_1^\infty \frac{-d}{dp}(1-p)^k dk = -p \frac{d}{dp} \frac{-(1-p)}{\ln(1-p)}$

$E[P(\text{Heads})] = 1$

Fred estimates $f = h/(h+t)$

Assuming $h = 1$;



$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - $

$= -p \frac{d}{dp} \frac{(1-p)}{p} = p\left(\frac{1}{p^2}\right) = \boxed{\frac{1}{p}}$

@ $p = 0.5 \approx 2$

Exercise 2.35. a) $E[P(k)] = \sum_1^\infty k(1-p)^{k-1} \cdot p = \int_1^\infty k(1-p)^{k-1} p \, dk = p \int_1^\infty \frac{-d}{dp}(1-p)^k dk = -p \frac{d}{dp} \frac{-(1-p)}{\ln(a)}$

$= -p \frac{d}{dp} \frac{(1-p)}{p} = p\left(\frac{1}{p^2}\right) = \boxed{\frac{1}{p}}$

@ $p = \frac{1}{6} \approx 6$

b) $\boxed{\text{Similar to part a}}$

c) $\boxed{\text{Similar to part a.}}$

d) The sum of $E[P(k \mid \text{Before clock})] + E[P(k \mid \text{After clock})] - 1 = \boxed{11 \text{ rolls.}}$

e) $\boxed{\text{The answer of part d is different from part a because the dice roller, Fred, must consider the random probability of the clock.}}$

Exercise 2.36. Fred has brothers Alf and Bob.

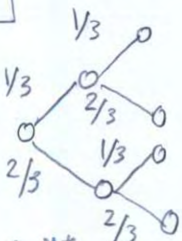
The opportunity Fred is older than Alf would be $\frac{\text{Probability (Fred>Alf)}}{\text{Total Probability}} = \frac{FAB, FBA, BFA}{FAB, FBA, BFA, AFB, ABF, BAF} = 1/2 = 50\%$

This opportunity is equivalent to Fred's age being greater than Bob's age.

If Fred is older than Alf and Bob: $P(F>B \mid F>A) = FBA, BFA, FAB$ $\boxed{= 2/3}$

Exercise 2.37. $P(\text{Truth}) = 1/3$; $P(\text{Lie}) = 2/3$

$P(\text{Truth} \mid \text{Person \#2}) = \frac{P(\text{Person \#2} \mid \text{Truth}) P(\text{Truth})}{P(\text{Person \#2})} = \frac{(1/3)(1/3)}{(1)} = \boxed{\frac{1}{9}}$



Exercise 2.38. Binomial Distribution Method: $P(\text{3-bits}) + P(\text{2-bits}) = 3f^2(1-f) + f^3$

where $P(N, n) = \sum_{n=(N+1)/2} \binom{N}{n} f^n (1-f)^{N-n}$

Sum rule Method: $P(r) = \sum P(s) \cdot P(r \mid s)$

$P(\text{error}) = P(\text{error}) \cdot P(\text{error} \mid r = 000) + P(\text{error}) P(\text{error} \mid r = 111)$

$+ P(\text{error}) \cdot P(\text{error} \mid r = 001) + P(\text{error}) \circ P(\text{error} \mid r = 010)$

$+ P(\text{error}) \cdot P(\text{error} \mid r = 011) + P(\text{error}) \circ P(\text{error} \mid r = 100)$

$+ P(\text{error}) \cdot P(\text{error} \mid r = 101) + P(\text{error}) \circ P(\text{error} \mid r = 110)$

$= 2P(\text{error}) \circ P(\text{error} \mid r = 000) + 6 P(\text{error}) \cdot P(\text{error} \mid r = XXY)$

$\Rightarrow$

Exercise 2.29 $P(k) = (1-p)^{k-1} p$; $H(\lambda) = \sum p(x) \ln \frac{1}{p(x)} = -\sum_{k=1}^{n} (1-p)^{k-1} p \left[(k-1)\log(1-p) + \log p\right]$

$$= -p\log p \sum_{1}^{\infty} (1-p)^{k-1} - p\log(1-p) \sum_{1}^{n} (1-p)^{k-1}(k-1)$$

"Infinite geometric progression" $\sum^{\infty} (a_n)^{k-1} = \frac{1}{1-a_n}$; $(0 + (1-p) + (1-p)^2 + \cdots)$

$$= -p\log p \cdot \left(\frac{1}{p}\right) - p\log(1-p)\left(\frac{1-p}{1-(1-p)}\right)$$

$$= \underline{-p\log p - (1-p)\log(1-p)}$$

If the coin had a bias of f, then the entropy would be reassigned as $f = p$ to become $H(\lambda) = \underline{-f\log f - (1-f)\log(1-f)}$.

Exercise 2.39 $p_n \simeq \begin{cases} \frac{0.1}{n} & \text{for } n \in 1, \ldots, 12367 \\ 0 & n > 12367 \end{cases}$

"Zips Law (1949)" $H(\lambda) = \sum p(\lambda) \log \frac{1}{p(\lambda)} = \sum_{1}^{12367} \frac{0.1}{n} \log \frac{n}{0.1} = \boxed{2.92 \text{ bits per word}}$

Chapter 3:

Exercise 3.1. $P(\{5,3,9,3,8,4,7\} | \text{Dice A}) = P(5|\text{Dice A}) \cdot P(3|\text{Dice A}) \cdot P(9|\text{Dice A}) \cdot P(3|\text{Dice A}) P(8|\text{Dice A}) P(4|\text{Dice A}) P(7|\text{Dice A})$

$$= 9.84 \times 10^{-8}$$

$P(\{5,3,9,3,8,4,7\} | \text{Dice B}) = P(5|\text{Dice B}) P(3|\text{Dice B}) P(9|\text{Dice B}) P(3|\text{Dice B}) P(8|\text{Dice B}) P(4|\text{Dice B}) P(7|\text{Dice B})$

$$= 5.0 \times 10^{-8}$$

$$P(\text{Dice A} | \{5,3,9,3,8,4,7\}) = \frac{P(\{5,3,9,3,8,4,7\} | \text{Dice A})}{P(\{5,3,9,3,8,4,7\} | \text{Dice A}) + P(\{5,3,9,3,8,4,7\} | \text{Dice B})}$$

$\boxed{= 0.6636\,66\%}$ chance Dice A.

Exercise 3.2. a) $P(\{3,5,4,8,3,9,7\} | \text{Dice C}) = \boxed{9.77 \times 10^{-14}}$

$P(\{3,5,4,8,3,9,7\} | \text{Dice A}) = \boxed{66\%}$

b) $P(\{3,5,4,8,3,9,7\} | \text{Dice B}) = \boxed{33\%}$

c) $P(\{3,5,4,8,3,9,7\} | \text{Dice C}) < \boxed{1\%}$

Exercise 3.3. $x = 1\text{cm}$ to $x = 20\text{cm}$. Exponential Distribution $p(\lambda) = \frac{e^{-x/\lambda}}{\lambda}$

Number of steps = 20.

$$P(\lambda) = \int_{1}^{20} \frac{e^{-x/\lambda}}{\lambda} dx = e^{-1/\lambda} - e^{-20/\lambda}$$

$$P(x|\lambda) = \frac{e^{-x/\lambda}}{\lambda P(\lambda)} \quad \text{for } x \in \{1 \ldots 20\}$$

$$P(\lambda | \{1 \ldots 20\}) = \frac{P(\{1 \ldots 20\} | \lambda) P(\lambda)}{P(\{1 \ldots 20\})} = \frac{1}{P(\{1 \ldots 20\})} \frac{e^{-\sum_{1}^{20} x/\lambda}}{(\lambda P(\lambda))^N}$$

Lambda (λ) describes the length between particles on the detector, and if collected per second, the rate between particles.

Exercise 3.4: $P('O') = 60\%$; $P('AB') = 1\%$; $P(\text{Scene} \mid \text{Person}, 'AB') = P('AB') = \boxed{1\%}$

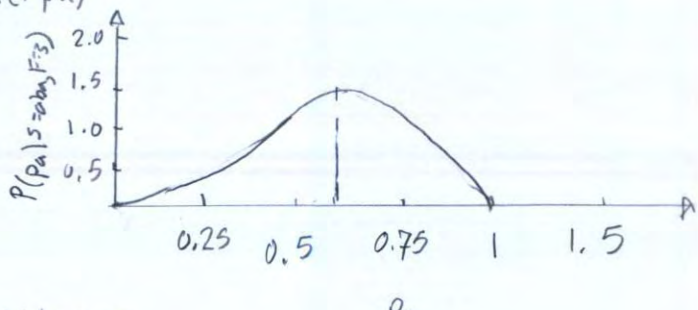
$$P(\text{Scene} \mid \text{Each Person}, \text{Blood}) = 2 \cdot P('AB')P('O') = \boxed{2 \times 60\%}$$

$$\frac{P(\text{Scene} \mid \text{Person}, 'AB')}{P(\text{Scene} \mid \text{Each Person}, \text{Blood})} = \frac{1}{2 \times 0.6} = \boxed{0.83}$$

Exercise 3.5: $P(p_a \mid s = aba, F = 3)$

$$P(p_a \mid s, F, H_1) = \frac{p_a^{F_a}(1-p_a)^{F_b}}{P(s \mid F, H_1)} = \frac{p_a^{F_a}(1-p_a)^{F_b}}{\int_0^1 p_a^{F_a}(1-p_a)^{F_b}\,dp_a} = \frac{p_a^{F_a}(1-p_a)^{F_b}}{\frac{\Gamma(F_a+1)\Gamma(F_b+1)}{\Gamma(F_a+F_b+2)}}$$

$$P(p_a \mid s = aba, F=3) = \frac{p_a^2(1-p_a)^1}{\frac{\Gamma(2+1)\Gamma(1+1)}{\Gamma(2+1+2)}} = \frac{5!}{3!\,2!} p_a^2(1-p_a)^1$$



Most probable p_a; $\frac{dP(p_a \mid s = aba, F=3)}{dp_a} = 0$

$\boxed{p_a = 2/3}$

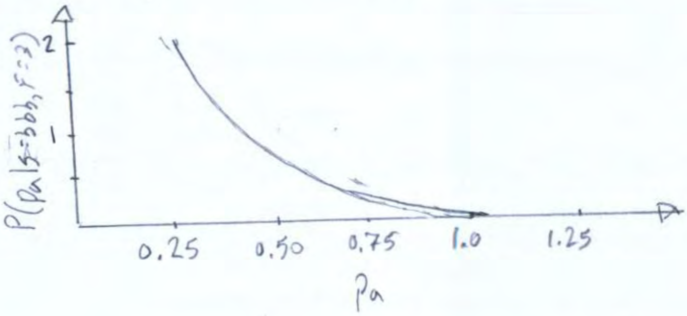
Mean value of p_a under this distribution:

$$E[P(p_a \mid s = aba, F=3)] = \int_0^1 p_a \cdot 10\, p_a^2(1-p_a)\,dp_a = \boxed{0.5}$$

$P(p_a \mid s = bbb, F=3) = \boxed{5 \cdot (1-p_a)^3}$

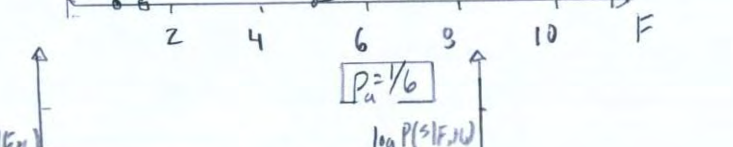
Most probable value: $\boxed{p_a = 1}$

Mean value of p_a: $\boxed{1.25}$

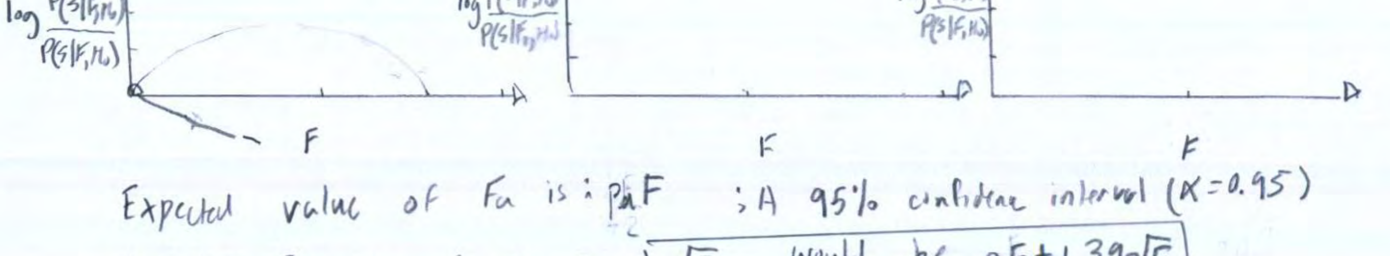


Exercise 3.6. $\log \frac{P(s \mid F, H_0)}{P(s \mid F_0, H_0)}$

$$\log \frac{P(s \mid F, H_0)}{P(s \mid F, H_0)} = \log \frac{P(s \mid F, H_1)\, P(H_1)}{P(s \mid F, H_0) \cdot P(H_0)} = \log \frac{F_a!\,F_b!}{(F_a+F_b+1)!} \Big/ p_0^{F_a}(1-p_0)^{F_b}$$



Exercise 3.7

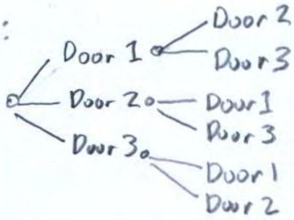


Expected value of F_a is: $p_a F$; A 95% confidence interval ($\alpha = 0.95$)

Standard Deviation of x is: $\boxed{\sqrt{\frac{F}{2}}}$ would be $p_a F \pm 1.39\sqrt{F}$.

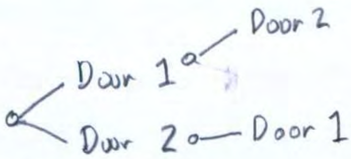
Exercise 3.8 Graphically:

State 1:



P(choice 1) = 1/3 P(choice 2) = 1/2

State 2:



P(choice 1) = 1/2 P(choice 2) = 1

The outlooks of P(choice 1) · P(choice 2) is better through switching doors, i.e. switching to Door #2.

Equation:

State 1: $P(\{1, 2, 3\} \mid H) = \frac{1}{3}$

State 2: $P(H_1 \mid D=3) = \frac{P(D=3 \mid H_1)\, P(H_1)}{P(D=3)} = \frac{(\frac{1}{2})(\frac{1}{3})}{(\frac{1}{2})}$

$P(H_2 \mid D=3) = \frac{P(D=3 \mid H_2)\, P(H_2)}{P(D=3)} = \frac{1 \cdot \frac{1}{3}}{(\frac{1}{2})}$

$P(H_3 \mid D=3) = \frac{P(D=3 \mid H_3)\, P(H_3)}{P(D=3)} = \frac{0 \cdot \frac{1}{3}}{\frac{1}{2}}$

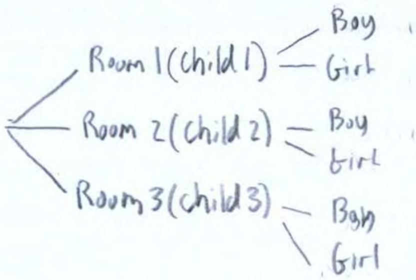
Through switching to door #2, the contestant will have the greatest chance of winning.

A realization occurred that the graphical method does not incorporate a normalizing constant, but arrives to similar answers because of exact multiplicative/divisor.

Exercise 3.9. If the contestant is not choosing, then the outcomes are supposedly equal for switching (or staying) in Door #1.

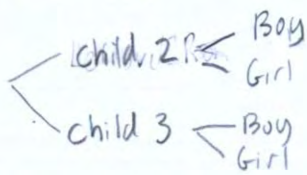
Exercise 3.10. Graphical:

State 1:



P(choice 1) = 1/3 = Girl

State 2:



P(choice 2) = 1/2

The probability of the there being two boys and a girl, or two girls and a boy are equally likely.

Equation

State 1: $P(H_1) = P(H_2) = P(H_3) = Girl = \frac{1}{3}$

State 2:

$P(H_1 \mid C=B) = \frac{P(C=B \mid H_1)\, P(H_1)}{P(C=B)} = \frac{0 \cdot \frac{1}{3}}{\frac{1}{2}}$

$P(H_2 \mid C=B) = \frac{P(C=B \mid H_2)\, P(H_2)}{P(C=B)} = \frac{\frac{1}{2}(\frac{1}{3})}{\frac{1}{2}}$

$P(H_3 \mid C=B) = \frac{P(C=B \mid H_3)\, P(H_3)}{P(C=B)} = \frac{\frac{1}{2}(\frac{1}{3})}{\frac{1}{2}}$

Bayes theorem shows similar outcomes to graphical analysis

Exercise 3.11 $P(\text{murder}|\text{Priors}) = \dfrac{P(\text{priors}|\text{murder}) \cdot P(\text{murder})}{P(\text{priors})} = \dfrac{\frac{1}{1000}(1)}{9} = \boxed{\dfrac{1}{9000}}$

Exercise 3.12 $P(\text{Black}) = P(\text{White}) = \boxed{1/2}$

$P(\text{Black} | \text{Additional White}) = 0$; $P(\text{White} | \text{Additional White}) = \boxed{1}$

Posterior (White) = 1 ; $\text{Posterior} = \dfrac{\text{likelihood} \times \text{prior}}{\text{evidence}}$; $\boxed{1 = \dfrac{P(\text{Additional}|\text{White})P(W)}{P(\text{Additional})}}$

$1 = \dfrac{1 \cdot 1/2}{1} = \boxed{1/2}$.

Exercise 3.13. $\text{Posterior} = \dfrac{\text{likelihood} \times \text{Prior}}{\text{evidence}} = \dfrac{1 \quad \times 10^6}{10^6} = 1$

Exercise 3.14 Sample space = {HH, HT, TH, TT}

Probability of two heads = $\boxed{1/4}$

Exercise 3.15 n(Heads) = 140 ; n(tails) = 110   Eqn 3.22 $\dfrac{P(H_1|s,F)}{P(H_0|s,F)} = \dfrac{P(s|F,H_1)P(H_1)}{P(s|F,H_0)P(H_0)}$

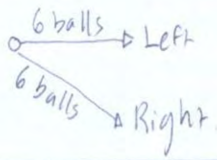
$\dfrac{P(H_1|s,F)}{P(H_0|s,F)} = \dfrac{140!\,110!}{(140+110+1)!} \Big/ (1/2)^{140}(1/2)^{110}$ $= \dfrac{F_a!\,F_b!}{(F_a+F_b+1)!} \Big/ P_0^{F_a}(1-P_0)^{F_b}$

$= 0.4767 \simeq 48\%$

The likelihood of an unbiased coin for the provided evidence is 48% ; suggesting, the null hypothesis ($H_0 = H_1$) does not have sufficient evidence for bias.

Chapter 4:

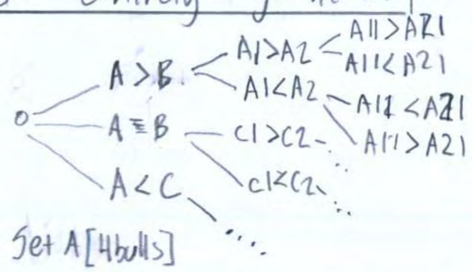
Exercise 4.1. n = 12 ; weight {1...11} ≠ weight {12}



a) Information is measured in states, that describe probability of the system

b) When the ball of different mass is identified, the information is entirely gathered

c)



Set A [4 balls]

Set B [4 balls]

Set C [4 balls]

State 1   State 2   State 3

Best Case = Worst Case = $\boxed{3}$

• Shannon Information $= \sum_i \log \frac{1}{P(x)} = \log(3 \cdot 2^2) = \boxed{\log 12}$

d) i) State of a flipped coin = log 2
ii) State of two flipped coins = $\log 2^2$
iii) outcome of a four sided dice = log 4

e) 6:6 ; $\boxed{\log 2}$
4:4 ; $\boxed{\log 3}$

Exercise 4.2: $H(x,y) = P(x,y)\log\frac{1}{P(x,y)} = P(x)P(Y)\log\frac{1}{P(x)P(Y)} = P(X)\cdot\log\frac{1}{P(x)} + P(Y)\log\frac{1}{P(Y)}$

$\boxed{= H(X) + H(Y)}$

Example 4.3: The number of guesses: $64 \to 32 \to 16 \to 8 \to 4 \to 2 \to 1$; $\boxed{6 \text{ guesses.}}$

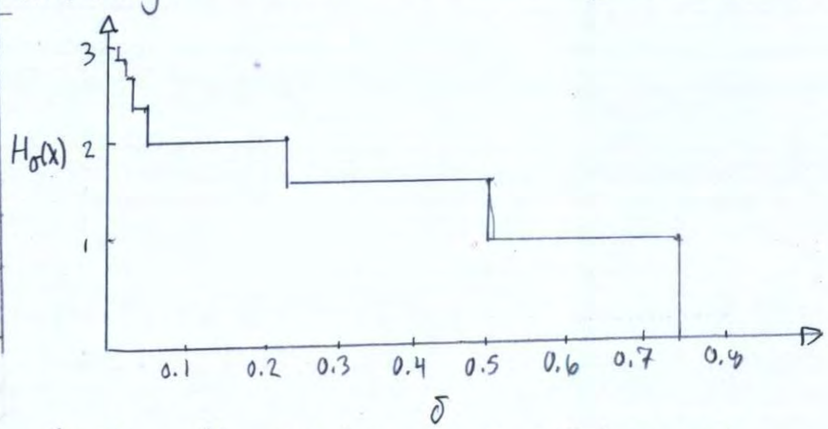
Exercise 4.4: Shannons information provider the number of bits representations of decimal(0 to 255) and ASCII decimal(0 to 127).
Decimal(0 to 255): $\log_2(255) = 7.99$ Decimal(0 to 127): $\log_2(127) = 6.99$
The reduction of physical memory is achieved through removing redundancy and expressing values in a compact fashion.

Exercise 4.5: If the outcomes are greater than 2^l, where l = # of bits; then yes, a compressing algorithm would duplicate the bits during decompression.

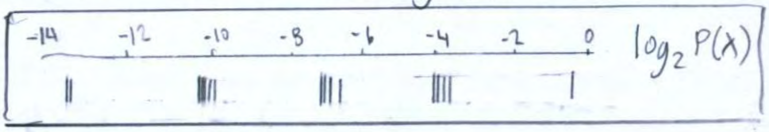
Example 4.6:

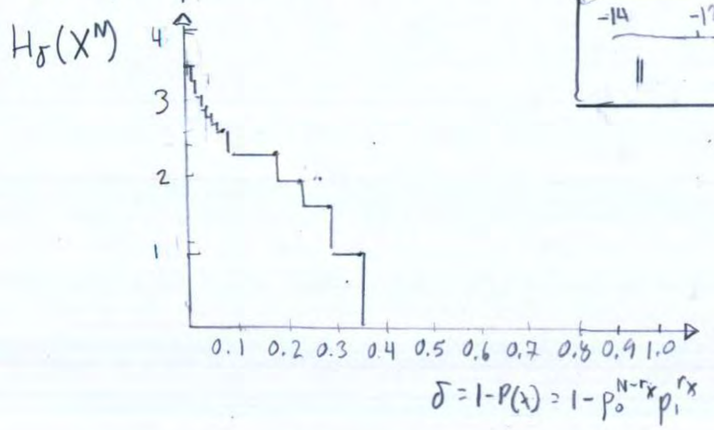
A_x	P_x	Bit code	δ
a	$\frac{1}{4}$	000	00
b	$\frac{1}{4}$	001	01
c	$\frac{1}{4}$	010	10
d	$\frac{3}{16}$	011	11
e	$\frac{1}{64}$	100	—
f	$\frac{1}{64}$	101	—
g	$\frac{1}{64}$	110	—
h	$\frac{1}{64}$	111	—

Bit ensemble: $\boxed{3}$
Lower bit ensemble: $\boxed{2}$
Loss of information (δ): $\boxed{1/16}$
Probability of correct information: $\boxed{93.75\%}$



Example 4.7. $X = (X_1, X_2 \ldots X_n)$ where $X_n \in \{0,1\}$ with probabilities $p_0 = 0.9$, $P_1 = 0.1$
$P(x) = p_0^{N-r_x} p_1^{r_x}$; where r_x is the number of 1's having P_1.





Exercise 4.8: $\boxed{\text{Cusps represent the point where shannons information changes by 1}}$

Exercise 4.9: The second group is correct because weighing six balls does not maximize information; however, are wrong due to their statement, "no, weighing six against six conveys no information at all." $H_4 = \log(3)$ and $H_6 = \log(4)$; information gained less than a bit.

Exercise 4.10: $n = 39$ balls: Raw Information Content $[H_0]$ $= \log_2\left(\frac{1}{1/39}\right) = \boxed{5.29}$ Essential Bit Content $[H_\delta]$ $= \log_2\left(\frac{1}{1/3}\right) = \boxed{1.58}$

Exercise 4.11: A strategy for analyzing the 'two-sided balance problem' is a determination of probability $(P(x))$, then plotting $\delta = 1 - P(x)$ vs $H_\delta = \log_2\left(\frac{1}{P(x)}\right)$. The whole values of $H_\delta(x)$ represent bits of information to investigate further. Information is best minimized with a 16, 8, 4, then 2 sequence.

Exercise 4.12. The minimum number of weights needed is four $(\{10, 5, 3, 2\})$.

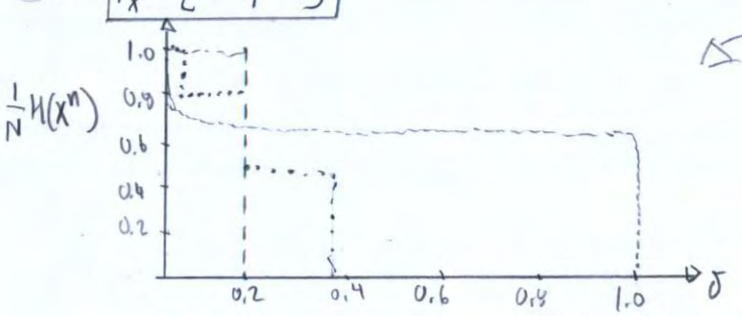
Exercise 4.13 a) Yes, a rotation of sets of four balls generates a compare and contrast Venn Diagram to identify the unique ball.

b) If N balls are weighed, then the labels require a rotation of the pans identification.

Exercise 4.14: a) A worst case for two balls of heavier or lighter mass is six weighings. Three 'odd' balls worst case is also six weighings.

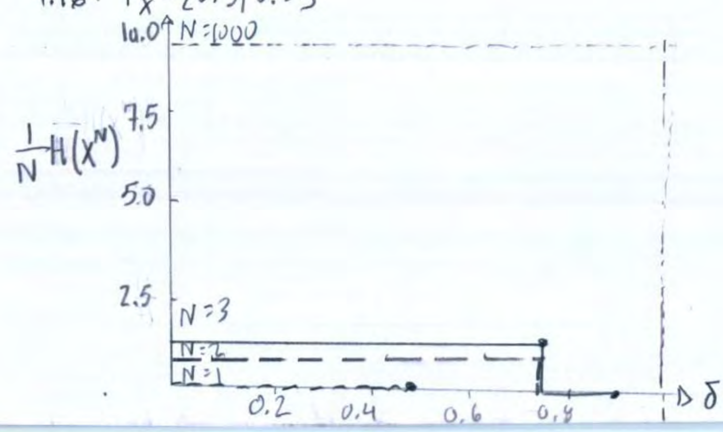
b) The knowledge of ball weights is irregardless of the process to find the odd balls in the set.

Exercise 4.15: $\boxed{P_x = \{0.2, 0.8\}}$



Note: The book rounded and normalized.

Exercise 4.16: $P_x = \{0.5, 0.5\}$



Exercise 4.17. 'Asymptotic Equipartition' principle is similar to Boltzmann Entropy and Gibbs Entropy because each is dependent upon the finite distributions of the system.

Exercise 4.18. $P(x) = \frac{1}{Z}\frac{1}{x^2+1}$ $x \in (-\infty, \infty)$; The normalizing constant Z represents the sum total of the Cauchy partition $\sum_{-\infty}^{\infty} \frac{1}{x^2+1} = \boxed{\pi}$

Mean: $E[x] = \int_{-\infty}^{\infty} x P(x)\,dx = \int_{-\infty}^{\infty} \frac{x}{Z(x^2+1)}\,dx = \frac{1}{Z}\int_{0}^{\infty}\frac{du}{u+1} = \boxed{\text{undefined.}}$

Variance: $E[x^2] = \int_{-\infty}^{\infty} x^2 P(x)\,dx = \int_{-\infty}^{\infty} \frac{x^2}{Z(x^2+1)}\,dx = \boxed{\text{undefined.}}$

$Z = x_1 + x_2$; where x_1, x_2 = independent random variables

$$P(z) = P(x_1, x_2) = P(x_1)\cdot P(x_2) = \frac{1}{Z^2}\int_{-\infty}^{\infty}\frac{dx_1}{x_1^2+1}\int_{-\infty}^{\infty}\frac{dx_2}{x_2^2+1} = \frac{1}{\pi^2}\int_{-\infty}^{\infty}\frac{dx_1}{(x_1^2+1)([z-x_1]^2+1)}$$

$$= \frac{1}{\pi^2}\left[\int\frac{Ax+B}{(x_1^2+1)}dx_1 + \int\frac{Cx_1+D}{([z-x_1]^2+1)}dx_1\right] ; (Ax_1+B)([z-x_1]^2+1) + (Cx_1+D)(x_1^2+1) = 1$$

$$A = \frac{2}{z^3+4z} ; B = \frac{z}{z^3+4z}$$

$$C = -\frac{2x}{z^3+4z} ; D = \frac{3z}{z^3+4z}$$

$$= \frac{1}{\pi^2}\left[\frac{1}{z^3+4z}\left(\int\frac{2x+z}{x^2+1}dx - \int\frac{2x-3z}{([z-x_1]^2+1)}\right)\right] = \frac{z}{\pi}\frac{1}{z^2+4}$$

N-samples from the Cauchy-Distribution of $z = x_1 + x_2$ is similar to a Cauchy-Distribution having similar expectation and variance as $P(x_1)$ or $P(x_2)$.

Exercise 4.19. $P(x \geq a) \leq e^{-sa} g(s)$ and $P(x \leq a) \geq e^{-sa}\cdot g(s)$ where $g(s) = \sum P(x) e^{sx}$

if $t = \exp(sx)$; $x = \frac{1}{s}\log(t)$; $P(x \geq a) = P(\frac{1}{s}\log t \geq a)$; $P(t \leq e^{-sa}) = e^{-sa}\sum P(x)$

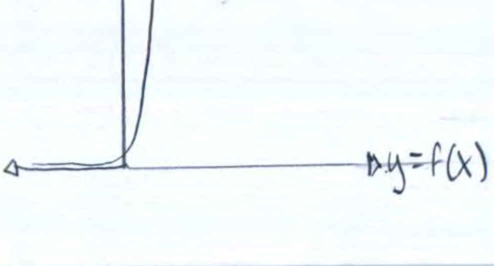
$P(t \leq e^{-sa}) \leq e^{-sa} g(s)$

$P(t \geq e^{-sa}) \geq e^{-sa} g(s)$

Exercise 4.20. $f(x) = x^{x^{x^{x^{x\cdots}}}}$ where $x > 0$

$\text{inverse}(P(x)) = \text{inv}(x^{x^{x^x}})$; $x = f(x)^{n(\frac{1}{x})}$

$[f(x)]^{(\frac{1}{x})} = x$



Chapter 5:

Example 5.1 : $\boxed{\{0,1\}^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}}$

Example 5.2 : $\boxed{\{0,1\}^+ = \{0, 1, 00, 01, 10, 11, 000, 001, \ldots\}}$

Example 5.3 : $A_X = \{a, b, c, d\}$

$P_X = \{1/2, 1/4, 1/8, 1/8\} = \{1000, 0100, 0010, 0001\}$

$c^+(acdbac) = 100000100001010010000001$

Example 5.4: $C_1 = \{0, 101\}$ is a prefix code because 0 is not the prefix of 101 and 101 is not the prefix of 0.

Example 5.5: $C_2 = \{1, 101\}$ is not a prefix code.

Example 5.6: $C_3 = \{0, 10, 110, 111\}$ is a prefix code

Example 5.7: $C_4 = \{00, 01, 10, 11\}$ is a prefix code

Exercise 5.8: C_2 is not uniquely decodable. because $c^+(x) = c^+(y)$

Example 5.9: The exercise 4.1 is capable of being assigned as a ternary code because each binary weighing amounted to three weighings.

Example 5.10: $A_X = \{a, b, c, d\}$

$P_X = \{1/2, 1/4, 1/8, 1/8\}$

$x = (acdbac)$

Entropy $= H(X) = \sum P(x) \log \frac{1}{P(x)} = 1.75$ bits

Length $= L(C, X) = \sum p(x) l(x) = 1.75$ bits

$c^+(x) = 0110111100110$

C_3 is a prefix and uniquely decodable.

Example 5.11: $L(C_4, X) = \sum P(x) l(x) = \boxed{2 \text{ bits}}$

Example 5.12: C_5: $A_X = \{a, b, c, d\}$

$\{0, 1, 00, 11\}$

$L(C_5, X) = \sum P(x) l(x) = \frac{1}{2}\log_2 1 + \frac{1}{4}\log_2 1 + \frac{1}{8}\log_2 2 + \frac{1}{8}\log_2 2$

$= 1.25$ bits

$H(X) = \sum P(x) \log \frac{1}{P(x)} = 1.75$ bits

Although, the sequence is not uniquely decodable.

Example 5.13: C_6:

a_i	$c(a_i)$	p_i	$h(p_i)$	l_i
a	0	1/2	1.0	1
b	01	1/4	2.0	2
c	011	1/8	3.0	3
d	111	1/8	3.0	3

$L(C_6, X) = \sum p_i l_i = 1.75$ bits

$H(X) = 1.75$ bits

- C_6 is not a prefix code because $c(a)^+ \in c(b)^+ \in c(c)^+$

C_6 is uniquely decodable because of the overlap of prefixes.

Exercise 5.14 Kraft Inequality: For any $C(x)$ over a binary alphabet $\{0,1\}$ the codewords must satisfy: $\sum_{i=1}^{I} 2^{-l_i} \leq 1$

where $I = |A_x|$

If codeword $x = a_1 \cdot a_2 \cdot a_3 \cdots a_n$ $\boxed{\text{Equation}}$

$$= s_1 \cdot s_2 \cdot s_3 \cdots s_n$$

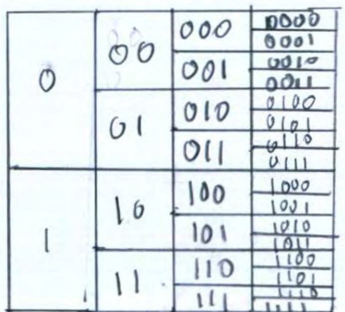
$$= \sum 2^{-l_{i1}} \cdot \sum 2^{-l_{i2}} \cdot \sum 2^{-l_{i3}} \cdots \sum 2^{-l_{iN}}$$

$$= \sum \sum \sum \cdots \sum 2^{-l_{iN}}$$

$$= \sum 2^{-l} A_l \leq \sum 1 \leq N l_{max}$$

$$S^N \leq N l_{max}$$

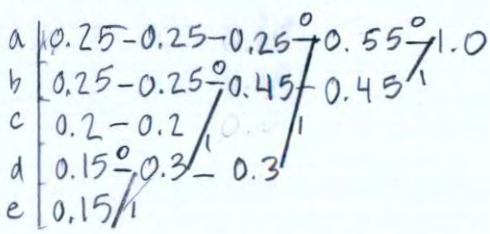
$\boxed{\text{Graphically}}$



$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \sum_{i=1}^{4} 2^{-l} A_l \leq \sum 1 \leq N l_{max}$$

Example 5.15. $A_x = \{a, b, c, d, e\}$

$P_x = \{0.25, 0.25, 0.2, 0.15, 0.15\}$



Huffman Algorithm: ① Two least probable codewords are selected because they have the longest length. ② Combine these two symbols into a single symbol

$\boxed{\text{Huffmans algorithm provides a method to discover the optimal codelength.}}$

Exercise 5.16. $\boxed{\text{The proof Huffman's codeword is the minimum is represented by an ensemble of size = 3: where, } A_x = \{a, b, c\}\text{, and } P_x = \{1/2, 1/4, 1/4\}\text{. A case is } C = \{0, 10, 11\} \text{ having } L = 1.5 \text{ bits, and other examples not following Huffman's algorithm show } L > 1 \text{ bit.}}$

Example 5.17. $\boxed{\text{Huffman's algorithm of Figure 2.1 generated codeword disparities} \sim 1 \text{ bit to achieve a lossless relationship.}}$

Example 5.18: $A_x = \{a, b, c, d, e, f, g\}$

$P_x = \{0.01, 0.24, 0.05, 0.20, 0.47, 0.01, 0.02\}$

$\boxed{\text{The Huffman algorithm produced a bit-length of 1.97}}$

a_i	P_i	Huffman
a	0.01	000000
b	0.24	01
c	0.05	0001
d	0.20	001
e	0.47	1
f	0.01	000001
g	0.02	00001

Exercise 5.19: $C = \{00, 11, 0101, 111, 1010, 100100, 0110\}$ is not uniquely decodable because the second and fourth element are similar.

Exercise 5.20: $C = \{00, 012, 0110, 0112, 100, 201, 212, 22\}$ is uniquely decodable; in that no two indices have similar prefixes.

Exercise 5.21: $A_X = \{0, 1\}$
$P_X = \{0.9, 0.1\}$

Huffman Code		Expected Length	Entropy
X^2	{1, 01, 000, 001}	1.29 bits	0.94 bits
X^3	{1, 01, 000, 001, 1110, 0011, 0111, 1111}	1.22 bits	1.41 bits
X^4	{1, 011, 0101, 001, 000, 000111, 000110, 0001011, 000000, 000011, 000001, 0000001, 0000000, 00000000, 000000001, 000000000}	2.00 bit	2.01 bit

Note: An unusual problem because $H(X^n) < L(C, X^n)$, which contradicts the upper limit of bit assignment being entropy.

Exercise 5.22: $\{P_1, P_2, P_3, P_4\}$; Length $= \sum P(x) \cdot l_i$; $2 = p_1(x) \cdot l_1 + p_2(x) l_2 + p_3(x) l_3 + p_4(x) \cdot l_4$

$= [p_1(x) + p_2(x) + p_3(x) + p_4(x)]\, l$; if $l_1 = l_2 = l_3 = l_4$

$1 = p_1(x) + p_2(x) + p_3(x) + p_4(x)$

$A_{1x} = \{00, 01, 10, 11\}$
$P_{1x} = \{1/2, 1/4, 1/8, 1/8\}$
$A_2 x = \{0, 1, 00, 11\}$
$P_{2x} = \{1/4, 1/4, 3/8, 3/8\}$

Exercise 5.23: $Q = \{\vec{p}_1, \vec{p}_2\} = \{(1/2, 1/4, 3/16, 1/16), (3/4, 1/8, 3/16, 1/16), (7/8, 1/16, 3/64, 1/64)\}$

$$\vec{p}_1 = \vec{\mu}_1 q^{(1)} + \vec{\mu}_2 q^{(2)} + \vec{\mu}_3 q^{(3)} = [\mu_1\ \mu_2\ \mu_3] \begin{bmatrix} q^{(1)} \\ q^{(2)} \\ q^{(3)} \end{bmatrix} = \boxed{[\mu_1\ \mu_2\ \mu_3] \begin{bmatrix} 1/2 & 1/4 & 3/16 & 1/16 \\ 3/4 & 1/8 & 3/16 & 1/16 \\ 7/8 & 1/16 & 3/64 & 1/64 \end{bmatrix}}$$

Exercise 5.24. A simple explanation for winning the game twenty one questions is routine. The sequence of questions best eliminate large categories of information to deduce an answer. An example statement, "Does the object breathe?", would eliminates three of five biological kingdoms of classification. Another question may be, "Is the object inanimate?" The astringent method is to question the largest information categories. A routine for twenty one questions helps produce positive outcomes.

Exercise 5.25: $p = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$; $Length = \sum_i P(x) \cdot l_i = \frac{1}{2} l_1 + \frac{1}{4} l_2 + \frac{1}{8} l_3 + \frac{1}{8} l_4$

$$= 2^{-1} l_1 + 2^{-2} l_2 + 2^{-3} l_3 + 2^{-3} l_4$$

If $l_1 = 1$; $l_2 = 2$; $l_3 = 3$; $l_4 = 3$, then $Length = \boxed{1.75 \text{ bits.}}$

$$Entropy = \sum_i P(x) \log_2 \left(\frac{1}{P(x)}\right) = \frac{1}{2}(1) + \frac{1}{4} \cdot 2 + \frac{1}{8} 3 + \frac{1}{8} 3 = 1.75 \text{ bits}$$

Exercise 5.26: An ensemble described by the Huffman algorithm is of lowest expected length as compared to entropy.

Exercise 5.27. $A_x = \{a, b, c, d, e, f, g, h, i, j, k\}$; $P_x = \{1/11, 1/11, 1/11, 1/11, 1/11, 1/11, 1/11, 1/11, 1/11, 1/11, 1/11\}$

$= \{111, 1011, 0011, 1101, 0101, 1001, 0001, 110, 010, 100, 000\}$

$Length = \sum_i P(x) l_i = 3.55$ bits ; $Entropy = 3.46$ bits ; $Length - Entropy = \boxed{0.09 \text{ bits.}}$

Exercise 5.28: $Length(A_x) = I$

$Probability(A_x) = \frac{1}{I}$ $\quad I \neq 2^i$ where $i \in \mathbb{Z}$ $\quad$ Prove $f^+ = 2 - \frac{2^{l+}}{I}$ where $l+ \equiv \log_2 I$

$$f^+ = P(x) \cdot l(x) = \frac{1}{I}[2I - 1] = 2 - \frac{1}{I} = 2 - \frac{2^{\log_2 I}}{I} = 2 - \frac{2^{l+}}{I}$$

$$L = \sum_0^n P(x) l_i = \log_2 I + 1 - \frac{1}{I} = \log_2 I - 1 + 2 - \frac{2^{l+}}{I}$$

$$= l^+ - 1 + f^+$$

$$\frac{d\Delta L}{dI} = \frac{d}{dI}\left[L - H(x)\right] = \frac{d \log_2 I}{dI} - \frac{d1}{dI} + \frac{d2}{dI} - \frac{d}{dI}\frac{1}{I} - \frac{d}{dI} P(x) \log_2 I$$

$$= \frac{\ln 2}{I} + \frac{1}{I^2} - P(x)\frac{\ln 2}{I} = \frac{\ln 2}{I}(1 - P(x)) + \frac{1}{I^2}$$

Exercise 5.29. $P_x = \{0.99, 0.01\}$; Huffmans Code will efficiently compress a sparse binary source by evaluating the data regions with long codewords, then leaving the rest as shortened codewords. This is efficient because high probability $\times$ low length is a smaller expected length.
The proposed solution requires n codewords at the length of the ensemble.

Exercise 5.30, The strategy to finding the poisoned glass is similar to the "weighing" or "two balance" problem. A 1/3 mixture is conducted against 1/3, then if either group is absent of poison, the remaining 1/3 is poisoned. This routine bubbles down to 3^n glasses, where n is the amount of tests.
An optimal test criteria is $\log_3(\#\text{glasses})$, but is expected to be $\frac{1}{\#\text{glasses}}$.

Exercise 5.31 C_3:

a_i	$c(a_i)$	p_i	$h(p_i)$	l_i
a	0	1/2	1.0	1
b	10	1/4	2.0	2
c	110	1/8	3.0	3
d	111	1/8	3.0	3

$$P(1bit|C_3) = \frac{\sum_i P(C_3|1bit)P(1bit)}{P(C_3)} = \frac{\frac{1}{2}(0)+\frac{1}{4}(1)+\frac{1}{8}(2)+\frac{1}{8}(3)}{\frac{1}{2}(1)+\frac{1}{4}(2)+\frac{1}{8}(3)+\frac{1}{8}(3)}$$

$\boxed{= 1/2}$

Exercise 5.32: The Huffman algorithm generates r mod(q-1) codewords, where r is the ensemble size and q is the number of leaves combined in the tree. An optimal coding algorithm requires r mod(q-1) + 1; such that, erroneous (ensemble) values are inserted to compensate for the sub-par combinations.

Exercise 5.33: Metacode: a construct from several symbol codes that assign different-length codewords to alternative symbols.

The optimal binary codewords require $\sum_i 2^{-l} \leq 1$, so a metacode of K symbol codes does not fit the case $\frac{1}{K}\sum_i 2^{-l} \leq 1$ and is suboptimal

Chapter 6: Stream Codes

Exercise 6.1: $h(x|H) = \log_2\left(\frac{1}{P(x|H)}\right)$; $P(x|H) = \frac{1}{2^{h(x|H)}}$

$$P_{Tot}(x_i|H) = \sum_i P(x_i|h_i) = \sum_0^n \frac{1}{2^{h(x_i|h)}} = \frac{1}{2^0} + \frac{1}{2^1} + \sum_2^n \frac{1}{2^{h(x_i|h)}}$$

$$\boxed{= 1.5 + \sum_2^n \frac{1}{2^{h(x_i|h)}}}$$

Base 10 to is two

Base-2 bits minimum.

Exercise 6.2: Huffman-With-Header:

Header: $p \in \{p_1, p_2, p_3 \dots p_n\}$

$a_i \in \{a_1, a_2, a_3 \dots a_n\}$

$l_i \in \{l_1, l_2, l_3 \dots l_n\}$

$\boxed{\text{Expected Length}: L(C, X) = \sum_{i=1}^{16} p_i \cdot l_i \leq H(x) + 1}$

Arithmetic Code using Laplace Model:

$$P_L(a|x_1 \dots x_{n-1}) = \frac{F_a + 1}{\sum_{a'}(F_{a'} + 1)}$$

Expected length: $L(C, X) = \sum_{i=1}^n P_L(a|x_1 \dots x_n) \cdot \boxed{F \leq H(x) + 1}$

Arithmetic Code using Dirichlet Model:

$$P_D(a|x_1 \dots x_{n-1}) = \frac{F_a + \alpha}{\sum(F_{a'} + \alpha)}$$

Expected Length: $L(C, X) = \sum_{i=1}^n P_D(a|x_1 \dots x_{n-1}) \cdot F \leq H(x) + 1$

Exercise 6.3: $\{P_0, P_1\} = \{0.99, 0.01\}$

a) Random value: $2^{16} - 1$

Rescaled to: 00000000000000001111111111111111

Emitted Value: 11

b) $H_2(p) = H(p, 1-p) = p\log_2 \frac{1}{p} + (1-p)\log_2 \frac{1}{(1-p)}$

$H_2(0.01) = 0.01\log_2\left(\frac{1}{0.01}\right) + 0.99\log_2\left(\frac{1}{0.99}\right) = \boxed{0.081 \text{ bits}}$

1000 bits of Arithmetic coding $1000 \times H_2(0.01) = \boxed{81 \text{ bits}}$

Exercise 6.4: A uniquely decodable compression prefix requires uniqueness, and if not unique, a (pointer, bit) to symbolize (where, why). The (pointer, bit) increases size for strings length for prefixes which are duplicates

Exercise 6.5: Encode $\underbrace{000000000000}_{12\ \text{zeros}}1\underbrace{00000000000}_{11\ \text{zeros}}$

Lempel-Ziv Algorithm:

source substrings	λ	0	00	000	0000	001	00000	000000
s(n)	0	1	2	3	4	5	6	7
s(n) binary	000	001	010	011	100	101	110	111
(pointer, bit)		(,0)	(1,0)	(10,0)	(11,0)	(010,1)	(100,0)	(110,0)

Exercise 6.6: Decode 0010101110110010010001101101010000011

(pointer, bit)		(,0)	(0,1)	(01,0)	(11,1)	(011,0)	(010,0)	(100,0)	(110,1)	(0101,0)	(0001,1)
s(n) binary	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010
s(n)	0	1	2	3	4	5	6	7	8	9	10
Source Substring	λ	0	1	00	001	000	10	0100	101	0000	01

Exercise 6.7. Length[N]; Weight[K]; K 1's; N-K 0's. N = 5, K = 2

An arithmetic coding algorithm for repetive occurences are best described by cumulative probability. For every reoccurring value in the sequence, a probability is determined by assigning a probability to another reoccurrence. If the probability is greater than 50% (0.5), then a 1-bit is assigned, and less, a 0-bit.

In the case: length is 5, the number of 1's is 3, then Laplace or Dirichlet model's are fit. Laplace's model $P(1|x_1 \ldots x_n) = \frac{F_1}{F_1 + F_0 + 1}$ describes a multiplicative probability from the beta distribution. The $P(1, x_1 \ldots x_n) = P(1|1) \cdot P(1|11) \cdot P(1|111) \cdot P(1|1110) \cdot P(1|11100)$

$= \left(\frac{1}{2}\right) \cdot \left(\frac{2}{3}\right) \cdot \left(\frac{3}{4}\right)\left(\frac{3}{5}\right)\left(\frac{3}{6}\right) = \frac{3}{40}$

$\approx 1\ 1\ 1\ 1\ 1$

{11100, 11010, 11001, 101001, 100101, 100011, 010011, 001011, 000111}

Exercise 6.8. A selection of K objects from N describes the binomial coefficient model with a probability of $\binom{N}{K} = \frac{N!}{K!(N-K)!}$.
The number of required bits is $\log_2\binom{n}{k} \approx N H_2(K/N)$ bits.
A selection is made by the probability of occurring objects; a 1-assignment for K/N, and 0-assignment for $(N-K)/N$. The repetive process continues for 1's as $(K-k)/(N-n)$ probabilities and $1-(K-k)/(N-n)$ 0's.

Exercise 6.9: Source $[X] \xrightarrow{F_0} 0$, $f_1 = 0.01 \rightarrow 1$

Find $X = X_1 X_2 X_3$; $P(X \mid X_1X_2X_3) = \frac{P(X_1X_2X_3 \mid f_1, 3) \cdot p(f_1)}{P(X_1X_2X_3 \mid 3)}$

$= \frac{F_A + 1}{F_1 + F_0 + 2} = \frac{1}{5}$

$1 - H_2(0.01) = 0.2 - 0.081$
$= 0.119$

$E(F_A) = \sum_0^{1000} P(F_A)\, l(F_A)$

where $P(F_A) = p_1^{F_A}(1-p_1)^{F-F_A}$

$= 1000 \cdot p_1 = 10$ bits

$Var(F_A) = \sum_0^{1000} P(F_A) \cdot l(F_A)^2$

$= 1000 \cdot p_1(1-p_1) = \frac{99}{10} = \boxed{9.9 \text{ bits}}$

Exercise 6.10: An arithmetic coding algorithm to generate random bit strings of length N with density f is:

```
int u = 0.0; Doub Ri = Σ_1^{i-1} P(Xn=ai | X1...Xn-1)
int v = f; Doub Qi = Σ_1^{i} P(Xn=ai | X1...Xn-1)
int N = 10;
Doub p = v-u;
For(int i=0; i<N; i++) {
    v = u + p · Ri(Xi | X1...Xi-1)
    u = u + p · Qi(Xi | X1...Xi-1)
    p = v-u;
}
```

The algorithm describes the N-length interval in terms of the lower and upper cumulative probabilities. This process is akin to cumulating multiplicative probabilities.

Exercise 6.11. Encode the string 0100001000100010101000001 using the modified Lempel-Ziv algorithm.

Source substrings	λ	0	1	00	001	000	10	001	01	010	0000	01
s(n)	0	1	2	3	4	5	6	7	8	9	10	11
s(n) Binary	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011
(pointer, bit)		(,0)	(0,1)	(01,0)	(11,1)	(001,0)	(010,0)	(100,)	(001,1)	(1000,0)	(0101,0)	(1000,)
Child [0,1]		X|X	X|	X|	|	X|	|	|	X|	|	|	|
New (pointer bit)			(0,)	(01,0	(11,1)	(000,0)	(010,0)	(000,)	(001,1)	(1000,0)	(0101,0)	(1000,)

Exercise 6.12 $\boxed{\text{If stringlength is odd, then the modified Lempel-Ziv algorithm is capable of being a 'complete' algorithm, because each branch of the binary tree has two leafs. Although, an even length string is 'incomplete'; due to the fact, branches are left without both children of similar prefix.}}$

Exercise 6.13: $\boxed{\text{A string of repetitive values has low entropy (say a binary string of zeros), but would not compress well by Lempel-Ziv's algorithm because of the redundancy.}}$

Exercise 6.14: $P(x) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(\frac{-\sum x_n^2}{2\sigma^2}\right)$; $r = (\sum x_n^2)^{1/2}$

Estimate mean and variance of r^2.

Note: $\int \frac{1}{(2\pi\sigma^2)^{N/2}} x^4 \exp\left(\frac{-x^2}{2\sigma^2}\right) dx = 3\sigma^4$.

$$\boxed{E[r^2]} = \int_0^\infty \frac{r^2}{(2\pi\sigma^2)^{1/2}} \exp\left(\frac{-r}{2\sigma^2}\right) dr = \frac{\sqrt{2\sigma^2\pi}}{(2\pi\sigma^2)^{N/2}} \left(\frac{1}{2\left(\frac{1}{2\sigma^2}\right)}\right)^{1/2} = \frac{\sigma^2}{(2\pi\sigma^2)^{(N-1)/2}} = \sigma^2$$

$$\boxed{Var[x^2]} = E[x^2] - E[x]^2 = \int_0^\infty \frac{r^4}{(2\pi\sigma^2)^{1/2}} \exp\left(\frac{-r}{2\sigma^2}\right) dr - (\sigma^2)^2 = 3\sigma^4 - \sigma^4 = 2\sigma^4$$

$$\boxed{Shell}: r^2 = \sigma^2 ; r = \sigma ; P(x_{shell}) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2}\right)$$

$\boxed{\text{Probability Density}}$ $P(x=0) = \frac{1}{(2\pi\sigma^2)^{1/2}}$

$\boxed{\text{Probability Shell per Probability Density}}$: $P(shell)/P(x=0) = \exp\left(\frac{-1}{2}\right)$

@ $N = 1000$; $P(Shell)/P(x=0) = \exp\left(-\frac{1000}{2}\right)$

Exercise 6.15: $A = \{a, b, c, d, e, f, g, h, i, j\}$

$P = \{\frac{1}{100}, \frac{2}{100}, \frac{4}{100}, \frac{5}{100}, \frac{6}{100}, \frac{8}{100}, \frac{9}{100}, \frac{10}{100}, \frac{25}{100}, \frac{30}{100}\}$

Optimal Binary Coding constructs a given set of symbol probabilities to a code which matches Shannon Information content.

Using Huffman Coding: {11111, 11110, 1110, 0111, 0110, 110, 0101, 0100, 10, 00}

Expected Length $= \sum^{10} P(x) \cdot l(x) = \boxed{2.64 \text{ bits}}$

Exercise 6.16: $y = x_1 x_2$; $X : A_x = \{a, b, c\}$; $P_x = \{\frac{1}{10}, \frac{3}{10}, \frac{6}{10}\}$;

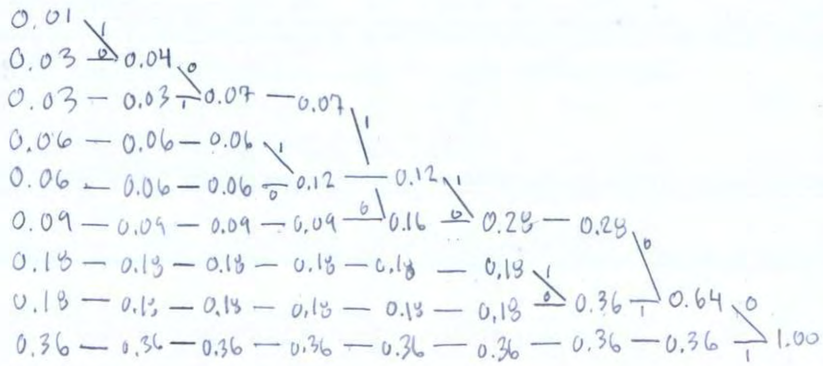
$P(y) = P(x_1) \cdot P(x_2)$

x_2 \ x_1	0.1	0.3	0.6
0.1	0.01	0.03	0.06
0.3	0.03	0.09	0.18
0.6	0.06	0.18	0.36

$H(y) = \sum_{i=1}^{9} P(y) \log_2 \frac{1}{P(y)}$

$\boxed{= 2.59 \text{ bits.}}$

Optimal Binary Code:



code	length
000101	6
000100	6
00011	5
00001	5
0010	4
0000	4
011	3
010	3
1	1

Expected Length $= \sum P(x)\cdot l(x) = \boxed{2.73 \text{ bits}}$

Exercise 6.17 $P=\{0.1, 0.9\}$; $E[X]=Np$; $Var[X]=Np(1-p)$; $SD[X]=\sqrt{\frac{9}{100}N}$

$= \frac{N}{10}$ $\quad = \frac{9N}{100}$ $\quad = \frac{3\sqrt{N}}{10}$

If $N=1000$ → $\boxed{E[X]=100 \;;\; Var[X]=90 \;;\; SD[X]=3\sqrt{10}}$

Exercise 6.18 $L(p)=\sum p_n l_n$; $H(p)=\sum p_n \log_2 \frac{1}{p_n}$; Show average information rate per second is

$p_n = \frac{1}{Z}2^{-\beta l_n}$; where $Z=\sum 2^{-\beta l_n}$; $\beta = \frac{H(p)}{L(p)}$

Average Information Per second $= \dfrac{\text{Information per symbol}}{\text{Average Duration per symbol}}$

$$= \frac{I(p)}{L(p)} = \frac{\sum \log_2 \frac{1}{p_n}}{\sum p_n l_n} = \frac{\sum \log Z\, 2^{\beta l_n}}{\sum \frac{1}{Z}2^{-\beta l_n}\cdot l}$$

Maximal Information per second $$= \frac{\sum \frac{1}{Z}2^{-\beta l_n}\cdot \sum \frac{1}{Z}2^{-\beta l_n} l - \sum \log Z\, 2^{\beta n} \sum \frac{1}{Z}2^{-\beta l_n} l}{\left(\sum \frac{1}{Z}2^{-\beta l_n}\cdot l\right)^2}$$

$$= \frac{I(p)}{L(p)} - I(p) = \boxed{I(p)\left(\frac{1}{L(p)} - 1\right)}$$

Exercise 6.19: $L(p) = \sum_{1}^{52} P(x)\,\ell = \frac{1}{52}\sum_{1}^{52} \ell = \frac{52(52+1)}{2\cdot 52} = \boxed{27 \text{ bits}}$

Exercise 6.20: 13 cards from 52 card deck. Bids: 1♣, 1♢, 1♡, 1♠, 1NT, 2♣, 2♢... 7♡, 7♠, 7NT

a) If $\binom{52}{13}$ describes the number of combinations, then $\log_2\binom{52}{13}$ is the amount of bits to describe a hand.

b) Shannon Information: $I(p) = \log_2\left(\frac{1}{p}\right)$; $P = \frac{\binom{52}{4}\binom{52}{13}}{\binom{52}{13}} = \frac{\text{Prob suit} \cdot \text{Prob Number}}{\text{Total Prob.}}$

$$= \sum_{i=0}^{n} \log_2\left(\frac{1}{p_n}\right); \; P = \frac{\binom{52-2n}{4}\binom{52-2n}{13-n}}{\binom{52-2n}{13-n}}$$

$$= \sum_{i=0}^{13} \log_2\left(\frac{1}{p_n}\right); \; P = \binom{52-2n}{4} \Big/ \binom{52}{13}$$

$\boxed{= 322 \text{ bits}}$

Exercise 6.21: a)

	Two Buttons	Three Buttons
Arabic	0□ → 9□	00□ - 99□
Roman :	MI□, XD, CD, ID	MX□, MC□, MI□, MM□, XX□, XC□, XI□, CC□, CI□, II□

A complete code satisfies the Kraft Inequality.

$S^N = \left[\sum_i 2^{-\ell}\right]^N \le N \cdot l_{max}$; where N: length of string. ℓ = Length of sequence.

Yes, the 'arabic' and 'roman', two (or three) button sequences are complete.

b) The sample space of the 'arabic' and 'roman' microwave are not 100% similar. A demonstration of a four button sequence for the 'arabic' microwave shows (999□) 9 min 99sec is not possible for the 'roman' microwave; While, the 'roman' by definition acheives larger numbers, including (MMM□) 30 min.

c) An implicit probability distribution over timer to which each of the codes is matched.

d) The implicit probability distribution for which the microwaves are best matched is lower cooking times, more specifically, less than $\boxed{10 \text{ min}}$.

e) $E[X] = \sum_{0}^{10} P(x)\cdot \ell(x) = \frac{10}{10}(3) = 3$ symbols

Maximum Number of symbols is 3 symbols for a 'plausible' sequence less than 10 min.

f) A more efficient cooking-time-encoding system would be greater base-counting system, per say, base-16.

Exercise 6.22:
- $C_b(5) = 101$ is not uniquely decodable because of the lack of terminating characters.
- An option for mapping $n \in \{1, 2, 3 ..\}$ to $C(n) \in \{0,1\}^+$ that is uniquely decodable would be to end (or begin) each representation by a binary flag e.g. 00000, or 11111
- Alternative codes for integers are purposed for large file systems, and describe Base-16, Base-32 or Base-64

Chapter 7: Codes for Integers:

Exercise 7.1: a)

n	$C_u(n)$
1	1
⋮	⋮
256	$\boxed{0000 \ldots -1}$

length - 256

n	l_b	$C_\alpha(n)$
1	⋮	1
⋮	⋮	⋮
256	8	$\boxed{\text{Answer}}$

$$C_\alpha(n) = C_U[l_b(n)]C_B$$
$$= (256) \cdot 8 \cdot 1$$
$$\boxed{= 2048 \text{ bits}}$$

b) 1Kb = 1024 bytes; # bits = $\log_2(100\text{Kb}) = \log_2(102400) = \boxed{16.64 \text{ bits}}$

Exercise 7.2: $C_\alpha(n)$ is a unary code for $l_b(n)$, and is 1000011
$C_\beta(n)$ is a headless binary representation of n, and is 1101100.
$C_\beta(n)$ is a headless binary representation of n, and is 1100001111010111
$C_\gamma(n)$ is a sequence of codes involving C_β, l_b, and C_B, shown as 97 pairs of bits.
$C_3(n)$ is an end-of-file symbol, and for n characters is 35 groups of three bits.
$C_7(n)$ is another end-of-file symbol, and is 25 sets of four bits.
$C_{15}(n)$ is described as 25 sets of four bits at 100 total bits.
The shortest end-of-file notation is $C_{15}(n)$ at 100 total bits.

Exercise 7.3: A C_{2048} code requires 11-bits to end the file and another 11-bits to describe a value of 2047 characters, this would be less than Elias's code for 2047 characters.

Chapter 8: Dependent Random Variables

Exercise 8.1: $X \equiv (U, V)$, $Y \equiv (V, W)$; $H(X,Y) = \sum_{xy \in A_x A_y} P(x,y) \log \frac{1}{P(x,y)} = P(X) \cdot \log \frac{1}{P(X)} + P(y) \log \frac{1}{P(y)}$

$$= P(U,V) \cdot \log \frac{1}{P(U)P(V)} + P(U,W) \cdot \log \frac{1}{P(V)P(W)}$$

$$= H(U) + H(V) + H(V) + H(W) = H(V) + 2H(V) + H(W)$$

$$H(X|Y) = \sum P(X,Y) \log \frac{1}{P(X|Y)} = \sum P(x,y) \log \frac{P(X,y)}{P(y)}$$

$$= H(v) + 2H(u) + H(w)$$

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

$$= H(U) + \cancel{H(V)} - \cancel{H(V)} + 2H(u) + H(W)$$

$$\boxed{= 3H(u) + H(W)}$$

Exercise 8.2. $H(X|y=b_k) = \sum P(x|y=b_k) \log \frac{1}{P(x|y=b_k)} = H(X) + H(Y|X) - H(Y)$

If $P(y|x) = P(y)$, then $H(x|y) = H(x) + \sum P(y,x) \log \frac{1}{P(y,x)} - \sum P(y) \log \frac{1}{P(y)}$

$= H(x) + \sum P(y) P(y|x) \log \frac{P(y)}{P(y|x)} - \sum P(y) \log \frac{1}{P(y)}$

$= H(x) + 0 + \sum P(y) \log P(y) = \boxed{H(x)}$

Exercise 8.3. $H(x,Y) = \sum P(x,y) \log \frac{1}{P(x,y)} = \sum P(x)P(y|x) \log \frac{1}{P(x)P(y|x)} = \sum P(x)P(y|x) \log \frac{1}{P(x)} + \sum P(x)P(y|x) \log \frac{1}{P(y|x)}$

$= H(x) + \sum_x P(x) \sum_y P(y|x) \log \frac{1}{P(y|x)}$

$\boxed{= H(X) + H(Y|X)}$

Exercise 8.4. $I(X;Y) \equiv H(X) - H(X|Y)$; $I(X;Y) = D_{KL}(P(X,y) || P(x)P(Y))$

$= \sum P(X,y) \log \frac{P(X,y)}{P(X)P(Y)}$

$= \sum P(y,x) \log \frac{P(X,y)}{P(x)P(y)} = \boxed{I(Y;X)}$

$I(X;Y) = \sum P(X,y) \log \frac{P(X,y)}{P(x)P(Y)} = \sum P(X)P(X|Y) \log \frac{P(X|Y)}{P(X)}$

$= \sum_x P(x) \sum_y P(X|Y) \log P(X|Y) + \sum P(x)P(X|Y) \log \frac{1}{P(x)}$

$= -H(X|Y) + H(X) = \boxed{H(X) - H(X|Y)}$

Exercise 8.5. $D_H(X,y) = H(X,Y) - I(X;Y)$

Axiom 1: $D_H(x,y) = \sum P(x,y) \log \frac{1}{P(x,y)} - \sum P(x,y) \log \frac{P(X,y)}{P(X)P(Y)}$

$= \sum P(x,y) \log \frac{P(x)P(y)}{P(x,y)^2} = \boxed{\sum P(x,y) \cdot \log \frac{1}{P(x|y)P(Y|X)}}$

$= \emptyset$

Axiom 2: $D_H(X,X) = \sum P(X,X) \log \frac{1}{P(X,X)} - \sum P(X,X) \log \frac{P(X,X)}{P(X)P(X)}$

$= \sum P(X,X) \log \frac{1}{P(X|X)P(X|X)} = \boxed{0}$

Axiom 3: $D_H(X,Y) = \sum P(x,y) \log \frac{1}{P(x|y)P(Y|X)} = \sum P(y,X) \log \frac{1}{P(y|x)P(X|Y)}$

$\boxed{= D_H(Y,X)}$

$$H(X|Y) = \sum_y H(X|y) = 7/4 + 7/4 + 2 + 0 = \boxed{11/8}$$

$$H(Y|X) = H(X,y) - H(X) = 3.375 - 1.75 = \boxed{13/8}$$

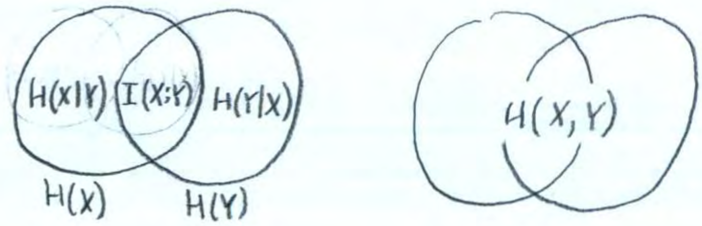
$$I(X;Y) = H(X) - H(X|Y) = 1.75 - 1.375 = \boxed{3/8}$$

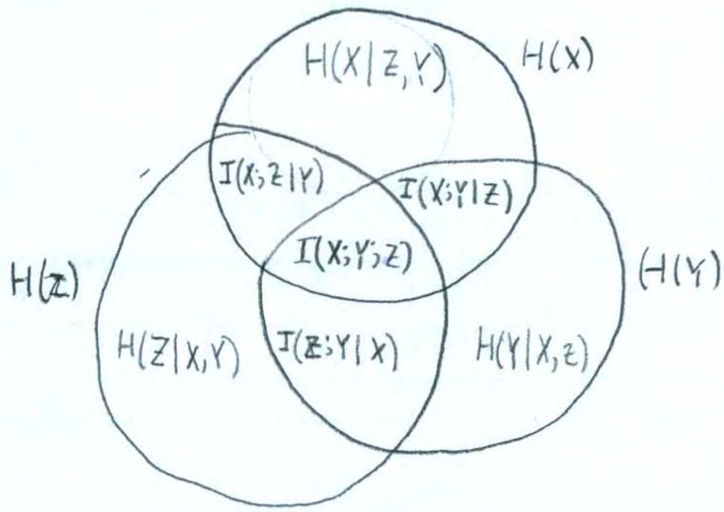
Exercise 8.7. $A_X = A_Y = A_Z = \{0,1\}$; $P_X = \{p, 1-p\}$; $P_Y = \{q, 1-q\}$; $Z = (X+Y) \bmod 2$

a) If $q = 1/2$, $P_Z = \{1/2, 1/2\}$; $I(Z;X) = H(Z) - H(Z|X) = 1 - 1 = \boxed{0}$

b) $P_Z = \{pq + (1-p)(1-q), p(1-q) + q(1-p)\}$; $I(Z;X) = \boxed{H(Z) - H(Z|X)}$

Exercise 8.8.





The triple Venn Diagram is misleading because $I(X;Y)$ is not listed, nor is $I(X;Z)$ and $I(Y;X)$.

$pq + 1 - p - q + pq$

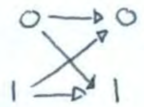
Exercise 8.9: W = state of the world ; d = Data gathered ; r = processed data

$W \to d \to r$; $P(w,d,r) = P(w) \cdot P(d|w) P(r|d)$

$$I(W;R) = \sum_{x,y} P(w,r) \log \frac{1}{P(w|r)} = \sum P(w,r) \log \frac{P(w,r)}{P(w)P(r)}$$

Chapter 9: Communication over a Noisy Channel:

Example 9.1: $f = 0.15$; $P_X: \{p_0 = 0.9, p_1 = 0.1\}$; $y = 1$; Binary Symmetric Channel:



Observation of 1: $P(x=1|y=1) = \dfrac{P(y=1|x=1)P(x=1)}{P(y=1|x=0)P(x=0) + P(y=1|x=1)P(x=1)}$

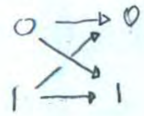
$\boxed{= 17/44}$

$P(x=0|y=1) = \dfrac{P(y=1|x=0)P(x=0)}{P(y=1|x=0)P(x=0) + P(y=1|x=1)P(x=1)}$

$\boxed{= 27/44.}$

The input $x=0$ is more probable than $x=1$.

Exercise 9.2: $f = 0.15$; $P_X = \{p_0 = 0.9, p_1 = 0.1\}$; $y = 0$; Binary Symmetric channel:



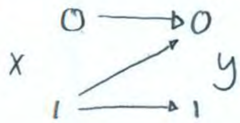
Observation of 0; $P(x=1|y=0) = \dfrac{P(y=0|x=1)P(x=1)}{P(y=0|x=0)P(x=0) + P(y=0|x=1)P(x=1)}$

$\boxed{= 1/52}$

Example 9.3: $f = 0.15$; $P_X = \{p_0 = 0.9, p_1 = 0.1\}$;

Z-channel:

$A_X = \{0,1\}$; $A_Y = \{0,1\}$



Observation of $y=1$; $P(x=1|y=1) = \dfrac{P(y=1|x=1)P(x=1)}{P(y=1|x=1)P(x=1) + P(y=1|x=0)P(x=0)}$

$\boxed{= 1.0}$

Exercise 9.4: $f = 0.15$; $P_X = \{p_0 = 0.9, p_1 = 0.1\}$

Z-channel:

$A_X = \{0,1\}$; $A_Y = \{0,1\}$

Observation of $y=0$; $P(x=1|y=0) = \dfrac{P(y=0|x=1)P(x=1)}{P(y=0|x=1)P(x=1) + P(y=0|x=0)P(x=0)}$

$\boxed{= 1/61.}$

Example 9.5: $f = 0.15$; $P_X: \{p_0 = 0.9, p_1 = 0.1\}$; $P(y=0) = 0.78$, $P(y=1) = 0.22$

$I(X;Y) = H(Y) - H(Y|X) = P(y)\log\frac{1}{P(y)} - P(Y|X)\log\frac{1}{P(Y|X)}$

$= 0.22\log_2\left(\frac{1}{0.22}\right) + (1-0.22)\log_2\left(\frac{1}{1-0.22}\right) - 0.15\log_2\left(\frac{1}{0.15}\right) - (1-0.15)\log_2\left(\frac{1}{1-0.15}\right)$

$= 0.76 - 0.61 \boxed{= 0.15 \text{ bits}}$

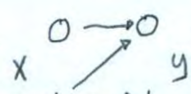
Example 9.6: $P(y=1) = 0.085$; $I(X;Y) = H(Y) - H(Y|X)$

$= 0.085\log_2\left(\frac{1}{0.085}\right) + (1-0.085)\log_2\left(\frac{1}{1-0.085}\right) - [0.90 H_2(0) + 0.1 H_2(0.15)]$

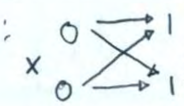
$= 0.42 - (0.1 \times 0.61) \boxed{= 0.36 \text{ bits.}}$

Exercise 9.7: $I(X,Y)$; Binary Symmetric Channel:  ; $P_X = \{p_0 = 0.5, p_1 = 0.5\}$; $f = 0.15$

$$I(X,Y) = H(X) - H(X|Y) = H_2(0.5) - H_2(0.15)$$
$$= 0.5\log\frac{1}{0.5} + (1-0.5)\log\frac{1}{(1-0.5)} - \left[0.15\log\frac{1}{0.15} + (1-0.15)\log\frac{1}{(1-0.15)}\right]$$
$$= 1.00 - 0.61 = \boxed{0.39 \text{ bits}}$$

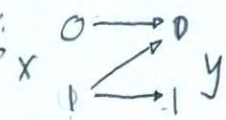
Exercise 9.8 $I(X;Y)$; Z channel:  ; $P_X = \{p_0 = 0.5, p_1 = 0.5\}$; $f = 0.15$

$$I(X;Y) = H(X) - H(X|Y) = H_2(0.5) - H_2(0.15)$$
$$= \log\left(\frac{1}{0.5}\right) - \left[0.15\log\frac{1}{0.15} + (1-0.15)\log\frac{1}{(1-0.15)}\right]$$
$$= 1.00 - 0.61 = \boxed{0.39 \text{ bits}}$$

Exercise 9.9 Binary Symmetric Channel:  ; $f = 0.15$; $P_X = \{p_0 = 0.9, p_1 = 0.1\}$

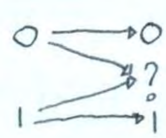
$$I(X;Y) = 0.15 \text{ bits};\ C(Q_{BSC}) = H_2(0.5) - H_2(0.15) = \boxed{0.39 \text{ bits.}}$$
$$\frac{dI(X;Y)}{dp_1} = \frac{d}{dp_1}\left[H_2((1-f)p_1 + (1-p_1)f) - H_2(f)\right] = \boxed{0.39 \text{ bits}}$$

Example 9.10: $C(Q_{NTW}) = p_1\log_2\frac{1}{p_1} + p_2\log_2\frac{1}{p_2} + p_3\log_2\frac{1}{p_3} = 3\left(\frac{1}{3}\log_2\frac{1}{(1/3)}\right) = \boxed{\log_2 9.}$

Example 9.11: Z Channel:  ; $f = 0.15$; $I(X;Y) = H(Y) - H(Y|X)$
$$= H_2(p_1(1-f)) - (p_0 H_2(0) + p_1 H_2(f))$$
$$= H_2(p_1(1-f)) - p_1 H_2(f) = \boxed{0.445}$$

Exercise 9.12: $C(Q_{BSC}) = H_2(0.5) - H_2(f)$
$$\boxed{= 1 - H_2 f}$$

Exercise 9.13: $f = 0.15$; $C_{BEC} = H_2(f) - fH_2(f) = 1 - f.$
$$C_{BEC,General} = \boxed{H_2(p_1) - f \cdot H_2(p)}$$

Exercise 9.14: Q, $N=2$; $f = 0.15$; Binary Erasure Channel: 

$P(y=0|x=0) = 1-f$ $\quad P(y=0|x=1) = 0$

$P(y=?|x=0) = f$ $\quad P(y=?|x=1) = f$

$P(y=1|x=0) = 0$ $\quad P(y=1|x=1) = 1-f$

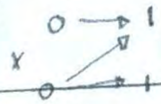
> Q is an extended channel of discrete memoryless channel; moreover, a graphical depiction of input/output for large N-bit counts.

$|A_X|^2 = |00, ?0, 10, 0?, ??, 1?, 01, ?1, 11|$

$|A_Y| = |00, 01, 10, 11|$

Q	00	01	10	11
00	0.72	0	0	0
?0	0.13	0	0.13	0
10	0	0	0.72	0
0?	0.13	0.13	0	0
??	0.02	0.02	0.02	0.02
1?	0	0	0.13	0.13
01	0	0.72	0	0
?1	0	0.13	0	0.13
11	0	0	0	0.72

Columns 00 and 11 show the least overlap, which, aids the analysis for cross-signals being of two categories. The decoder that would fit the extended channel requires $\{00, ?0, 0?\}$ to be considered 0, and $\{1?, ?1, 11\}$ to be assigned 1.

Exercise 9.15: Z-channel:  $; f = 0.15; C(Q_Z) = 0.685.$

a) p_1^* is less than 0.5 because of maximization of information that was evaluated to be 0.445.

b) $I(X;Y) = H(Y) - H(Y|X) = H_2(p_1(1-f)) - p_1 H_2(f)$.

$$\frac{dI(X;Y)}{dp_1} = (1-f)\left[\log_2(1-p_1(1-f)) - \log_2 p_1(1-f)\right] - H_2(f)$$

$$= (1-f)\log_2 \frac{1-p_1(1-f)}{p_1(1-f)} - H_2(f) = 0$$

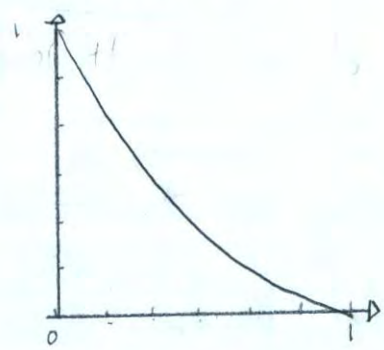
$$\frac{1-p_1(1-f)}{p_1(1-f)} = 2^{H_2(f)/(1-f)}$$

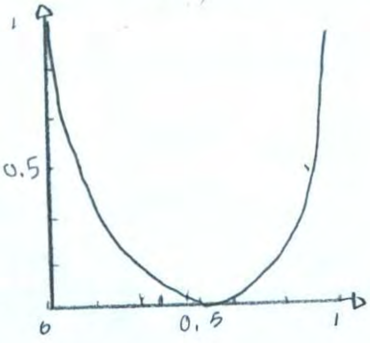
$$1-p_1(1-f) = p_1(1-f)\cdot 2^{H_2(f)/(1-f)}$$

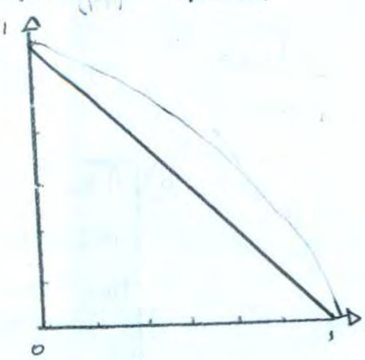
$$\boxed{p_1 = \frac{1/(1-f)}{1+2^{H_2(f)/(1-f)}}}$$

c) $\lim_{f \to 1} p_1^* = \lim_{f \to 1} \frac{1/(1-f)}{1+2^{H_2(f)/(1-f)}} = \lim_{f \to 1} \frac{-1/(1-f)^2}{\log \frac{1}{F}\, 2^{H_2(f)/(1-f)}/(1-f)^2} = \boxed{\frac{1}{e}}$

Exercise 9.16







$I(X;Y)' = (1-f)\log_2 \frac{1-p_1(1-f)}{p_1(1-f)} - H_2(f)$

$\boxed{\text{Z Channel}}$

$I(X;Y)' = 1 - p\log_2\frac{1}{p} - (1-p)\log_2\frac{1}{(1-p)}$

$\boxed{\text{Binary Symmetric Channel}}$

$I(X;Y)' = 1 - p$

$\boxed{\text{Binary Erasure Channel}}$

Exercise 9.17 The capacity of a five-input, ten-output channel described below is the maximum number of bits transmitted. Columns 1 & 3 provide 2 bits of error-free transmission.

$$\begin{array}{c} \\ 0\\1\\2\\3\\4\\5\\6\\7\\8\\9 \end{array}
\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\
\begin{bmatrix}
0.25 & 0 & 0 & 0 & 0.25 \\
0.25 & 0 & 0 & 0 & 0.25 \\
0.25 & 0.25 & 0 & 0 & 0 \\
0.25 & 0.25 & 0 & 0 & 0 \\
0 & 0.25 & 0.25 & 0 & 0 \\
0 & 0.25 & 0.25 & 0 & 0 \\
0 & 0 & 0.25 & 0.25 & 0 \\
0 & 0 & 0.25 & 0.25 & 0 \\
0 & 0 & 0 & 0.25 & 0.25 \\
0 & 0 & 0 & 0.25 & 0.25
\end{bmatrix} \end{array}
\qquad
\begin{array}{c} 0 \\ \begin{bmatrix} 0.25\\0.25\\0.25\\0.25\\0\\0\\0\\0\\0\\0 \end{bmatrix} \end{array}
+
\begin{array}{c} 3 \\ \begin{bmatrix} 0\\0\\0\\0\\0\\0\\0.25\\0.25\\0.25\\0.25 \end{bmatrix} \end{array}
= 2 \text{ bits of Information}$$

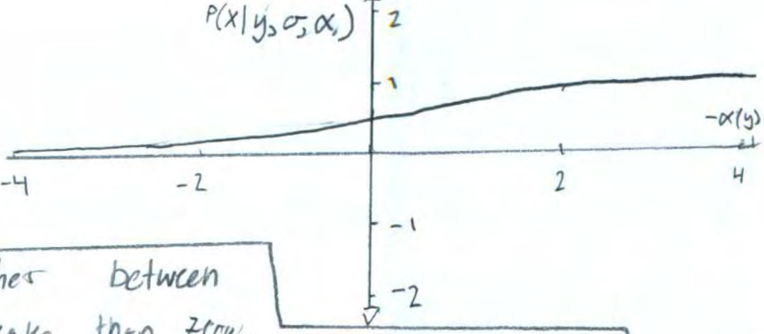
Exercise 9.18 $x \in \{-1,+1\}$; A_y = Output ; $Q(y|x,\alpha,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-x\alpha)^2}{2\sigma^2}}$; α = signal amplitude

a) $\ln Q(x|y,\alpha,\sigma) = \ln \frac{Q(y|x=1,\alpha,\sigma)}{Q(y|x=-1,\alpha,\sigma)} = \frac{(y-\alpha)^2}{(y+\alpha)^2} = \alpha(y)$

Rules of page 36 ; $q = 1-p$; $a = \ln p/q$

$p = q e^{-a} = (1-p)e^{-a}$; $p(1+e^{-a}) = 1$; $p = \frac{1}{1+e^{-a}}$;

$Q(y|x=1,\alpha,\sigma) = \boxed{\frac{1}{1+e^{-a}}}$



b) The optimal decoder distinguishes between inputs. If the input is greater than zero, then an output is assigned to 1, and when the output is less than zero, the output becomes zero.
If the answer is expressed as a function in terms of the signal-to-noise ratio α^2/σ^2, the error function

$$\Phi(z) \equiv \int_{-\infty}^{-\alpha x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2} = \frac{1 \pm \operatorname{erf}\left(\frac{\alpha x}{\sqrt{2}\sigma}\right)}{2}$$

Exercise 9.19. a) Total Possible Twos = (4x5-sized Twos + 8x10-sized Twos) x (Black or white)

= [(4-length x-direction) x (5-length y-direction) + (8-length x-direction) x (10 length y-direction)] x 2

$$= [13 \times 12 + 9 \times 7] \times 2 = \boxed{438}$$

$$P(y|x=2) = \frac{P(x=2|y)\,P(y)}{P(x=2)} = \frac{219 \times 2}{2^{32}} = \boxed{1.02 \times 10^{-7}}$$

$$H(y|x=2) = \sum_i P(y|x=2) \log \frac{1}{P(y|x=2)} = \boxed{2.37 \times 10^{-6}}$$

Exercise 9.20 $n = 24$; $P(\text{Same Birthdays}|n=24) = \boxed{\frac{365P24}{365^{24}}}$

$$P(\text{Distinct Birthdays}|n=24) = 1 - P(\text{same Birthdays}|n=24)$$

$$= 1 - 0.46$$

$$\boxed{= 0.54}$$

$$E[P(\text{Same Birthday}|n=24)] = \sum_{i=1}^{24} \frac{x}{365} = \frac{24(24-1)}{2} \cdot \frac{1}{365} = \boxed{\frac{276}{365}}$$

Exercise 9.21: The probability for error of twenty-four people $A_x = \{1, 2 \ldots 24\}$ being conveyed to 365 mappings $A_x = \{1, 2 \ldots 365\}$ is similar to the birthday problem. Error would occur when the same mapping happens to be assigned i.e. same birthday for twenty four people in 365 days time. The error is $\frac{365P24}{365^{24}} = 0.44$

The capacity of the channel is I_{max}, such that the most information is transmitted. I_{max} is $H_2(1/365) - H(24/365)$; or 0.28 bits. Rate of communication is defined by bits/sec, including 0.28 bits/sec.

Exercise 9.22: K rooms; Q people

a) $P(\text{Same Room} \mid Q \text{ people}) = \frac{KPQ}{K^Q} = \boxed{\text{Error}}$

b) $P(\text{Order of Same Room} \mid Q \text{ people}) = \boxed{\frac{KCQ}{K^Q}}$

If $q = 364$, and $K = 1$, then Error-a =

Chapter 10: The Noisy-Channel Coding Theorem:

Exercise 10.1: The inequality $I(s;\hat{s}) \geq NR(1 - H_2(p_b))$ withstands for complex correlations among bit errors because of the capacity-C for error-free communication. A complex correlation lowers the maximum capacity, since samples are not independent, and in-turn, lowers information.

Exercise 10.2: $\frac{dI(X;Y)}{dp_i} = \frac{d}{dp_i}\left[NR(1 - H_2(Q_{j|i}))\right] = \boxed{NR(1 - Q_{j|i})\frac{dQ_{ji}}{dp_i}}$

Exercise 10.3: $\frac{d^2I(X;Y)}{d^2p_i} = \frac{d^2}{dp_i^2}\left[NR(1 - H_2(p))\right] = \boxed{NR}$

Exercise 10.4: Ternary Confusion Channel: $A_x = \{0, ?, 1\}$, $A_y = \{0, 1\}$

$0 \to 0$, $? \to 0$, $? \to 1$, $1 \to 1$

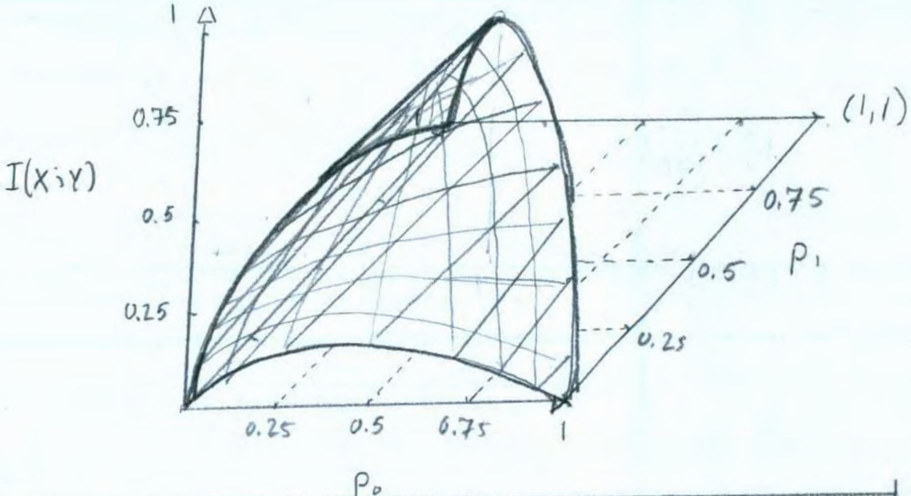
$P(y=0|x=0) = 1$ $P(y=0|x=?) = 1/2$ $P(y=0|x=1) = 0$

$P(y=1|x=0) = 0$ $P(y=1|x=?) = 1/2$ $P(y=1|x=1) = 1$

$$I(X;Y) = H(Y) - H(Y|X) = H_2(p_0 \cdot 1 + p_? \cdot 1/2 + p_1 \cdot 0) - p_?$$

where $\vec{p} = (p_0, p_?, p_1) = (p_0, (1-p_0-p_1), p_1)$

$$\boxed{\text{Plot of } I(X;Y) = \left[(p_0 + (1-p_0-p_1)/2)\log_2 \frac{1}{p_0+(1-p_0-p_1)/2} + (1-p_0-(1-p_0-p_1)/2)\log_2 \frac{1}{1-p_0-(1-p_0-p_1)/2}\right] - (1-p_0-p_1)}$$



Exercise 10.5: $\frac{dI(X;Y)}{dp_i} = \lambda$ for all i; when $I(X;Y)$ is maximized, then the value λ is greater than zero. If $p_i = 0$, then $\frac{dI(X;Y)}{dp_i} \le \lambda$, and λ becomes the upper-limit.

```
template <typename F>
Void dInformation ( F information, int iter, Double step){
    Double capacity[iter], max;
    for(int i=0; i<iter; i++) {
        capacity[i] = [information(i+step)-information(i)]/step;
        if(capacity[i] > max)
            max = capacity[i];
    }
    }
    Std::cout << "The maximum capacity is: " << max;
}
```

Exercise 10.6: $(1-Q(y|x)) = \prod_i (1-Q(y|x_i)) = 0$ for $Q(y|x_i) \stackrel{!}{=} \boxed{0}$

Exercise 10.7: $\frac{d^2 I(X;Y)}{d^2 p_i} = \frac{d^2}{d^2 p_i}\left[NR(1-H_2(Q(y|x)))\right] \stackrel{!}{=} \boxed{NR\,\frac{dQ(y|x)}{dp}}$

Exercise 10.8: $P(y) = \sum_X P(X)\,Q(y|X) = \sum Q(y) \stackrel{!}{=} \boxed{1}$

Example 10.9: A channel: $P(y=0|x=0)=0.7$; $P(y=0|x=1) \stackrel{!}{=} \boxed{0.1;}$

$\boxed{P(y=?|x=0)=0.2;\ P(y=?|x=1)=0.2;}$ ← Removing this channel forms a symmetric channel.

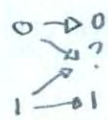
$P(y=1|x=0)=0.1$; $P(y=1|x=1) \stackrel{!}{=} 0.7$

Exercise 10.10: $I(X;Y) = H(Y) - H(Y|X) = H(Y) - h(p)$; $C = \max_{P(X)} I(X;Y) \stackrel{!}{=} \boxed{1-h(p)}$

Exercise 10.11: Yes, nonsymmetric channels exist having uniform input distributions. A proof for the model is channels have $I(J-1)$ degrees of freedom; while, optimal input distributions $(I-1)$, so a nonsymmetric channel having an optimal input distribution does exist with $I(J-1)-(I-1) = I(J-1)+1$ degrees of freedom.

Exercise 10.12:

$$Q = \begin{bmatrix} 1-q & 0 \\ q & q \\ 0 & 1-q \end{bmatrix}$$

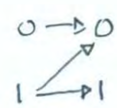


$I(X;Y) = H(Y) - H(Y|X) = \sum p(y)\log\frac{1}{p(y)} - \sum p(Y|X)\log\frac{1}{P(Y|X)}$

$= p_y \log\frac{1}{p_y} + (1-p_y)\log\frac{1}{(1-p_y)} - q\,p_{Y|X}\log\frac{1}{p_{Y|X}} - q(1-p_{Y|X})\log\frac{1}{1-p_{Y|X}}$

$\boxed{= H_2(p_y) - qH_2(p_y)}$

$$Q = \begin{bmatrix} 1 & q \\ 0 & 1-q \end{bmatrix}$$



$Q_{(2,1)} = Q_{(1,1)} \cdot Q_{(1,1)} = \begin{bmatrix} 1-q & 0 \\ q & q \\ 0 & 1-q \end{bmatrix}$

$I(X;Y) = H(Y) - H(Y|X) = H_2(p_1(1-f)) - p_1 H_2(f)$

$C_z = \frac{dI(X;Y)}{dp_1} \stackrel{!}{=} \boxed{(1-f)\log_2\frac{1-p_1(1-f)}{p_1(1-f)} - H_2(f) \geq \frac{1}{2}(1-q)}$

The inequality is presented because of the independence of the (2,1) code to represent an erasure channel.

Exercise 10.13: $N=20$: The argument for minimum number of trips is for indistinguishable wires is a single trip of two steps.

$N=1{,}000$: does not change the amount of trips:

The information content of the greedy approach is $\log[P(r|N,g_r)] = \log\left[\frac{N!}{\prod(r!)^{g_r} g_r!}\right]$, and when maximized using a Lagrange Multiplier $\log\left[\frac{N!}{\prod(r!)^{g_r} g_r!} + \lambda \sum g_r r\right]$,

$$g_r = \frac{e^{\lambda r}}{r!}$$

Chapter 11: Error-Correcting Codes & Real Channels

1. $I(X;Y) = \sum P(x)P(y|x)\log\frac{P(y|x)}{P(y)}$; Lagrangian Constraint: λx^2; $-\mu$.

$$\frac{dI(X;Y)}{dx} = \frac{d}{dx}\left[\int P(x)P(y|x)\ln\frac{P(y|x)}{P(y)}dx - \lambda\int x^2 P(x)\,dx - \mu\int P(x)\,dx\right]$$

$$= \int P(y|x)\ln\frac{P(y|x)}{P(y)}dx - \lambda x^2 - \frac{d\mu}{dx} = 0$$

$$\int P(y|x)\ln\frac{P(y|x)}{P(y)}dx = \lambda x^2 + \mu'$$

$$\int P(y|x)\ln P(y|x)\,dx - \int P(y|x)\ln P(y) = \lambda x^2 + \mu'$$

For this term to become λx^2 or μ', then a Gaussian is required.

Exercise 11.2. $I(X;Y) = H(Y) - H(Y|X) = \frac{1}{2}\log_2(2\pi e(P+N_0W)) - \frac{1}{2}\log_2(2\pi e(N_0W))$

$$= \frac{1}{2}\log_2\left(1+\frac{P}{N_0W}\right) = \boxed{\frac{1}{2}\log\left(1+\frac{v}{\sigma^2}\right)}.$$

The engineering perspective denotes a Power/Noise relationship: While, the book seems to skip steps.

A logarithm table shows:

$$\int x^m \ln x\,dx = x^{m+1}\left(\frac{\ln x}{m+1} - \frac{1}{(m+1)^2}\right);\ m \neq -1$$

This integral seems not work because of the coefficients, but may help somebody else.

Exercise 11.3: $P(\text{Communication}) = P(C')\cdot P(C)\cdot P(Q)\cdot P(D)\cdot P(D') = \boxed{P(C')\cdot P(Q')\cdot P(D')}$

Exercise 11.4: $N = 100$; $b = 0.2$; $f = 0.5$; Capacity of a Binary Symmetric Channel $[C] = 1 - H_2(p)$

Capacity of a Binary Symmetric Channel with error and dependent bursts. $[C] = 1 - \left(\frac{H_2(p) + Nb}{N}\right)$

$$= 1 - \left(\frac{0.2\log\frac{1}{0.2} + (1-0.2)\log\frac{1}{(1-0.2)} + 100\cdot 0.2}{100}\right)$$

$$= 1 - 0.207 = \boxed{0.793}$$

Capacity of a Binary Symmetric Channel with independent bursts $[C] = 1 - H_2(f\cdot b) = 1 - 0.469$ $\boxed{= 0.531}$

The interleaving $[C = 0.531]$ has lower capacity than the dependent bursts $[C = 0.793]$. This means a compact disc that interleaves information to compensate for temporary bursts; in turn, lowers capacity.

Exercise 11.5: Signal-to-noise v/σ^2

a) Exercise 11.2 described $\boxed{C = \frac{1}{2}\log\left(1 + \frac{v}{\sigma^2}\right)}$

b) $X \in \{\pm\sqrt{v}\}$ is constraint; $\boxed{C' = \frac{\sqrt{v}/\sigma^2}{1 + \left(\frac{\sqrt{v}}{\sigma}\right)^2}}$

c) $y \to y' = \begin{cases} 1 & y > 0 \\ 0 & y \le 0 \end{cases}$; $C'' = \frac{1}{1 + \left(\frac{\sqrt{v}}{\sigma}\right)^2} - \frac{2\left(\sqrt{v}/\sigma\right)^2}{\left(1 + (\sqrt{v}/\sigma)^2\right)^2} = \frac{\left(1 + (\sqrt{v}/\sigma)^2\right) - 2\left(\sqrt{v}/\sigma\right)^2}{\left(1 + (\sqrt{v}/\sigma)^2\right)^2}$

$$\boxed{= \frac{1 - \left(\sqrt{v}/\sigma\right)^2}{\left(1 + (\sqrt{v}/\sigma)^2\right)^2}}$$

d)

Exercise 11.6: N = Code Length; R = Rate; K = Bits where $R = K/N$

A maximum (worst case) fraction of errors has probability ($p = \frac{N-K}{N} = 1-R$); irrespective of the type of binary error-correcting algorithm.

Exercise 11.7: Binary Erasure Channel: $A_x = \{0,1\}$, $A_y = \{0, ?, 1\}$

(Spielman, 1996) Encoding: A check bit is assigned to a $\oplus$(XOR) gate with the neighboring message bits.

Decoding: Given the value of a check bit and all but one of the message bits on which it depends, set the missing message bit to be the XOR of the check bit and its known message bits.

Probability of Error: $Pr(\mathcal{E}_s) \leq \binom{K}{s}\binom{\beta K}{as/2}\left(\frac{as}{2\beta K}\right)^{as}$

K-Ways of choosing S-nodes

βK-ways of choosing $as/2$-nodes i.e a T-subset.

The as neighbors of vertices in S is $\left(\frac{as}{2\beta K}\right)^{as}$; βK = check bits
a = average node degree.

This encoding algorithm pairs successive bits with a XOR-Gate.

Exercise 11.8: See Exercise 11.7, but probability of error is adjusted to

$$Pr\begin{cases} 1-f & y = x \\ ? & y \neq x \end{cases}$$

Exercise 11.9: The website listed at "Jetstor" prescribed 10 lessons about RAID. Levels found in RAID technology are 0, 1, 2, 3, 4, 5, 6, 10, 50, and 0+1. A minimum systems implements two drives; while, higher-levels describe seperation of HAMming codes, blocks of discs, parity to blocks, storage systems, layers, mirrors, and stripping fault-tolerance.

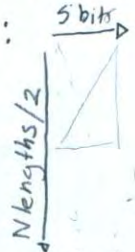
Exercise 11.10: A RAID system of three disks is best implemented using a (7,4) Hamming Code because the code cross-references bits. If three discs were lost then a digital fountain code would enable recovery.

Codes:

Exercise 12.1: Worst case:

$\boxed{N \times S \text{ comparisons.}}$

Average Case:



$\boxed{\text{Unsure - not books answer}}$

Exercise 12.2. Baysian Probability: $P(H_0 \mid \text{Likelihood}) = \dfrac{P(\text{Likelihood} \mid H_0) \cdot P(H_0)}{P(H_1)}$

$$= (2)^r \cdot \frac{1}{(1/2)^r} = \boxed{2^{2r}}$$

where r = bits evaluated correct.

Exercise 12.3: $P(\text{No Collision} \| M, S) = \boxed{\dfrac{{}_SP_m}{S^M}}$

$$P(\text{collision} \| M, S) = 1 - \frac{{}_SP_m}{S^M} = \boxed{1 - \frac{S \cdot (S-1) \cdots (S-M+1)}{S^M}}$$

$M = 1$

$$P(\text{collision} \| M, S) = 1 - \binom{S}{M} = 1 - \frac{S \cdot (S-1) \cdots (S-M+1)}{S^M}$$

$$= 1 - \frac{S \cdot (S-1) \cdots (S - \frac{S}{100} + 1)}{S^{S/100}}$$

$$= 1 - \exp\left(\frac{-\cancel{S}(S-1)/2}{\cancel{S}/100}\right) = \boxed{1 - \exp(-50(S-1))}$$

Example 12.4: $139 + 1254 + 238 = 1631$; $1631 \% 9 \neq 0$: $\boxed{\text{Test Passed}}$

Exercise 12.5: A correct casting-out-nines match gives in favor of the hypothesis that the addition has been done correctly by a probability of 9:1.

Example 12.6: $P(A \neq B \mid H_F) = \left(\frac{1}{2}\right)^M$

Exercise 12.7: The scale of parity checking is different than error correction. A parity check has error detection within the 32 bits of extra information. Error correction requires evaluation of all the bits and tends to be more rigorous than a local parity check.

Example 12.8: M = Bits ; Bet #1 = N_1 ; Bet #2 = N_2 ; $P(\text{Not Distinct} \mid N_1, N_2) = \frac{N_1 N_2}{2^m}$

$$N_1 N_2 = 2^{M+1}$$

If $N_1 = N_2$; then $N_{12} = \sqrt{2^{M+1}} = \boxed{2^{\frac{M}{2}+\frac{1}{2}}}$

Exercise 12.9: Population $= 9\times10^9$ people ;

Common Address: House #, Street #, City, State, Zip Code.

$$\approx \left(\tfrac{1}{9}\right)^X \cdot \left(\tfrac{1}{9}\right)^y \cdot \left(\tfrac{1}{27}\right)^Z \cdot \left(\tfrac{1}{27}\right)^B \cdot \left(\tfrac{1}{9}\right)^C ; \; X+y+Z+B+C = N.$$

$$\approx \left(\tfrac{1}{9}\right)^N$$

$$\frac{9\times10^9 \text{ people}}{(1/9)^N} = 1 ; \quad \boxed{N = \log_{10}(9\times10^9 - 1/9)}$$

Exercise 12.10: If a person writes 10^9 words within their lifetime, then in an alphabet of [a-z], [0-9], the probability a string matches is $\frac{10^9}{37^n} = 1$; $n = 9/\log_{10} 37. \approx 6$

Exercise 12.11: a) 3×10^9 nucleotides from a four letter alphabet {A, C, G, T}

$$\frac{N}{4^L} = 1 \; ; \; L = \log_2(N)/\log_2(4) \; ; \; L > \log(3\times10^9)/\log_2 4 = \boxed{15.7 \text{ AA.}}$$

b) GCCCCC CACCCCT GCCCCC

A repeated subsequence of DNA effects binding, but not the information content.

Chapter 13: Binary Codes:

Example 13.1: The (7,4) Hamming Code has distance (d=3) because the codewords differ in at least 3 bits.

Example 13.2: Weight enumerator functions A(W) relate the number of codewords that have weight W.

Example 13.3: Generator Matrix G[KxN] ; # 1's per row: d_o

The value d_o is a constant, so the sequence of codes are very bad!

Exercise 13.4: $$p = \frac{1}{K}\sum_{k=2}^{K}\binom{N}{k} f^k (1-f)^{N-k} = \frac{d}{N}\binom{N}{2} f^2 (1-f)^{N-2} + \cdots + \frac{1}{K}\binom{N}{K} f^K (1-f)^{N-K}$$

Exercise 13.5: A 'bad' distance code has d/N tend to zero as blocklength N increases; which, is defined by $d = N-K$. The Hamming code when concatenated is an example of a 'bad' code because the ratio d/N is 'bad' distance, but described as 'very good' to order.

Exercise 13.6: $P(\text{block error}) \leq \sum_{w>0} A(w)[\beta(f)]^{w}$

$$\log_2 \langle A(w)\rangle \simeq NH_2(w/N) - M$$

$$\simeq N[H_2(w/N) - (1-R)] \text{ for any } w>0$$

$$P(\text{block error}) \leq \sum_{w>0} \langle A(w)\rangle [\beta(f)]^{w} = \sum_{w>0} e^{N[H_2(w/N)-(1-R)]} \cdot [\beta(f)]^{w}$$

$$\log P(\text{block error}) \leq \sum_{w>0} N[H_2(w/N) - (1-R)] + \prod_{w>0} \log[\beta(f)^{w}]$$

$$\frac{\log P(\text{block error}) - \prod_{w>0} \log[\beta(f)^{w}] - \sum N[H_2(w/N)]}{w} + 1 \boxed{\leq R_{ub}}$$

Exercise 13.7: [1110100] & [0111010] & [1011001] are parallel to themselves, and not orthoganol.

Example 13.8: $G = [1\,1\,1]$; $H = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$; $G^{\perp} = H = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$; $\boxed{G^{\perp} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}}$

Exercise 13.9: $G_{(1,3)} = [0\,0\,0]$; $H = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$; $G^{\perp} = H = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$; $G^{\perp} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$

$h_1 = [1\,1\,1]$; $h_2 = [0\,0\,0]$;

Exercise 13.10: A 'good' distance code has d/N tend to a constant greater than zero. Low-density parity-check codes have d_0/N approach zero because of the parity-check. Also, low-density parity-check codes are good; such that, $H_2(f) = 1-R > 0$.

Exercise 13.11: P; $G = [I_K | P^T] = H$; $\boxed{[I_K | P^T] = [P | I_m]}$

Exercise 13.12: (8,4) Self-Dual Code

$$G = [I_4 | P^T] = \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{array}\right]$$

(7,4) Hamming Code

$$H = [P | I_4] = \left[\begin{array}{cccc|cccc} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

The (8,4) Self-Dual Code is similar to (7,4) Hamming Codes by an inversion of the bits; where, the parity-check represents the parity-check bits of the sixteen codewords of the (7,4) Hamming Code.

Exercise 13.13: $\boxed{[n,n,1]\text{: Whole space } F_q^n;\ [n,n-1,2]\text{: Parity Code; } [n,1,n]\text{: Repetition Code in addition to, Reed-Solomons } F_q,\ n \leq q \text{ are maximum distance Seperable codes.}}$

Exercise 13.14: t = Codeword; (N,K) Code C; y = recieved signal; Gaussian Channel

$P(y|t)$ = Assumed channel Mode

$\boxed{\text{The codeword Decoding Problem}}$: a task of inferring which codeword t was transmitted given the recieved signal.

$\boxed{\text{The bitwise Decoding Problem:}}$ the task of inferring for each transmitted bit t_n how likely it is that bit was a one rather than a zero.

Prove the optimal bitwise-decoder is closely related to the probability of error of the Optimal code-word decoder:

P_b = Optimal bitwise decoder

$= \sum P(u) P(n_B|u,N)$

P_B = Block Error Probability of the Maximum likelihood decoder

$= P(u)$

$$\boxed{\text{Then, } P_B > P_b > \text{constant} \times P_B \quad \text{where constant} \propto \text{average distance error}}$$

$$= \frac{d_{min}}{N}.$$

Exercise 13.15: $(15,11)$ Hamming Code has a minimum distance $d = N-K = 15-11 = \boxed{4}$

$(31,26)$ Hamming Code has a minimum distance $\boxed{d=5}$.

Exercise 13.16: $A(w)$ = Average Weight Enumerator Function

$R = 1/3$ of a Random Linear Code

$N = 540$

$M = 360$.

If $w=1$, $\langle A(w)\rangle = \binom{N}{w} 2^{-M} = \binom{5}{1} 2^{-360} = \boxed{\frac{5}{2^{360}}}$

Exercise 13.17:

$$G = \begin{bmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 & \cdot & 1 & 1 & \cdot & 1 \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 & \cdot & 1 & 1 & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & 1 & 1 & \cdot & 1 & \cdot & 1 & 1 \\ \cdot & \cdot & \cdot & 1 & \cdot & 1 & 1 & \cdot & \cdot & 1 & 1 & \cdot & 1 & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 & \cdot & \cdot & 1 & 1 & \cdot & 1 & \cdot \end{bmatrix}$$

The minimum distance is $\min(N-K, G(:,15)) = 10$.

Also, the weight enumerator function of G is

$$\boxed{\langle A(w) \rangle = \binom{15}{w} 2^{-10}}$$

Exercise 13.18:

$$H_1 = \left[\begin{array}{ccccc|ccccc|ccccc} 1 & \cdot & \cdot & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & 0 & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & 1 & 1 & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & 0 & 1 & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & 0 & \cdot & 1 & 1 \end{array}\right]$$

The minimum distance for $\vec{H}$ is $\min(N-K, G(:,15)) = 5$.

Steps to Reduce Matrix:

1. $R_2 \leftarrow R_2 - R_1$
2. $R_3 \leftarrow R_3 - R_2$
3. $R_4 \leftarrow R_4 - R_3$
4. $R_5 \leftarrow R_5 - R_4$
5. $R_5 \leftarrow \frac{1}{2} R_5$
6. $R_7 \leftrightarrow R_8$
7. $R_8 \leftrightarrow R_1$
8. $R_9 \leftrightarrow R_1$
9. $R_5 \leftarrow R_5 - R_1/2$
10. $R_5 \leftarrow R_5 + R_9/2$
11. $R_4 \leftarrow R_4 - R_9$
12. $R_5 \leftarrow R_5 - R_8/2$
13. $R_4 \leftarrow R_4 + R_8$
14. $R_3 \leftarrow R_3 - R_8$
15. $R_5 \leftarrow R_5 + R_7/2$
16. $R_4 \leftarrow R_4 - R_7$
17. $R_3 \leftarrow R_3 + R_7$
18. $R_2 \leftarrow R_2 - R_7$
19. $R_5 \leftarrow R_5 - R_6/2$
20. $R_4 \leftarrow R_4 + R_6$
21. $R_3 \leftarrow R_3 + R_6$
22. $R_2 \leftarrow R_2 + R_6$
23. $R_1 \leftarrow R_1 - R_6$
24. $R_4 \leftarrow R_4 + R_5$
25. $R_3 \leftarrow R_3 - R_5$
26. $R_2 \leftarrow R_2 + R_5$
27. $R_1 \leftarrow R_1 - R_5$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

The reduction of the matrix has ten independent rows.

$\langle A(w) \rangle = \binom{15}{10} \cdot 2^{-m}$

Exercise 13.19: $R_c = \prod_{c=1}^{c} \frac{N_c - M_c}{N_c}$; $R_{11} = \prod_{c=1}^{11} \frac{12-M_c}{12} = \left(\frac{12-1}{12}\right)\left(\frac{12-2}{12}\right) \cdots \left(\frac{12-11}{12}\right)$ $\boxed{= 0.0537}$

$$R_{10} = \prod_{c=1}^{10} \frac{12-M_c}{12} = \left(\frac{12-1}{12}\right)\left(\frac{12-2}{12}\right) \cdots \left(\frac{12-10}{12}\right)$$

$$\vdots$$

$$R_1 = \prod_{c=1}^{1} \frac{12-M_c}{12} = \left(\frac{12-1}{12}\right)$$

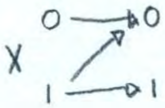
$$R_c = [R_1, \ldots, R_{10}, R_{11}]$$

$\boxed{R_{c2} = \prod_{c=1}^{c} \left(\frac{N_c - M_c}{N_c}\right)^2 \text{ would converge at a higher asymptotic rate.}}$

Exercise 13.20: The number of typical noise vectors is $2^{NH_2(f)}$. Roughly, an equivalent amount of distinct syndromes z exist, because $Hn = z$.

For a given f, the largest possible of rate $R = 1 - \frac{M}{N}$.

Exercise 13.21: Z-channel:



$P(y=0 \mid x=0) = 1$ $\quad$ $P(y=0 \mid x=1) = f$

$P(y=1 \mid x=0) = 0$ $\quad$ $P(y=1 \mid x=1) = 1-f$

Loss in Communication Rate $= \frac{\text{Maximum Rate}}{\text{Capacity of the channel}} \boxed{= \prod_{c=1}^{c} \frac{N_c - M_c}{N_c}}$

Exercise 13.22: 'Bad' Code: a code family that cannot achieve arbitrarily small probability of error, or that can achieve arbitrarily small probability of error only by decreasing the information rate to zero.

'Very' Bad Distance Code: if d tends to a constant as N increases

If 'very' bad distance code, d is a constant, and the probability of block error $P(\text{block error}) \cong \binom{d}{d/2} f^{d/2} \cdot (1-f)^{d/2}$ cannot achieve small values, implying a 'bad' code.

Exercise 13.23: Puncturing: Prior distance: $N - K = M$

Post distance: $N' - K = M'$; $M' < M$.

Shortening: Prior Distance: $N-K=M$

Post Distance: $N-K \neq M'$; $M = M'$

Intersection: Prior Distance: $N-K=M$

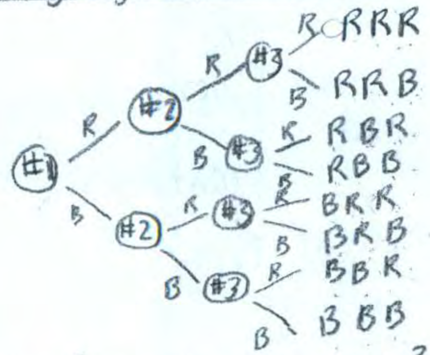
Post Distance: $N'-K'=M'$; $M \neq M' \,||\, M=M'$

> If a 'bad' distance code has d/N tending to zero, then Puncturing and Intersections may shift the code to a 'good' distance, but not always because of the codeword definition. Intersecting a codeword may change a 'very' good code to a 'bad' code.

Exercise 13.24: Three Players: Player #1 [50% Red, 50% Blue]

Player #2 [50% Red, 50% Blue]

Player #3 [50% Red, 50% Blue]



$RR = \frac{2}{8}$; $BR = \frac{6}{8}$; $BB = \frac{2}{8}$

RR: $\frac{1}{8}$ R, $\frac{3}{8}$ B; BR: $\frac{1}{2}$ R, $\frac{1}{2}$ B; BB: $\frac{3}{8}$ R, $\frac{1}{8}$ B

> • The best odds are to guess the opposite color when hats of similar color occur, but if a player notices other colors, then a pass should occur.

Seven Players: Player #1-7 [50% Red, 50% Blue]

> ◦ The best odds are when a player notices hats of similar color, and should guess the opposite. Otherwise, pass.

RRRRRRR

$\left.\begin{matrix} \text{RRRRRRB} \\ \text{RRRRRBR} \\ \vdots \\ \text{BRRRRRR} \end{matrix}\right\} 7$

$\left.\begin{matrix} \text{BRRRRRB} \\ \vdots \\ \text{BBRRRRR} \end{matrix}\right\} 6$

$\left.\begin{matrix} \text{BBRRRRB} \\ \vdots \\ \text{BBBRRRR} \end{matrix}\right\} 5$

$\left.\begin{matrix} \text{BBBRRRB} \\ \vdots \\ \text{BBBBRRR} \end{matrix}\right\} 4$

$\left.\begin{matrix} \text{BBBBRRB} \\ \vdots \\ \text{BBBBBRR} \end{matrix}\right\} 3$

$\left.\begin{matrix} \text{BBBBBRB} \\ \text{BBBBBBR} \end{matrix}\right\} 2$

BBBBBBB 1

Exercise 13.25:

$$H_{EG,\ell}^{orth} = \begin{cases} [H_1^T \ H_2^T \dots H_e^T | 1] & \text{for odd column weights} \\ [H_1^T \ H_2^T \dots H_e^T | 1] & \text{for even column weights} \end{cases}$$

Self-Dual:

$$S = \begin{bmatrix} H^T & 0 \\ 0 & H \end{bmatrix} = H^T H^T = 0$$

Assuming real-values, random samples of randomly sized n-dimensional matrices produces a space.

related to orthogonal subspaces.

$\boxed{\text{The amount of low-density parity-check codes that are orthogonal is a ratio of the n-dimensional matrix growth.}}$

Exercise 13.26: $E_b/N_0 = \dfrac{\bar{x}_n^2}{2\sigma^2 R}$; $P(R|\sigma^2, R) = e^{-w E_b/N_0}$

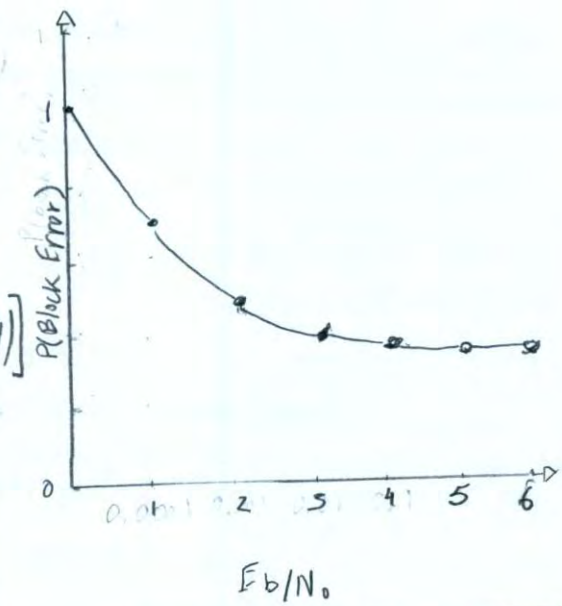
$$P(\text{block Error}) = \sum_{w=0}^{N} \binom{N}{w} 2^{-N} e^{-w E_b/N_0}$$

$$= 2^{-N} \sum_{w=0}^{N} \binom{N}{w} \left(e^{-E_b/N_0}\right)^w$$

$$= 2^{-N} \left(1 + e^{-E_b/N_0}\right)^N$$

$$= 2^{\left[-N\left(1 - \log_2\left(1 + e^{-E_b/N_0}\right)\right)\right]}$$

Assuming, N=1, for the plot.



Chapter 15: Further Exercises on Information Theory:

Exercise 15.1: $A_x = \{0, 1\}$, $P_x = \{0.995, 0.005\}$, X^{100} block Code, where $x \in X^{100}$ contains (≤ 3) 1's.

a) If a binary symmetric channel, $p_b = \sum_{n=(N+1)/2}^{N} \binom{N}{n} f^n (1-f)^{N-n}$

$$\cong 2^N \frac{1}{\sqrt{\pi N/2}} f[f(1-f)]^{(N-1)/2}$$

$$\cong \frac{1}{\sqrt{\pi N/8}} f[4f(1-f)]^{(N-1)/2}$$

$$0.005 = \frac{f[4f(1-f)]^{(N-1)/2}}{\sqrt{\pi \cdot N/8}}$$

$$(N-1)/2 \cong \frac{\log(0.005) + \log \frac{\sqrt{\pi N/8}}{f}}{\log 4f(1-f)}$$

$$\boxed{\text{If } f = 0.1\text{, then } N = 6}$$

b) Probability of a string x containg r 1's and $N-r$ 0's where $r=3$, and $N=100$.

$$P(x) = p_1^r \circ p_0^{N-r} = 0.995^3 \cdot (0.005)^{97} \cong 0$$

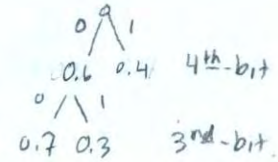
Probabability of a string being ignored $\boxed{\cong 1}$

Exercise 15.2: $P_x = \{0.1, 0.2, 0.3, 0.4\}$; $C = \{0001, 001, 01, 1\}$

a. Entropy of the fourth bit transmission:

$$H(X) = \sum p(x) \log_2 \frac{1}{p(x)} = 0.4 \cdot \log_2 \left(\frac{1}{0.4}\right) = \boxed{5.3 \text{ bits}}$$

b. Conditional Entropy of the fourth bit given the third:



$P(4^{th}\text{bit} | 3^{rd}\text{bit}) = 0.18$

$$H(X | Y_3) = \sum P(X|y) \cdot \log \frac{1}{P(X|y)} = 0.18 \cdot \log_2 \frac{1}{0.18}$$

$$\boxed{= 0.45}$$

c. Entropy of the Hundreth bit:

$$P_{100} = 0 \; ; \; H(P_{100}) = 0 \cdot \log \frac{1}{0} = \boxed{0}$$

d. Conditional Entropy of the hundreeth bit given the ninety-ninth bit:

$$\boxed{H(X|Y) = 0}$$

Exercise 15.3: $\boxed{\text{Probability Space}}$

1+1 = 2	2+2 = 4	3+3 = 6	4+4 = 8	5+5 = 10	6+6 = 12
1+2 = 3	2+3 = 5	3+4 = 7	4+5 = 9	5+6 = 11	
1+3 = 4	2+4 = 6	3+5 = 8	4+6 = 10		
1+4 = 5	2+5 = 7	3+6 = 9			
1+5 = 6	2+6 = 8				
1+6 = 7					

$\boxed{\text{Range of Probability Space}}$ 2 - 12

$\boxed{\text{Most Probable outcome}}$ 7

$\boxed{\text{Divide-and-conquer}}$ ① Guess mid-range : 7.
② If above seven, then guess nine else, guess four.
③ Continue guessing mid-range.

Exercise 15.4: A coin can purpose for straws when flipped multiple times.

Exercise 15.5: $\boxed{\text{Likelihood of suit: } 52C5 / 13C1 = 2000\%}$

Exercise 15.6:

Exercise 15.7: $H(x) = \sum p(x) \log \frac{1}{p(x)}$; $H(p) = \infty$ when $(0 < p_n < p$ and $n \gg 1)$

Exercise 15.8: $A_x = \{a, b, c, d\}$; $P_x = \{1/2, 1/4, 1/8, 1/8\}$

$$T_{N\beta} \equiv \{x \in A_X^N : |\frac{1}{N} \log_2 \frac{1}{P(x)} - H| < \beta\} ; N = 8, \beta = 0.1$$

$$\equiv \{x \in A_x^8 : |\frac{1}{8} \log_2 \frac{1}{P(x)}| < 0.1\} ; P(x) = 2^{-0.8} = 0.57 = 57\%$$

Words defined by boundary: $P(x) \cdot 4^8 = \boxed{37,641 \text{ words}}$

Exercise 15.9: $A_s = \{a, b, c, d, e\}$, $P_s = \{1/3, 1/3, 1/9, 1/9, 1/9\}$

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 2/3 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1/3 & 0 \end{bmatrix} ; A_x = A_y = \{0, 1, 2, 3\} : R = 3/4 ; \epsilon > 0$$

P(x,y)	0	1	2	3	P(y)
0	1	0	0	0	1
1	0	0	2/3	0	2/3
2	0	1	0	1	2
3	0	0	1/3	0	1/3
P(x)	1	1	1	1	4

In this difficult problem, the probability of error (p_b) is supposed to be less than error (ϵ). Two methods exist, Shannon's Noisy coding theorem and a lossy compressor to minimize error. A channel shown may build a decoder from A_x to A_y relationships.

Exercise 15.10: $C = \{0000, 0011, 1100, 1111\}$; $p_0 = \{1/2, 1/8, 1/8, 1/4\}$

Assuming $f = 0.1$ and binary symmetric channel

$$P(y=0000 \mid x=0000) = \frac{P(\text{Binary symmetric Channel})\, P(x=0000)}{\sum^{4} P(y=0000 \mid x')\cdot p(x')}$$

Where $P(\text{Binary Symmetric channel}) = \boxed{f^d (1-f)^{n-d}}$

Exercise 15.11: $Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{bmatrix}$; $C = \max_{P_x} I(X;Y) = \frac{5}{3}$

Optimal Input Distribution: $\boxed{\{1, 1/3, 1/3\}}$

Exercise 15.12: Q_{8bits} (input) ; Q_{8bits} (output) ; $C = \max_{P_x} I(X;Y) = H(X) - H(X|Y)$

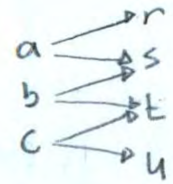
$= \log_2 8 - 1$

$= \log_2 4$

$\boxed{\text{A lossless encoder and decoder is the } (X, 5) \text{ Hamming code, where } X = 8 \text{ bits.}}$

Exercise 15.13: $X \in \{a, b, c\}$; $y \in \{r, s, t, u\}$

$Q = \begin{bmatrix} 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 \end{bmatrix}$



$C = \max_{P_x} I_c = (1/2 + 1/2 + 1/2 + 1/2) - (1/2 + 1/2) = \boxed{1}$

Exercise 15.14: $X_{10} = \left(\sum_{n=1}^{9} n x_n\right) \bmod 11$; Books ISBN: 817525766-0

$X_{10} = (1\cdot 8 + 2\cdot 1 + 3\cdot 7 + 4\cdot 5 + 5\cdot 2 + 6\cdot 5 + 7\cdot 7 + 8\cdot 6 + 9\cdot 6) \bmod 11$

$= (242) \bmod 11$

$\boxed{= 0}$

Show a code detects all errors of ten digits:

1-010-00000-4

1-010-00080-4

X_{10} of 1-010-00080-4 produces a non-zero modulus.

Show that this code can be used to detect adjacent transposes:

(1-010-00000-4) mod 11 = (40) mod 11 = 7

(1-100-00000-4) mod 11 = (39) mod 11 = 6

a digit was switched by an adjacent position

The other pairs of non-adjacent digits relate through the modulus increase or decrease.

Modulus-10 coding would not work so well because specific changes, alteration of an even indexed digit by five counts, and shift of index by five, -do not evaluate akin to modulus-11.

Exercise 15.15:

$$Q = \begin{bmatrix} 1-f & f & 0 & 0 \\ f & 1-f & 0 & 0 \\ 0 & 0 & 1-g & g \\ 0 & 0 & g & 1-g \end{bmatrix} \quad ; \quad P_x = \left\{ \frac{p}{2}, \frac{p}{2}, \frac{1-p}{2}, \frac{1-p}{2} \right\}$$

$$H(Y) = [1-f+1-f+1-g+1-g] = \left[\frac{p}{2} + \frac{p}{2} + \frac{(1-p)}{2} + \frac{(1-p)}{2}\right] = 1 \; ; \; H(Y|X) = \left[2\left(\frac{1-p}{2}\right) + 2\left(\frac{p}{2}\right)\right] = 1$$

Optimal Input Distribution : $\frac{d}{dp} I(X;Y) = \frac{d}{dp}\left[H(Y) - H(Y|X)\right]$

$$= \frac{d}{dp}\left[H_2(p) - p\left(H_2(f) - H_2(g)\right)\right]$$

$$0 = \log\left(\frac{1-p}{p}\right) - H_2(f) + H_2(g)$$

$$p = \frac{1}{1 + 2^{H_2(f) - H_2(g)}}$$

If $f = 1/2$ and $g = 0$, $p = \frac{1}{1 + 2^{H_2(0.5) - H_2(0)}} = \boxed{\frac{1}{3}}$

Exercise 15.16:

Error-Detecting codes

- Parity-check
- Check sum
- Cyklic Redundancy Check

Both

- Error Detection
- Extra-bits of Data in Transmission (Parity)

Error-correcting codes

- correct single-bit errors
- correct single-bit flips
- correct multi-bit errors
- Hamming-Code
- Reed-Solomons

Exercise 15.17:

$$X^{(A)} \xrightarrow[R_A]{\text{encode}} t^{(A)} \to C$$

$$X^{(B)} \xrightarrow[R_B]{\text{encode}} t^{(B)} \to C$$

$P(X^{(A)}, X^{(B)})$:

$X^{(B)}$ \ $X^{(A)}$	0	1
0	0.49	0.01
1	0.01	0.49

① $H(X^{(A)} | X^{(B)}) = H(X^{(A)}) + H(X^{(A)}, X^{(B)})$

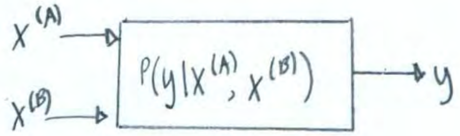
② $H(X^{(A)}, X^{(B)}) = H(X^{(A)} | X^{(B)}) - H(X^{(A)})$

③ $R_A = 1 - H(X^{(A)})$; $H(X^{(A)}, X^{(B)}) = H(X^{(A)} | X^{(B)}) - 1 + R_A$

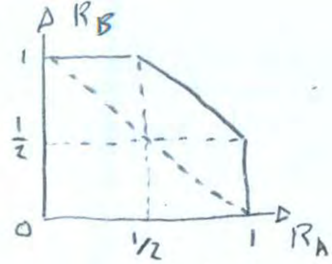
$$= H(X^{(B)} | X^{(A)}) - H(X^{(B)}) - 1 + R_A$$

$$= H(X^{(B)} | X^{(A)}) - 2 + R_A + R_B$$

Exercise 15.18:



$x^{(B)}$ \ $x^{(A)}$	0	1
0	0	1
1	1	2



① $H(X^{(A)}|X^{(B)}) = H(X^{(A)}) + H(X^{(A)}, X^{(B)})$

② $H(X^{(A)}, X^{(B)}) = H(X^{(A)}|X^{(B)}) - H(X^{(A)})$

③ $R_A = 1 - H(X^{(A)})$; $H(X^{(A)}|X^{(B)}) = [H(X^{(B)}|X^{(A)}) - 2] + (R_A + R_B)$

$R_B = 1 - H(X^{(B)})$;

How can users A and B use this channel so that their messages can be deduced from the recieved signals?

∘ The input rate of each user should sum less than one

How fast can A and B communicate?

∘ A maximum rate for $X^{(A)}$ or $X^{(B)}$ is 1 unit i.e. $1 = \frac{M}{N}$.

Can reliable communication be achieved at rates (R_A, R_B) such that $R_A + R_B > 1$?

∘ No, the maximum rate of convoluted signals is not independent of each other.

Exercise 15.19: Conditional Distribution: $Q(y^{(A)}, y^{(B)}|x)$; $R_A[A]$, $R_B[B]$, (R_0, R_A, R_B)

$C_A[A]$, $C_B[B]$

Transmission Time $[\phi_A]$ of Channel A

Transmission Time $[\phi_B]$ of Channel B.

Where $1 = \phi_A + \phi_B$

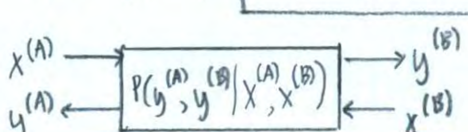
$$H(X^{(A)}|X^{(B)}) = [H(X^{(B)}|X^{(A)}) - 2] + (R_A + R_B)$$

$$\boxed{= [H(X^{(B)}|X^{(A)}) - 2] + \left[\frac{C_A}{1 - H_2(p_A)} + \frac{C_B}{1 - H_2(p_B)}\right]}$$

Exercise 15.20: Binary Symmetric Channel ; Noise $f_A < f_B$, C_A & C_B.

$$H(X^{(A)}|X^{(B)}) = \boxed{[H(X^{(B)}|X^{(A)}) - 2] + \left[\frac{C_A}{1 - H_2(f_A)} + \frac{C_B}{1 - H_2(f_B)}\right]}$$

Exercise 15.21:



Simultaneous rates $1 = R_A + R_B$

Limitations:

$y^{(A)}$:

$x^{(B)}$ \ $x^{(A)}$	0	1
0	0	0
1	1	0

$y^{(B)}$:

$x^{(B)}$ \ $x^{(A)}$	0	1
0	0	1
1	0	0

Chapter 16: Message Passing:

Exercise 16.1: Probability $= 100\% = P(\text{Path } 1) + P(\text{Path } 2) + P(\text{Path } 3) + P(\text{Path } 4) + P(\text{Path } 5)$

$$= \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5}$$

Probability of flipping a coin at n-junctions: $P(\text{Junction}) = \left(\frac{1}{2}\right)^n$

The distributions of regular paths is different than flipping a coin at each junction.

Exercise 16.2: The path from A to B requires a message passing algorithm. Being that the structure is not a line, then a Rule-set B message passing algorithm is best implemented.

Exercise 16.3: The communication channel in Figure 16.11a has asymptotic properties because of an expected value, described by

$$E[X] = \sum_{A}^{B} F_s(X, X_s) = \sum_{A}^{B} \prod_{i=1}^{L\text{-steps}} F_s(x_i, X_s),$$ also known as sum-product algorithm

Exercise 16.4: $I(x,y) = \sum_{u=0}^{X} \sum_{v=0}^{y} f(u,v)$ is an "integral image", which, is also described by rule-set A message passing algorithm, because every summation is a line of data.

A sum of image intensities across the range of (x_1, y_1) to (x_2, y_2) is Intensity $= \sum_{x_1}^{x_2} \sum_{y_1}^{y_2} f(x_i, y_i)$

Chapter 17: Communication over constrained Noiseless Channel

Example 17.1: 00100101001010100010 is an example of a valid string.

Example 17.2: A valid string for the model is 00111001110011000011

Example 17.3: 100100110110011010 1 is the valid string

Exercise 17.4: $H_2(f) = f \cdot \log_2 \frac{1}{f} + (1-f)\log_2 \frac{1}{(1-f)}$; $H_2(f) \cong -f\left[(f-1) - \frac{(f-1)^2}{2}\right] - (1-f)\left[f - \frac{(f)^2}{2}\right]$

$$\cong \cancel{f} - \cancel{f^2} + \frac{f(f-1)^2}{2} - \cancel{f} + \cancel{f^2} + \frac{(1-f)f^2}{2}$$

$$\cong \frac{\left(\frac{1}{2}+\delta\right)\left(\delta-\frac{1}{2}\right)^2}{2} + \frac{\left(\frac{1}{2}-\delta\right)\left(\frac{1}{2}+\delta\right)^2}{2}$$

$\cong$

$$\boxed{\cong -\frac{(\sigma-1/2)(\sigma+1/2)}{2}}$$

Exercise 17.5: $f=0.5$ 1s; C_2 Code; 1's appear 1/3 of the string and 0's 2/3 of the characters.

s	t
0	0
1	10

A sparse source density of $f=0.38$ produces 1s $[f/(1+f) = 0.2753]$ or 27.5% of the valid character length.

Exercise 17.6: Fibonacci Series: $F_n = F_{n-1} + F_{n-2}$; $\lim_{n\to\infty} \frac{F_n}{F_{n-1}} = \lim_{n\to\infty} \frac{\frac{x^{n+1}-(x')^{n+1}}{x-x'}}{\frac{x^n-(x')^n}{x-x'}}$

$$= x\lim_{n\to\infty} \frac{x^n}{x^n-(x')^n} - x'\lim_{n\to\infty} \frac{(x')^n}{x^n-(x')^n}$$

$$= x = \boxed{\frac{1+\sqrt{5}}{2}}$$

Exercise 17.7: The relationship of Channel C to channels A and B occurs from an accumulator or bit inversion. Channel A is generated from Channel C by inversion $[1\to0; 0\to1]$ and an accumulator of duplicate bits. While, Channel B is Channel C when digits of B are run through an accumulator, and C, an inversion, along with, an accumulator.

Exercise 17.8: Sample Space {00000000, 00000010 ...... 11111110}; Rate = $\frac{\text{source}}{\text{transmitted}} = 5/8$

The rate achieved by mapping an integer number of source bits to N=16 transmitted bits is $R = n/16$, where n-max is 10.

Exercise 17.9: Optimal Transition Probability [Q]

If $Ae^{(R)} = \lambda e^{(R)}$; $(A-\lambda)e^{(R)} = (a_n\lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0)e^{(R)}$

$$= (a_nA^n + a_{n-1}A^{n-1} + \dots + a_1A + a_0)e^{(R)}$$

$$= 0; \quad Q_{s's} = \frac{a_1Ae^{(R)}}{(a_nA^n + a_{n-1}A^{n-1} + \dots)e^{(R)}} = \boxed{\frac{a_1Ae^{(R)}}{\lambda e^{(R)}}}$$

Similarly for $e^{(L)T}A = \lambda e^{(L)T}$; $Q_{s's} = \boxed{\frac{e_s^{(L)}a_1A}{\lambda e_s^{(L)}}}$

Exercise 17.10: Optimal Transition Probability Matrix: $\boxed{Q_{s'|s} = \frac{e_{s'}^{(L)} \cdot A_{s's}}{\lambda e_s^{(L)}}}$

Conditional Entropy: $H(Y|X) = \boxed{\sum P(y,x) \log \frac{1}{P(Y|X)}}$

Invariant Distribution: $\boxed{P(s) = \alpha \cdot e_s^{(L)} \cdot e_s^{(R)}}$

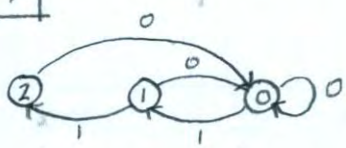
$$H(s_n|s_{n-1}) = -\sum P(s)_{n-1} \cdot P(s_n|s_{n-1}) \cdot \log P(s_n|s_{n-1})$$

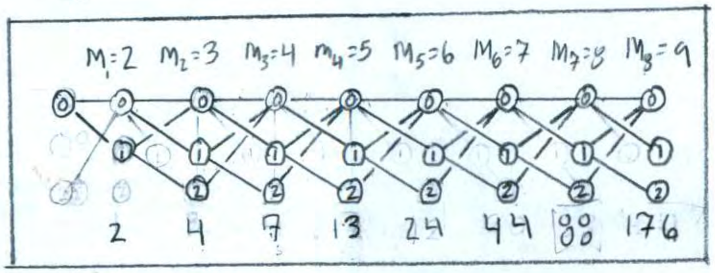
$$= -\sum \alpha e_s^{(L)} \cdot e_s^{(R)} \frac{e_{s'}^{(L)} A_s}{\lambda e_s^{(L)}} \log \frac{e_s^{(L)} A_s}{\lambda e_s^{(L)}}$$

$$= -\sum \alpha e_s^{(R)} \frac{e_s^{(L)} A_s}{\lambda} \left[\log e_{s'}^{(L)} + \cancel{\log A_s} - \log \lambda - \log e_s^{(L)}\right]$$

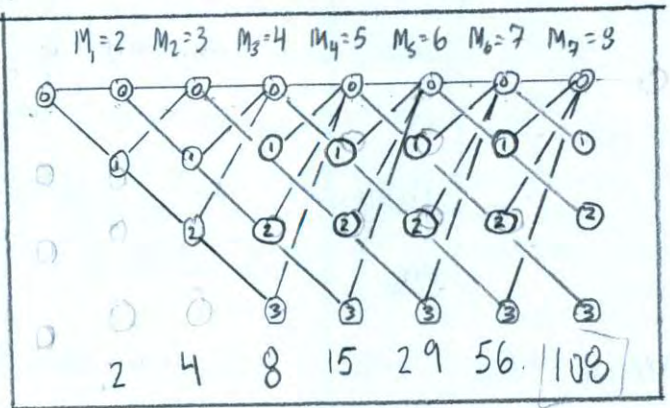
$$\boxed{= \log \lambda}$$

Exercise 17.11: $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$





$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$



- The capacity of State Diagram #1 is $C = \log_2 \lambda_1 = \log_2(1.839) = 0.879$.
- The capacity of State Diagram #2 is $C \cong \log_2(1.927) = 0.946$
- Matrix Q is shown in figure 17.9, but channel A run-length-limited plot is $Q = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$.

Exercise 17.12: Run-Length 0's permitted; Run-length 1's is $L >> 1$

- The capacity is near 100% because the channel is almost unrestrained.
- A series expansion involving L is $\sum_{l=1}^{L+1} 2^{-\beta l} = 2^{-\beta} \frac{(2^{-\beta})^{L+1} - 1}{2^{-\beta} - 1}$

where the geometric series $\sum_{n=0}^{N} ar^n = \frac{a(r^{N+1}-1)}{r-1}$

is utilized. If the capacity is near one, then the summation is $1 \cong 2^{-\beta} \frac{(2^{-\beta})^{L+1} - 1}{2^{-\beta} - 1}$ and $\beta \cong 1 - 2^{-(L+2)}/\ln 2$.

- An optimal matrix Q is close to $\begin{bmatrix} 0 & 0 \\ 1 & 1/2 \end{bmatrix}$, but is adjusted as the number of ones increase.

Exercise 17.13: 'Morse' Channel - is an example of a constrained channel with variable symbol durations.

the dot	d
the dash	D
the short space	s
the long space	S

The capacity of this channel is:

...when equal duration:

$$1 = \frac{4}{2^{\beta l}}; \quad \beta = \log_2 \frac{4}{l}$$

.... symbol durations are 2, 4, 3 and 6

$$1 = p_1 + p_2 + p_3 + p_4$$

$$= \frac{1}{2^{2\beta}} + \frac{1}{2^{4\beta}} + \frac{1}{2^{3\beta}} + \frac{1}{2^{6\beta}}$$

$$1 = \frac{2^{13\beta} + 2^{11\beta} + 2^{12\beta} + 2^{9\beta}}{2^{15\beta}}$$

$$\boxed{\beta = 0.597}$$

Exercise 17.14: If Morse code utilized a probability distribution similar to Figure 2.1, then one would notice Morse's design relates to the most probable letters.

A •—	F ••—•	K —•—	P •——•	U ••—	Z ——••
B —•••	G ——•	L •—••	Q ——•—	V •••—	
C —•—•	H ••••	M ——	R •—•	W •——	
D —••	I ••	N —•	S •••	X —••—	
E •	J •———	O ———	T —	Y —•——	

Exercise 17.15: width [L]; 3 directions per step: $H = \log 3$ per step; Total Entropy $\left[H_{Tot} = N\log 3\right]$

Free Energy

$\boxed{\text{Connection Matrix}}$

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ & & \ddots & & \ddots & & \ddots & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$\boxed{\text{Entropy of the Polymer}}$: $10\log 3$.

$\boxed{\text{The change in entropy associated with the polymer entering the tube is}}$ $10\log 3 - 10\log 1 = 10\log 3$.

or $L\log 3$

Chapter 18: Crosswords and Codebreaking:

Exercise 18.1: Capacity of Word: English, as a constraint.

$$\beta = \frac{H(p)}{L(p)} = \frac{\sum p_n \log \frac{1}{p_n}}{\sum p_n \cdot l_n}$$

$$= \frac{0.0913 \cdot \log\frac{1}{0.0913} + 0.0596\log\frac{1}{0.0596} + 0.0133\log\frac{1}{0.0133} + 0.0335\log\frac{1}{0.0335} + 0.0599\log\frac{1}{0.0599} + 0.0567\cdot\log\frac{1}{0.0567} + 0.0313\log\frac{1}{0.0313}}{1\cdot 0.0913 + 2\cdot 0.0596 + 3\cdot 0.0133 + 4\cdot 0.0335 + 5\cdot 0.0599 + 6\cdot 0.0567 + 7\cdot 0.0313}$$

$$= \frac{1.13}{1.24} = 0.91 \cong \boxed{91\% \text{ capacity}}$$

Exercise 18.2: The capacity of the two dimensional channel is

$$\beta = \frac{\sum H(p)}{\sum L(p)} = \frac{\sum_1^N \sum_1^N p_i \log\frac{1}{p_i}}{\sum_1^N \sum_1^N p_i l_i}; \quad \boxed{\lim_{N\to\infty} \beta = \frac{1}{2}}$$

Exercise 18.3: $\boxed{\begin{array}{l}\text{The information leaked by pressing "Q" and the output is a different letter is } I(X) = \log\frac{1}{P(X,X|H_1)} = \log\frac{A(A-1)}{(1-m)} = \log\frac{26\cdot 25}{(1-0.076)} = 9.46 \text{ bits.} \\ \text{The crib needed to confirm alignment relates to } \log\frac{P(X,y|H_1)}{P(X,y|H_0)} = M\log mA + N\log\frac{(1-m)A}{A-1} \text{ being positive odds.}\end{array}}$

o Although the number of characters in the crib depend on the number of matches

Chapter 19: Why have Sex? Information Acquisition and Evolution:

Exercise 19.1: The results of sexual population depend on population size. When the population (N) is small ($<2G$), less than 1 bit of information changes and a single generation represents an exact DNA replica, however, large populations show DNA changes proportional to $\sqrt{G}$ bits per generation. The limit to evolution is a model where population N is less than $\sqrt{G}$.

Exercise 19.2: a) The model for gene crossover assumes normalized fitness $f = F/G$ is greater than $1/2$ because a single mutation has probability $\leq 1/2$. A fixed m suggests a Gaussian mean and variance for normalized fitness given by: $\delta(t) = \frac{1}{2\sqrt{mG}}(1 - c \cdot e^{-2m \cdot t})$ for $f \leq 1/2$.

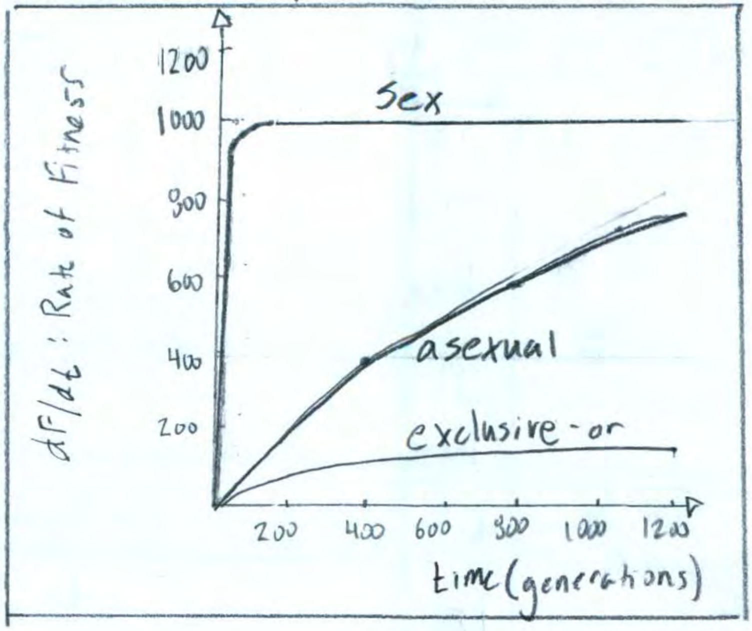
b) If crossovers occur exclusively at hot-spots located every d bits along the genome, a best fit model is $F = \frac{N}{d}$. A d-value less than or equal to two generates a similar model to part a, due to the fact, normalized fitness f becomes $1/2$.

Exercise 19.3: The fitness function as a sum of exclusive-ors of pairs of bits

$$f = \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} (X_i \wedge X_j)}{G}$$

and a graph comparing a simple additive function:

The evolving fitness of an exclusive-or of+ pair of bits is different than sexual and ascxual species, as shown by the plot.



Sex Function:

$$f(t) = \frac{1}{2}\left[1+\sin\left(\frac{\sqrt{\frac{2}{\pi+2}}}{\sqrt{G}}(t+)\right)\right]\cdot G$$

Asexual Function

$$f(t) = \frac{G}{2\sqrt{mG}}\left(1-e^{-2mt}\right)$$

Exclusive-or Function

$$f(t) = \frac{X}{2\sqrt{G}}$$

Output Function $\frac{dF}{dt} = \sqrt{\frac{2}{\pi+2}}\sqrt{f(t)\cdot G}$ where $G=1000$; $m=0.4$

Exercise 19.4: An assumption of 2×10^5 generations for human/chimpanzee divergence, then an error-correcting machinery is calculated to be correct more than the genome in this time.

Exercise 19.5: $f=\exp(-\Delta E/kT)$; $f \simeq 10^{-4}$; $\Delta E = 1.64\times10^{-20}$ J @ 298K.

The cellular processing contains a kinetic proofreading, which implies other pathways to regulate binding energy. This Kinetic proofreading lowers errors in identification by adjusting the thermodynamic limits. (Hopfield 1974)

Exercise 19.6: If a person learns 10 new words per day at a length of five characters, with a retrieval efficiency of 80%, then the brain stores 20-bits in 80 years.

Chapter 20: K-means Clustering

Exercise 20.1: Steps to K-means:

① Assignment Step: Find the closest mean for each datapoint.

② Update Step Adjust the mean according to point assignment, normalized by the total point assignment.

③ Repeat assignment and update until averages no longer move i.e. converge.

An energy definition describing distance would be the distance formula, and this function is bounded. The assignment step minimizes the closest "energy" function, the update step adjusts means to lower "energy", and a repeat promotes the process. Each step is a Lyapunov State.

Exercise 20.2: $\lim_{\beta \to \infty} r_k^{(n)} = \lim_{\beta \to \infty} \frac{\exp(-\beta d(m^{(k)}, x^{(n)}))}{\sum_{k'} \exp(-\beta d(m^{(k')}, x^{(n)}))}$ has a range between 0 and 1.

The means with no assigned points in the update step require zero change.

Exercise 20.3: $[Var(x_1), Var(x_2)] = (\sigma_1^2, \sigma_2^2)$ where $\sigma_1^2 > \sigma_2^2$, $K = 2$, $N \gg 1$

Assume $m^{(1)} = (m, 0)$ and $m^{(2)} = (-m, 0)$

① Assignment $r_1(x) = \frac{e^{-\beta(x_1 - m)^2/2}}{e^{-\beta(x_1+m)^2/2} + e^{-\beta(x_1-m)^2/2}} = \frac{1}{1 + \exp(-2\beta m x_1)}$

② Update $m^{(1)} = \frac{\sum r_1^{(n)} \cdot x^{(n)}}{\sum r_1^{(n)}} = \frac{\int_0^{\infty} r_1^{(n)}(x)\, x^{(n)}\, dx}{\int_0^{\infty} r_1^{(n)}(x)\, dx} = 2\beta m \int_0^{\infty} r_1^{(n)} x^{(n)} dx = 2\beta m \int_0^{\infty} \frac{1}{2}(1 + \beta m x_1)\, x_1$

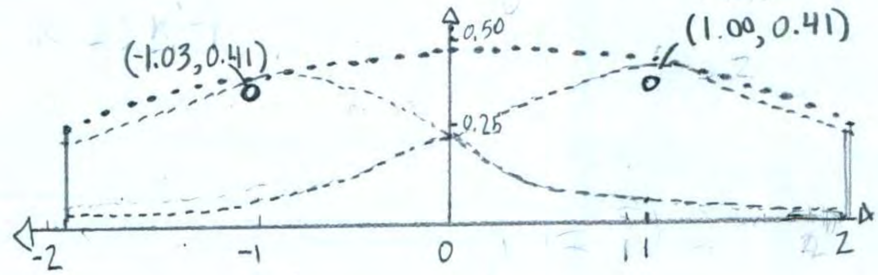
$= \sigma_1^2 \beta^2 m^2$

If β parameter adjusts the mean, then large values of β report a non-converging mean. When β is less than one, the mean converges upon multiple iterations.

Exercise 20.4: Hard K-Means

① Generated 40 points from [-2 to +2]

② Fit Gaussians to the points, then added the values



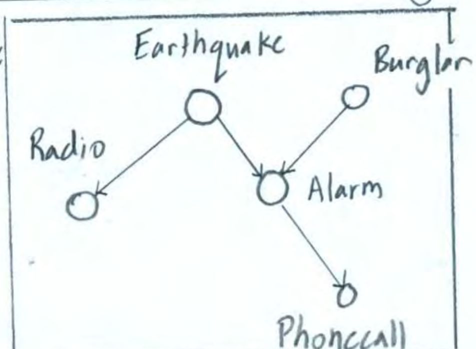
③ ΔX for Hard K-means: 2.03

This value is greater than the distance of the Gaussian means that generated the distance.

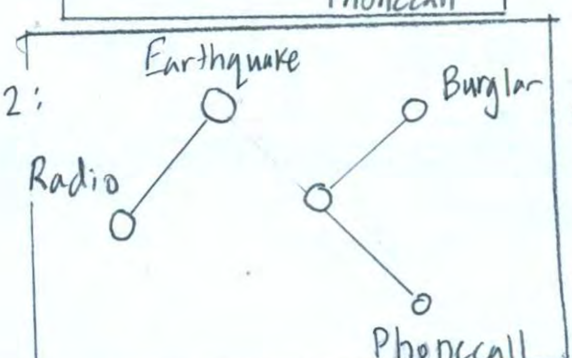
Soft K-means: If β is larger than 1, then the soft K-means converges to two means, while a β value less than one produces a single mean at zero.

Chapter 21: Exact Inference by Complete Enumeration

Example 21.1:



Exercise 21.2:



$$P(e) = \frac{P(b,e \mid a=1)\, P(a=1)}{P(a=1 \mid b,e) \cdot P(b)} = \frac{P(b,e=1 \mid a=1) \cdot 0.002}{P(a=1 \mid b,e=1) \cdot 0.001}$$

If $b=1$, then $P(b=1, e=1 | a=1) = 0.0005$

and $P(a=1 | b=1, e=1) = 0.9901099$

then $\boxed{P(e) = 1.01 \times 10^{-3}}$

If $b=0$, then $P(b=0, e=1 | a=1) = 0.0055$

and $P(a=1 | b=0, e=1) = 0.01099$

then $\boxed{P(e) = 1.00}$

Exercise 21.3: $\Pi \in \{\Pi_1, \Pi_2, \ldots, \Pi_{21}\}$; $\sigma \in \{\sigma_1, \sigma_2, \ldots, \sigma_{10}\}$; $\mu \in \{\mu_1, \mu_2, \ldots, \mu_{10}\}$

Total parameters [K]: 20 ; $\boxed{10^K = 10^{20}}$

Chapter 22: Maximum Likelihood and Clustering:

Example 22.1: Log Likelihood: $\ln P(x | \mu, \sigma) = \boxed{-N \cdot \ln(\sqrt{2\pi}\sigma) - \sum_n (x_n - \mu)^2 / (2\sigma^2)}$

Derivative of log Likelihood [μ]: $\dfrac{d \ln P(x|\mu,\sigma)}{d\mu} = +2\sum_n (x_n - \mu) / (2\sigma^2)$

$= 0$

$\mu = \dfrac{X}{n} = \boxed{\bar{x}}$

Example 22.2: Second Derivative of log Likelihood: $\dfrac{d^2 \ln P(x|\mu,\sigma)}{d^2\mu} = -2n / 2\sigma^2 = \boxed{-\dfrac{n}{\sigma^2}}$

$\boxed{\sigma_\mu = \sigma / \sqrt{N}}$; $\boxed{\mu = \bar{x} \pm \sigma_\mu}$

Example 22.3: $\dfrac{d \ln P(\{x_n\}_{n=1}^N | \mu, \sigma^2)}{d \ln \sigma} = -\dfrac{N}{\sigma} + \dfrac{S}{\sigma^3} = 0$; $\boxed{\sigma_{MAX} = \sqrt{\dfrac{S}{N}} = \sqrt{\dfrac{\Sigma (x-\mu)^2}{N}}}$

$\dfrac{d^2 \ln P(\{x_n\}_{n=1}^N | \mu, \sigma^2)}{d^2 \ln \sigma} = -\dfrac{2S}{\sigma^2} = -2N$; $\boxed{\sigma = \sqrt{\dfrac{1}{2N}}}$

Exercise 22.4: $\dfrac{\partial \ln P(\{x_n\}_{n=1}^N | \mu, \sigma^2)}{\partial \mu} = 2\Sigma (x_n - \mu) / 2\sigma^2$; $\boxed{\bar{x} = \mu}$

$\dfrac{\partial \ln P(\{x_n\}_{n=1}^N | \mu, \sigma^2)}{\partial \sigma} = -\dfrac{N}{\sigma} + \dfrac{S}{\sigma^3} = 0$; $\boxed{\sigma_N = \sqrt{\dfrac{S}{N}}}$

Exercise 22.5: $P(x|\mu_1, \mu_2, \sigma) = \left[\sum_{k=1}^{2} p_k \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)\right]$

Prior Probability $\{p_1 = 1/2, p_2 = 1/2\}$; $\theta \equiv \{\{\mu_k\}, \sigma\}$; Bayes.

$$P(k_n = 1 | x_n, \theta) = \frac{1}{1 + \exp[-(w_1 x_n + w_0)]} \; ; \; P(k_n = 2 | x_n, \theta) = \frac{1}{1 + \exp[(w_1 x_n + w_0)]}$$

$\boxed{\text{Posterior Probability}}$: $P(\{x_n\}_{n=1}^{N} | \{\mu_k\}, \sigma) = \prod_n P(x_n | \{\mu_k\}, \sigma)$

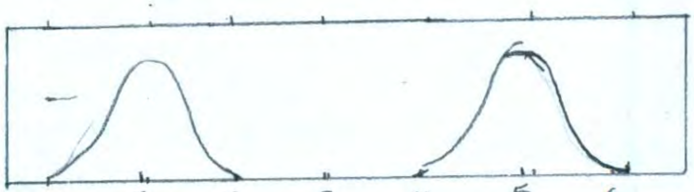
$\boxed{\text{Maximum Log Likelihood}}$ $\frac{\partial}{\partial \mu} L = \sum_n p_{kn} \frac{(x_n - \mu_k)}{\sigma^2}$ where $p_{kn} = P(k_n = k | x_n, \theta)$

$$\frac{\partial^2}{\partial \mu} L = -\sum_n p_{kn} \frac{\mu}{\sigma^2}$$

$\boxed{\text{Newton-Raphson Method}}$ $\mu' = \mu - \frac{\partial L}{\partial \mu} \cdot \frac{\partial \mu}{\partial^2 L} = \mu + \sum p_{kn} \frac{(x_n - \mu_k)}{\cancel{\sigma^2}} \cdot \frac{\cancel{\sigma^2}}{\sum p_{kn} \cdot \mu}$

$$\mu' = \frac{\sum p_{kn} \cdot x_n}{\sum p_{kn}}$$

$\boxed{\text{Contour Plot}}$:



$\mu_1 = 1, \mu_2 = 5, \sigma = 1$

Exercise 22.6: If $r_k \approx 1$, then $m^{(k)} \approx x^{(n)} + \epsilon$; $\sigma_k^2 = \frac{\sum r_k^{(n)} (x^{(n)} - x^{(n)} + \epsilon)^2}{I R^{(k)}}$

$$\lim_{\epsilon \to 0} \sigma_k^2 = \frac{\sum r_k \epsilon^2}{I R^{(k)}} = \boxed{0}$$

Exercise 22.7: $\boxed{\text{A version of K-means that models a mixture of Gaussians not axis-aligned involves a non-diagnolized covariance.}}$

$$\sigma_i^{2(k)} = \begin{pmatrix} \mathrm{Cov}(x_1, x_1) & \cdots & \mathrm{Cov}(x_1, x_k) \\ \vdots & \cdots & \vdots \\ \mathrm{Cov}(x_k, x_1) & \cdots & \mathrm{Cov}(x_k, x_k) \end{pmatrix}$$

$\boxed{\text{① Assignment Step}}$

$$r_k^{(n)} = \frac{\pi_k \frac{1}{\prod_i \sqrt{2\pi}\,\sigma_i^{(k)}} \exp\left(-\sum_i (m_i^{(k)} - x_i^{(n)})^2 / 2(\sigma_i^{(k)})^2\right)}{\sum_{k'} \pi_{k'} \frac{1}{\prod_i \sqrt{2\pi}\,\sigma_i^{(k')}} \exp\left(-\sum_i (m_i^{(k')} - x_i^{(n)})^2 / 2(\sigma_i^{(k')})^2\right)}$$

$\boxed{\text{② Update Step}}$

$$m^{(k)} = \frac{\sum r_k^{(n)} x^{(n)}}{R^{k}}$$

$$\sigma_i^{2(k)} = \frac{\sum_k \sum_k r_k^{(n)} (x_i^{(n)} - m_i^{(n)})^2}{R^{(k)}}$$

$$\pi_k = \frac{R^{(k)}}{\sum R^{(k)}}$$

where $R^{(n)} = \sum r_k^{(n)}$

Exercise 22.8: a) $P(r|\lambda) = \frac{\lambda^r}{r!} \exp(-\lambda)$; $\frac{dP(r|\lambda)}{d\lambda} = \left[\frac{\lambda^{r-1}}{(r-1)!} - \frac{\lambda^r}{r!}\right] \exp(-\lambda) = 0$

$$r = \lambda$$

$$\underline{\lambda_{max} = 9}$$

Error bars of λ : $\frac{d^2 P(r|\lambda)}{\partial^2 \lambda} = \frac{e^{-x} \cdot \lambda^{r-2} (r^2 - r(2\lambda+1) + \lambda^2)}{r!}$

$$= \frac{e^{-\lambda} \cdot \lambda (\lambda^2 - 6\lambda + 6)}{6} = 0$$

$$\boxed{\lambda = 3 \pm \sqrt{3}}$$

Error

b) $b = 13$ photons/min ; $r = \lambda + b$; $r = 9$;

Maximum Likelihood Estimate for $\lambda = 9$.

An biased estimator over or underestimates the data, While an unbiased estimator is exact. This data shows a biased value, with alternatives suggested.

Exercise 22.9: N_a, N_b. N tosses; Beta Distribution: $P(x|\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} x^{\alpha-1}(1-x)^{\beta-1}$

Probability Heads | p

Maximum Likelihood | $\frac{dP(p|\alpha,\beta)}{dp} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\left[(\alpha-1)p^{\alpha-2}(1-p)^{\beta-1} - (\beta-1)p^{\alpha-1}(1-p)^{\beta-2}\right]$

$$= 0; \frac{(\alpha-1)}{(\beta-1)}(1-p) = p; \frac{(\beta-1)}{\beta-\alpha} = p$$

Where $\beta = N_b$ and $\alpha = N_a$

Maximum a posteriori : $\frac{((N+1)-\alpha-1)}{(N+1)-\alpha-\alpha} = \frac{N-\alpha}{N-2\alpha+1} = p_{N+1}$

Maximum Likelihood of Logit | $\frac{da}{dp} = \frac{1}{p} + \frac{1}{1-p} = 0$

$$p-1 = p$$

$$p = 1/2$$

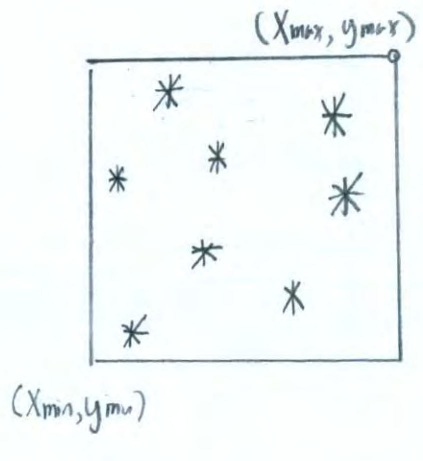
The logit function demonstrates a likelihood of the next toss being 50% heads

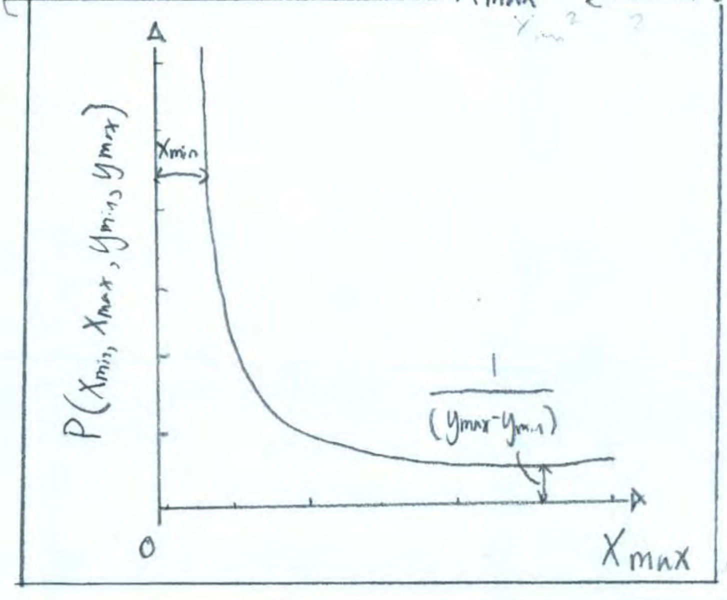
Exercise 22.10: Uniform Distribution | $P(a,b) = \frac{1}{b-a}$

2-D Uniform Distribution | $P(a,b|N) = \left(\frac{1}{b-a}\right)^N$

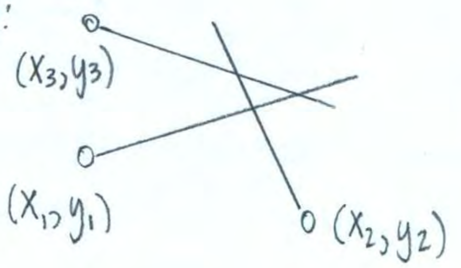
Maximum Likelihood Estimate | $P(x_{min}, y_{min}, x_{max}, y_{max}) = \frac{1}{(x_{max}-x_{min})} \circ \frac{1}{(y_{max}-y_{min})}$

$X_{max} = \max X_i$





Exercise 22.11:



True Bearings: θ_n

Gaussian Noise: σ

Maximum Likelihood of a Gaussian: $\vec{X} = \theta_n \pm \sigma = X_1 + X_2$

$$\vec{X} = \left(\frac{X_1+X_2+X_3}{3} \pm \sqrt{\frac{(X_2+X_3)^2+(X_1+X_2)^2+(X_3+X_1)^2}{3}},\ \frac{y_1+y_2+y_3}{3} \pm \sqrt{\frac{(y_1+y_2)^2+(y_2+y_3)^2+(y_1+y_3)^2}{3}} \right)$$

The maximum likelihood is better because the cocked hat generates an isosceles triangle, which considers three degrees of error.

Exercise 22.12: $P(X|W) = \frac{1}{Z(w)} \exp\left(\sum W_k f_k(x)\right)$

$$\frac{d \ln P(X|W)}{dw} = \frac{d}{dw}\left[-N \ln Z(w) + \sum \sum W_k f_k(x^{(n)})\right]$$

$$= -N \frac{\partial}{\partial w} \ln Z(w) + \sum f_k(X) = -N\left[\frac{1}{Z(w)} \sum \frac{\partial}{\partial w} \exp\left(\sum W_k f(x)\right)\right] + \sum f_k(x)$$

$$= -N\left[\sum P(X|W) f(X)\right] + \sum f_k(X)$$

$$\boxed{\frac{1}{N} \sum f_k(x) = \sum P(x|W) F_k(x)}$$

Exercise 22.13: $H = \sum P(x) \log \frac{1}{P(x)}$; $\langle f_k \rangle_{P(x)} = F_k$; $P(X)_{Maxent} = \frac{1}{Z} \exp\left(\sum_k W_k f_k(x)\right)$

$$H_\lambda = \sum P(X) \log \frac{1}{P(X)} + \lambda_1\left(\sum P(x) - 1\right) + \lambda_2\left(\sum f_k P(X) - F_k\right)$$

$$\frac{dH_\lambda}{dp} = -\sum_k - \sum_k \ln P(X) + \lambda_1 + \lambda_2 \sum f_k = 0$$

$$\boxed{P(x) = e^{-1 + \lambda_1 + \lambda_2 \sum f_k}}$$

Solving for λ_2: $F_k = \sum f_k P(x) = \sum f_k \, e^{-k+\lambda_1+\lambda_2 \sum f_k}$

$$= e^{-k+\lambda_1} \sum f_k \, e^{\lambda_2 \sum f_k}$$

$$F_k e^{k-\lambda_1} = \sum f_k \, e^{\lambda_2 \sum f_k}$$

$$e^{k-\lambda_1} = \frac{1}{F_k} \sum f_k \, e^{\lambda_2 \sum f_k}$$

$$\sum \left(1 - \frac{f_k}{F_k}\right) e^{\sum f_k} = \sum \left(1 - \frac{f_k}{F_k}\right) x^{k_i} = 0$$

$x = e^{\lambda_2}$; $\boxed{\lambda_2 = \ln x}$

Solving for λ_1: $\boxed{\lambda_1 = 1 - \ln \sum e^{\lambda_2 f_k}}$

Therefore, $P(x) = e^{-k+\lambda_1+\lambda_2 \sum f_k}$

$$= e^{-k} \circ e^{1-\ln \sum e^{\lambda_2 f_k}} \circ e^{\lambda_2 \sum f_k}$$

$$= \frac{e^{\ln x \sum f_k}}{\sum e^{\ln x \, f_k}} = \boxed{\frac{1}{Z} \exp\{\sum w_i f_k)}$$

Exercise 22.14: Gaussian Distribution: $P(w) = \left(\frac{1}{\sqrt{2\pi}\,\sigma_w}\right)^k \exp\left(-\sum_1^k w_i^2 / 2\sigma_w^2\right)$

Thin shell radius: $\text{Shell} = \left[\int_0^\infty r^{k-1} P(r)\, dr\right] \times \text{Surface Area}$

$$= \left[\left(\frac{1}{\sqrt{2\pi}\,\sigma_w}\right)^k \int_0^\infty r^{k-1} \exp\left(-r^2/2k\sigma_w^2\right) dr\right] \times \text{Surface Area.}$$

$$= E[r^{k-1}] \circ S_n \circ V_n$$

$$= E[r^{k-1}] \circ (k+1) \circ V_{k+1} \circ V_k$$

$$\boxed{= E[r^{k-1}] \circ (k+1) \circ \frac{\pi^{(k+1)/2}}{\Gamma(\frac{k+1}{2}+1)} \circ \frac{\pi^{k/2}}{\Gamma(\frac{k}{2}+1)}}$$

Two Gaussians that differ in radius by 1% : $\text{Shell} = \boxed{\left[E[r^{k-1}] \circ (k+1)^2 \circ \frac{\pi^{(k+1)/2}}{\Gamma(\frac{k+1}{2}+1)^2} \frac{\pi^{k/2}}{\Gamma(\frac{k}{2}+1)^2}\right] \times \left[E\left[\left(\frac{99}{100} r\right)^{k-1}\right]\right]}$

Exercise 22.15: N datapoints, N distributions, n-Gaussians $N(x_n | \mu, \sigma_n)$

$\boxed{\text{Maximum Likelihood}}$: $\mu = \frac{\Sigma p(x) x}{\Sigma p(x)}$

$\boxed{\text{The reliable } \mu \text{ is the weighted average}}$:

$$\mu = \frac{\frac{+27.020}{24.25}(27.020) + \frac{3.570}{24.25}(3.570) + \frac{8.191}{24.25}(8.191) + \frac{9.898}{24.25}(9.898) + \frac{9.603}{24.25}(9.603) + \frac{9.945}{24.25}(9.945) + \frac{10.056}{24.25}(10.056)}{24.25}$$

$= 1.01$.

$\boxed{\text{Maximizing the likelihood generated a systematic means to "mean."}}$

Exercise 22.16: $\boxed{\text{Noise level}}$: $\sigma_v = 10$; $\boxed{\text{Gaussian Prior}}$: $P(\omega_i | \alpha) = \text{Normal}(0, 1/\alpha)$

where $\alpha = 1/\sigma_W^2$

$\log \sigma_W \in \{-1, 1\}$

Scenario #1: $\{d_1, d_2, d_3, d_4\} = \{2.2, -2.2, 2.8, -2.8\}$

a) Maximum Likelihood W: $\ln P(w, \log \alpha | d) = -\frac{(\vec{w} - \vec{d})^2}{2 \log \alpha^2} - \ln \sqrt{2\pi}$

$$= \frac{-(W-2.2)^2 - (W+2.2)^2 - (W-2.8)^2 - (W+2.8)^2}{Z} - \ln\sqrt{2\pi}$$

$$\frac{d \ln P(W, \log \alpha | d)}{dW} = -(W-2.2) - (W+2.2) - (W-2.8) - (W+2.8) - \ln\sqrt{2\pi}$$

$$= -4W + 10 - \ln\sqrt{2\pi}$$

$$= 0$$

$$\boxed{W = 2.27}$$

Maximum Likelihood α:

$$\frac{d \ln P(W, \log \alpha | d)}{d\alpha} = -\frac{4W + 10}{\log \alpha} - \ln\sqrt{2\pi}$$

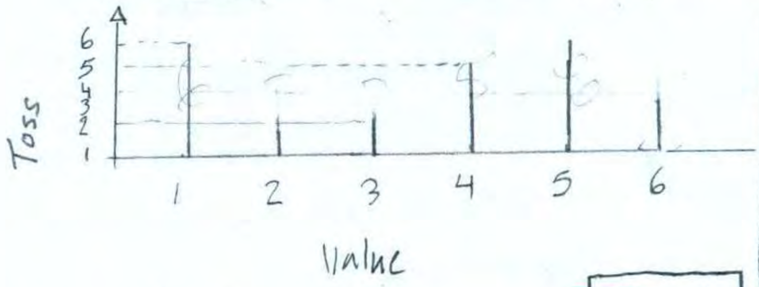
$$\boxed{\alpha = 2.718 = e}$$

Note: A common method is approximating W as a vector $\{W_1, W_2, W_3, W_4\}$, which is similar to a regression of $\ln P(w, \log \alpha | d)$.

Chapter 23: Useful Probability Distributions:

Exercise 23.1:

Toss #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
value	5	6	1	1	1	4	1	1	3	4	2	5	4	3	6	5	4	6	4	5	5	2	5	4	6	5	5	1



This distribution is a telltale example of a fit which is not Gaussian-like. Why? Because the dice has bounded values.

Exercise 23.2: $u = x^{1/3}$; Improper Distribution: $\boxed{\frac{1}{x} = \frac{1}{\sqrt[3]{u}}}$

Exercise 23.3: Dirichlet Distribution:

$$P(p|\alpha m) = \frac{1}{Z(\alpha m)} \prod_{i=1}^{I} p_i^{\alpha m_i - 1} \cdot \delta\left(\sum p_i - 1\right) \equiv \text{Dirichlet}^{(I)}(p|\alpha m)$$

$$\text{Where } Z(\alpha m) = \prod_i \Gamma(\alpha m_i)/\Gamma(\alpha)$$

Dice (p_R, p_B); Six component Probability Vector $(p_1, p_2, p_3, p_4, p_5, p_6)$

where $p_R = p_1 + p_2$ and $p_B = p_3 + p_4 + p_5 + p_6$

Hyperparameters $\{u_1, u_2, u_3, u_4, u_5, u_6\}$

$$P(p_R, p_B) = \iiiint\!\!\iint \text{Dir}(\vec{p}|\vec{u})\, \delta(p_R - (p_1 + p_2))\, \delta(p_B - (p_3 + p_4 + p_5 + p_6))\, d\vec{p}$$

$$\boxed{= P(u_1 + u_2, u_3 + u_4 + u_5 + u_6)}$$

Exercise 23.4: Gamma Distribution: $P(x|s,c) = \Gamma(x; s,c) = \frac{1}{Z}\left(\frac{x}{s}\right)^{c-1} \cdot \exp\left(-\frac{x}{s}\right)$

where $Z = \Gamma(c)\, s$

Maximum Likelihood Parameter s:

$$\frac{d \ln P(x|s,c)}{ds} = \frac{d}{ds}\left[(c-1)\ln\left(\frac{x}{s}\right) - \frac{x}{s} - \ln \Gamma(c) \cdot s\right] = 0$$

$$= (c-1)s^2 + x - s; \quad \boxed{s = \frac{1 \pm \sqrt{1 - 4(c-1)x}}{2(c-1)}}$$

Maximum Likelihood Parameter c:

$$\frac{d \ln P(x|s,c)}{dc} = \frac{d}{dc}\left[\ln\left(\frac{x}{s}\right) - \frac{\Gamma'(c)}{\Gamma(c)}\right] = \boxed{0}$$

No closed Solution — "Digamma"

Chapter 24: Exact Marginalization:

Exercise 24.1: A maximum likelihood estimate of standard deviation $[\sigma_n]$ is an approximation from a predefined model, commonly Gaussian. The true standard deviation is larger, depends on an mean that is unbiased, and requires a distribution to be true.

Exercise 24.2:

$$P(\mu|D) = \int_0^{\infty} P(\mu|\sigma^2, D) \circ P(\sigma|D)\, d\sigma$$

$$\text{where } \; P(D|\mu,\sigma) = \prod_i^n P(D_i|\mu,\sigma) \circ P(\mu,\sigma) \circ P(\sigma)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\Sigma(X_i-\mu+\bar{X}-X)^2/2\sigma^2} \circ \left(\frac{1}{\sqrt{\sigma}}\right) e^{-(\mu-\mu_0)^2/2k\sigma^2} \circ \left(\frac{1}{\sigma^2}\right)^{\alpha_0-1} e^{(-\beta_0 \tau)}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \circ \left(\frac{1}{\sigma^2}\right)^{\alpha_0+\frac{n}{2}-1} \circ e^{-(\beta_0+\frac{1}{2}\Sigma(X_i-\bar{X})^2)/\sigma^2} \circ \left(\frac{1}{\sigma}\right) e^{-[(\mu-\mu_0)^2 k + n(\bar{X}-\mu)^2]/2\sigma^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma^2}\right)^{\alpha_0+\frac{n}{2}-1} \circ e^{-(\beta_0+\frac{1}{2}\Sigma(X_i-\bar{X})^2)/\sigma^2} \circ \left(\frac{1}{\sigma}\right) e^{-(k+n)(\mu-\mu_0)^2/2\sigma^2} \circ e^{\frac{-nk}{2(n+k)\sigma^2}(\bar{X}-\mu_0)^2}$$

$$P(\sigma|D) = \left(\frac{1}{\sigma^2}\right)^{\alpha_0+\frac{n}{2}-1} \circ e^{(\beta_0+\frac{1}{2}\Sigma(Y_i-\bar{Y})^2+\frac{nk}{2(n+k)}(\bar{X}-\mu_0)^2)}$$

$$= \text{Gam}(\alpha,\beta); \text{ where } \alpha = \alpha_0 + \frac{n}{2}$$

$$\beta = \beta_0 + \frac{1}{2}\Sigma(X_i-\bar{X})^2 + \frac{nk}{2(n+k)}(\bar{X}-\mu_0)^2$$

$$P(\mu|D) = \int_0^{\infty} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \cdot \left(\frac{1}{\sigma^2}\right)^{\alpha-1} \cdot e^{-\beta/\sigma^2} \circ \sqrt{\frac{(n+k)}{2\pi\sigma^2}}\, e^{-\frac{(n+k)}{2\sigma^2}(\mu-\mu_0)^2} d(1/\sigma^2)$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \sqrt{\frac{(n+k)}{2\pi}} \int_0^{\infty} \left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{1}{2}-1} \circ e^{-\beta/\sigma^2 - \frac{(n+k)}{2\sigma^2}(\mu-\mu_0)^2} \circ d(1/\sigma^2)$$

$$\boxed{= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \sqrt{\frac{(n+k)}{2\pi}}\, \Gamma(\alpha+1/2) \left(\beta + \frac{n+k}{2}(\mu-\mu_0)^2\right)^{-\alpha-1/2}}$$

$$\boxed{= \frac{\beta^{-1/2}}{\Gamma(\alpha)} \sqrt{\frac{(n+\kappa)}{2\pi}}\, \Gamma(\alpha+1/2) \left(1 + \frac{1}{2\alpha} \frac{(\mu-\mu_0)^2}{\frac{\beta}{(n+\kappa)\alpha}}\right)^{-(2\alpha+1)/2}}$$

Normalization constant   Student t-distribution

Scale parameter: $\beta/(n+\kappa)\alpha$

Degrees of Freedom: 2α

Exercise 24.3: $P(x|\mu,\sigma) = \prod_i^n P(x|\mu,\sigma) \circ P(\mu,\sigma) \circ P(\sigma)$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma^2}\right)^{\alpha_0 + \frac{n}{2} - 1} \circ e^{-(\beta_0 + \frac{1}{2}\sum(x_i - \bar{x})^2)/\sigma^2} \circ \left(\frac{1}{\sigma}\right) \circ e^{-(\kappa+n)(\mu-\mu_0)^2/2\sigma^2} \circ e^{\frac{-n\kappa}{2(n+\kappa)\sigma^2}(\bar{x}-\mu_0)^2}$$

$$P(\sigma|x) = Gam(\alpha, \beta)\text{; Where } \alpha = \alpha_0 + \frac{n}{2}\text{; } \beta = \beta_0 + \frac{1}{2}\sum(x-\bar{x})^2 + \frac{nR}{2(n+\kappa)}(\bar{x}-\mu_0)^2$$

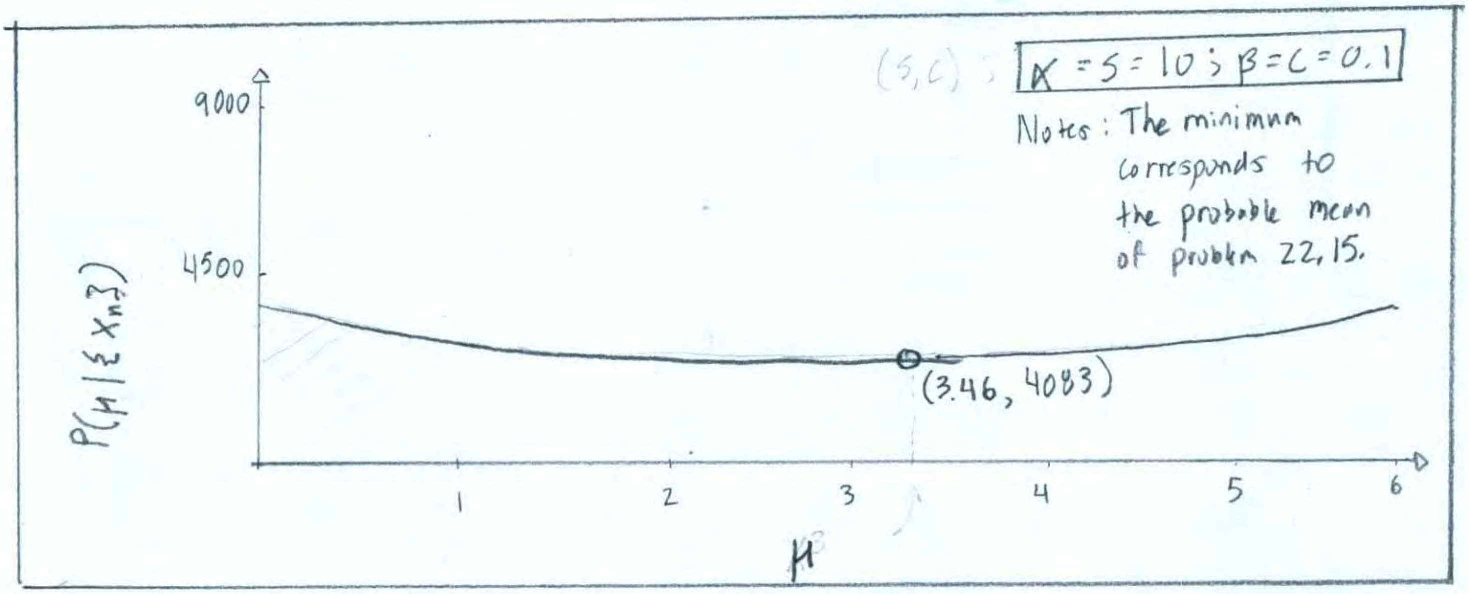
$$P(\mu|x) = \frac{\Gamma(\alpha+1/2)}{\Gamma(\alpha)} \sqrt{\frac{(n+\kappa)}{2\pi\beta}} \left(1 + \frac{1}{2\alpha} \frac{(\mu-\mu_0)^2}{\frac{\beta}{(n+\kappa)\alpha}}\right)^{-(2\alpha+1)/2} \quad \boxed{\text{From 24.2}}$$

$$\boxed{P(\mu|\{x_n\}) = \frac{\Gamma(\alpha+1/2)}{\Gamma(\alpha)} \sqrt{\frac{(n+\kappa)}{2\pi\beta}} \circ \left(1 + \frac{1}{2\alpha} \frac{\sum_i(\mu-\mu_0)^2}{\frac{\beta}{(n+\kappa)\alpha}}\right)^{-(2\alpha+1)/2}}$$

Where $\{x_n\} = \{x_A, x_B, x_C, x_D, x_E, x_F, x_G\}$

$= \{-27.020, 3.570, 8.191, 9.898, 9.603, 9.945, 10.056\}$

A plot of the students t-distribution:



t	Likelihood	Posterior Probability
0000000	0.0262	0.296
0001011	0.00041	0.0047
0010111	0.0037	0.0423
0100110	0.015	0.1691
0101101	0.00041	0.0047
0110001	0.00010	0.0012
1000101	0.015	0.1691
1001110	0.0037	0.0423
1010010	0.00041	0.1641
1011001	0.0037	0.0423
1100011	0.00010	0.0012
1101000	0.00041	0.0047
1110100	0.0037	0.0423
1111111	0.000058	0.0007

Example Calculation of $P(y_n \mid t=0000000) = \dfrac{P(y_1|t=0)\,P(y_2|t=0)\,P(y_3|t=0)\,P(y_4|t=0)\cdots P(y_7|t=0)}{P(y_1|t=1)\cdot P(y_2|t=1)\,P(y_3|t=1)\,P(y_4|t=1)\cdots P(y_7|t=1)}$

$= 0.8 \cdot 0.8 \cdot 0.1 \cdot 0.8 \cdot 0.8 \cdot 0.8 \cdot 0.8$

$\boxed{= 0.0262}$

Example Calculation of $P(t=0000000 \mid y) = \dfrac{P(t=0000000 \mid y)}{\sum P(t \mid y)}$

$\boxed{= 0.296}$

n	$P(y\mid t=1)$	$P(y\mid t=0)$	$P(t=1\mid y)$	$P(t=0\mid y)$
1	0.2	0.8	0.25	0.75
2	0.2	0.8	0.25	0.75
3	0.9	0.1	0.66	0.33
4	0.2	0.8	0.25	0.75
5	0.2	0.8	0.25	0.75
6	0.2	0.8	0.25	0.75
7	0.2	0.8	0.25	0.75

Most probable codeword of t is 0000000, but the bitwise decoding generates 0010000. This means the normal signal does not correspond to the decoder.

Chapter 25: Exact Marginalization of Trellises:

Exercise 25.1: Gaussian Channel: $\frac{P(y_n|t_n=1)}{P(y_n|t_n=0)} = \exp\left(\frac{2xy_n}{\sigma^2}\right)$

Binary Symmetric Channel: $P(y=0|x=0) = 1-f$ $\quad P(y=0|x=1) = f$

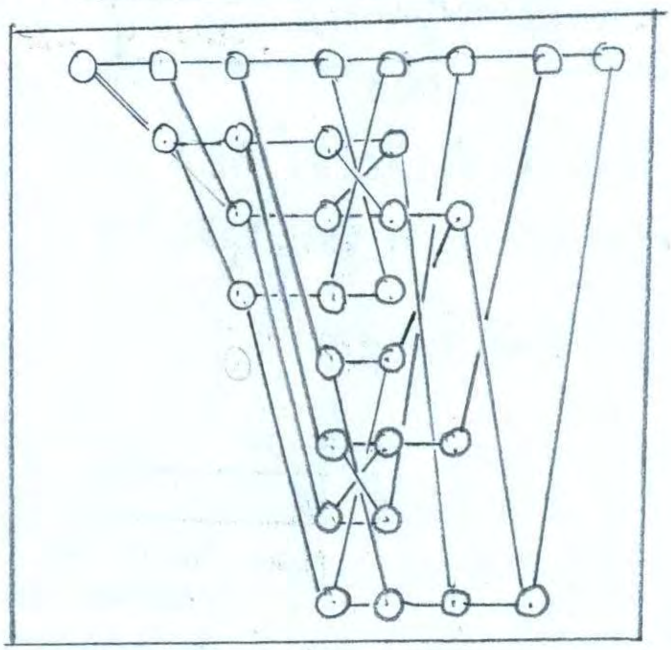
$P(y=1|x=0) = f$ $\quad P(y=1|x=1) = 1-f$

$$\frac{P(y_n|t_n=1)}{P(y_n||t_n=0)} = 1$$

A Gaussian channel is equivalent to a Binary symmetric channel because the case of either $y_n=1$ or $y_n=0$ describe to a Baysian Probability equal to one.

Exercise 25.2:

s	t	s	t	s	t	s	t
0000	0000000	0100	0100110	1000	1000101	1100	1100011
0001	0001011	0101	0101101	1001	1001100	1101	1101000
0010	0010111	0110	0110001	1010	1010010	1110	1110100
0011	0011100	0111	0111010	1011	1011001	1111	1111111



Example 25.3: The Hamming Trellis in figure 25.1c describes likelihoods over the sixteen codewords in a (7,4) code.

Exercise 25.4: Normalized Likelihood (0.2, 0.2, 0.9, 0.2, 0.2, 0.2, 0.2)

The most probable codeword is shown in the table below.

Exercise 25.5: $x_i = \sum_j^n w_i x_j = w_1 \circ x_1 + w_2 \circ x_2 + \cdots + w_n x_n$

$$= P(y_1) + P(y_2) + \cdots + P(y_n)$$

$$\boxed{= P(y_1, y_2, \ldots, y_n)}$$

Exercise 25.6: The constant of proportionality is two paths per k-nodes i.e 2^k.

Exercise 25.7: $\beta_j = \sum_{i:1}^n w_{ij} \beta_i = \sum_i^n \sum_j^n w_{ij} \beta_i$; $\frac{\beta_j}{\sum_j^n w_j} = \sum_i^n w_i \beta_i$

$$\frac{\beta_j}{\sum_j P(y_n)} = \sum_i^n w_i \beta_i$$

$$\frac{\beta_j}{P(y)} = \boxed{P(\beta_j \mid t=j)} \propto \beta_i.$$

Exercise 25.8: $\frac{1}{Z} r_n^{(t)} = \frac{r_n^{(t)}}{r_n^{(0)} + r_n^{(1)}} = \frac{r^{(t)}}{\sum_0^1 r^{(i)}} = \frac{P(t,y)}{P(y)} = \boxed{P(t \mid y)}$

Exercise 25.9: Simple Parity Code: P_3

n	$P(y_n \mid t_n)$	
	$t_n = 0$	$t_n = 1$
1	1/4	1/2
2	1/2	1/4
3	1/8	1/2

Min-Sum Algorithm: $\min\left(\sum -\log P(y_n \mid t_n)\right)$

Sum-Product Algorithm: $\sum\left(\prod P(y_n \mid t_n)\right)$

t	$t_1\ t_2\ t_3$	$\sum -\log(P(y_n \mid t_n))$
000	(1/4)(1/2)(1/8)	6.0
001	(1/4)(1/2)(1/2)	5.6
010	(1/4)(1/4)(1/8)	7.0
011	(1/4)(1/4)(1/2)	5.0
100	(1/2)(1/2)(1/8)	5.0
101	(1/2)(1/2)(1/2)	3.0
110	(1/2)(1/4)(1/8)	6.0
111	(1/2)(1/4)(1/2)	4.0

Maximum A Posteriori is $t = 101$.

The bitwise decoding problem is $\frac{1}{Z} r_n^{(t)}$ where the decoder fits each bits probability to $P(t_1) \cdot P(b_2) \cdot P(t_3) / Z$

Chapter 26: Exact Marginalization in Graphs:

Exercise 26.1: A normalized marginal is part/total., so the normalized marginal probability is $P_n(x_n) = \frac{Z_n(x_n)}{\sum Z_n(x_n)} = \boxed{\frac{P(x_n)}{\sum P(x_n)}}$.

Exercise 26.2: $P_n(x_n) = \frac{P_n(x_n)}{Z} = \boxed{\frac{f_1(x) \cdot f_2(x) f_3(x) f_4(x_1, x_2) f_5(x_2, x_3)}{\sum P_1(x_1) + P_2(x_2) + P_3(x_3)}}$

Exercise 26.3: Marginal Functional: $Z_n(x_n) = \prod_{m \in M(n)} r_{m \to n}(x_n)$

If the graph is tree-like, then $Z_n(x_n) = \prod_{m \in M(n)} r_{m \to n}(x_n)$

$$\boxed{= \prod_{m \in M(n)} \left(\prod_{n \in N(m)} q_{n \to m}(x_{n'}) \right)}$$

Exercise 26.4: More complicated marginal functionals ~~is~~ are related to sum-product algorithms of tree-like graphs because intermediate messages are generated and evaluated by the algorithm. For example, counting people in a forest without radio will arrive to complicated interaction when the beginning and end messages are not "found", but the sum-product algorithm resolves these situations, as described in a tree-like graph.

Exercise 26.5:

$$r_{m\to n}(x_n) = \sum_{x}\left(f_m(x_m)\prod_{n} q_{n\to m}(x_n)\right)$$

$$\boxed{= f_1(x_1)\cdot f_2(x_2)\, f_3(x_3)}$$

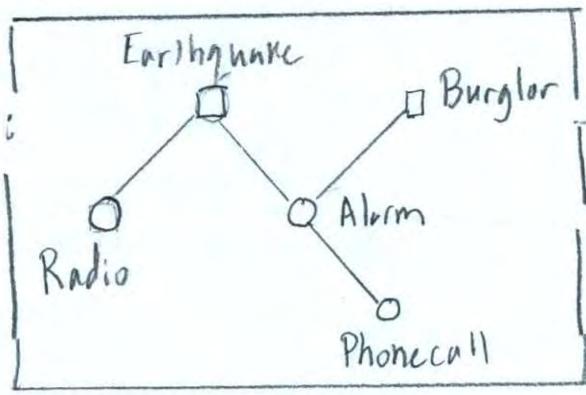
Exercise 26.6:

$$\boxed{r_{m\to n}(x_n) = \frac{f_1(x_1)\cdot f_2(x_2)\cdot f_3(x_3)}{r_{m\to 1}(x_1) + r_{m\to 2}(x_2) + r_{m\to 3}(x_3)}}$$

Exercise 26.7: $\displaystyle q_n(x_n) = \prod_{m\in M(n)} r_{m\to n}(x_n)$; $\displaystyle \phi_m(x_m) = \frac{f(x_m)}{\prod_{n\in N(m)} r_{m\to n}(x_n)}$

$$r_{m\to n} = \frac{\sum_{x}\left(\phi_m(x_m)\prod_{n} q_n(x_n)\right)}{= \sum_{x}\left(f_m(x_m)\prod_{n} q_{n\to m}(x_n)\right)}$$

Exercise 26.8:



$$\boxed{P(B)\cdot P(E)\cdot P(A|B,E)\cdot P(P|A)\cdot P(R|E)}$$

Chapter 27: Laplace's Method:

Exercise 27.1: Poisson Distribution: $P(r|\lambda) = \exp(-\lambda)\dfrac{\lambda^r}{r!}$

Improper Prior : $P(\lambda) = 1/\lambda$

a) $Z_P = \int P^*(\lambda)\,dx$; $\ln Z_P = \int \ln P^*(\lambda_0) + \frac{\partial^2}{\partial\lambda^2}\ln P^*(\lambda)\frac{(\lambda-\lambda_0)^2}{2} + \dots\, d\lambda$

Normalizing Constant of Poisson Distribution; λ^r

$$\ln Z_P = \int \ln P^*(\lambda_0) + A\int\frac{(\lambda-\lambda_0)^2}{2}d\lambda = \int \ln P^*(r|\lambda_0) + A\frac{(\lambda-\lambda_0)^3}{3} + \dots$$

$$Z_P = P^*(r|\lambda_0)\cdot\frac{A\lambda^r}{r!} = \boxed{\frac{A\lambda^r}{r!}\exp(-\lambda)}$$

b. $Z_p = \int p^*(\lambda)\,dx$; $\ln Z_p = \int \ln P^*(\ln \lambda_0)\, d\ln\lambda - A \int \log \lambda/\lambda_0 \, d\ln\lambda + \cdots$

$$= \int \ln P^*(\ln \lambda_0)\, d\ln\lambda - A_0 \frac{\lambda_0}{\lambda}\left[\ln \lambda/\lambda_0 - 1\right] + \cdots$$

$$\boxed{\ln Z_p = \ln P^*(\ln \lambda_0) - A_0 \frac{\lambda_0}{\lambda}\left[\ln \lambda/\lambda_0 - 1\right]}$$

Exercise 27.2: $Z(u_1, u_2) = \int_{-\infty}^{\infty} f(a)^{u_1} (1 - f(a))^{u_2} \, da$ where $f(a) = 1/(1+e^{-a})$

$$\ln Z(u_1, u_2) = \int_{-\infty}^{\infty} \ln f(a_0)^{u_1} \cdot (1 - f(a_0))^{u_2} \cdot da_0 + A \int_{-\infty}^{\infty} f(a_0 - a_0)^{u_1} \cdot (1 - f(a - a_0))^{u_2} \, da_0 + \cdots$$

where $A = \frac{\partial^2}{\partial a^2} \ln Z(u_1, u_2)$

$$= \ln \Gamma(u_1)\Gamma(u_2) - \ln \Gamma(u_1 + u_2)$$

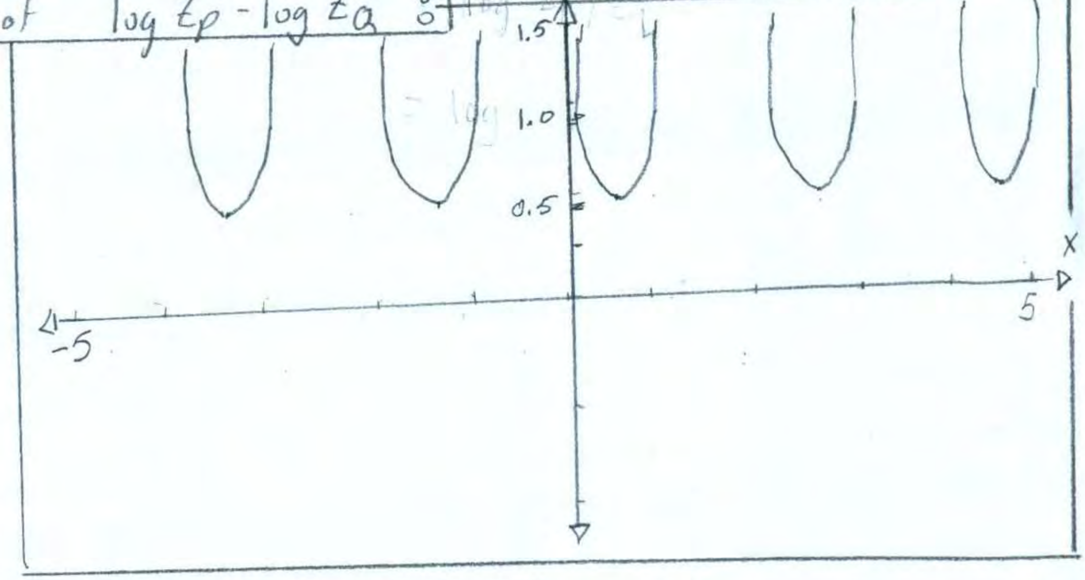
$$= \ln \frac{\Gamma(u_1)\Gamma(u_2)}{\Gamma(u_1 + u_2)}$$

$$\boxed{Z(u_1, u_2) = \frac{\Gamma(u_1)\Gamma(u_2)}{\Gamma(u_1 + u_2)}}$$

$(u_1, u_2) = (1/2, 1/2)$; $Z(1/2, 1/2) = \frac{\Gamma(1/2)\cdot\Gamma(1/2)}{\Gamma(1)} = \frac{\sqrt{\pi}\sqrt{\pi}}{1} = \boxed{\pi}$

$(u_1, u_2) = (1, 1)$; $Z(1, 1) = \frac{\Gamma(1)\cdot\Gamma(1)}{\Gamma(1)} = \boxed{1}$

The error of $\log Z_p - \log Z_a$



$\text{Error}(x) = \log((x-1)! \cdot (-x)!) - \log(1)$

Exercise 27.3: N datapoints $\{(x^{(n)}, t^{(n)})\}$; $y(x) = w_0 + w_1 x$; $t^{(n)} \sim \text{Normal}(y(x^{(n)}), \sigma_v^2)$

Assume Gaussian priors on w_0 and w_1

$$Z_p = \int N^*(y(x)|w, \sigma^2)\, dx$$

$$\ln Z_p = \int \ln N^*(y(x)|w, \sigma^2)\, dx$$

$$= \underbrace{\int \ln N^*(y(x_0)|w, \sigma^2)\, dx}_{1^{st} \text{ term Taylor exp.}} + \underbrace{\left(\frac{1}{2\sigma^2}\right) \frac{\partial^2}{\partial x^2} \ln N^*(y(x_0)|w,\sigma^2) \int (y(x-x_0) - t^{(n)})^2 dx}_{2^{nd} \text{ term Taylor exp.}} + \cdots$$

$$\ln Z_p = -\frac{1}{2}\int (y^*(x_0) - t^{(n)})^2 dx - \frac{A}{2\sigma^2}\int (y^*(x-x_0) - t^{(n)})^2 dx + \cdots$$

$$\boxed{Z_p = e^{-\frac{1}{2}\int (y^*(x_0) - t^{(n)})^2 dx - \frac{A}{2\sigma^2}\int (y^*(x-x_0) - t^{(n)})^2 dx + \cdots}}$$

An unnormalized predictive distribution for $t^{(N+1)}$ given $x^{(N+1)}$

∘ Solving for w_1:

$$\frac{d}{dw_1}(\Sigma y(x)-t) ; w_1 = \frac{2w_0 - 9}{52}$$

∘ Solving for w_0:

$$\frac{d}{dw_0}(\Sigma y(x)-t) ; w_0 = \frac{4w_1+29}{3}$$

∘ Plugging w_1 into w_0:

$$w_0 = \frac{368}{17}$$

$$w_1 = \frac{31}{48}.$$

∘ $P(X|H_2) = \frac{1}{2\pi} e^{-\frac{1}{2}[(\frac{368}{17}+\frac{31}{48}(-8)-8)^2+(\frac{368}{17}+\frac{31}{48}(-2)-10)^2+(\frac{368}{17}+\frac{31}{48}(6)-11)^2] - 1/2}$

$\cong 0.$

Note: Without normalization, the Evidence function eq. probability is non-adjusted.

Exercise 28.3: $n=30$ rolls of a six-sided dice.

\# of each face: $\vec{F} = \{3,3,2,2,9,11\}$

Dirichlet Formula: $P(p|\alpha m) = \frac{1}{Z(\alpha m)}\prod_{i=1}^{I} p_i^{\alpha m_i - 1}\cdot\delta(\Sigma p_i - 1) \equiv Dir^{(I)}(p|\alpha m)$

where $Z(\alpha m) = \prod_i \Gamma(\alpha m_i)/\Gamma(\alpha)$.

Prior Distribution: $P(0,1) = 1$

$$\frac{P(H_0|p)}{P(H_1|p)} = \frac{P(0,1)\cdot P(p|\alpha m)}{P(0,1)\cdot P(p|m)} = \frac{\Gamma(5\alpha)}{\Gamma(6\alpha)} \frac{(1/6)^{6\alpha-1}}{[(\frac{3}{30})^2(\frac{2}{30})^2(\frac{1}{30})^2(\frac{11}{30})]^{30\alpha-6}}$$

with α = choice of parameter.

Chapter 29. Monte Carlo Methods:

Exercise 29.1: $Q(x)$ = Simpler Density; $P(x)$ = Density; $\hat{\Phi}$ = "Importance" $= \dfrac{\sum_i w_r \phi(x^{(r)})}{\sum_i w_r}$

where $w_r \equiv \dfrac{P^*(x^{(r)})}{Q^*(x^{(r)})}$

$$\lim_{R\to\infty} \hat{\Phi} = \lim_{R\to\infty} \frac{\sum_i w_r \phi(x^{(r)})}{\sum_r w_r} = \phi(x^{(r)})$$

$$\lim_{R\to\infty} \sigma^2 = \lim_{R\to\infty} \int d^N x\, P(x)\left(\phi(x) - \Phi\right)^2 = 0$$

$\hat{\phi}$ is a biased estimator for small R, but an assessment of numerator and denominator show individually, unbiased estimation. In better terms, a bias is formed when a data ratio has occurred.

Exercise 29.2: A multi-modal sampler $[Q(x)]$ is a better estimate because $P^*(x)$ is multi-modal. The evolution of estimator $\hat{\phi}(x)$ at small R is more biased than large R, upon a multi-modal $P^*(x)$ function.

Exercise 29.3: Random Walk: $\langle x^2 \rangle = \sqrt{\frac{1}{T}\sum_{t=1}^{T} (x)^2} = \sqrt{T \epsilon^2}$ where $\epsilon = \frac{x}{T}$

$= \sqrt{T}\,\epsilon$

Exercise 29.4: Gibbs Sampling: $x_1^{(t+1)} \sim P(x_1 | x_2^{(t)}, x_3^{(t)}, \ldots, x_K^{(t)})$

$x_2^{(t+1)} \sim P(x_2 | x_1^{(t+1)}, x_3^{(t)}, \ldots, x_K^{(t)})$

$x_3^{(t+1)} \sim P(x_3 | x_1^{(t+1)}, x_2^{(t+1)}, \ldots, x_K^{(t)})$

A single variable-update of Gibbs sampling describes $P(x_1|x_1)$ with updates: $x_1^{(t+1)} \sim P(x_1 | x_1^{(t)})$

$x_1^{(t+2)} \sim P(x_1 | x_1^{(t+1)})$

$x_1^{(t+3)} \sim P(x_1 | x^{(t+2)})$

And a property $x^{(t+1)} = x'$, so a Metropolis algorithm when single-valued.

Exercise 29.8: Base Transitions: B_1 & B_2

Concatenation of B_1 & B_2: $T(x_1, x_2) = \int d^N x_1' \, B_2(x_2', x_1') B_1(x_1', x_2)$

Detailed Balance Property: $T(x_1, x_2) P(x_2) \neq T(x_2, x_1) P(x_1)$ not satisfied.

Exercise 29.9: Gibbs sampling satisfies the detailed balance property. Two evident methods exist, equations and graphs.

A sampling of multiple parameters would be:

$$x_1^{(t+1)} \sim P(x_1 | x_2^{(t)}, x_3^{(t)}, \ldots, x_K^{(t)})$$

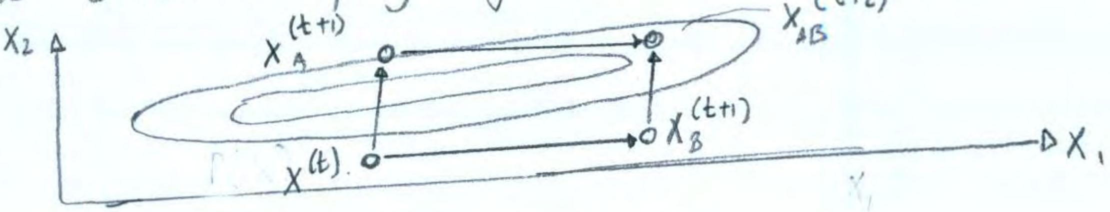
$$x_2^{(t+1)} \sim P(x_2 | x_1^{(t+1)}, x_3^{(t+1)}, \ldots, x_K^{(t)})$$

$$\vdots$$

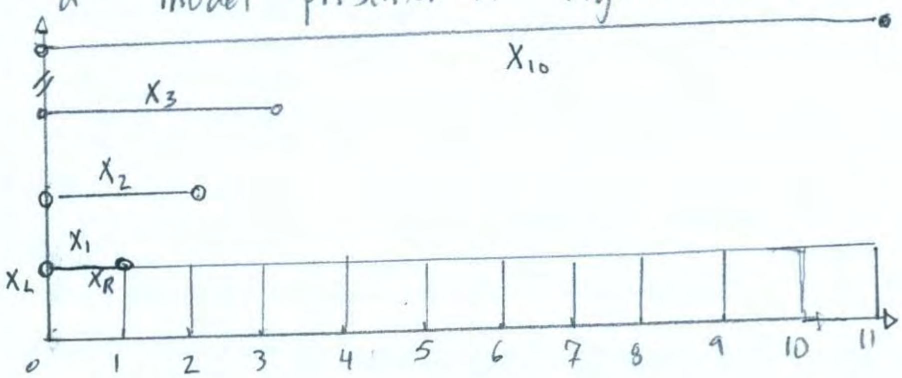
These equations fit the detailed balance property:

$$T(x_1^{(t+1)}, x_2^{(t+1)}) P(x_2^{(t+1)}, x_1^{(t+1)}) = T(x_2^{(t+1)}, x_1^{(t+1)}) P(x_1^{(t+1)}, x_2^{(t+1)})$$

Also, the Gibbs Sampling algorithm fits an invariant graph:



Exercise 29.10: The total iterations necessary is 10 for a model presented in Figure 29.17:

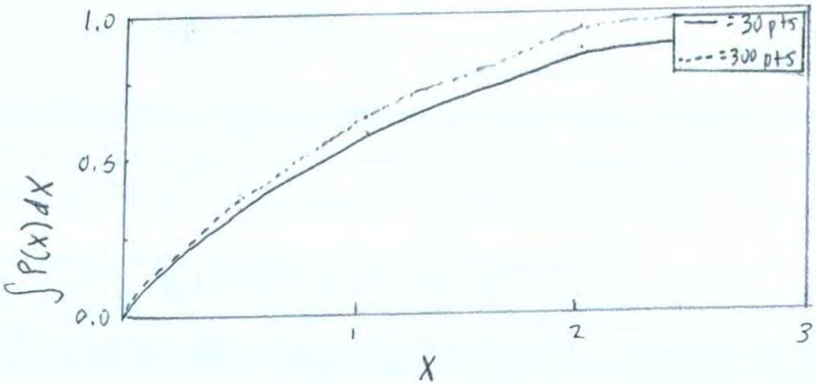


$$\lim_{x_L \to \infty} P^*(x_L) = u' \quad ; \quad \lim_{x_R \to \infty} P^*(x_R) = u'$$

$$= \left[\int dx\, \frac{Z_Q}{Z_P} \exp\left(-\frac{x^2}{2}\left(\frac{2}{\sigma_p^2} - \frac{1}{\sigma_q^2}\right)\right)\right] - 2 + 1$$

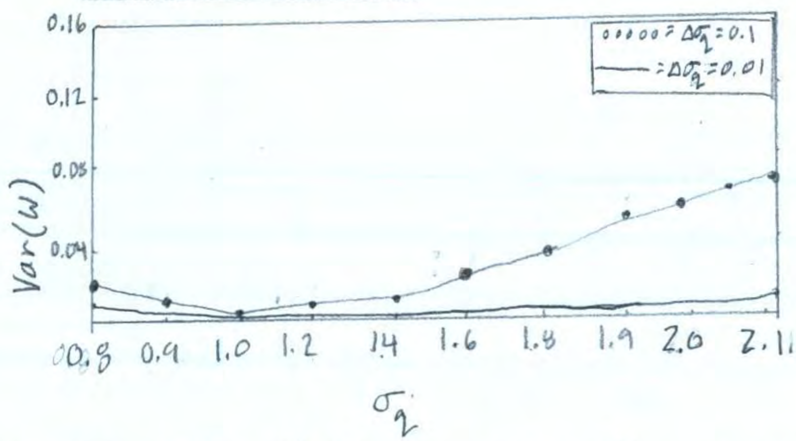
$$= \frac{\sigma_q}{\sqrt{2\pi\sigma_n^2}} \sqrt{2\pi} \left(\frac{2}{\sigma_p^2} - \frac{1}{\sigma_q^2}\right)^{-1/2} - 1 = \frac{\sigma_q^2}{\sigma_p (2\sigma_q^2 - \sigma_p^2)^{1/2}} - 1$$

Plot of Normalization Constant:



The normalization constant Z approaches a value of 1 for a Gaussian function, when sampled.

Plot of Variance:



The variance estimate for sampled weights does not converge, describing an ever increasing error when sampling. a large variance

Exercise 29.14: $P'(x) = \begin{cases} 1/19 & x' \in \{1,2,3,\dots,19\} \\ 0 & \text{otherwise.} \end{cases}$; The correct acceptance rule for Fred's proposal density is 100% or 1 because Freds "idea" to shift the sample space.

Exercise 29.15: Prior: $N(0,1)$; $P(\beta) = \Gamma(\beta; b_\beta; c_\beta) = \frac{1}{\Gamma(c_\beta)} \cdot \frac{\beta^{c_\beta - 1}}{b_\beta^{c_\beta}} \cdot \exp\left(-\frac{\beta}{b_\beta}\right)$

$$\mu^{(t+1)} \sim P(x \mid \mu^t; \sigma^2) \cdot P(x \mid 0, 1)$$

$$\sigma^{(t+1)} \sim \Gamma\left(\frac{1}{\sigma^{2,t}}; b_\beta; c_\beta\right) \cdot \Gamma\left(\frac{1}{\sigma^{2,t}}; b_\beta; c_\beta\right)$$

$$\mu^{(t+2)} \quad \vdots \quad \vdots \quad \vdots$$

$$\sigma^{(t+2)}$$

Exercise 20.20: The Monte Carlo method manages volume sampling without measuring volumes by computing a ratio between regions, adjusting lower or upper bounds.

Exercise 20.21: A Bayesian Monte Carlo Method requires an importance sampling distribution, such as: $f_p = \int \frac{f(x)\,p(x)}{q(x)} q(x)\,dx$

$$\approx \frac{1}{T} \sum \frac{f(x^{(t)})\,p(x^{(t)})}{q(x^{(t)})}$$

with an expected value of $E[f] = \int \left[\int f(x)\,p(x)\,dx\; p(f|D)\,df \right]$

$$= \sum_{i=1}^{n_f} \sum_{j=1}^{n_x} f(x)\,p(x)\,P(f|D)$$

and a variance of sampling $V[f] = \int \left[\int f(x)\,p(x)\,dx - \int f(x')\,p(x')\,dx' \right]^2 p(f|D)\,df$

$$= \sum_{i=1}^{n_f} \left[\sum_{j}^{n_x} f(x)\,p(x) - \sum_{j}^{n_{x'}} f(x')\,p(x') \right]^2$$

Chapter 30: Efficient Monte Carlo Methods:

Exercise 30.1: If $X_i^{(t+1)} = \mu + \alpha(X_i^{(t)} - \mu) + (1-\alpha^2)^{1/2} \sigma v$

then $X_i^{(t+2)} = \mu + \alpha(X^{(t+1)} - \mu) + (1-\alpha^2)^{1/2} \sigma v$

$$= \mu + \alpha^2 (X^{(t)} - \mu) + (\alpha + 1)(1-\alpha^2)^{1/2} \sigma v$$

Exercise 30.2: If $P(x,y) = \frac{\exp(-x^2y^2 - x^2 - y^2)}{Z}$, then random overrelaxation Gibbs Sampling would require abundant time to satisfy a solution. The assertion to randomize to order may also generate a regional variance to the overrelaxed samples.

Then Gibbs sampling could leak bits per iteration by sampling a value above 1/2 in a Gaussian function. These odds are: $1 - \frac{P(1/2|0,1)}{\sum P(x|0,1)} = 30.85\%$

Exercise 30.7: Hamiltonian Monte Carlo:

$$P(x) = \frac{e^{-E(x)}}{Z} \quad \text{-or-} \quad P(x,p) = \frac{e^{-E(x)} \cdot e^{-K(p)}}{Z_H}$$

The algorithm extracts 1 bit per information when $1/P(x) \geq 2$. Thus, the Hamiltonian Monte Carlo samples $\left[e^{E(x)} \cdot e^{K(p)} \cdot Z_H\right]$ bits, which a function formed as can evaluate greater than 2.

Exercise 30.8: Importance Sampling: $w_r = P^*(x^{(r)})/Q^*(x^{(r)})$

dumb Metropolis Sampling: $a = P^*(x')/P^*(x)$

In a modern 64-bit computer, floating point arithmetic is faster than integer arithmetic. The add/subtract, in addition to, multiply/divide are better with a floating point. A 32-bit computer was a 50:50 split for choosing floats or integers in computation.

Exercise 30.9: Crossover Operator: $\frac{P^*(x')\, P^*(y')}{P(x)\, P^*(y)}$

Individual Variation: $x^{(r)} \rightarrow x^{(r)\prime}$

The crossover operation is more efficient when $p^*(x') \cdot p^*(y)$ does not factor or synthesize to a singular probability. Otherwise crossover is equally as efficient, such as for a Gaussian.

Chapter 31: Ising Models

Exercise 30.1: Entropy : $S = \sum p(x)[\ln 1/p(x)]$

where $p(x) = \frac{1}{Z(\beta)} \exp[-\beta E(x)]$

a) $S = \sum p(x) \ln[1/p(x)] = \sum \frac{e^{-\beta E(x)}}{Z(\beta)} \ln[Z(\beta) e^{\beta E(x)}]$

$= \left(\frac{\sum e^{-\beta E(x)}}{Z(\beta)}\right) \ln Z(\beta) + \left(\frac{\sum e^{-\beta(x)}}{Z(\beta)}\right) \ln e^{\beta E(x)}$

$= \ln Z(\beta) + \beta E(x)$

b) $S = -\frac{\partial F}{\partial T} = -\frac{\partial}{\partial T}[-kT \ln Z] = k \ln Z$ "Boltzmann's Equation"

Exercise 31.2: A Monte Carlo simulation has an entropy. The entropy description is $S = \sum p(x)[\ln 1/p(x)]$ for the probability $p(x)$ sampled by Monte Carlo. Heat Capacity is obtained by $C_v = \frac{\partial}{\partial \beta}\left[\frac{S - \ln Z(\beta)}{\beta}\right]$

$= -\frac{S}{\beta^2} - \frac{E\beta + \ln Z(\beta)}{\beta^2}$

$= -\frac{2[S + \ln Z(\beta)]}{\beta^2}$

where $S = \sum p(x)[\ln 1/p(x)]$

$\ln Z(\beta) = \ln \sum \exp(-\beta E(x))$

Exercise 31.3: Spins [J] effect energy [E] by the function $E = \sum_{(m,n)} J \cdot x_m x_n$, and an entirely "low spin" state is proposed by assigning $J = \begin{cases} 1 & x_m x_n > 0 \\ -1 & x_m x_n < 0 \end{cases}$

Chapter 32: Exact Monte Carlo Sampling:

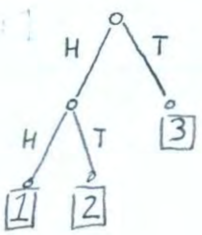
Exercise 32.1: No relationship exists between the equilibriation time and the time taken for the trajectories to coalesce. A proof sampling one integer in time is not equivalent to the time to coalesce.

Exercise 32.2: If Fred increases T_0 each run, then the sample would equilibrate to coalescence, differently. The driven shifts of T_0 may generate sample coalescence at different times $(T-T_0)$.

Exercise 32.3: (Holmes and Mallick [1998]) A perfect sampling is generalized by "Intersection-Related Accidents in Texas," in that, linear regression fits a smoothed relative risk, while perfect sampling choppy data.

Exercise 32.4: Surname growth function $P_i = P_{0i} e^{r_i t}$ where P_{0i} is the initial number of members, and r_i is the rate of growth per surname, would arrive to a single surname when $\lim_{t \to \infty} \frac{P_i}{\sum_n P_{i,n}}$ is 1. The time required is dependent upon each surnames rate, and initial population.

Exercise 32.5: Sample Space: 1, 2, 3
HH HT T



Flip #1

Flip #2

Exercise 32.6: A scientist comprehends the time to coalesce conveys information about Z because a nonconverging series requires a lot of time. Generally, time is related to the precision of Z for a Gaussian function, including the opportunity to utilize Z again for another ensemble

Exercise 33.4: $\tilde{F}(Q_K, Q_\theta) = \sum_i \int d^D\theta\, Q_K(k) Q_\theta(\theta) \ln \dfrac{Q_K(k)\, Q_\theta(\theta)}{P(X, \theta | H)}$

$$= \sum_i \int d^D\theta\, Q_K(k) Q_\mu(\mu) Q_\sigma(\sigma) Q_\pi(\pi) \ln \frac{Q_K(k) Q_\mu(\mu) Q_\sigma(\sigma) Q_\pi(\pi)}{P(X, \theta | H)}$$

$$\frac{d\tilde{F}(Q_K, Q_\theta)}{dQ_K} = \sum_i \int d^D\theta\, Q_K(k) Q_\mu(\mu) Q_\sigma(\sigma) Q_\pi(\pi) \ln \frac{Q_K(k) Q_\mu(\mu) Q_\sigma(\sigma) Q_\pi(\pi)}{P(X, \theta | H)} = 0$$

$$Q_K(k) = \frac{P(X, \theta | H)}{Q_K(\,) Q_\mu(\mu) Q_\sigma(\sigma) Q_\pi(\pi))}$$

$$= \frac{r_K}{Q_\mu(\mu) Q_\sigma(\sigma) Q_\pi(\pi)}$$

Update Step of Soft k-means: $\theta = \dfrac{\sum r_K X}{\sum r_K}$

(Mackay, 2001) listed plots similar to Exercise 33.3 and 33.4.

Exercise 33.5: $P(\vec{x}) = P(X_1, X_2 | \sigma_1^2, \sigma_2^2)$; $e^{(1)} = (1,1)$; $e^{(2)} = (1,-1)$

Objective Function: $G = \int dx\, P(x) \ln \dfrac{P(x)}{Q(x, \sigma^2)}$

$$= \int dx\, P(x|0, \sigma_q) \ln \frac{P(x|0, \sigma_q)}{Q(x|0, \sigma_p)}$$

$$= \int dx\, P(x|0, \sigma_q) \ln \left[\ln \frac{\sigma_p}{\sigma_Q} - \frac{1}{2} x^2 \left(\frac{1}{\sigma_Q^2} - \frac{1}{\sigma_p^2} \right) \right]$$

$$= \frac{1}{2} \left(\ln \frac{\sigma_p^2}{\sigma_Q^2} - 1 + \frac{\sigma_Q^2}{\sigma_p^2} \right)$$

Since F is a seperable function in σ_1^2 and σ_2^2, then

$$F = \frac{1}{2} \left(\ln \frac{\sigma_1^2}{\sigma_2^2} - 1 + \frac{\sigma_Q^2}{\sigma_1^2} + \ln \frac{\sigma_2^2}{\sigma_Q^2} - 1 + \frac{\sigma_Q^2}{Q_2^2} \right)$$

Gibbs Inequality: the relative distribution differences is indicated by a threshold (≥ 0).

Joint Distribution: $P(X, y)$

Seperable Distribution: $Q(X, y) = Q_x(X)\, Q_y(y)$

Objective Function: $G(Q_x, Q_y) = \sum P(X, y) \log_2 \dfrac{P(X, y)}{Q_x(X)\, Q_y(Y)}$

$G(Q_x, Q_y) = 0$; when $Q_x(X)\, Q_Y(Y) = P(X, y)$

Alternate Objective function: $F(Q_x, Q_y) = \sum Q_x(X) Q_Y(Y) \log_2 \dfrac{Q_x(X)\, Q_Y(Y)}{P(X, y)}$

$P(X, y)$

y \ X	1	2	3	4	$P(y)$
1	1/8	1/8	0	0	1/4
2	1/8	1/8	0	0	1/4
3	0	0	1/4	0	1/4
4	0	0	0	1/4	1/4
$P(X)$	1/4	1/4	1/4	1/4	1

$F(Q_x, Q_y)$

y \ X	1	2	3	4
1	-1/16	-1/16	∞	∞
2	-1/16	-1/16	∞	∞
3	∞	∞	-1/8	0
4	∞	∞	∞	-1/8

The three minima in $F(Q_x, Q_y)$ are -1/16, -1/8, and -1/8, as seen in the right table.

Chapter 34: Independent Component Analysis and Latent Variable Modelling:

Exercise 34.1: If $X_j^{(n)}$ is $X_j^{(n)} = Gs + n$, then

$$P(\{X^{(n)}, s^{(n)}\}_{n=1}^{N} \mid G, H) = \prod_{n=1}^{N}\left[\left(\prod_j \delta\left(X_j^{(n)} - \sum G_{ji} s_i^{(n)}\right)\right)\left(\prod_i p_i(s_i^{(n)})\right)\right]$$

$$= \prod_{n=1}^{N}\left[\left(\prod_j \delta\left(Gs + n - \sum G_{ji} s_i^{(n)}\right)\right)\left(\prod p_i(s_i^{(n)})\right)\right]$$

$$P(X^{(n)} \mid G, H) = \int d^I s^{(n)}\, P(X^{(n)} \mid s^{(n)}, G, H)\, P(s^{(n)} \mid H)$$

$$= \int d^I s^{(n)} \prod_j \delta\left(Gs + n - \sum G_{ji} s_i^{(n)}\right) \prod_i p_i(s_i^{(n)})$$

$$x = V'(D^{-1})^{0.5} V \cdot x'$$

③ $W = I(n,n)$ //Unmixing Matrix

$u = x \cdot W$ // Estimated Source Signals

$r = corr([u\ s])$ //Correlation between estimated source u and signal s.

maxiter = 10000;

eta = 1;

hs(maxiter): //Values of function

gs(maxiter); //Gradient Magnitude

④
```
for iter = 1:maxiter
    u = xW;   //Estimated signal
    U = tanh(u)  // Estimate max entropy
    detW = abs(det(W))  //Solve W, find h
    h = ((1/N) · sum(sum(U)) + 0.5 · log(detW)
    g = inv(W') - (2/N) · x' · U;  //Find matrix of gradient
    W = W + eta · g  //Update W
    hs(iter) = h  //Record h.
    gs(iter) = norm  //Record magnitude of gradient
end
```

Requirements: ∘ Statistically independent underlying signals

∘ Non-Gaussian

∘ Zero mean and fixed variance.

Exercise 34.3: a) An altered ICA algorithm to handle substantial Gaussian Noise requires an additive term in Ax to become $Ax + n$.

b.) A modified ICA algorithm to measure greater amounts of latent variables is a larger S-matrix

The algorithm projects a Dirichlet$(\pi_1, \ldots, \pi_K \mid \alpha)$, but Rasmussen denotes an intractable integral. A better method is, not Gibbs Sampling, but a Gibbs sweeps to simulate n - components, in a single step. (e.g. sum regions of the dataset, repetitively)

Again, the method to Baysian inference for an infinite Gaussian mixture is summing regions repetitely, using Gibbs sweeps.

Chapter 35: Random Inference Topics:

Example 35.1: Does the leading coefficient of a natural number correspond to another natural number? No, because of invariance per unit.

Exercise 35.2: Uniform Distribution : $P(a,b) = \frac{1}{b-a}$

Part 1 : $P(0, 180^\circ) = \frac{1}{180}$

Part 2 : $P(a_1, a_2, b_1, b_2) = P(a_1, b_1)\,P(a_2, b_2)$

$$P(0, 180, 0, 180) = \frac{1}{180} \cdot \frac{1}{180} = \frac{1}{324}$$

Exercise 35.3: Predictions about successive dates:

① $P(X) > 0$

② ~~P(X) corresponds to interval of sports season~~

③ ~~P(X) corresponds to a person was recording~~

④ ~~P(X) relates to time of activity or length of game.~~

— = False assumptions.

A multivariate Beta Distribution is the Dirichlet:

$$P(X, \alpha) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$$

Hypothesis Likelihood: $$\frac{P(X, \alpha_{A\to B})}{P(X, \alpha_{B\to A})} = \frac{\Gamma(\alpha_0)}{\prod_i \Gamma(\alpha m_i)} \cancel{\prod_{i=1}^{K} x^{\alpha_i - 1}} \cdot \frac{\prod_i \Gamma(F_i + \alpha)}{\Gamma(F_i + \alpha)} \cdot \frac{1}{\cancel{\prod_i x^{\alpha_i - 1}}}$$

$$= \frac{\Gamma(\alpha_0)}{\prod_i \Gamma(\alpha m_i)} \cdot \frac{\prod \Gamma(F_i + \alpha)}{\Gamma(F_i + \alpha)}$$

$$= \frac{(765+1)(235+1)}{(950+1)(50+1)} = 3.8$$

Exercise 35.6: Poisson Distribution:

$$P(r, \lambda) = \frac{\lambda^r}{r!} e^{-\lambda}$$

$\lambda(t) = \exp(a + b\sin(\omega t + \phi))$; $t = 0 \ldots T$; N = photons

Inference of a, b, ω, and ϕ

$$P(D \mid N, \lambda) = \left[\prod^{N} \lambda(t)\right] \exp\left[-\sum \lambda(t)\right] \cdot \Delta t^{N}$$

Exercise 35.7: If "_" is the tenth digit,

then $891.10.0 \rightarrow 891.\_10.0$

$P(x) = P(891.10.0) \cdot P(4^{th}\text{ digit "\_"})$

$= \left(\frac{1}{43}\right) \cdot \left(\frac{1}{10}\right)^3 = 2.33 \times 10^{-5} = 0.002\%$

then $891.10.0 \rightarrow 891.1\_0.0$

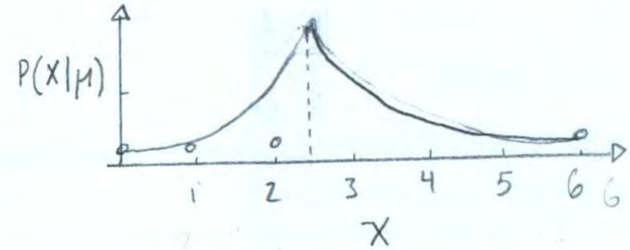
$P(\lambda) = P(891.10.0) \cdot P(4^{th}\text{ digit "\_"})$

$= \left(\frac{1}{43}\right)\left(\frac{1}{10}\right)^4 = 2.33 \times 10^{-6} = 0.0002\%$

Exercise 35.8: Bi-exponential Distribution: $P(X|\mu) = \frac{1}{Z} \exp(-|X-\mu|)$

where $Z = 2$.

Assuming $\{X_n\} = \{0, 0.9, 2, 6\}$, then $\mu = 2.23 \pm 1.96 \cdot Z/2$



Chapter 36: Decision Theory

Exercise 36.1: $P(d_n) = \text{Normal}(d_n; \mu_n, \sigma^2 + \sigma_n^2)$ - Prior

$P(X_n|d_n) = \text{Normal}(d_n; \mu_n', \sigma_n^{2'})$ - Posterior

where $\mu_n' = \dfrac{d_n/\sigma^2 + \mu_n/\sigma_n^2}{1/\sigma^2 + 1/\sigma_n^2}$ and $\dfrac{1}{\sigma_n^{2'}} = \dfrac{1}{\sigma^2} + \dfrac{1}{\sigma_n^2}$

Expected Utility with no prospecting $= \displaystyle\int P(X|d_n) \cdot P(d_n)\, dd_n$

$$= \frac{1}{2\pi} \cdot \frac{1}{\sqrt{(\sigma^2+\sigma_n^2)\cdot\sigma_n^{2'}}} \int e^{-\frac{(X_n-\mu_n')^2}{2\sigma_n'^2}} \cdot e^{-\frac{(d_n-\mu_n')^2}{2(\sigma^2+\sigma_n^2)}}\, dd_n$$

$$= \frac{\sigma^2 \cdot \sigma_n^2}{2\pi} \sqrt{\frac{\pi}{2}} \frac{\sigma_n' \sqrt{(\sigma^2+\sigma_n^2)}\, e^{-\frac{(\mu_n'-\mu_n)^2}{2(\sigma_n'^2+\sigma^2+\sigma_n^2)}}}{\sqrt{\sigma_n'^2+\sigma^2+\sigma_n^2}}$$

$$\times \operatorname{erf}\left(\frac{-\mu_n'(\sigma^2+\sigma_n^2) - \mu_n(\sigma_n^{2'}) + X(2\sigma_n'^2+\sigma^2)}{\sqrt{2\sigma_n^{2'}(\sigma^2+\sigma_n^{2'})(\sigma_n'^2+\sigma^2+\sigma_n^{2'})}}\right)$$

Exercise 36.2: $n_a = \underset{n}{\operatorname{argmax}}\ \mu_n = E[U|n]$

$E[U|\text{optimal } n] = \max_n \mu_n$, Exercise 36.3: $\frac{1}{\sigma_n^{2\prime}} = \frac{1}{\sigma^2} + \frac{1}{\sigma_n^2} = \frac{\sigma_n^2 + \sigma^2}{\sigma_n^2 \sigma^2}$; $s^2 = \frac{\sigma_n^2 \sigma^2}{\sigma_n^2 + \sigma^2}$

Exercise 36.4: Three Doors: 1, 2, and 3; Select #1 - 1 of 3 doors; Select.

Select #2 - 1 of 2 doors; Re-Select

$\boxed{\text{Monte-Hall Problem}}$

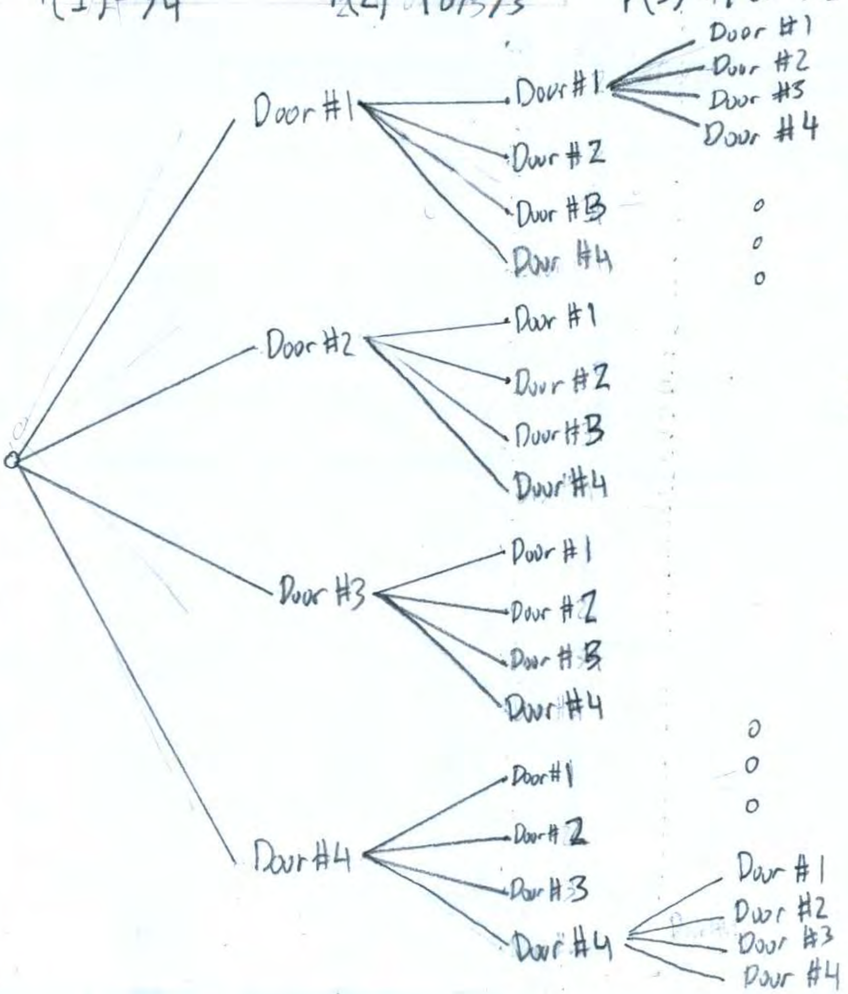
$\boxed{\text{Modified Monte-Hall Problem}}$

Four Doors: 1, 2, 3 and 4. Select #1 - 1 of 4 doors; Select

Select #2 - 1 of 3 doors; Re-select

Select #3 - 1 of 2 doors; Re-select

$P(1) = 1/4$ $P(2) = 1$ or $1/3$ $P(3) = 1$ or $1/2$



$P(\text{Total}) = P(1) \cdot P(2) \cdot P(3)$

Best odds to switch.

Exercise 36.5: A Utility Function: $E[U|a] = \int d^K x \; U(x, a) P(x|a)$

Choice (or preference): $1 = \sum P(x) = P(x_1) + P(x_2) + \dots + P(x_n)$

Proof by Contrapositive: Choice $\Rightarrow$ Utility Function

Choice $\Rightarrow \neg$ Utility Function

Exercise 36.6: N = Partners; d_n = desirability;

a, b, c.) $E[U|a] = \int d^k x \cdot U(x, a) \cdot P(x|a) = \sum_{i=1}^{N} U(x,a) \cdot P(x|a)$

$$E[U \mid \text{New Partners}] = P(d_n < d)\, d + \sum_{d}^{N} U(x, a) \cdot P(x|a)$$

$$E[U \mid \text{No New Partners}] = \begin{cases} -d & \text{if } d \geq d_n \\ -d_n & \text{if } d \leq d_n \end{cases}$$

$$E[U \mid \text{New Partner}] - E[U \mid \text{No New Partners}] = \begin{cases} \sum_{d}^{N} U(x,a)\, P(x|a) & \text{if } d \geq d_n \\ \sum_{N}^{d} U(x,a)\, P(x|a) & \text{if } d \leq d_n \end{cases}$$

The strategy is to compare the sum of past desirabilities to the current desirability value. Is the mean desirability rising?

Exercise 36.7: Philosophy justifies the regret table. The animal could regret choice when taunted, or offered by another animal, so regret exists as a semi-quantitative topic.

Freds Utility Table:

	Action	
	Lottery	Cash
No Win	$-£f_1$	$£(1-f_1)$
Win	$£8f_1$	$£(1-f_1)$

Fred's Regret Table:

	Action	
	Lottery	Cash
No Win	1	0
Win	0	8

The investments choice is lottery.

A minimax regret community will invest into the lowest net regret, which is a high risk lottery fund. The investment justification is lowest regret.

	Action	
Outcome	Buy	Don't Buy
No Wins	9/10	0
Wins	0	7.2

Exercise 36.8: I = # horses ; t = time of bet; $m(t)$ = money ; b = probability vector.

If horse i wins, then the return is $m(t+1) = b_i o_i m(t)$

also the bookies odds $\sum_i \frac{1}{o_i} = 1$

and horse odds p_i

Optimal Betting Strategy:

	Action	
	Bet	Don't Bet
No Wins	$-m(t)$	0
Wins	$(b_i o_i - 1)m(t)$	0

	Action	
	Bet	Don't Bet
No Wins	$m(t)$	0
Wins	0	$(b_i o_i - 1)m(t)$

A minimax regret community would bet money a time.

Covers betting strategy: Return is a Lagrangian.

$$= \sum p_i \log(m(t)) + \lambda \sum b_i$$

$$= \sum p_i \log(b_i o_i m(t)) + \lambda \sum b_i$$

Maximum Expected Return:

$$\frac{d\,\text{Return}}{d b_i} = \frac{p_i}{b_i} + \lambda = 0 \; ; \; \lambda = -\frac{p_i}{b_i} = -1$$

$$; \sum b_i = 1$$

Optimal Betting Strategy is $b_i = p_i\, m(t)$

Growth of Money: $\ln E[x] = \sum W(b,p)$

$$E[x] = 2^{nW(b,p)}$$

If a single bet occurs, then the expectation is $2^{W(b,p)}$.

Cover's aimed at a clear solution.

Exercise 36.9: Dice Sample Space → Number 7 is the largest likelihood.

No, I would not accept the bet because of a 2.75% chance number 7 first appears as the sum.

I would not accept the second bet, either.

The guys third bet sucks too.

Chapter 37: Baysian Inference and Sampling Theory

Exercise 37.1: Bayesian Inference: $P(p_{A+}, p_{B+} | \{F_i\}) = \dfrac{P(\{F_i\} | p_{A+}, p_{B+}) P(p_{A+}, p_B)}{P(\{F_i\})}$

$$= \sum_{1}^{12} \binom{n}{k} p_A^{k} (1-p_A)^{n-k}$$

$P(p_{A+} < p_B | n=12) = 0.03$

H_0: Null Hypothesis = p_A is not biased

H_1: Alternative Hypothesis = p_B is biased.

A significance level (95%, p-value = 0.05) suggests a non-biased coin.

Note: The model follows the books convention.

Exercise 37.2: Poisson Distribution: $P(r, \lambda) = \frac{\lambda^r}{r!} e^{-\lambda r}$

Friday	Saturday
12 vehicles/hr	18 vehicles/hr
λ_F	λ_S

a. $\lambda_S > \lambda_F$

b. 1.5.

Exercise 37.3:

	Treatment	
Carrier	A	B
No	29	7
Yes	1	3

a. The probability that treatment A is more effective than treatment B:

```
#include <iostream>
#include <cmath>
int main(){
    int FAp, fAm, fBp, fBm , FAp, FAm, FBp, FBm, NA, NB, fp, fm;
    fAp = 1 ;
    fAm = 29;
    fBp = 3 ;
    fBm = 7;
    NA = fAp + fAm;
    NB = FBp + fBm;
    fp = (FAp + fBp) | (NA + NB);
    fm = (FAm + fBm) | (NA + NB);
    FAp = fp * NA;
    FAm = fp * NA;
    FBp = fp * NB;
    FBm = fp * NB;
```

Doub Chisq = $\frac{pow(FAp - FAp, 2)}{FAp} + \frac{pow(fAm - FAm, 2)}{FAm} + \ldots$

$\frac{pow(fBm - FAm, 2)}{FAm} + \frac{pow(fAp - FAp, 2)}{FAp}$

```
if (Chisq < 3.84)
    cout << "A significant difference exists between treatment A
                                    and treatment B" << endl.

cout << "The effectiveness of treatment A vs treatment B is: "/
```

$$<< \left[\frac{FAP}{N_A+N_B}\right]\left[1-\frac{FBP}{NA+NB}\right] / 2\left[\left[\frac{FAP}{N_A+N_B}\right]-\left[\frac{FAP}{N_A+N_B}\right]^2 / 2\right] << endl;$$

```
cout << "The probability P(pA+ < 10pB+) is: " << endl
Doub PAPB = 0;
for(int i=0; i<40; i++)
```

$$PAP_B += \frac{fact(4)}{fact(i)\,fact(4-i)} \circ \left(\frac{1}{30}\right)^4$$

```
cout << PAPB << endl;
cout << "The probability P(pB+ < 10pA+) is: " << endl
PAPB = 0;
for(int i=0; i<40; i++)
```

$$PAPB += \frac{fact(4)}{fact(i)\,fact(4-i)} \left(\frac{3}{10}\right)^4$$

```
cout << PAPB << endl;
```

Chapter 39: The single neuron as a classifier

Exercise 39.1: The function $y(x;w) = 1/(1+e^{-w \cdot x})$ is a logistic or sigmoid seen in Chapter 22: Maximum Likelihood and clustering.

Exercise 39.2: $G(w) = -\sum\left[t^{(n)} \ln y(x^{(n)};w) + (1-t^{(n)}) \ln(1-y(x^{(n)};w))\right]$

$$\frac{\partial G(w)}{\partial w_j} = -\sum \frac{t^{(n)} x_j^{(n)}}{y(x^{(n)};w)} - \frac{(1-t^{(n)})x_j^{(n)}}{1-y(x^{(n)};w)} = -\sum t^{(n)} - t^{(n)} y(x^{(n)};w) - y(x^{(n)};w) + t^{(n)} y(x^{(n)}_j;w)$$

$$= -\sum_{n=1}^{N} -(t^{(n)} - y^{(n)})\, x_j^{(n)} = 0$$

Exercise 39.3: $M(W) = G(W) + \alpha E_W(W)$; $E_W(W) = \frac{1}{2}\sum W_i^2$; $G(W) = -\sum[t^{(n)} \ln y(x;w) + (1-t^{(n)})\ln(1-y(x;w))]$

$$\frac{dM(W)}{dw} = \sum_{n=1}^{N} -(t^{(n)} - y^{(n)}) X_j^{(n)} + \alpha \sum_{i=1}^{N} W_i$$

The hyperparameter α is a 'weight decay' regulizer because the value shifts the objective function by a proportional amount from $[0 - \infty]$.

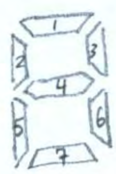
Exercise 39.4: Gaussian Distribution: $P(X|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$

$$P(z|s) = \prod_{i=1}^{I} \text{Normal}(z;0,\sigma^2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-(X_1-\mu)^2/2\sigma_1^2 - (X_2-\mu)^2/2\sigma_1^2}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\sum(X_i-\mu)^2/2\sigma_i^2}$$

$$P(s|z) = \frac{P(z|s)P(s)}{\sum P(z|s)} = \frac{\frac{1}{2\pi\cancel{\sigma_x^2}} e^{-W^TX+\theta}}{\frac{1}{\sqrt{2\pi\cancel{\sigma_x}}}\left[e^0 + e^{-WX+\theta}\right]}$$

$$= \sqrt{2\pi}\left[\frac{1}{1+e^{-WX+\theta}}\right] \text{ where } -\sum(X_i-\mu)^2/2\sigma_i^2 = -WX$$

Exercise 39.5:  $C(2) =$  $C(3) =$  $C(8) =$ 

a) Probability of s given X is $2/2^7 = 2^{-6}$

$$P(s=2|X) = \frac{1}{1+e^{-W^TX+\theta}} = \frac{e^{-W^TX+\theta}}{1+e^{-W^TX+\theta}} = \frac{e^{-W^TX+\theta}}{\sum e^{-W^TX+\theta}} = 0.1$$

$$e^{-W^TX+\theta} = 9 ; -W^TX+\theta = \ln 9$$

$$W = \frac{+\ln 1/9 - \theta}{X}$$

b. $P(s|x) = 1/10 = \frac{e^{a_s}}{\sum e^{a_{s'}}}$ An alphabet less susceptible to confusion must consider probability of activation.

Exercise 39.6: (3,1) error-correcting code: $x^{(1)} = (1,0,0)$; $x^{(2)} = (0,0,1)$; $\{P_1, P_2\}$

f = Noise level ; r = recieved vector ;

$$P(s|r) = \frac{p(r|s)\,p(s)}{p(r)}$$

$$= \frac{e^{-w_0 - \sum w_n r_n}}{\sum e^{-w_0 - \sum w_n r_n}}$$

$$= \frac{1}{1 + \exp^{-w_0 - \sum w_n r_n}}$$

Coefficient w_0 : $\frac{dz_i}{dw_0} = -1$

Coefficient w_1 : $\frac{dz_i}{dw_1} = -r_1$

Coefficient w_2 : $\frac{dz_i}{dw_2} = -r_2$

Coefficient w_3 : $\frac{dz_i}{dw_3} = -r_3$

The probability $P(s|r)$ is the activation function of a nueron from 0 to 1.

Chapter 40: Capacity of a Nueron

Exercise 40.1 : $\sum_{k=0}^{N} \binom{N}{k}$ is the binomial coefficient derived by the ratio of permutations.

Exercise 40.2 : $\sum_{k=0}^{N} \binom{N}{k} = 2^N$

Exercise 40.3: $P(\text{Single Hemisphere}) = P(1)\,P(2)\,P(3) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$

Exercise 40.4: $2N$ points split by a straight line $w^T x$ in general positions is not impossible when $K \approx N$, nor for points not in general positions.

The total number of regions obtained from $2N$ points is $2T(2N-1, K-1) = 2\sum_{k=0}^{K-1}\binom{2N-1}{k}$

$$= 2^{2N} - 2\binom{2N-1}{K} {}_2F_1(1, K-2N+1; K+1; -1)$$

where ${}_2F_1$ is the hypergeometric function.

Exercise 40.5: Four points on a single hemisphere relates to the probability of an upper and lower hemisphere per point,

$$P(\text{single hemisphere}) = P(1)\cdot P(2)\cdot P(3)\cdot P(4) = \frac{1}{2^4} = \frac{1}{16}$$

Exercise 40.6: $K=3$ dimensions; $\{x\} = \{(1,0,0), (0,1,0), (0,0,1), (1,1,1)\}$

a. $\{t\} = \{1,1,1,1\}$; $W = (1, 1, 1, \frac{1}{3})$

b. $\{t\} = \{1,1,0,0\}$; $W = (1, 1, 0, 0)$

c. An unrealizable $\{t\}$; $\{t\} \propto Re(WX) + Im(WX)$

Exercise 40.7: Three lines in a plane create seven regions.

Exercise 40.8: The total sensory information in my life is thirty two years of data at 32Hz and 1280x1040 pixels. While the true amount saved is less than 90%. A comparison to Shakespeare alphanumeric, average number of pages, and a 100% retrivable, the brain is likely not full, yet.

Exercise 40.9: Mackay and McCulloch (1952)

The paper describes maximum bit content of a nueron (4.3 bits per pulse) where the initial data began at capacity, also known as pulse modulation.

Rate of Acquisition Theory: 4.96×10^6 bits/sec.

Chapter 41: Learning as Inference:

Exercise 41.1: $M(w) = M(w_{MP}) + \frac{1}{2}(w - w_{MP})^T A (w - w_{MP}) + \cdots$

$$A_{ij} = \frac{\partial^2}{\partial w_i \partial w_j} M(w) = \frac{\partial^2}{\partial w \partial w_j}\left[M(w_{MP}) + \frac{1}{2}(w - w_{MP})^T A (w - w_{MP}) + \cdots\right]$$

$$= \frac{\partial^2}{\partial w_i \partial w_j}\left[G(w) + \alpha E_w(w) + \frac{1}{2}(w - w_{MP})^T A (w - w_{MP}) + \cdots\right]$$

$$= \frac{\partial^2}{\partial w_i w_j}\left[-\sum_n t^{(n)} \ln y(x;w) + (1 - t^{(n)}) \ln(1 - y(x;w)) + \alpha E_w + \cdots\right]$$

where $y(x;w) = \frac{1}{1 + e^{-\sum wx}}$

$$= \sum y(x;w)(1 - y(x;w)) t^{(n)2} + \alpha \sum \delta_{ij}$$

Exercise 41.2: $P(t^{(N+1)} = 1 \mid x^{(N+1)}, D, \alpha) = \int d^K w \, y(x;w) \cdot \frac{1}{\sqrt{2\pi A_{ij}}} \exp\left(-\frac{(x - w_{MP})^2}{2 A_{ij}}\right)$

$$= \int da \, f(a) \, \text{Normal}(a_{MP}, s^2)$$

Exercise 41.3: A Gaussian is neither heavy-tailed or light-tailed, but expected Gaussians tend be an overestimate, so are most accurate in Z-tests ($p < 0.05$).

Chapter 42: Hopfield Networks

Exercise 42.1: The η-value is not important because the coefficient scales the set of weights linearly, the activation function; synchrounous and asynchronous.

Exercise 42.2: Chapter 39: Single Neurons as a Classifier

Exercise 42.3 Chapter 33: Variational Methods.

Exercise 42.3: If a synchronous update for a Binary Hopfield network is $a_i = \sum_i w_{ij} X_j$, then any combination which sum to zero, $a_i = 0$, fails to converge.

Exercise 42.4: Similar to Exercise 42.3, a synchronous Binary Hopfield Network where $a_i = \sum_i w_{ij} X_j$ sums to zero from initial conditions, or weights fails to converge.

Exercise 42.5: $X_j^{(t)} = \langle 1, 0, 1, 1, -1 \rangle$ "Input vector"

$$w_{ij} = \begin{bmatrix} X & 0 & 1 & 1 & -1 \\ 0 & X & 0 & 0 & 0 \\ 1 & 0 & X & 1 & -1 \\ 1 & 0 & 1 & X & -1 \\ -1 & 0 & -1 & -1 & X \end{bmatrix}$$

"Weight Matrix"

$w_{ij} = \eta \sum X_i^{(n)} X_j^{(n)}$

where $\eta = 1$

$$a_{ij} = \begin{bmatrix} X & 0 & 1 & 1 & 0 \\ 0 & X & 0 & 0 & 0 \\ 1 & 0 & X & 1 & 1 \\ 1 & 0 & 1 & X & 1 \\ 1 & 0 & 1 & 1 & X \end{bmatrix}$$

"Activation Matrix"

$a_i = \sum_i w_{ij} X_j$

$$\text{Updated } x = x^{(t+1)} = \Theta(\langle X, 0, 1, 1, 0 \rangle)$$
$$= \langle X, 1, 1, 1, 0 \rangle$$

Exercise 42.6: $\tilde{F}(x) = \left(-\frac{1}{2} \sum_i J_{mn} x_m x_n - \sum_i h_n \bar{x}_n\right) - \frac{1}{\beta} \sum_i H_2^{(n)}(q_n)$

$$\frac{d\tilde{F}(x)}{dt} = -\sum_i J X_m - \sum_i h_n = -\sum w_i X_i + X_0$$

$$y = \begin{bmatrix} 0 & 1 & 0 & -1 & 1 \\ 1 & 0 & 1 & -1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ -1 & -1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

– "Output or Threshold"

$y = \text{sigmoid}(a)$

$$e = \begin{bmatrix} 0 & -1 & 0 & 1 & -1 \\ -1 & 0 & -1 & 1 & -1 \\ 0 & -1 & 0 & -1 & -1 \\ 1 & 1 & -1 & 0 & -1 \\ -1 & -1 & -1 & -1 & 0 \end{bmatrix}$$

"Errors"

$e = t - y$

$$gw = \begin{bmatrix} 0 & -2 & 0 & 2 & -2 \\ -2 & 0 & -1 & 1 & -1 \\ 0 & -1 & 0 & 1 & 1 \\ 2 & 1 & 1 & 0 & -2 \\ -2 & -1 & 1 & -2 & 0 \end{bmatrix}$$

"Gradient"

$gw = x' * e$

"Symmetrical Gradient"

$$gw = \begin{bmatrix} 0 & -4 & 0 & 4 & -4 \\ -4 & 0 & -2 & 2 & -2 \\ 0 & -2 & 0 & 2 & 2 \\ 4 & 2 & 2 & 0 & -4 \\ -4 & -2 & 2 & -4 & 0 \end{bmatrix}$$

$gw = gw + gw'$

Exercise 42.7: Error Function: $\phi(-z) \simeq \frac{1}{\sqrt{2\pi}} \frac{e^{-z^2/2}}{z}$

$$\frac{d\phi(-z)}{dz} = \frac{1}{\sqrt{2\pi}} \cdot \frac{e^{-z^2/2}(z^2+1)}{z^2}$$

If $z = -\frac{I}{\sqrt{IN}}$, then $0 = e^{I^2/2IN}\left(\frac{I^2}{IN}+1\right)$

$$e^{I/2N} = e^{I/2N} \cdot \left(\frac{I}{N}\right)$$

$$\frac{I}{2N} = \frac{I}{2N} + \ln\left(\frac{I}{N}\right)$$

$$I = N_{max}$$

Exercise 42.8: $x^{(1)} = \langle 2, 1, 0, -1, 2 \rangle$ "Input Vector"

$$W^{(1)} = \begin{bmatrix} 0 & 2 & 0 & -2 & 4 \\ 2 & 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & -1 & 0 & 0 & -2 \\ 4 & 2 & 0 & -2 & 0 \end{bmatrix}$$

"Weight Matrix"

$$W_{ij} = X^T * X$$

$$a_{ij} = \begin{bmatrix} 0 & 4 & 0 & -4 & 8 \\ 4 & 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ -4 & -1 & 0 & 0 & 2 \\ 8 & 2 & 0 & 2 & 0 \end{bmatrix}$$

"Activation Matrix"

$$a_{ij} = W_{ij} \cdot X$$

$$W = \begin{bmatrix} 0 & 8 & 0 & -12 & 0 \\ 8 & 0 & 2 & -4 & 6 \\ 0 & -2 & 0 & -2 & -2 \\ -12 & -4 & -2 & 0 & 0 \\ 0 & 6 & -2 & 0 & 0 \end{bmatrix}$$

"Make Step"

W = W + eta * (gW - alpha * W)

Where eta = 1 ; alpha = 1

Exercise 42.9: Binary Memories: m and n $(m_i, n_i \in \{-1, +1\})$

Hopfield Network: $w_{ij} = \begin{cases} m_i m_j + n_i n_j & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases}$

Biases: $b_i = 0$

State: $X = m$

Activation: $a_i = \sum w_{ij} X = m_i m_j + n_i n_j$

$= \mu m_i + \nu n_i$

New state: $X = m + 2d$ where perturbation d is $d_i \{-1, 0, +1\}$

Overlap: $O_{mn} = \sum_{i=1}^{I} m_i n_i$

Activation: $a_i = \sum W_{ij} X = \sum m_i n_i \cdot m + \sum m_i n_i \cdot 2d$

$= -|O_{mn}| - 2$

Exercise 42.10: $X' = \begin{cases} x_i^{(n)} & i' \neq i \\ 0 & i' = i \end{cases}$

$w_{ii}' = \begin{cases} w_{ii'} & i' \neq i \\ 0 & i' = i \end{cases}$

$$N_{max} \cong \frac{2I}{4\ln 2I + 2\ln(1/\epsilon)} = \frac{2I}{4 + 4\ln I + 2\ln(1/\epsilon)} = \frac{I}{2\ln I + 2\ln(1/\epsilon)}$$

An input of $2I$ reduces N_{max} to an input equal to I.

Exercise 42.11: Half of the board: $\frac{N \times N}{2}$.

d-number of neighbors: $\frac{N \times N \times \cdots \times N}{d}$

The arrangement is a minimum structure of every other cell is dead, alive, dead, alive....

Exercise 42.12: P(Entire Board) = P(10 empty squares) + P(54 filled squares)

$$P(10 \text{ Empty Squares}) = 1 - \sum_{4}^{64} \frac{(l+1)}{2^l} = 1 - \frac{3}{4} = 0.25$$

If weights halve per move, then the number of moves with 10 empty squares is two moves per piece, but the board position is not ten empty pieces, as favored by the question.

Chapter 43: Boltzmann Machines:

Exercise 43.1: $Z(W) = \sum_{x} P(x|W) = \sum e^{-\frac{x^T W x}{2}}$

$$\ln Z(W) = \ln\left(\sum_{x} P(x|W)\right)$$

$$\frac{\partial}{\partial W_{ij}} \ln Z(W) = \sum_{x} \left(\frac{x^T W x}{2}\right)' P(x|W) = \sum (x^T x) P(x|W)$$

$$= \langle x_i x_j \rangle_{P(x|W)}$$

Exercise 43.2: $P'(x|W, V, \ldots) = \frac{1}{Z} \exp\left(\frac{1}{2} \sum_{ij} W_{ij} x_i x_j + \frac{1}{6} \sum V_{ijk} x_i x_j x_k + \cdots\right)$

$$\log P'(x|W, V, \ldots) = \frac{1}{2} \sum_{ij} W_{ij} x_i x_j + \frac{1}{6} \sum V_{ijk} x_i x_j x_k + \cdots + \ln Z$$

$$\frac{\partial}{\partial V_{ijk}} \log P'(x|W, V, \ldots) = \frac{1}{6} \sum x_i x_j x_k$$

Mutual Information: $I(T;Y) \equiv H(T) - H(T|Y) = \sum_{y,t} P(y) P(t|y) \log \frac{P(t)}{P(t|y)}$

Term #1: $P(y=0) P(t=0|y=0) \log \frac{P(t=0)}{P(t=0|y=0)}$

Term #2: $P(y=1) P(t=0|y=1) \log \frac{P(t=0)}{P(t=0|y=1)}$

Term #3: $P(y=0) P(t=1|y=0) \log \frac{P(t=1)}{P(t=1|y=0)}$

Term #4: $P(y=1) P(t=1|y=1) \log \frac{P(t=1)}{P(t=1|y=1)}$

Classifier A: Term #1: 0 #2: 0 #3: 0 #4: 0

$I(T;Y) = 0$

Classifier B: Term #1: $\frac{-3}{100}$ #2: $\frac{2}{100}$ #3: 0 #4: $\frac{-3}{100}$

$I(T;Y) = +4/100$

Classifier C: Term #1: $-\frac{3}{100}$ #2: $\frac{2}{100}$ #3: 0 #4: 0

$I(T;Y) = +1/100$

Classifier D: Term #1 $\frac{-3}{100}$ #2 $\frac{2}{100}$ #3 0 #4 $\frac{7}{100}$ #5 0

$I(T;Y) = \frac{+8}{100}$

Classifier E: Term #1 $\frac{-4}{100}$ #2 $\frac{1}{100}$ #3 0 #4 $\frac{-3}{100}$ #5 $\frac{-1}{100}$

$I(T;Y) = +\frac{4}{100}$

A rejection class generated greater amounts of mutual information.

Exercise 43.3: Boltzmann Machine: $P(x|W) = \frac{1}{Z(W)} \exp\left[\frac{1}{2} x^T W x\right]$

Bernadd & Bonfill (Exploration of Boltzmann Machines via Bars and Stripes)

The study describes an analysis of Boltzmann Machines in a gradient descent to evaluate the "Bars and Stripes" data sets. When hidden units are fit, the convergence time is longer because a "normalized" factor requires more computation. A robust approximation of 3x3, and 5x5 matrices were reproduced.

Chapter 44: Supervised Learning in Multilayer Networks

Exercise 44.1: y = output ("Guess"); t = test set; e = error rate

False positive rate e_+ and false negative rate e_-

t \ y	0	1
0	90	0
1	10	0

Classifier A.

t \ y	0	1
0	80	10
1	0	10

Classifier B.

t \ y	0	1
0	78	12
1	0	10

Classifier C

Alternative format:

t \ y	0	?	1
0	74	10	6
1	0	1	6

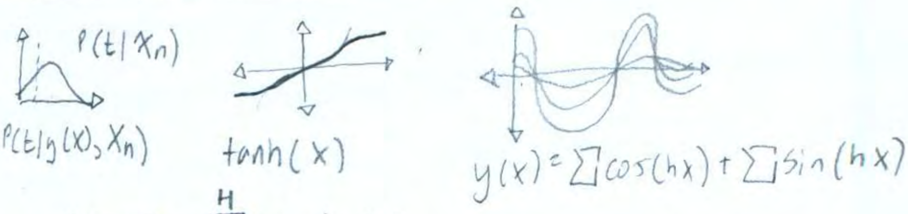
Classifier D

t \ y	0	?	1
0	78	6	6
1	0	5	5

Classifier E

Chapter 45: Gaussian Processes:

Exercise 45.1:



Example 45.2: $y(X;W) = \sum_{h=1}^{H} w_h \phi_h(X)$

$\boxed{\text{Basis Function}}$ $\boxed{\text{Nonlinear Radial Basis Function}}$

$$\phi_h(x) = \exp\left[-\frac{(X-c_h)^2}{2r^2}\right]$$

Example 45.3: $\boxed{\text{Nonlinear Basis Function}}$

$$y(X;W) = \sum_{h=1}^{H} w_h^{(2)} \tanh\left(\sum_i w_{hi}^{(1)} x + w_{h0}^{(1)}\right) + w_0^{(2)}$$

Example 45.4:

$$Q_{nn'} = S\int_{h_{min}}^{h_{max}} dh\, \phi_h(x^{(n)})\, \phi_h(x^{(n')}) = S\int_{h_{min}}^{h_{max}} dh \exp\left[\frac{-(x^{(n)}-h)^2}{2r^2}\right]\exp\left[\frac{-(x^{(n')}-h)^2}{2r^2}\right]$$

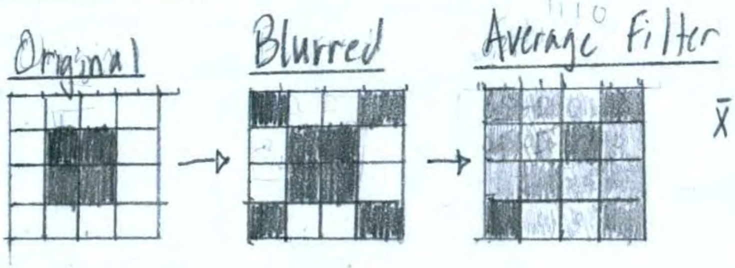
$$= \sqrt{\pi r^2} \exp\left[-\frac{(x^{(n')}-x^{(n)})^2}{4r^2}\right]$$

Example 45.5: Covariance Matrix: $C(X,X';\theta) = \theta_1 \exp\left[-\frac{1}{2}\sum_{i=1}^{I}\frac{(x_i-x_i')^2}{r_i^2}\right] + \theta_2$

$$C(X;X';\theta)^2 = \theta_1^2 \exp\left[-\sum\frac{(x_i-x_i')^2}{r_i^2}\right] + const.$$

Chapter 46: Deconvolution

Exercise 46.1:



$$\bar{X} = \frac{\sum_{j}^{4} X_j}{4}$$

Chapter 47: Low-Density Parity-Check Codes

Exercise 47.1: Log-probability Ratios

$$l_{mn} \equiv \ln\frac{q_{mn}^0}{q_{mn}^1};\ \delta q_{mn} = q_{mn}^0 - q_{mn}^1 = \alpha_{mn} p_n^0 r_{mn}^0 - \alpha_{mn} p_n^1 r_{mn}^1$$

$$= X_{mn} \frac{p_n^0}{2}(1+\delta r_{mn}) - \frac{p_n^1}{2}(1-\delta r_{mn})$$

Exercise 47.2: A linear code describes rate R and block length N, with $K = RN$ source bits and $M = (1-R)N$ parity-check bits; The number of distinct linear codes is $N_1 = 2^{MK} = 2^{N^2R(1-R)}$. The number of low-density parity-check matrices is $N_2 = \frac{\binom{N}{k}^M}{M!}$ which is much smaller than N_1.

Exercise 47.3: M columns of weight 2 or αM columns, where $\alpha > 1$, then the words weight $\log(1-R) \cdot N = \log M$.

Exercise 47.4: Weight Enumerator Function: $A(W)$; $\bar{X} = P_0 A(w=0) + P_1 A(w=1)$

Chapter 48: Convolutional Codes and Turbo Codes:

Exercise 48.1: A recursive filters input state source of $s = \{0,0,1,1,1,0,0,0\}$ generates the same output for a nonrecursive and recursive function

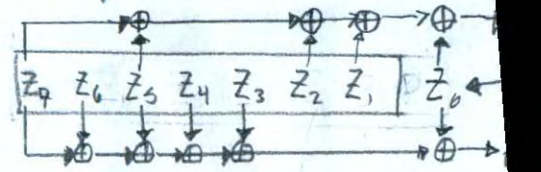
Exercise 48.2: From Figure 48.7, the most probable path for the first bit being flipped is 1111.

Exercise 48.3:

Exercise 48.3:

	0	1	2	3	4	5	6	7
0		⊕		⊕				
1	⊕		⊕					
2		⊕			⊕	⊕		
3	⊕				⊕			
4				⊕		⊕		
5			⊕		⊕		⊕	⊕
6						⊕		⊕
7						⊕	⊕	

Nonsystematic Nonrecursive



Convolutional Code!!!

Chapter 49: Repeat-Accumulate Codes:

Exercise 49.1: Power Laws: $P(\tau) \propto \tau^{-p}$

The power law founded in statistics relates two quantities by a single quantity. An appicable power law investigation of codes is, "How does repeat-accumulate codes, or Gallagher Codes relate to Shannon's limit by iteration amounts?" The capacity of a communication channel refers to the maximum rate of error-free data theoretically achieved subject to random errors. Both repeat-accumulate and Gallagher Codes approach Shannon's limit scaled to millions (or billions) of iterations without an error in the communication channel. Gallagher Codes implement a parity bit per iteration and repeat-accumulate codes permutations of bits per iteration, which fit Shannon's mathematical limit by $\text{Rate} = \frac{\text{Capacity}}{1-\text{Entropy}} = \frac{I(p_a;p_b)}{1-H_2(p_b)}$; where $I(p_a;p_b)$ = Mutal Information, $H_2(p_a)$ = Binary Entropy Function

Chapter 50: Digital Fountain Codes:

Exercise 50.1: $K = 700$-pages ; N = Page-inspections

a) If $N = K$, then 1 = Pages inspected + Pages not inspected.

Pages not inspected = 1 - Pages inspected

$$= 1 - \frac{1}{K}$$

$$= \frac{K-1}{K}$$

$$\text{Total Pages not Inspected} = \left(\frac{K-1}{K}\right)^n$$

$$= 0.37$$

b) If $N > K$, then Total Pages not inspected < 0.37

c) Total Pages not inspected $= \left(\frac{K-1}{K}\right)^N = \left(\frac{K-1}{K}\right)^{K\ln(K/\delta)}$

$$\lim_{\delta \to 0}(\text{Total Pages not inspected}) = \lim_{\delta \to 0}\left(\frac{K-1}{K}\right)^{K\ln(K/\delta)}$$

Exercise 50.2: Ideal Soliton Distribution:

$(t = 0; d = h_0(d))$;

If $(d > 1)$, then $E[d] = h_0(d)d/K$

$\vdots$

$(t = t^{(n)}; d > 1)$ $\quad E[d-1] = h_t(t)\frac{d}{K-t}$

If $(h_t(t) = 1)$, then $h_0(1) = 1$ and $h_0(2) = K/2$

If $(h_t(2) = (K-t)/2)$, then $h_0(d) = 3$.

Thus, the ideal soliton distribution is: $P(d, K, t) = \begin{cases} h_0(d) = \frac{1}{K}; & d = 1 \\ h_t(d) = \frac{1}{d(d-1)} & d > 1 \end{cases}$

Exercise 50.3:

Digital Fountain	Raid.
∘ Encoder/Decoder	∘ Parity bits/drives
∘ LT Codes	∘ Reed-Solomans Erasure Code.

Exercise 50.4: Goal: $h_t(1) = 1 + S$

$$P(d, K, t) = \begin{cases} h_0(d) = 1 + S & d = 1 \\ h_t(d) = \frac{K}{d(d-1)} + \frac{S}{d} & d > 1 \end{cases}$$

$$\sum h_0(d) = 1 \quad ; \quad \sum h_0(d) d = d$$

Exercise 50.5: $$\tau(d) = \begin{cases} \frac{S}{K}\frac{1}{d} & d = 1 \dots (K/S) - 1 \\ \frac{S}{K}\ln(S/\delta) & d = K/S \\ 0 & d > K/S \end{cases}$$

The spike at $d = K/S$ in equation 50.4 describes a replacement for the tail of equation 50.6.

Per se, $$\int_{d=\frac{K}{S}}^{\infty} \frac{K}{d(d-1)} + \frac{S}{d}\, d \approx \int_{d=\frac{K}{S}}^{\infty} \frac{K\,dd}{d-1} - \int_{d=\frac{K}{S}}^{\infty} \frac{K\,dd}{d} + S\int_{d=\frac{K}{S}}^{\infty} \frac{dd}{d}$$

$$= K\left[\log(1-d) - \log(d)\right] + S\log(d)\Big|_{d=\frac{K}{S}}^{\infty}$$

= Does not converge.

Where a discrete approximation fits a model in order to end computation.

Exercise 50.6: The decoding process requires completion for decoding times, the non-convergent tail $(K>S)$ approximates final computation. If high-weight packets $(K>S)$ are removed in computation, then an infinite amount of packets are not computed.

Exercise 50.7: $K' = K + 2\ln(S/\delta)S$

Luby (2002) $K' = K + O(\sqrt{K}\, 2\ln(K/\delta))$

$$K' = K\beta = K \cdot \left(\sum \rho(i) + \tau(i)\right)$$

$$= K + \sum_{i=1}^{K/R-1} \frac{R}{i} + R\ln(R/\delta)$$

$$\leq K + R\,H(K/R) + R\ln(R/\delta)$$

Luby demonstrated a K-maximum limit to encoding fountains.

Exercise 50.8: $K = 10{,}000$ packets

$$\mu(d) = \frac{\rho(d) + \tau(d)}{Z} \Rightarrow \frac{\rho(d)+\tau(d)}{\sum \rho(d) + \tau(d)} = \begin{cases} \frac{1}{K} + \frac{S}{K}\frac{1}{d} & d = 1 \\ \dfrac{\frac{1}{d(d-1)} + \frac{S}{K}\frac{1}{d}}{\sum \frac{1}{d(d-1)} + \frac{S}{K}\frac{1}{d}} & d = 2,3 \cdots (K/S)-1 \\ \dfrac{\frac{1}{d(d-1)} + \frac{S}{K}\ln(S/\delta)}{\sum \frac{1}{d(d-1)} + \frac{S}{K}\ln(S/\delta)} & d = \frac{K}{S} \\ \dfrac{\frac{1}{d(d-1)}}{\sum \frac{1}{d(d-1)}} & d > \frac{K}{S} \end{cases}$$

Exercise 50.9: A model where data streams broadcasted to cars involve active (on) and inactive (off) states where each vehicle gathers data at unique times. A quantifiable model involves 5% more data transmitted, rather than a (packet x error %) or and Reed - Solomons parity codes.

Exercise 50.10: A digital fountain is suboptimal because the code sends an abundant amount of information, typically (>5%).

Exercise 50.11: The probability $K \times K$ has full rank by Luby (2002) is $[\ K' = K + 2\ln(S/\delta)S\ ]$. where K = file length and S = degree-one checks; an amount of degree checks should be zero when full rank.

An optimal decoder suffices $K' = K + \Delta$ where $\Delta = 2\ln(S/\delta)S$; the probability of abundant data.

Exercise 50.12: Richardson and Urbanke (2001b)

A low-density parity check code employed to a digital fountain code arises to optimal states. The matrix structure for a low-density parity check begins with a packet P, identity I, and many sparse matrices S.

$$\left(\begin{array}{c|c|c} \text{Packet} & \text{Sparse Data} & \text{Identity} \\ \hline \text{Sparse} & \text{Identity} & \emptyset \end{array}\right)$$

The sparse data is a random parity check to change sub-optimal digital fountains to near optimal.

Exercise 50.13: $\begin{pmatrix} \text{Packet} \\ + \\ \text{Noise} \end{pmatrix} = \begin{pmatrix} \text{Packet} \end{pmatrix} + \begin{pmatrix} \text{Noise} \end{pmatrix}$