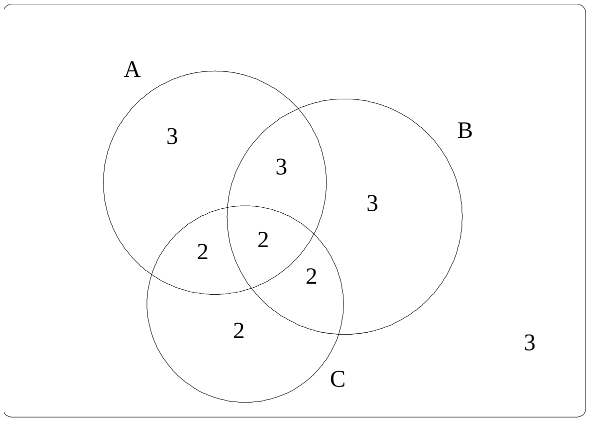
BIOINFORMATICS & COMPUTATIONAL GENETICS MSc
PROBABILITY AND STATISTICS

SOLUTIONS

1. Events A="even numbered", B="blue ball", C="number less than 5".



(1) $A \cap B \cap C$ - "blue ball 2 or 4" : Prob = 2/20.
(2) $B \cap C'$ - "blue ball 5,6,7,8,9,19" : Prob = 6/20.
(3) $A \cup B \cup C$ - "at least one of blue, or even, or less than 5" : Prob = 17/20.
(4) $A \cap (B \cup C)$ - "even, and either blue or ess than 5" : Prob = 7/20.
(5) $A' \cap B' \cap C'$ - "red 5,7,9" - Prob = 3/20.

2. Table 1:

$$\left.\begin{aligned}
\mathrm{P}(E) &= \mathrm{P}(E \cap F) + \mathrm{P}(E \cap F') &&= \frac{6}{24} + \frac{2}{24} = \frac{8}{24} = \frac{1}{3} \\
\mathrm{P}(F) &= \mathrm{P}(E \cap F) + \mathrm{P}(E' \cap F) &&= \frac{6}{24} + \frac{12}{24} = \frac{18}{24} = \frac{3}{4} \\
\mathrm{P}(F|E) &= \frac{\mathrm{P}(E \cap F)}{\mathrm{P}(F)} &&= \frac{6/24}{8/24} = \frac{3}{4} = \mathrm{P}(F) \\
\mathrm{P}(F|E') &= \frac{\mathrm{P}(E' \cap F)}{\mathrm{P}(E')} &&= \frac{12/24}{16/24} = \frac{3}{4} = \mathrm{P}(F)
\end{aligned}\right\} \Longrightarrow \text{INDEPENDENT}$$

Table 2:

$$\left.\begin{aligned}
\mathrm{P}(E) &= \mathrm{P}(E \cap F) + \mathrm{P}(E \cap F') &&= \frac{6}{28} + \frac{4}{28} = \frac{10}{28} = \frac{5}{14} \\
\mathrm{P}(F) &= \mathrm{P}(E \cap F) + \mathrm{P}(E' \cap F) &&= \frac{6}{28} + \frac{12}{28} = \frac{18}{28} = \frac{9}{14} \\
\mathrm{P}(F|E) &= \frac{\mathrm{P}(E \cap F)}{\mathrm{P}(F)} &&= \frac{6/28}{10/28} = \frac{3}{5} \neq \mathrm{P}(F) \\
\mathrm{P}(F|E') &= \frac{\mathrm{P}(E' \cap F)}{\mathrm{P}(E')} &&= \frac{12/28}{18/28} = \frac{2}{3} \neq \mathrm{P}(F)
\end{aligned}\right\} \Longrightarrow \begin{matrix}\text{NOT}\\ \text{INDEPENDENT}\end{matrix}$$

Check also that $\mathrm{P}(E \cap F) = \mathrm{P}(E)\mathrm{P}(F|E)$ etc. for both tables.

3. $\mathrm{P}(F) = \mathrm{P}(A \cap B \cap C \cap D \cap E) = \mathrm{P}(A)\mathrm{P}(B|A)\mathrm{P}(C|A \cap B)\mathrm{P}(D|A \cap B \cap C)\mathrm{P}(E|A \cap B \cap C \cap D)$

(a) A, B, C, D, E independent, $\mathrm{P}(A) = \mathrm{P}(B) = \mathrm{P}(C) = \mathrm{P}(D) = \mathrm{P}(E) = 0.90$. so

$$\mathrm{P}(F) = (0.9)^5 = 0.5905$$

(b) Theorem of Total Probability:

$$\mathrm{P}(B) = \mathrm{P}(B|A)\mathrm{P}(A) + \mathrm{P}(B|A')\mathrm{P}(A') = (0.95 \times 0.9) + (0.75 \times 0.1) = 0.93.$$

(c) $\mathrm{P}(F) = \mathrm{P}((A \cup B) \cap C \cap D \cap E) = \mathrm{P}(A \cup B)\mathrm{P}(C)\mathrm{P}(D)\mathrm{P}(E)$ where

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$$

(i) A, B independent, so $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B) \implies \mathrm{P}(A \cup B) = 0.9 + 0.9 - 0.9^2 = 0.99$,

$$\implies \mathrm{P}(F) = 0.99 \times (0.9)^3 = 0.7217$$

(ii) A, B not independent, $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B|A) \implies \mathrm{P}(A \cup B) = 0.9 + 0.93 - 0.90 \times 0.95 = 0.975$,

$$\implies \mathrm{P}(F) = 0.975 \times (0.9)^3 = 0.7108$$

4. Given that $\mathrm{P}(S) = 0.001, \mathrm{P}(T|S) = 1.0, \mathrm{P}(T|S') = 0.1$.

(i) Theorem of Total Probability:

$$\mathrm{P}(T) = \mathrm{P}(T|S)\mathrm{P}(S) + \mathrm{P}(T|S')\mathrm{P}(S') = (1.0 \times 0.001) + (0.01 \times 0.999) = 0.1009.$$

(ii) Let $I =$ "incorrect diagnosis"; then $I = (T \cap S') \cup (T' \cap S)$,

$$\implies \mathrm{P}(I) = \mathrm{P}(T \cap S') + \mathrm{P}(T' \cap S) = \mathrm{P}(T|S')\mathrm{P}(S') + \mathrm{P}(T|S)\mathrm{P}(S) = (0.1 \times 0.999) + (0.0 \times 0.001) = 0.0999.$$

(iii) Bayes Theorem:

$$\mathrm{P}(S|T) = \frac{\mathrm{P}(T|S)\mathrm{P}(S)}{\mathrm{P}(T)} = \frac{1.0 \times 0.001}{0.1009} = 0.0099$$

5. First let $E =$ "sequence of n heads", $F=$ "coin is normal". Want

$$\mathrm{P}(F|E) = \frac{\mathrm{P}(E|F)\mathrm{P}(F)}{\mathrm{P}(E)} = \frac{\mathrm{P}(E|F)\mathrm{P}(F)}{\mathrm{P}(E|F)\mathrm{P}(F) + \mathrm{P}(E|F')\mathrm{P}(F')}$$

and here $\mathrm{P}(F) = 1/2$, $\mathrm{P}(E|F) = (1/2)^n$, $\mathrm{P}(E|F') = 1$, so

$$\mathrm{P}(F|E) = \frac{1}{1 + 2^n}$$

Now let $E =$ "sequence of k heads and $n - k$ tails. Now

$$\mathrm{P}(E|F) = \binom{n}{k}(1/2)^n \qquad \mathrm{P}(E|F') = \binom{n}{k}p^k(1-p)^{n-k}$$

so

$$\mathrm{P}(E) = \mathrm{P}(E|F)\mathrm{P}(F) + \mathrm{P}(E|F')\mathrm{P}(F') = \binom{n}{k}(1/2)^n \times 1/2 + \binom{n}{k}p^k(1-p)^{n-k} \times 1/2$$

and

$$P(F|E) = \frac{1}{1 + 2^n p^k (1-p)^{n-k}}$$

6. Let A, B and C represent the events that the prize lies behind the selected door, the opened door, and the remaining door respectively, and let H_B represent the event that the host opens door the second door. Want to compare the conditional probabilities of A and C given H_B. Now,

$$P(A|H_B) = \frac{P(H_B|A)P(A)}{P(H_B)} = \frac{P(H_B|A)P(A)}{P(H_B|A)P(A) + P(H_B|B)P(B) + P(H_B|C)P(C)}.$$

Here $P(A) = P(B) = P(C) = 1/3$, and $P(H_B|A) = 1/2$ (host can choose to open B or C), $P(H_B|B) = 0$ (host cannot open B if the prize is behind it) and $P(H_B|C) = 1$ (host must open B, as A is already selected). Hence

$$P(A|H_B) = \frac{1/2 \times 1/3}{1/2 \times 1/3 + 0 \times 1/3 + 1 \times 1/3} = 1/3.$$

Similarly, $P(C|H_B) = 2/3$, so it is **always** probabilistically advantageous to SWITCH doors as $P(A|H_B) < P(C|H_B)$.

7. (a) X="operating (0), not operating (1)": $\mathbb{X} = \{0, 1\}$
(b)X="number of defective components": $\mathbb{X} = \{0, 1, 2, ..., 100\}$.
(c)X="number of failed components": $\mathbb{X} = \{0, 1, 2, ...\}$.
(d)X="time at which 100th component fails": $\mathbb{X} = \{x : x \geq 0\}$.

8. (a) OK
(b) Not a probability distribution ($f_X(2) < 0$)
(c) OK
(d) Not a probability distribution (doesn't sum to 1).

9. (a) OK
(b) OK
(c) Not a probability distribution (integral is not finite)
(d) OK

10. $f_X(0) = F_X(0) = 1/2$, and for $x = 1, 2, 3, ...$

$$f_X(x) = F_X(x) - F_X(x-1) = (1 - (1/2)^{x+1}) - (1 - (1/2)^x) = (1/2)^{x+1}$$

Finally

$$P[X \text{ is even }] = P[X = 0] + P[X = 2] + P[X = 4] + ... = (1/2) + (1/2)^3 + (1/2)^5 + ...$$

as $f_X(x) = P[X = x]$. This is an infinite geometric progression with common term $(1/2)^2$. Hence

$$P[X \text{ is even }] = 1/2 \times \frac{1}{1 - (1/2)^2} = 2/3$$

11. Need $\int_0^1 f_X(x)\ dx = 1 \Longrightarrow \int_0^1 c(1-x)x^2 dx = 1 \Longrightarrow c\left[\ x^3/3 - x^4/4\right]_0^1 = 1 \Longrightarrow c = 12.$

CDF obtained by integrating the density function:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_0^x 12(1-t)t^2 dt = 4x^3 - 3x^4 \qquad 0 \le x \le 1$$

and hence $\mathrm{P}[1/4 < X < 3/4] = F_X(3/4) - F_X(1/4) = \dfrac{176}{256}$

12. (a) By integration, $c = 1/(\alpha - \beta)$, that is, need

$$\int_0^1 f_X(x)\ dx = 1$$

(b) By differentiation

$$f_X(x) = \alpha \sec^2 x \exp\{-\alpha \tan x\} \qquad 0 \le x \le 2\pi$$

(c) By differentiation

$$f_X(x) = \frac{2\beta(\beta^2 - x^2)}{(\beta^2 + x^2)^2} \qquad 0 \le x \le \beta$$

13. X_1=sum of scores; takes values on $\{2, 3, ..., 12\}$. X_2=largest of the two scores; takes values on $\{1, 2, 3, 4, 5, 6\}$. Joint distribution specified by

		X_1											
		2	3	4	5	6	7	8	9	10	11	12	Tot
	1	p_1	0	0	0	0	0	0	0	0	0	0	p_1
	2	0	p_2	p_1	0	0	0	0	0	0	0	0	p_1+p_2
	3	0	0	p_2	p_2	p_1	0	0	0	0	0	0	p_1+2p_2
X_2	4	0	0	0	p_2	p_2	p_2	p_1	0	0	0	0	p_1+3p_2
	5	0	0	0	0	p_2	p_2	p_2	p_2	p_1	0	0	p_1+4p_2
	6	0	0	0	0	0	p_2	p_2	p_2	p_2	p_2	p_1	p_1+5p_2
Tot		p_1	p_2	p_1+p_2	$2p_2$	p_1+2p_2	$3p_2$	p_1+2p_2	$2p_2$	p_1+p_2	p_2	p_1	1

where $p_1 = 1/36$, $p_2 = 2/36$. Marginal for X_1 given by column totals, so that

x	2	3	4	5	6	7	8	9	10	11	12
$f_{X_1}(x)$	p_1	p_2	p_1+p_2	$2p_2$	p_1+2p_2	$3p_2$	p_1+2p_2	$2p_2$	p_1+p_2	p_2	p_1

Marginal for X_2 given by row totals, so that

x	1	2	3	4	5	6
$f_{X_2}(x)$	p_1	p_1+p_2	p_1+2p_2	p_1+3p_2	p_1+4p_2	p_1+5p_2

Conditional for X_1, given $X_2 = 4$ given by the fourth row of the table, divided by the fourth row total.

x	0	1	2	3	4	5	6	7	8	9	10
$f_X(x)$	0.107	0.268	0.302	0.201	0.088	0.026	0.001	0.000	0.000	0.000	0.000

14. $X \sim Bin(10, 0.2)$, so $\mathbb{X} = \{0, 1, 2, ..., 10\}$, $f_X(x) = \text{P}[X = x] = \binom{10}{x}(0.2)^x(0.8)^{10-x}$

15. X="number of heads", so $X \sim Bin(5, 0.5)$
(i) $\text{P}[X = 0] = (0.5)^5 = 0.03125$
(ii) $\text{P}[X \geq 4] = \text{P}[X = 4] + \text{P}[X = 5] = \binom{5}{4}(0.5)^5 + \binom{5}{5}(0.5)^5 = 0.1875$
(iii) $\text{P}[\text{Odd number of heads}] = \text{P}[X = 1] + \text{P}[X = 3] + \text{P}[X = 5] = 0.5$

16. X="number of defectives", so $X \sim Bin(50, 0.01)$
(i) $\text{P}[X \leq 2] = \text{P}[X = 0] + \text{P}[X = 1] + \text{P}[X = 2]$, so

$$\text{P}[X \leq 2] = \binom{50}{0}(0.01)^0(0.99)^{50} + \binom{50}{1}(0.01)^1(0.99)^{49} + \binom{50}{2}(0.01)^2(0.99)^{48} = 0.9862$$

(ii) If X_1 is the Bernoulli random variable corresponding to component 1, want $\text{P}[X_1 = 1|X = 3]$.

$$\text{P}[X_1 = 1|X = 3] = \frac{\text{P}[X = 3|X_1 = 1]\text{P}[X_1 = 1]}{\text{P}[X = 3]} = \frac{\text{P}[Y = 2]\text{P}[X_1 = 1]}{\text{P}[X = 3]}$$

where Y is the random variable corresponding to the total number of defectives in components 2,...,49. Now $Y \sim Bin(49, 0.01)$, so

$$\text{P}[X_1 = 1|X = 3] = \frac{\binom{49}{2}(0.01)^2(0.99)^{47} \times 0.01}{\binom{50}{3}(0.01)^3(0.99)^{47}} = \frac{3}{50}$$

17. $X \sim Poisson(\lambda) \Longrightarrow p_0 = e^{-\lambda}, \; p_x = \dfrac{\lambda}{x}p_{x-1}$

x	0	1	2	3	4	5	6	7	8	9	10
$f_X(x)$	0.007	0.034	0.084	0.140	0.175	0.175	0.146	0.104	0.065	0.036	0.018

18. Must check pdf is non-negative, and $\displaystyle\int_0^\infty f_X(x)\,dx = 1$; density decays exponentially from maximum value λ at $x = 0$.
Can show $F_X(x) = \text{P}[X \leq x] = 1 - e^{-\lambda x}$ for $x \geq 0$, so

$$\text{P}[X \leq 1] = 1 - e^{-2} = 0.8647 \qquad \text{P}[X > 4] = 1 - \text{P}[X \leq 4] = e^{-8} = 3.35 \times 10^{-4}$$

19. To compute expectation and variance, need to integrate

$$\mathrm{E}_{f_X}[X] = \int_{-\infty}^{\infty} x f_X(X)\ dx \qquad \mathrm{E}_{f_X}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(X)\ dx.$$

This is achieved by changing variables in the integrals from x to $t = (x-\mu)/\sigma$. The result is finally obtained by noting that

$$\int_{-\infty}^{\infty} e^{-t^2/2}\ dt = \sqrt{2\pi} \qquad \text{(as the integrand is proportional to the standard normal pdf)}$$

$$\int_{-\infty}^{\infty} te^{-t^2/2}\ dt = 0 \qquad \text{(as the integrand is an ODD function)}$$

$$\int_{-\infty}^{\infty} t^2 e^{-t^2/2}\ dt = \sqrt{2\pi} \qquad \text{(using a further substitution } u = t^2\text{)}$$

Hence $\mathrm{E}_{f_X}[X] = \mu \qquad \mathrm{E}_{f_X}[X^2] = \mu^2 + \sigma^2 \Longrightarrow \mathrm{Var}_{f_X}[X] = \sigma^2$.

20. $F_Y(y) = \mathrm{P}[Y \leq y] = \mathrm{P}[\sigma X + \mu \leq y] = \mathrm{P}[X \leq (y-\mu)/\sigma] = \Phi\left(\dfrac{y-\mu}{\sigma}\right)$, and so by differentiation

$$f_Y(y) = \frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}$$

where Φ and ϕ are the cdf and pdf of the standard normal distribution.

If $X \sim N(\mu, \sigma^2)$, $F_X(x) = \Phi\left(\dfrac{x-\mu}{\sigma}\right)$. Note also that

$$\mathrm{P}[X > x] = 1 - \mathrm{P}[X \leq x] \qquad \mathrm{P}[x_1 < X \leq x_2] = \mathrm{P}[X \leq x_2] - \mathrm{P}[X \leq x_1]$$

and for $x < 0$,

$$\Phi(x) = 1 - \Phi(-x).$$

Then

	$\mu = 0, \sigma^2 = 1$	$\mu = 0.5, \sigma^2 = 2.$
(i) $\mathrm{P}[X \leq 2.2]$	0.9861	0.8023
(ii) $\mathrm{P}[X > -0.75]$	0.7734	0.6368
(iii) $\mathrm{P}[-1.7 \leq X < -1]$	0.1141	0.0909
(iv) $\mathrm{P}[X > 1.96]$	0.0250	0.2327

$Y \sim \chi^2_1 \Longleftrightarrow Y = X^2$ where $X \sim N(0,1)$. Hence

$$\mathrm{P}[\ Y > y\ ] = \mathrm{P}[\ X^2 > y\ ] = 1 - \mathrm{P}[\ -\sqrt{y} < X < \sqrt{y}\ ] = 2(1 - \Phi(\sqrt{y}))$$