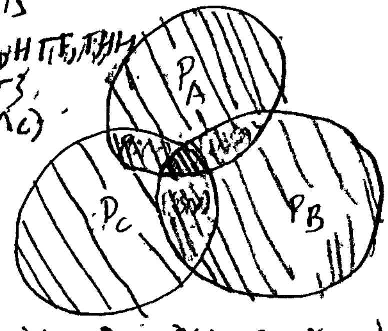
Chapter 1: 1a. Sample Space: {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

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. {~~HHT, 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) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) = P(B \cap C) + P(A \cap B \cap C)$

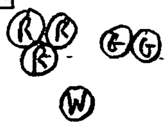
$= P(A \cup B) \cup P(C) = [\underbrace{P(A) + P(B) - P(A \cap B)}_{\text{Addition Law}}] \cup P(C)$

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

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

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

3. Semi



3 draws; RRR RRG RRW RGG GGW

RGR RWR GRG GWG WGR

GRR WRR GGR WGG WRG

n = 6 $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{6!}{3!(3!)} = \frac{6 \cdot 5 \cdot 4}{6} = \frac{120}{6} = \frac{20}{6 \cdot 6} = \frac{5}{9}$

k = 3

$\frac{6!}{1!(5!)} = 6$

Event A: 1 Draw

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

Event B: 2 Draw

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

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

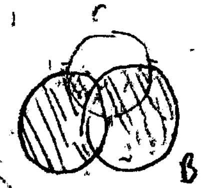
$- P(A_2 \cap A_3) - P(A_2 \cap A_n)$

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

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

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

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



6. 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.

(1,4) (4,1) (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)

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

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

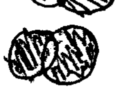
$A \cap (B \cup C) = (4,2),(4,3),(4,4),(4,5),(4,6)$

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

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

8. 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: $P(r|f,N) = \binom{N}{r} f^r (1-f)^{N-r}$

Binomial coefficient. Prob heads. Prob tails

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

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

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

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

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

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

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 $N \simeq 2 \frac{\log 10^{-15}}{\log 4f(1-f)}$ 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"

$$HG^T = [P\ I_3]\begin{bmatrix} I_4 \\ P \end{bmatrix} \cdot PI_4 + I_3P = \begin{bmatrix} 1&1&1&0&1&0&0 \\ 0&1&1&1&0&1&0 \\ 1&0&1&1&0&0&1 \end{bmatrix} \begin{bmatrix} 0&0&0&1 \\ 0&0&1&0 \\ 0&1&0&0 \\ 1&0&0&0 \\ 0&1&1&1 \\ 1&1&1&0 \\ 1&1&0&1 \end{bmatrix} = \boxed{\begin{bmatrix} 0&2&2&2 \\ 2&2&2&0 \\ 2&2&0&2 \end{bmatrix}}$$

Exercise 1.5 Refer to the (7,4) Hamming code

<u>(7,4) Hamming Code</u>

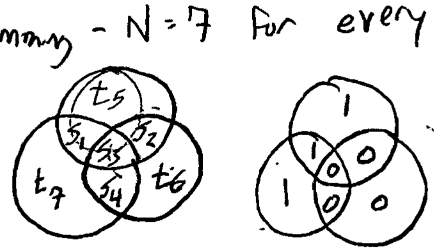
Block Code - rule for converting a sequence of bits s, of length K, to a sequence t of length N bits.

Note: Redundancy occurs when $N > K$.

Linear Code - When the N-K bits are a linear function of the original K bits, i.e. parity check bits.

(7,4) Hamming - $N = 7$ for every $K = 4$ source bits.

Pictorial:



t = transmitted, transmitted bits

s = source bits

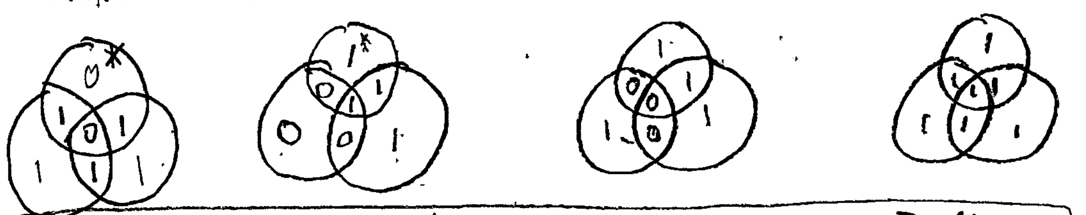
Parity-check bits are when each circle is even.

s	t	s	t	s	t	s	t
0000	0000000	0100	0100110	1000	1000101	1100	1100011
0001	0001011	0101	0101101	1001	1001110	1101	1101000
0010	0010111	0110	0110001	1010	1010010	1110	1110100
0011	0011100	0111	0111010	1011	1011001	1111	1111111

$t = G^T s$

Denote recieved $= G^T s + n$; Syndrome Vector $z = Hr$

a) $r = 1101011$ b $r = 0110110$ c $r = 0100111$ d $r = 1111111$



$z = (1,0,0)$ $z = (0,0,0)$ $z = (0,0,1)$ $z = (0,0,0)$

Exercise 1.6 a) Calculate p_B of (7,4) Hamming code as a function of noise level f and show that it goes as $21f^2$

$$P_B = P(\hat{s} \neq s) = \sum_{r=2}^{7} \binom{7}{r} f^2 (1-f)^{7-r} = \frac{7!}{5! \cdot 2!} f^2 (1-f)^5 = 21 f^2 (1-f)^5 \approx \boxed{21f^2}\, f^5 \ldots$$

b) $$p_b = \frac{1}{K} \sum_{k=1}^{K} P(\hat{s}_k \neq s_k) = \frac{3}{7} 21 f^2 \cdot f^5 \ldots = \boxed{9f^2} \cdot f^5 \ldots$$

Exercise 1.7: Permutation of XXXX ; X6 0,1 =

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

Exercise 1.8: Block Decoding Error

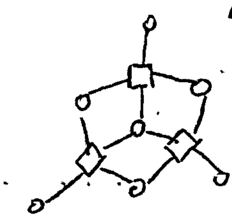
(7,4) Hamming Code

$$P_B = \sum_{r=2}^{n} \binom{7}{r} f^r (1-f)^{7-r}$$

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

Exercise 1.9: Prepare a bipartite graph. The (7,4) Hamming Code is 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 patterns 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: $P = \sum_{m=0}^{n} \binom{n}{m} f^m (1-f)^{n-m}$

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

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

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

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

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

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

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 $\boxed{P(n|f,N) = \binom{N}{n} f^{n}(1-f)^{N-n}}$

b. $E(P(n|f,N)) = \sum_{0}^{\infty} P(n|f,N)\cdot n = N\cdot f$; $Var(P(n|f,N)) = E([n-E(n)]^2) = E[n^2] - E[n]^2$

$= Nf(1-f)$

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

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

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

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

$= 1$

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

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

$\boxed{\text{Probability Distribution: } N=5,\ f_B = 1/5;\ Z = \frac{5}{4}(n_B-1)^2 \text{ ~~where~~ } n_B = 1,2,3,4,5 \text{ is } P(n_B) = \binom{N}{n_B} f^{n}(1-f)^{N-n}}$

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $f(X_1) \leq f(\lambda X_1 + (1-\lambda)X_2) \leq f(x_2)$

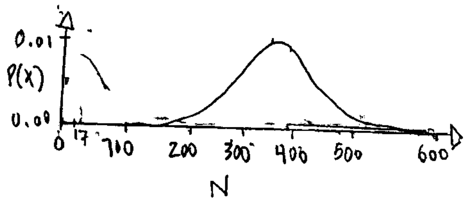
Example 2.15. Jensen's inequality: $E[f(x)] \geq f(E[X])$; $E[f(x)] = \bar{A} = 100m^2 \geq \bar{l}^2 = \boxed{10m^2}$

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

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



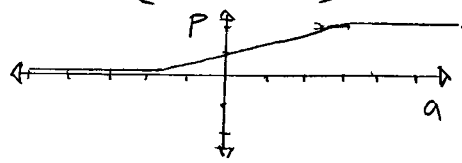
c) p_i, r_i = probabilities of Dice #1, #2
for a sum of $i = 1, 2, \ldots, 11, 12$ and options of 0-6 possible side labels.

$$\frac{1}{11}(x + x^2 + x^3 + \cdots + x^{12}) = (p_0x^0 + p_1x^1 + \cdots + p_6x^6)(r_0x^0 + r_1x^1 + \cdots + r_6x^6)$$

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

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

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



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

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

Exercise 2.18. $A_x=\{0,1\}$; Bayes Theorem: 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: 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 $f(a)=10$, $f(b)=5$, $f(c)=\boxed{10/7}$

$E[f(x)]=\sum p(x)\cdot f(x)=0.1\times10+0.2\times5+0.7\times\frac{10}{7}=\boxed{3.0}$

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

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

Exercise 2.23: $P_a=0.1$; $P_b=0.2$; $P_c=0.7$; $g(a)=0$; $g(b)=1$; $g(c)=0$; $E[g(x)]=0.2\cdot1.0\ \boxed{0.2}$

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

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

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

$\boxed{H(x)\ge0}$

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) $H(\vec{p}) = H(p_1, 1-p) + (1-p_1)H\left(\frac{p_2}{1-p_1}, \frac{p_3}{1-p_1}, \dots, \frac{p_I}{1-p_I}\right)$

Equation (2.44) $H(\vec{p}) = H[(p_1+p_2+\dots+p_m), (p_{m+1}+p_{m+2}+\dots+p_I)]$

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

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

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

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

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

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

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

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

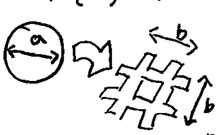
$= P(f)\cdot\log P(f) + f\cdot P(g)\log P(g) + (1-f)P(h)\cdot\log P(h)$

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

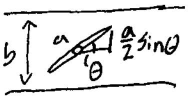
Exercise 2.29 $H(x) = \sum_{x=1}^{n} P(x)\log\frac{1}{P(x)} = \sum_{x=1}^{n}\binom{N}{x}(1/2)^x(1-1/2)^{N-x} - [x\cdot\log(1/2) - (N-x)\log(1/2) + \log\binom{N}{x}]$

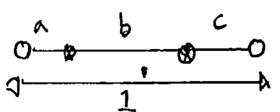
If $N=x$ because flips till heads, $= \boxed{\sum\left(\frac{1}{2}\right)^x[x\log 2]}$

Exercise 2.30 Urn = {w, ..., b...} P(Draw#2|White) = P(Draw#1; P(Draw #2|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: $a^2 = b^2 + c^2 - 2bc\cos A$ $\boxed{\text{Eqn 2}}$

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

$a = 1-b-c$

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

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

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

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

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

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

Exercise 2.34. $P(k = \text{tails}) = (1-p)^{k-1} \cdot p$ where $k = 1, 2, 3, \ldots n$ $\qquad \int a^x dx = a^x / \ln a$

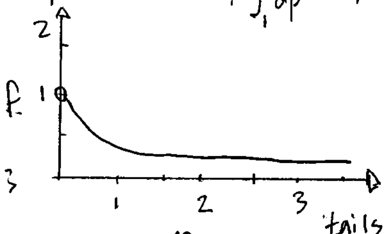
$E[P(k)] = \sum_1^{\infty} k(1-p)^{k-1} \cdot p = \int_1^{\infty} k(1-p)^{k-1} p \, dk = p \int_1^{\infty} \frac{-d}{dp}(1-p)^k dk = -p \frac{d}{dp} \frac{-(1-p)}{\ln(1-p)}$

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

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

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

Assuming $h = 1$;



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

@ $p = 0.5 \approx 2$

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

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

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

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

d) The sum of $E[P(k \mid \text{Before clock})] + E[P(k \mid \text{After clock})] - 1 = \boxed{11 \text{ rolls}}$ $\qquad$ @ $p = \frac{1}{6} \approx 6$

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.

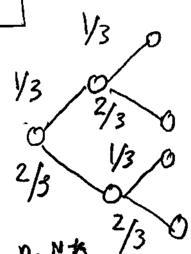
The opportunity Fred is older than Alf would be $\frac{\text{Probability (Fred > Alf)}}{\text{Total Probability}} = \frac{FAB, FBA, BFA}{FAB, FBA, BFA, AFB, ABF, BAF} = 1/2 = 50\%$

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: $P(3\text{-bits}) + P(2\text{-bits}) = 3f^2(1-f) + f^3$

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

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

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

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

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

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

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

$\Rightarrow$

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

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

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

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

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 $p(x) = \frac{e^{-x/\lambda}}{\lambda}$

Number of steps = 20.

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

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

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

$\boxed{\text{Lambda}(\lambda) \text{ describes the length between particles on the detector, and if collected per second, the rate between particles.}}$

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

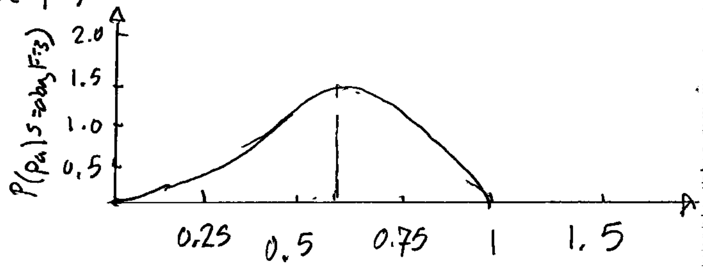
$P(\text{Scene} \mid \text{Each Person, Blood}) = 2 \cdot P('AB') P('O') = \boxed{2 \times 60\%}$

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

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

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

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



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

$\boxed{p_a = 2/3}$

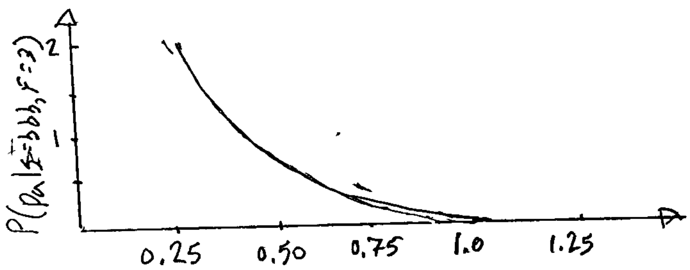
Mean value of p_a under this distribution:

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

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

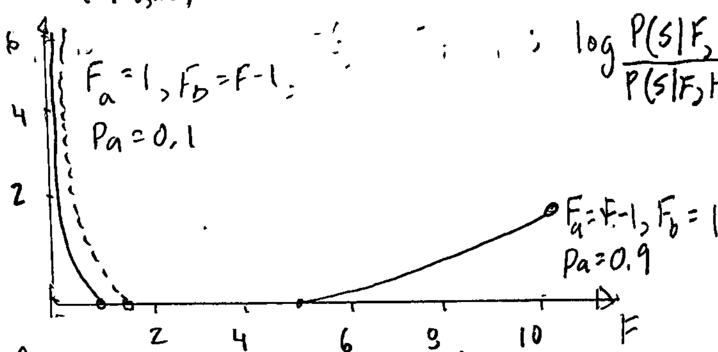
Most probable value; $\boxed{p_a = 1}$

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

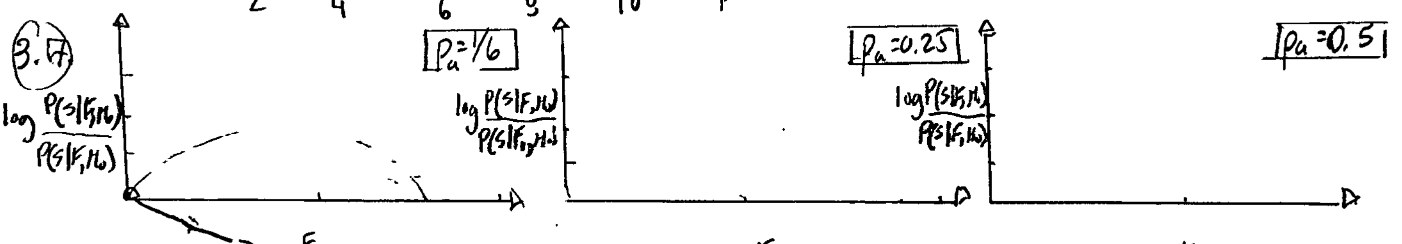


Exercise 3.6. $\log \frac{P(s \mid F, H_0)}{P(s \mid F_0, H_0)}$

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



Exercise 3.7

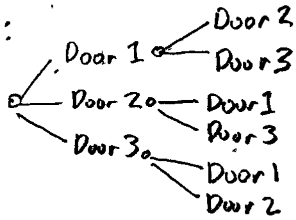


Expected value of F_a is: $p_a F$; A 95% confidence interval ($\alpha = 0.95$)

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

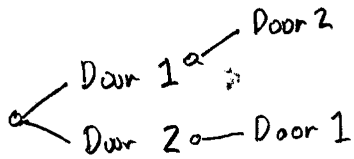
Exercise 3.8 Graphically:

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

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

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

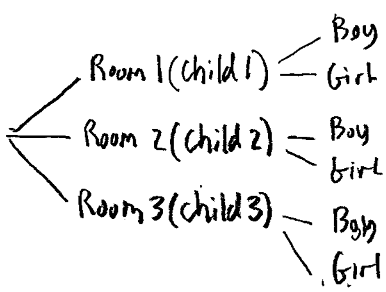
Through switching to door #2, the contestant will have the greatest chance of winning.

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) 'n Door #1.

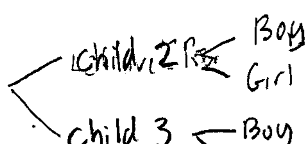
Exercise 3.10. Graphical:

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:

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

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

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

Bayes theorem shows similar outcomes to graphical analysis

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

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

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

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

Exercise 3.13. $\text{Posterior} = \frac{\text{likelihood} \times \text{Prior}}{\text{evidence}} = \frac{1 \quad \times 10^6}{10^6} = 1$ $\qquad 1 = \frac{1 \cdot 1/2}{1} = \boxed{1/2}$.

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 $\quad$ Eqn 3.22 $\frac{P(H_1|s,F)}{P(H_0|s,F)} = \frac{P(s|F,H_1)\,P(H_1)}{P(s|F,H_0)\,P(H_0)}$

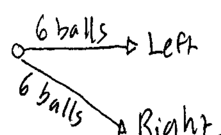
$$\frac{P(H_1|s,F)}{P(H_0|s,F)} = \frac{140!\,110!}{(140+110+1)!} \Big/ (1/2)^{140} (1/2)^{110} \qquad = \frac{F_a!\,F_b!}{(F_a+F_b+1)!} \Big/ p_0^{F_a}(1-p_0)^{F_b}$$

$= 0.4767 \approx 48\%$

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:

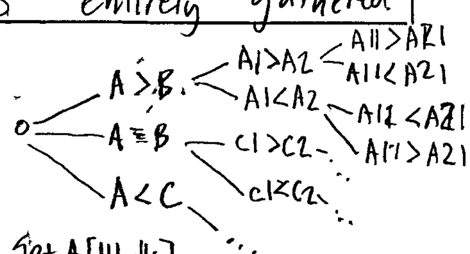
Exercise 4.1. n=12 ; weight {1...11} ≠ weight {12}



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 $\qquad$ State 2 $\qquad$ State 3

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 ; $\boxed{\log 3}$

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

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

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: $64 \to 32 \to 16 \to 8 \to 4 \to 2 \to 1$; $\boxed{6 \text{ guesses.}}$

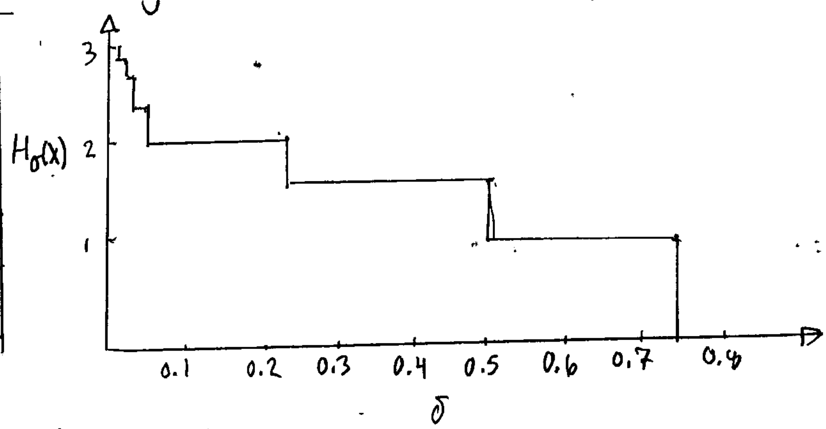
Exercise 4.4: Shannons information provides the number of bits representations of decimal(0 to 255) and ASCII decimal(0 to 127).
Decimal(0 to 255): $\log_2(255) = 7.99$ Decimal(0 to 127): $\log_2(127) = 6.99$
The reduction of physical memory is achieved through removing redundancy and expressing values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

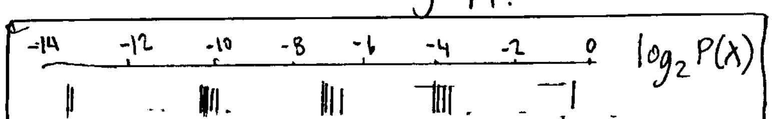
Example 4.6:

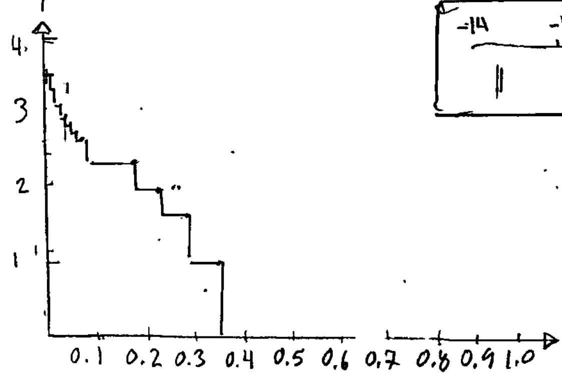
A_X	P_X	Bit code	δ
a	$\frac{1}{4}$	000	00
b	$\frac{1}{4}$	001	01
c	$\frac{1}{4}$	010	10
d	$\frac{3}{16}$	011	11
e	$\frac{1}{64}$	100	—
f	$\frac{1}{64}$	101	—
g	$\frac{1}{64}$	110	—
h	$\frac{1}{64}$	111	—

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. $X = (X_1, X_2 \dots X_n)$ 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.





$\delta = 1 - P(x) = 1 - p_0^{N-r_x} p_1^{r_x}$

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: $\{10, 5, 3, 2\}$.

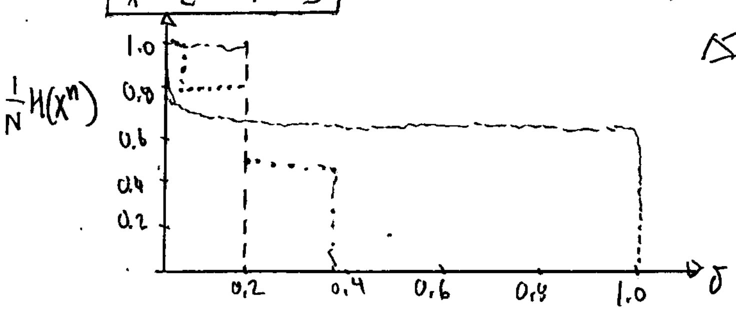
Exercise 4.13 a) Yes, a rotation of sets of four balls generates a compare and contrast Venn Diagram to identify the unique ball.

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.

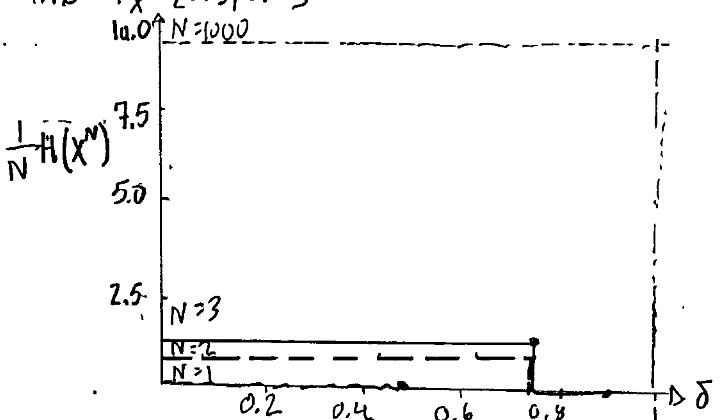
b) The knowledge of ball weights is irregardless of the process to find the odd balls in the set.

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

$$P(Z) = P(X_1, X_2) = P(X_1)\cdot P(X_2) = \frac{1}{Z^2}\int_{-\infty}^{\infty}\frac{dx_1}{x_1^2+1}\int_{-\infty}^{\infty}\frac{dx_2}{x_2^2+1} = \frac{1}{\pi^2}\int_{-\infty}^{\infty}\frac{dx_1}{(x_1^2+1)([Z-x_1]^2+1)}$$

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

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

$C = \frac{-2X}{Z^3+4Z}$; $D = \frac{3Z}{Z^3+4Z}$

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

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 $P(X_2)$.

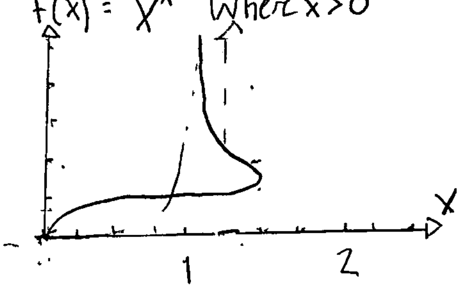
Exercise 4.19. $P(x \geq a) \leq e^{-sa}\cdot g(s)$ and $P(x \leq a) \geq e^{-sa}\cdot g(s)$ where $g(s) = \sum P(x)e^{sx}$

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

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

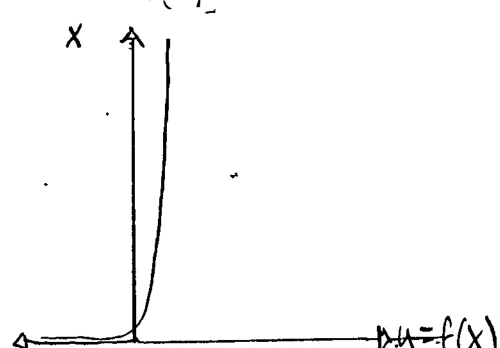
$P(t \geq e^{-sa}) \geq e^{-sa} g(s)$

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



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

$f(\cdot)^{(\frac{1}{X})} = X$



Chapter 5:

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

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

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

$c^+(x) = 0\,110\,111\,10\,0\,110$

C_3 is a prefix and uniquely decodable.

Example 5.11 : $L(C_4, X) = \sum P(x) l(x) = 2$ 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 $x = a_1 \cdot a_2 \cdot a_3 \cdots a_n$ [Equation]

$= S_1 \cdot S_2 \cdot S_3 \cdots S_n$

$= \sum 2^{-l_{i_1}} \cdot \sum 2^{-l_{i_2}} \cdot \sum 2^{-l_{i_3}} \cdots \sum 2^{-l_{iN}}$

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

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

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

[Graphically]

0	00	000	0000
			0001
		001	0010
			0011
	01	010	0100
			0101
		011	0110
			0111
1	10	100	1000
			1001
		101	1010
			1011
	11	110	1100
			1101
		111	1110
			1111
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

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

Example 5.15. $A_x = \{a, b, c, d, e\}$
$P_x = \{0.25, 0.25, 0.2, 0.15, 0.15\}$

a | 0.25 — 0.25 — 0.25 —0— 0.55 —0— 1.0
b | 0.25 — 0.25 —0— 0.45 — 0.45 —1
c | 0.2 — 0.2 —1
d | 0.15 —0— 0.3 — 0.3 —1
e | 0.15 —1

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

Huffmans algorithm provides a method to discover the optimal codelength.

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

Example 5.17. Huffman's algorithm of Figure 2.1 generated codeword disparities ~1 bit to achieve a lossless relationship.

Example 5.18: $A_x = \{a, b, c, d, e, f, g\}$
$P_x = \{0.01, 0.24, 0.05, 0.20, 0.47, 0.01, 0.02\}$

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	$\{1, 01, 000, 001, 1110, 0011, 01111, 11111\}$	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 $H(X^n) \ll L(C, X^n)$, which contradicts the upper limit of bit assignment being entropy.

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

$$= [P_1(x) + P_2(x) + P_3(x) + P_4(x)]\, l \text{ ; if } l_1 = l_2 = l_3 = l_4$$

$$1 = P_1(x) + P_2(x) + P_3(x) + P_4(x)$$

$$\boxed{\begin{aligned} A_{1X} &= \{00, 01, 10, 11\} \\ P_{1X} &= \{1/2, 1/4, 1/8, 1/8\} \\ A_{2}X &= \{0, 1, 00, 11\} \\ P_{2X} &= \{1/4, 1/4, 3/8, 3/8\} \end{aligned}}$$

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

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

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

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

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

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

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

$\text{Length} = \sum_i^{12} P(x) l_i = 3.55$ bits ; Entropy $= 3.46$ bits ; Length − Entropy $= \boxed{0.09 \text{ bits.}}$

Exercise 5.28: $\text{Length}(A_x) = I$

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

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

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

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

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

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

ln

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 & low length is a smaller expected length.
The proposed solution requires (n) codewords on 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

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

$\boxed{= 1/2}$

Exercise 5.32: The Huffman algorithm generates $r \bmod (q-1)$ codewords, where r is the ensemble size and q is the number of leaves combined in the tree. An optimal coding algorithm requires $r \bmod (q-1)+1$; 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 $\sum 2^{-l} \leq 1$, so a metacode of K symbol codes does not fit the case $\frac{1}{K}\sum 2^{-l} \leq 1$ 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 Base-2 is two 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:

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

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

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

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

Rescaled to: 00000000000000001 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

Emitted Value: $\boxed{1}$

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 to ~~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 0010110111101100100110000111010101000011

(pointer, bit)		(,0)	(0,1)	(01,0)	(11,1)	(011,0)	(010,0)	(100,0)	(110,1)	(0101,0)	(0000,1)
s(n) binary	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010
s(n)	0	1	2	3	401	050	96	0170	181	0090	010
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 probabilitys. For every reoccurring value in the sequence, a probability is determined by assigning a probabability 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 $P(1, x_1 \ldots x_n) = P(1|1) \cdot P(1|11) \cdot P(1|111) \cdot P(1|1110) \cdot P(1|11100)$

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

$\simeq 1 \; 1 \; 1 \; 1 \; 1$

{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: $\text{Source}[X] \xrightarrow{F_0} 0$, $f_1 = 0.01 \rightarrow F_1$

Find $X = X_1 X_2 X_3$; $P(X \mid X_1 X_2 X_3) = \dfrac{P(X_1 X_2 X_3 \mid F_1, B) \cdot p(F_1)}{P(X_1 X_2 X_3 \mid 3)}$

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

$1 - H_2(0.01) = 0.8 - 0.07$

$= 0.1$

$E(F_A) = \sum_{0}^{1000} P(F_A) \cdot l(F_A)$

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

$= 1000 \cdot p_1 = \cancel{10\ \text{bits}}$

$Var(F_A) = \sum_{0}^{1000} P(F_A) \cdot l(F_A)^2$

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

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^{i-1} P(Xn=ai | X1...Xn-1)
int v = f; ; Doub Qi = Σ^{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)	(00,0)	(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)	(001, 0)	(00, 0)	(000,)	(001, 1)	(1000, 0)	(0101, 0)	(1000,)

Exercise 6.12 If ~~string~~ length 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: A string of repetitive values has low entropy (say a sexbinary 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: $\int \frac{1}{(2\pi\sigma^2)^{N/2}} x^4 \exp\left(\frac{-x^2}{2\sigma^2}\right) dx = 3\sigma^4.$

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

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

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

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

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

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

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 $= \sum_1^{10} P(x) \cdot l(x) = \boxed{2.64 \text{ bits}}$

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

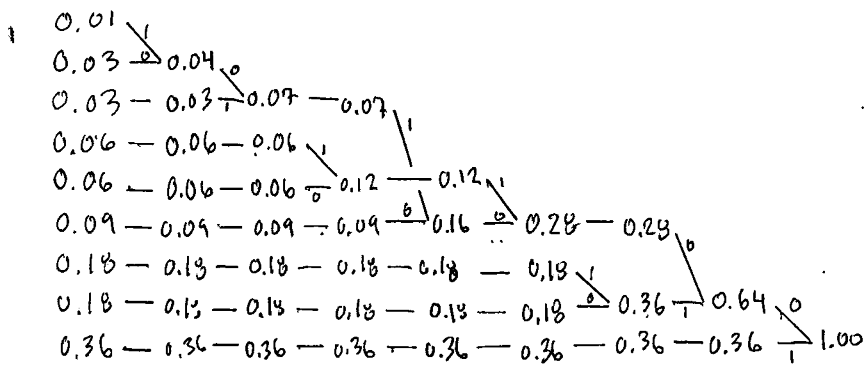
$P(y) = P(x_1) \cdot P(x_2)$

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 $= \sum P(x)\cdot \ell(x) = \boxed{2.73 \text{ bits}}$

Exercise 6.17 $P = \{0.1, 0.9\}$; $E[x] = Np = \frac{N}{10}$; $Var[x] = Np(1-p) = \frac{9N}{100}$; $SD[x] = \sqrt{\frac{9}{100}N} = \frac{3\sqrt{N}}{10}$

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

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

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

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

$$= \frac{I(p)}{L(p)} = \frac{\sum \log_2 \frac{1}{p_n}}{\sum p_n \ell_n} = \frac{\sum \log Z \cdot 2^{\beta \ell n}}{\sum \frac{1}{Z} 2^{-\beta \ell n} \cdot \ell}$$

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

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

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

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

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

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

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 (10) 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: