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 $\mathrm{B}(n, \frac{1}{5})$.

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

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

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

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

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

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:

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

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,

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

Exercise 2.8

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,

$$\begin{aligned}
&\int P(f_H|n_H, N) df_H = 1 \\
\iff &\frac{\binom{N}{n_H}}{P(n_H|N)} \int_1^0 f_H^{n_H}(1-f_H)^{N-n_H} df_H = 1 \\
\iff &P(n_H|N) = \binom{N}{n_H} \frac{\Gamma(n_H+1)\Gamma(N-n_H+1)}{\Gamma(N+2)} \\
&= \binom{N}{n_H} \frac{n_H!(N-n_H)!}{(N+1)!} \\
&= \frac{n_H!(N-n_H)!}{(N+1)n_H!(N-n_H)!} = \frac{1}{N+1}
\end{aligned}$$

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,

$$\begin{aligned}
\mathrm{E}[f_H] &= \int_0^1 f_H P(f_H|n_H, N) df_H \\
&= \frac{(N+1)!}{n_H!(N-n_H)!} \int_0^1 f_H^{n_H+1}(1-f)^{N-n_H} df_H \\
&= \frac{(N+1)!}{n_H!(N-n_H)!} \frac{(n_H+1)!(N-n_H)!}{(N+2)!} \\
&= \frac{n_H+1}{N+2}
\end{aligned}$$

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

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

Now,

$$\begin{aligned}
f(\sum_{i=1}^{I} p_i x_i) &= f(p_1 x_1 + \sum_{i=1}^{I} p_i x_i) \\
&\leq p_1 f(x_1) + [\sum_{i=2}^{I} p_i][f(\frac{\sum_{i=2}^{I} p_i x_i}{\sum_{i=2}^{I} p_i})] \quad (\because \text{Definition of convex function}) \\
&\leq p_1 f(x_1) + [\sum_{i=2}^{I} p_i][\frac{p_2}{\sum_{i=2}^{I} p_i} f(x_2) + \frac{\sum_{i=3}^{I} p_i}{\sum_{i=2}^{I} p_i} f(\frac{\sum_{i=3}^{I} p_i x_i}{\sum_{i=3}^{I} p_i})] \\
&= p_1 f(x_1) + p_2 f(x_2) + \frac{\sum_{i=3}^{I} p_i}{\sum_{i=2}^{I} p_i} f(\frac{\sum_{i=3}^{I} p_i x_i}{\sum_{i=3}^{I} p_i}) \\
&= \ldots \\
&\leq p_1 f(x_1) + p_2 f(x_2) + p_3 f(x_3) + p_4 f(x_4) + \ldots \\
&= \sum_{i=1}^{I} p_i f(x_i) \quad \blacksquare
\end{aligned}$$

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)

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

$P(S = s)$ 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.

Exercise 2.17

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:

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

Now, with tanh,

$$\begin{aligned}
p &= \frac{1}{1 + e^{-u}} \\
&= \frac{1}{2} \frac{2e^{\frac{u}{2}}}{e^{\frac{u}{2}} + e^{-\frac{u}{2}}} \\
&= \frac{1}{2} \left(\frac{e^{\frac{u}{2}} - e^{-\frac{u}{2}}}{e^{\frac{u}{2}} + e^{-\frac{u}{2}}} + 1 \right) \\
&= \frac{1}{2} \tanh\left(\frac{u}{2}\right) + \frac{1}{2}
\end{aligned}$$

For base 2,

$$\begin{aligned}
& b = \log_2 \frac{p}{q} \\
\iff & 2^b = \frac{p}{1 - p} \\
\iff & 2^b = \frac{p}{q} \\
\iff & (2^b + 1)p = 2^b \\
\iff & p = \frac{2^b}{2^b + 1} = \frac{1}{1 + 2^{-b}}
\end{aligned}$$

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

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

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

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

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:

$$\begin{aligned}\mathcal{E}\left[f(x)\right] &= \sum_x P(x)f(x) \\ &= 0.1 \cdot 10 + 0.2 \cdot 5 + 0.7 \cdot \frac{10}{7} \\ &= 3\end{aligned}$$

$$\begin{aligned}\mathcal{E}\left[\frac{1}{P(x)}\right] &= \sum_x P(x)\frac{1}{P(x)} \\ &= 1 + 1 + 1 = 3\end{aligned}$$

Exercise 2.22

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

Solution:

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

Exercise 2.23

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:

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

Exercise 2.24

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,

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

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:

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

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 $D_{\mathrm{KL}}(P||Q) \geq 0$ (Gibbs' inequality) with equality only if $P = Q$.

Solution:

$$\begin{aligned} \sum_x P(X) \log \frac{P(X)}{Q(X)} &= \mathrm{E}\left[-\log \frac{Q(X)}{P(X)}\right] \\ &\leq -\log \mathrm{E}\left[\frac{Q(X)}{P(X)}\right] \\ &= -\log \sum_x Q(X) \\ &= -\log 1 = 0 \end{aligned}$$

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

$$
\begin{aligned}
H(p) &= \sum_{i=1}^{I} p_i \log \frac{1}{p_i} \\
&= \sum_{i=1}^{m} (p_1 \sim p_m) \frac{p_i}{p_1 \sim p_m} \log \frac{1}{p_i} + \sum_{i=m+1}^{I} (p_{m+1} \sim p_I) \frac{p_i}{p_{m+1} \sim p_I} \log \frac{1}{p_i} \\
&= (p_1 \sim p_m) \sum_{i=1}^{m} \frac{p_i}{p_1 \sim p_m} \log \frac{p_1 \sim p_m}{p_i} + \sum_{i=1}^{m} (p_1 \sim p_m) \frac{p_i}{p_1 \sim p_m} \log \frac{1}{p_1 \sim p_m} \\
&+ \sum_{i=m+1}^{I} \frac{(p_{m+1} \sim p_I) p_i}{p_{m+1} \sim p_I} \log \frac{p_{m+1} \sim p_I}{p_i} + \sum_{i=m+1}^{I} \frac{(p_{m+1} \sim p_I) p_i}{p_{m+1} \sim p_I} \log \frac{1}{p_{m+1} \sim p_I} \\
&= (p_1 \sim p_m) H(\frac{p_1}{p_1 \sim p_m}, \ldots, \frac{p_m}{p_1 \sim p_m}) + (p_{m+1} \sim p_I) H(\frac{p_{m+1}}{p_{m+1} \sim p_I}, \ldots, \frac{p_I}{p_{m+1} \sim p_I}) \\
&+ H((p_1 \sim p_m), (p_{m+1} \sim p_I))
\end{aligned}
$$

Exercise 2.28

Solution:

Exercise

Solution: