Exercise 2.2

Are the random variables X and Y in the joint ensemble of figure 2.2 independent?

Solution:

No, $P(x, y) \neq P(x)P(y)$ since each row or column are not proportional to each other.

Exercise 2.4

An urn contains K balls, of which B are black and $W = K - B$ are white. Fred draws a ball at random from the urn and replaces it, N times.

(a) What is the probability distribution of the number of times a black ball is drawn, n_B?

(b) What is the expectation of n_B? What is the variance of n_B? What is the standard deviation of n_B? Give numerical answers for the cases $N = 5$ and $N = 400$, when $B = 2$ and $K = 10$.

Solution:

(a) $P(n_B) = \binom{N}{n_B}(\frac{B}{K})^{n_B}(\frac{K-B}{K})^{N-n_B}$

(b) $\frac{B}{K} = \frac{1}{5}$, $\frac{K-B}{K} = \frac{4}{5}$ so the distribution is $B(n, \frac{1}{5})$.

Hence, $E[n_b] = \frac{1}{5}n$, $Var[n_B] = \frac{1}{5}n(1 - \frac{1}{5})$, $Std[n_B] = \sqrt{Var[n_B]}$

For n = 5: $E[n_b] = 1$, $Var[n_B] = \frac{4}{5}$, $Std[n_B] = \sqrt{\frac{4}{5}}$

For n = 400: $E[n_b] = 80$, $Var[n_B] = 64$, $Std[n_B] = 8$

Exercise 2.5

An urn contains K balls, of which B are black and $W = K - B$ are white. We define the fraction $f_B \equiv \frac{B}{K}$. Fred draws N times from the urn, exactly as in exercise 2.4, obtaining n_B blacks, and computes the quantity

$$z = \frac{(n_B - f_B N)^2}{N f_B (1 - f_B)}. \qquad (2.19)$$

What is the expectation of z? In the case $N = 5$ and $f_B = \frac{1}{5}$, what is the probability distribution of z? What is the probability that $z < 1$? [Hint: compare z with the quantities computed in the previous exercise.]

Solution:

$$E[z] = \frac{1}{N f_b (1 - f_B)}(E[n_B^2] - 2E[f_B n_B N] + E[f_B^2 N^2]) = \frac{1}{N f(1 - f)}(Nf(1-f) + N^2 f^2 - 2N^2 f^2$$

With $N = 5$ and $f_B = \frac{1}{5}$,

$$z = \frac{(n_B - 1)^2}{\frac{4}{5}}$$

so

n_B	0	1	2	3	4	5
z	$\frac{5}{4}$	0	$\frac{5}{4}$	5	$\frac{45}{4}$	20

Hence,

$$P(z < 1) = P(n_B = 1) = \binom{5}{1} \cdot \left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^4 = \frac{256}{625} = 0.4096$$

Assuming a uniform prior on f_H, $P(f_H) = 1$, solve the problem posed in example 2.7 (p.30). Sketch the posterior distribution of f_H and compute the probability that the $N + 1$th outcome will be a head, for

(a) $N = 3$ and $n_H = 0$;

(b) $N = 3$ and $n_H = 2$;

(c) $N = 10$ and $n_H = 3$;

(d) $N = 300$ and $n_H = 29$.

Solution:

$$P(f_H|n_H, N) = \frac{P(n_H|f_H, N)P(f_H)}{P(n_H|N)} = \frac{P(n_H|f_H, N)}{P(n_H|N)} = \frac{\binom{N}{n_H} f_H^{n_H}(1 - f_H)^{N-n_H}}{P(n_H|N)}$$

Now,

$$\int P(f_H|n_H, N)df_H = 1$$

$$\iff \frac{\binom{N}{n_H}}{P(n_H|N)} \int_1^0 f_H^{n_H}(1 - f_H)^{N-n_H} df_H = 1$$

$$\iff P(n_H|N) = \binom{N}{n_H} \frac{\Gamma(n_H + 1)\Gamma(N - n_H + 1)}{\Gamma(N + 2)}$$

$$= \binom{N}{n_H} \frac{n_H!(N - n_H)!}{(N + 1)!}$$

$$= \frac{n_H!(N - n_H)!}{(N + 1)n_H!(N - n_H)!} = \frac{1}{N + 1}$$

so

$$P(f_H|n_H, N) = \frac{(N + 1)!}{n_H!(N - n_H)!} f_H^{n_H}(1 - f_H)^{N-n_H}$$

The graph looks like this: www.desmos.com/calculator/bjjlgnqg87
Now,

$$E[f_H] = \int_0^1 f_H P(f_H|n_H, N)df_H$$

$$= \frac{(N + 1)!}{n_H!(N - n_H)!} \int_0^1 f_H^{n_H+1}(1 - f)^{N-n_H} df_H$$

$$= \frac{(N + 1)!}{n_H!(N - n_H)!} \frac{(n_H + 1)!(N - n_H)!}{(N + 2)!}$$

$$= \frac{n_H + 1}{N + 2}$$

Thus,

(a) $\frac{1}{5}$

(b) $\frac{3}{5}$

(c) $\frac{4}{12} = \frac{1}{3}$

(d) $\frac{30}{302} = \frac{15}{151}$

Exercise 2.14

Prove Jensen's inequality.

Solution:

We want to prove

$$\sum_{i=1}^{I} p_i f(x_i) \geq f\left(\sum_{i=1}^{I} p_i x_i\right) \quad \text{for} \quad \sum p_i = 1 \quad \text{and } f \text{ convex.}$$

Now,

$$f\left(\sum_{i=1}^{I} p_i x_i\right) = f\left(p_1 x_1 + \sum_{i=1}^{I} p_i x_i\right)$$

$$\leq p_1 f(x_1) + \left[\sum_{i=2}^{I} p_i\right]\left[f\left(\frac{\sum_{i=2}^{I} p_i x_i}{\sum_{i=2}^{I} p_i}\right)\right] \quad (\because \text{Definition of convex function})$$

$$\leq p_1 f(x_1) + \left[\sum_{i=2}^{I} p_i\right]\left[\frac{p_2}{\sum_{i=2}^{I} p_i} f(x_2) + \frac{\sum_{i=3}^{I} p_i}{\sum_{i=2}^{I} p_i} f\left(\frac{\sum_{i=3}^{I} p_i x_i}{\sum_{i=3}^{I} p_i}\right)\right]$$

$$= p_1 f(x_1) + p_2 f(x_2) + \frac{\sum_{i=3}^{I} p_i}{\sum_{i=2}^{I} p_i} f\left(\frac{\sum_{i=3}^{I} p_i x_i}{\sum_{i=3}^{I} p_i}\right)$$

$$= \dots$$

$$\leq p_1 f(x_1) + p_2 f(x_2) + p_3 f(x_3) + p_4 f(x_4) + \dots$$

$$= \sum_{i=1}^{I} p_i f(x_i) \quad \blacksquare$$

Exercise 2.16

(a) Two ordinary dice with faces labelled 1,...,6 are thrown. What is the probability distribution of the sum of the values? What is the probability distribution of the absolute difference between the values?

(b) One hundred ordinary dice are thrown. What, roughly, is the probability distribution of the sum of the values? Sketch the probability distribution and estimate its mean and standard deviation.

(c) How can two cubical dice be labelled using the numbers $0, 1, 2, 3, 4, 5, 6$ so that when the two dice are thrown the sum has a uniform probability distribution over the integers $1 - 12$?

(d) Is there any way that one hundred dice could be labelled with integers such that the probability distribution of the sum is uniform?

Solution:

(a) let S = sum of values.

s	2	3	4	5	6	7	8	9	10	11	12
$P(S=s)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Let D = difference of values

d	0	1	2	3	4	5
$P(D=d)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{18}$

(b)

$$E[S_{100}] = 100\,E[s]$$
$$= 350 \quad (\because \text{Independent})$$
$$\text{Var}[S_{100}] = 100\,\text{Var}[s]$$
$$= 100\left(\frac{91}{6} - \left(\frac{7}{2}\right)^2\right) = \frac{875}{3} \approx 292$$

$P(S = s)$ vs s graph will be bell-curve-like between 100 and 600 with mean 350.

(c) Die 1 with $\{1, 2, 3, 4, 5, 6\}$ and die 2 with $\{0, 0, 0, 6, 6, 6\}$.

(d) Make rth die to be $\{0, 1, 2, 3, 4, 5\} \times 6^r$ so each combination will sum to a unique number.
This does not violate the CLT as variables are not identical and the Lindeberg Condition (no single variance dominate) is not satisfied.

If $q = 1 - p$ and $a = \ln p/q$, show that

$$p = \frac{1}{1 + \exp(-a)}. \tag{2.50}$$

Sketch this function and find its relationship to the hyperbolic tangent function $\tanh(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}$.

It will be useful to be fluent in base-2 logarithms also. If $b = \log_2 p/q$, what is p as a function of b?

Solution:

$$\frac{1}{1 + \exp(-a)} = \frac{1}{1 + \exp(-\ln\frac{p}{q})} = \frac{1}{1 + (\frac{p}{q})^{-1}}$$

$$= \frac{1}{1 + \frac{q}{p}} = \frac{p}{p + q} = p$$

Now, with tanh,

$$p = \frac{1}{1 + e^{-u}}$$

$$= \frac{1}{2} \frac{2e^{\frac{u}{2}}}{e^{\frac{u}{2}} + e^{-\frac{u}{2}}}$$

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

$$= \frac{1}{2} \tanh\left(\frac{u}{2}\right) + \frac{1}{2}$$

For base 2,

$$b = \log_2 \frac{p}{q}$$

$$\Longleftrightarrow 2^b = \frac{p}{1 - p}$$

$$\Longleftrightarrow 2^b = \frac{p}{q}$$

$$\Longleftrightarrow (2^b + 1)p = 2^b$$

$$\Longleftrightarrow p = \frac{2^b}{2^b + 1} = \frac{1}{1 + 2^{-b}}$$

Exercise 2.18

Let x and y be dependent random variables with x a binary variable taking values in $\mathcal{A}_X = \{0, 1\}$. Use Bayes' theorem to show that the log posterior probability ratio for x given y is

$$\log \frac{P(x=1|y)}{P(x=0|y)} = \log \frac{P(y|x=1)}{P(y|x=0)} + \log \frac{P(x=1)}{P(x=0)}. \qquad (2.51)$$

Solution:

$$\begin{aligned}
\log \frac{P(x=1|y)}{P(x=0|y)} &= \log \frac{P(y|x=1) \cdot P(x=1) \cdot P(y)}{P(y|x=0) \cdot P(x=0) \cdot P(y)} \\
&= \log \frac{P(y|x=1)}{P(y|x=0)} + \log \frac{P(x=1)}{P(x=0)}
\end{aligned}$$

Exercise 2.19

Let x, d_1 and d_2 be random variables such that d_1 and d_2 are conditionally independent given a binary variable x. Use Bayes' theorem to show that the posterior probability ratio for x given $\{d_i\}$ is

$$\frac{P(x=1|\{d_i\})}{P(x=0|\{d_i\})} = \frac{P(d_1|x=1)}{P(d_1|x=0)} \frac{P(d_2|x=1)}{P(d_2|x=0)} \frac{P(x=1)}{P(x=0)}. \qquad (2.52)$$

Solution:

$$\begin{aligned}
\frac{P(x=1|\{d_i\})}{P(x=0|\{d_i\})} &= \frac{P(d_1, d_2|x=1)P(x=1)}{P(d_1, d_2)} \cdot \frac{P(d_1, d_2)}{P(d_1, d_2|x=0)P(x=0)} \\
&= \frac{P(d_1|x=1)}{P(d_1|x=0)} \frac{P(d_2|x=1)}{P(d_2|x=0)} \frac{P(x=1)}{P(x=0)} \quad (\because d_1, d_2 \text{ independent})
\end{aligned}$$

Exercise 2.20

Consider a sphere of radius r in an N-dimensional real space. Show that the fraction of the volume of the sphere that is in the surface shell lying at values of the radius between $r - \epsilon$ and r, where $0 < \epsilon < r$, is:

$$f = 1 - \left(1 - \frac{\epsilon}{r}\right)^N. \tag{2.53}$$

Evaluate f for the cases $N = 2$, $N = 10$ and $N = 1000$, with (a) $\epsilon/r = 0.01$; (b) $\epsilon/r = 0.5$.

Solution:

$V_{\text{sphere}} \propto r^N$ so with a constant,

$$f = \frac{ar^N - a(r - \epsilon)^N}{ar^N} = 1 - (1 - \frac{\epsilon}{r})^N$$

(a)

N	2	100	1000
f ($\frac{\epsilon}{r} = 0.01$)	0.0199	0.0956	0.99996

(b)

N	2	10	1000
f ($\frac{\epsilon}{r} = 0.5$)	0.75	0.999	1.0

Exercise 2.21

Let $p_a = 0.1$, $p_b = 0.2$, and $p_c = 0.7$. Let $f(a) = 10$, $f(b) = 5$, and $f(c) = 10/7$. What is $\mathcal{E}[f(x)]$? What is $\mathcal{E}[1/P(x)]$?

Solution:

$$\mathcal{E}[f(x)] = \sum_x P(x)f(x)$$

$$= 0.1 \cdot 10 + 0.2 \cdot 5 + 0.7 \cdot \frac{10}{7}$$

$$= 3$$

$$\mathcal{E}\left[\frac{1}{P(x)}\right] = \sum_x P(x)\frac{1}{P(x)}$$

$$= 1 + 1 + 1 = 3$$

For an arbitrary ensemble, what is $\mathcal{E}[1/P(x)]$?

Solution:

$$\mathcal{E}\left[\frac{1}{P(x)}\right] = \sum_x P(x)\frac{1}{P(x)}$$
$$= \sum_{x \in \mathcal{A}_X} 1$$
$$= |x|$$

Let $p_a = 0.1$, $p_b = 0.2$, and $p_c = 0.7$. Let $g(a) = 0$, $g(b) = 1$, and $g(c) = 0$. What is $\mathcal{E}[g(x)]$?

Solution:

$$\mathcal{E}[g(x)] = 0.1 \cdot 0.1 + 0.2 \cdot 1 + 0.7 \cdot 0 = 0.2$$

Let $p_a = 0.1$, $p_b = 0.2$, and $p_c = 0.7$. What is the probability that $P(x) \in [0.15, 0.5]$? What is
$$P\left(\left|\log\frac{P(x)}{0.2}\right| > 0.05\right)?$$

Solution:

$x = b$ is the only possibility for $P(x) \in [0.15, 0.5]$ so 0.2.
For the second part,

$$\left|\log\frac{P(x)}{0.2}\right| > 0.05$$
$$\iff \log\frac{P(x)}{0.2} < -0.05, \ 0.05 < \log\frac{P(x)}{0.2}$$
$$\iff P(x) < 2^{-0.05} \cdot 0.2, \ 2^{0.05} \cdot 0.2 < P(x)$$
$$\iff P(x) < 0.1932, \ 0.2071 < P(x)$$

so

$$P\left(\left|\log\frac{P(x)}{0.2}\right| > 0.05\right) \iff p_a + p_b = 0.8$$

Exercise 2.25

Prove the assertion that $H(X) \leq \log(|\mathcal{A}_X|)$ with equality iff $p_i = 1/|\mathcal{A}_X|$ for all i. ($|\mathcal{A}_X|$ denotes the number of elements in the set $\mathcal{A}_X$.) [Hint: use Jensen's inequality (2.48); if your first attempt to use Jensen does not succeed, remember that Jensen involves both a random variable and a function, and you have quite a lot of freedom in choosing these; think about whether your chosen function f should be convex or concave.]

Solution:

$$H(X) = \text{E}\left[\log \frac{1}{P(X)}\right]$$
$$\leq \log(\text{E}\left[\log \frac{1}{P(X)}\right])$$
$$= \log(|\mathcal{A}_X|) \quad (\because \text{concave})$$

Equality when $\frac{1}{P(X)}$ is constant, i.e. p_i constant, i.e. $p_i = \frac{1}{|\mathcal{A}_X|}$

Exercise 2.26

Prove that the relative entropy (equation (2.45)) satisfies $D_{\text{KL}}(P||Q) \geq 0$ (Gibbs' inequality) with equality only if $P = Q$.

Solution:

$$\sum_x P(X) \log \frac{P(X)}{Q(X)} = \text{E}\left[-\log \frac{Q(X)}{P(X)}\right]$$
$$\leq -\log \text{E}\left[\frac{Q(X)}{P(X)}\right]$$
$$= -\log \sum_x Q(X)$$
$$= -\log 1 = 0$$

Equality iff $\frac{Q(X)}{P(X)}$ constant,
$\iff Q(X) = P(X) \quad (\because \sum_x Q(X) = \sum_x P(X) = 1)$

Exercise 2.27

Prove that the entropy is indeed decomposable as described in equations (2.43–2.44).

Solution:

Let $p_x \sim p_y = \sum_{i=1}^{x} p_i$.

$$H(p) = \sum_{i=1}^{I} p_i \log \frac{1}{p_i}$$

$$= \sum_{i=1}^{m} (p_1 \sim p_m) \frac{p_i}{p_1 \sim p_m} \log \frac{1}{p_i} + \sum_{i=m+1}^{I} (p_{m+1} \sim p_I) \frac{p_i}{p_{m+1} \sim p_I} \log \frac{1}{p_i}$$

$$= (p_1 \sim p_m) \sum_{i=1}^{m} \frac{p_i}{p_1 \sim p_m} \log \frac{p_1 \sim p_m}{p_i} + \sum_{i=1}^{m} (p_1 \sim p_m) \frac{p_i}{p_1 \sim p_m} \log \frac{1}{p_1 \sim p_m}$$

$$+ \sum_{i=m+1}^{I} \frac{(p_{m+1} \sim p_I) p_i}{p_{m+1} \sim p_I} \log \frac{p_{m+1} \sim p_I}{p_i} + \sum_{i=m+1}^{I} \frac{(p_{m+1} \sim p_I) p_i}{p_{m+1} \sim p_I} \log \frac{1}{p_{m+1} \sim p_I}$$

$$= (p_1 \sim p_m) H\left(\frac{p_1}{p_1 \sim p_m}, \ldots, \frac{p_m}{p_1 \sim p_m}\right) + (p_{m+1} \sim p_I) H\left(\frac{p_{m+1}}{p_{m+1} \sim p_I}, \ldots, \frac{p_I}{p_{m+1} \sim p_I}\right)$$

$$+ H((p_1 \sim p_m), (p_{m+1} \sim p_I))$$

Exercise 2.28

A random variable $x \in \{0, 1, 2, 3\}$ is selected by flipping a bent coin with bias f to determine whether the outcome is in $\{0, 1\}$ or $\{2, 3\}$; then either flipping a second bent coin with bias g or a third bent coin with bias h respectively. Write down the probability distribution of x. Use the decomposability of the entropy (2.44) to find the entropy of X. [Notice how compact an expression is obtained if you make use of the binary entropy function $H_2(x)$, compared with writing out the four-term entropy explicitly.] Find the derivative of $H(X)$ with respect to f. [Hint: $dH_2(x)/dx = \log((1 - x)/x)$.]

Solution:

Probability distribution:

x	0	1	2	3
$P(X = x)$	fg	$f(1 - g)$	$(1 - f)h$	$(1 - f)(1 - h)$

Using the decomposability,

$$H(X) = H_2(f) + f H_2(g) + (1 - f) H_2(h)$$

The derivative is,

$$\frac{dH(X)}{df} = \log \frac{1 - f}{f} + H_2(g) - H_2(h)$$

Exercise 2.29

An unbiased coin is flipped until one head is thrown. What is the entropy of the random variable $x \in \{1, 2, 3, \ldots\}$, the number of flips? Repeat the calculation for the case of a biased coin with probability f of coming up heads. [Hint: solve the problem both directly and by using the decomposability of the entropy (2.43).]

Solution:

Solving directly,

$$
\begin{aligned}
H(X) &= \frac{1}{2} \log 2 + \frac{1}{4} \log 4 + \frac{1}{8} \log 8 + \frac{1}{16} \log 16 + \ldots \\
&= \left(\frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \ldots \right) \log 2 \\
&= \log 2 \sum_{i=1}^{\infty} \frac{i}{2^i} \\
&= 2 \log 2 \quad \left(\because \sum_{i=1}^{\infty} i x^i = \frac{x}{(1-x)^2} \right) \\
&= 2
\end{aligned}
$$

With probability f,

$$
\begin{aligned}
H(X) &= f \log \frac{1}{f} + (1-f) f \log \frac{1}{(1-f)f} + \ldots \\
&= \sum_{i=1}^{\infty} -(1-f)^{i-1} f \log(f(1-f)^{i-1}) \\
&= \sum_{i=1}^{\infty} \left\{ -(1-f)^{i-1} f \log f + (1-i)(1-f)^{i-1} f \log(1-f) \right\} \\
&= \log \frac{1}{f} + \log(1-f) + f \log \frac{1}{1-f}
\end{aligned}
$$

Using decomposability,

$$
\begin{aligned}
H(X) &= H_2 f + (1-f) H(X) \\
&\Longleftrightarrow H(X) = \frac{H_2(f)}{f}
\end{aligned}
$$

which is equal to the directly calculated answer.

Exercise 2.30

An urn contains w white balls and b black balls. Two balls are drawn, one after the other, without replacement. Prove that the probability that the first ball is white is equal to the probability that the second is white.

Solution:

$$P(w_1) = \frac{w}{w + b} \tag{1}$$

$$P(w_2) = P(w_2 \wedge w_1) + P(w_2 \wedge b_1) \tag{2}$$

$$= \frac{w - 1}{w + b - 1} \times \frac{w}{w + b} + \frac{w}{w + b - 1} \times \frac{b}{w + b} \tag{3}$$

$$= \frac{w^2 - w + wb}{(w + b - 1)(w + b)} \tag{4}$$

$$= \frac{w}{w + b} \tag{5}$$

Exercise 2.31

A circular coin of diameter a is thrown onto a square grid whose squares are $b \times b$. $(a < b)$ What is the probability that the coin will lie entirely within one square? [Ans: $(1 - a/b)^2$]

Solution:

The center of the coin must be at least $\frac{a}{2}$ away from the edge of the grid. Considering the ratio within each grid,

$$\frac{(b - a)^2}{b^2} = (1 - \frac{a}{b})^2$$

Exercise 2.32

Buffon's needle. A needle of length a is thrown onto a plane covered with equally spaced parallel lines with separation b. What is the probability that the needle will cross a line? [Ans, if $a < b$: $2a/\pi b$] [Generalization – Buffon's noodle: on average, a random curve of length A is expected to intersect the lines $2A/\pi b$ times.]

Solution:

Let the angle between the lines and the needle be θ, then the chance of that needle crossing a line is $\frac{a \sin \theta}{b}$.

Taking the average for the angle, the probability is

$$\frac{1}{\pi}\int_0^{\pi}\frac{a\sin\theta}{b}d\theta = \frac{1}{\pi}\left[-\frac{a}{b}\cos\theta\right]_0^{\pi}$$
$$= \frac{1}{\pi}\left(\frac{a}{b}-\left(-\frac{a}{b}\right)\right)$$
$$= \frac{2a}{\pi b}$$

Exercise 2.33

Two points are selected at random on a straight line segment of length 1. What is the probability that a triangle can be constructed out of the three resulting segments?

Solution:

Assume $0 \le x \le y \le x + y \le 1$.
Then, the lengths of the three segments will be x, $y - x$, $1 - y$.
From the triangle condition,

$$x + (y - x) > 1 - y$$
$$(y - x) + (1 - y) > x$$
$$(1 - y) + x > y - x$$
$$\Longleftrightarrow y > \frac{1}{2}, \ x < \frac{1}{2}, \ y < x + \frac{1}{2}$$

We will now consider the graph where y and x can take. We have $x \ge 0$, $y \ge 0$, $y > x$, $x + y \le 1$ as the assumption so the total area for a valid segmentation is $\frac{1}{2}\cdot 1 \cdot \frac{1}{2} = \frac{1}{4}$.
The area within that further satisfies the above triangle condition is (by drawing the graph), $\frac{1}{2}\cdot\frac{1}{4}\cdot\frac{1}{2} = \frac{1}{16}$.
Hence, the probability we require is $\frac{\frac{1}{16}}{\frac{1}{4}} = \frac{1}{4}$.

Exercise 2.34

An unbiased coin is flipped until one head is thrown. What is the expected number of tails and the expected number of heads?
Fred, who doesn't know that the coin is unbiased, estimates the bias using $\hat{f} \equiv h/(h + t)$, where h and t are the numbers of heads and tails tossed. Compute and sketch the probability distribution of $\hat{f}$.
N.B., this is a forward probability problem, a sampling theory problem, not an inference problem. Don't use Bayes' theorem.

Solution:

$$E[H] = 1$$

$$E[T] = \frac{1}{2} \cdot 0 + \frac{1}{4} \cdot 1 + \frac{1}{8} \cdot 2 + \frac{1}{16} \cdot 3 + \ldots$$

$$= \sum_{i=1}^{\infty} \frac{i}{2^{i+1}}$$

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

Now, $\hat{f} = \frac{1}{1+t}$ so

t	0	1	2	3	4
$\hat{f}$	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$

The graph will be a discrete version of $y = 2^{-\frac{1}{x}+1}$.

Exercise 2.35

Fred rolls an unbiased six-sided die once per second, noting the occasions when the outcome is a six.

(a) What is the mean number of rolls from one six to the next six?

(b) Between two rolls, the clock strikes one. What is the mean number of rolls until the next six?

(c) Now think back before the clock struck. What is the mean number of rolls, going back in time, until the most recent six?

(d) What is the mean number of rolls from the six before the clock struck to the next six?

(e) Is your answer to (d) different from your answer to (a)? Explain.

Solution:

(a)

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{5}{6} \cdot \frac{1}{6} + 3 \cdot \left(\frac{5}{6}\right)^2 \cdot \frac{1}{6} + \ldots$$

$$= \sum_{i=1}^{\infty} i \left(\frac{5}{6}\right)^{i-1} \cdot \frac{1}{6}$$

$$= \frac{1}{6} \cdot \frac{1}{(1 - \frac{5}{6})^2} = 6$$

(b) Same as (a), so 6.

(c) Same as (a), so 6.

(d) $6 + 6 - 1 = 11$

(e) It is different because the probability that a one is struck between two rolls is higher when the gap between the two rolls is wider, so it is not equal to the distribution of the mean number of rolls.

Exercise 2.36

You meet Fred. Fred tells you he has two brothers, Alf and Bob.
What is the probability that Fred is older than Bob?
Fred tells you that he is older than Alf. Now, what is the probability that Fred is older than Bob? (That is, what is the conditional probability that $F > B$ given that $F > A$?)

Solution:

There are six possible ordering, FAB, FBA, AFB, ABF, BAF, BFA. Fred is older than Bob for FAB, FBA, AFB so the answer is $\frac{1}{2}$.
When it is given that Fred is older than Alf, the possible ordering reduces to FAB, FBA, BFA. Fred is older than Bob for FAB, FBA so now the answer is $\frac{2}{3}$.

Exercise 2.37

The inhabitants of an island tell the truth one third of the time. They lie with probability 2/3. On an occasion, after one of them made a statement, you ask another 'was that statement true?' and he says 'yes'.
What is the probability that the statement was indeed true?

Solution:

$$P(T|s) = \frac{P(T \wedge T)}{P(T \wedge T) + P(F \wedge F)} = \frac{\frac{1}{9}}{\frac{1}{9} + \frac{4}{9}} = \frac{1}{5}$$

Exercise 2.38

Compare two ways of computing the probability of error of the repetition code R_3, assuming a binary symmetric channel (you did this once for exercise 1.2 (p.7)) and confirm that they give the same answer.

Binomial distribution method. Add the probability that all three bits are flipped to the probability that exactly two bits are flipped.

Sum rule method. Using the sum rule, compute the marginal probability that $\mathbf{r}$ takes on each of the eight possible values, $P(\mathbf{r})$. $[P(\mathbf{r}) = \sum_s P(s)P(\mathbf{r}|s).]$ Then compute the posterior probability of s for each of the eight values of $\mathbf{r}$. [In fact, by symmetry, only two example cases $\mathbf{r} = (000)$ and $\mathbf{r} = (001)$ need be considered.] Notice that some of the inferred bits are better determined than others. From the posterior probability $P(s|\mathbf{r})$ you can read out the case-by-case error probability, the probability that the more probable hypothesis is not correct, $P(\text{error}|\mathbf{r})$. Find the average error probability using the sum rule,

$$P(\text{error}) = \sum_{\mathbf{r}} P(\mathbf{r})P(\text{error}|\mathbf{r}). \tag{2.55}$$

Solution:

Binomial distribution method.

$$\binom{3}{2} f^2(1-f) + f^3 = 3f^2 - 3f^3 + f^3 = 3f^2 - 2f^3$$

Sum rule method.

$$P(r) = \sum_s P(s)P(r|s)$$

When $r = 000$,

$$P(r) = \frac{1}{2}(1-f)^3 + \frac{1}{2}f^3$$

$$P(\text{error}|r) = \frac{P(r|\text{error}) \cdot P(\text{error})}{P(r)} = \frac{f^3 \cdot \frac{1}{2}}{\frac{1}{2}(f^3 + (1-f)^3)} = \frac{f^3}{f^3 + (1-f)^3}$$

When $r = 001$,

$$P(r) = \frac{1}{2}(1-f)^2 f + \frac{1}{2}(1-f)f^2 = \frac{1}{2}(1-f)f$$

$$P(\text{error}|r) = \frac{P(r|\text{error}) \cdot P(\text{error})}{P(r)} = \frac{f^2(1-f) \cdot \frac{1}{2}}{\frac{1}{2}f(1-f)} = f$$

Considering the symmetry for other rs,

$$P(\text{error}) = \sum_{\mathbf{r}} P(\mathbf{r})P(\text{error}|\mathbf{r})$$

$$= 2 \times \frac{f^3}{f^3 + (1-f)^3} \times (\frac{1}{2}(1-f)^3 + \frac{1}{2}f^3) + 6 \times f \times \frac{1}{2}(1-f)f$$

$$= f^3 + 3f^2(1-f)$$

$$= 3f^2 - 2f^3$$

which matches the result from the binomial distribution method.

Exercise 2.39

The frequency p_n of the nth most frequent word in English is roughly approximated by

$$p_n \simeq \begin{cases} \frac{0.1}{n} & \text{for } n \in \{1, \ldots, 12\,367\} \\ 0 & n > 12\,367. \end{cases} \qquad (2.56)$$

[This remarkable $1/n$ law is known as Zipf's law, and applies to the word frequencies of many languages (Zipf, 1949).] If we assume that English is generated by picking words at random according to this distribution, what is the entropy of English (per word)?

Solution:

$$H(X) = \sum_{n=1}^{12367} \frac{0.1}{n} \log \frac{n}{0.1}$$

$$= \sum_{n=1}^{12367} \frac{1}{10n} \log 10n$$

$$= \frac{1}{10} \sum_{n=1}^{12367} \frac{\log 10n}{n}$$

$$\approx \frac{1}{10} \int_1^{12367} \frac{\log_2 10n}{n} dn$$

$$= \frac{1}{10} \left[\frac{(\log_2 10n)^2}{2 \ln 2} \right]_1^{12367}$$

$$= \frac{\ln 2}{10} \left\{ \frac{(\log 123670)^2}{2} - \frac{(\log 10)^2}{2} \right\}$$

$$= 13.76 \ln 2$$

$$= 9.5377$$