1. The derived result shows that the pdf of Y is given by

$$f_Y(y) = f_X(g^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)$$

for $y \in \mathbb{R}$ and for strictly increasing g. Since $y = g(x)$ is one-to-one, we also have

$$\begin{aligned}
\mathbb{E}[Y] &= \int_{-\infty}^{\infty} y f_Y(y)\,\mathrm{d}y && \text{(from definition)}\\
&= \int_{-\infty}^{\infty} y f_X(g^{-1}(y))\,\mathrm{d}g^{-1}(y) && \text{(from the derived result and substitution)}\\
&= \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x
\end{aligned}$$

as desired.

2. The pdf will be obtained by the Divide-and-Conquer approach, which is to split the domain of X into several pieces and make $Y = g(X) = X^2$ piece-wise bijective. In particular, we split $y = g(x)$ as follows:

   - For $x < 0$, denote $y = g_1(x) = x^2$, and it follows that $g_1^{-1}(y) = -\sqrt{y}$ and that
   $$\frac{\mathrm{d}}{\mathrm{d}y}g_1^{-1}(y) = -\frac{1}{2\sqrt{y}}, \quad \text{where } y > 0.$$
   - For $x \geq 0$, denote $y = g_2(x) = x^2$, and it follows that $g_2^{-1}(y) = \sqrt{y}$ and that
   $$\frac{\mathrm{d}}{\mathrm{d}y}g_2^{-1}(y) = \frac{1}{2\sqrt{y}}, \quad \text{where } y > 0.$$

   Then combining the results will give the required pdf:

$$\begin{aligned}
f_Y(y) &= \sum_{i=1}^{2} f_X(g_i^{-1}(y))\left|\frac{\mathrm{d}}{\mathrm{d}y}g_i^{-1}(y)\right|\\
&= \frac{1}{2\sqrt{y}}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{(-\sqrt{y})^2}{2}} + \frac{1}{\sqrt{2\pi}}e^{-\frac{\sqrt{y}^2}{2}}\right) = \frac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}}
\end{aligned}$$

   for $y > 0$, and $f_Y(y) = 0$ elsewhere. Consequently,

$$\begin{aligned}
\mathbb{E}[Y] &= \int_0^{\infty} \frac{y}{\sqrt{2\pi y}}e^{-\frac{y}{2}}\,\mathrm{d}y\\
&= \sqrt{y}\left(\frac{-2}{\sqrt{2\pi}}e^{-\frac{y}{2}}\right)\Big|_0^{\infty} + \int_0^{\infty} \underbrace{\frac{1}{2\sqrt{y}}\frac{2}{\sqrt{2\pi}}e^{-\frac{y}{2}}}_{\text{density of } Y}\,\mathrm{d}y\\
&= 0 + 1 = 1
\end{aligned}$$

   and

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = \int_0^{\infty} \frac{y^2}{\sqrt{2\pi y}}e^{-\frac{y}{2}}\,\mathrm{d}y - 1\\
&= y^{\frac{3}{2}}\left(\frac{-2}{\sqrt{2\pi}}e^{-\frac{y}{2}}\right)\Big|_0^{\infty} + \underbrace{\int_0^{\infty} \frac{3\sqrt{y}}{2}\frac{2}{\sqrt{2\pi}}e^{-\frac{y}{2}}\,\mathrm{d}y}_{3\,\mathbb{E}[Y]} - 1\\
&= 0 + 3 - 1 = 2.
\end{aligned}$$

We avoid directly calling the properties of the $\chi^2(1)$ distribution or the gamma function as they are not mentioned in the first chapter.

3. (a) Our attempt is to switch the order of summations appropriately:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{j=0}^{\infty} j\,\mathbb{P}(X=j) && \text{(from definition)} \\
&= \sum_{j=0}^{\infty}\sum_{k=0}^{j-1} \mathbb{P}(X=j) \\
&= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \mathbb{P}(X=j)\mathbb{1}\{k<j\} \\
&= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \mathbb{P}(X=j)\mathbb{1}\{j>k\} && \text{(Fubini's theorem)} \\
&= \sum_{k=0}^{\infty} \mathbb{P}(X>k) = \sum_{k=0}^{\infty}(1-F_X(k)).
\end{aligned}$$

(b) The steps are almost the same as those in (a), except that summations are replaced by integrations:

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty t\,f(t)\,\mathrm{d}t && \text{(from definition)} \\
&= \int_0^\infty \int_0^t f(t)\,\mathrm{d}x\,\mathrm{d}t \\
&= \int_0^\infty \int_0^\infty f(t)\mathbb{1}\{x<t\}\,\mathrm{d}x\,\mathrm{d}t \\
&= \int_0^\infty \int_0^\infty f(t)\mathbb{1}\{t>x\}\,\mathrm{d}t\,\mathrm{d}x && \text{(Fubini's theorem)} \\
&= \int_0^\infty \int_x^\infty f(t)\,\mathrm{d}t\,\mathrm{d}x = \int_0^\infty (1-F_X(x))\,\mathrm{d}x.
\end{aligned}$$

(c) Denote $X^+ = \max\{X,0\} \geq 0$ and $X^- = \max\{0,-X\} \geq 0$. It follows that $X = X^+ - X^-$ and hence we have

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X^+] - \mathbb{E}[X^-] \\
&= \int_0^\infty (1-\mathbb{P}(X^+\leq x))\,\mathrm{d}x - \int_0^\infty (1-\mathbb{P}(X^-\leq y))\,\mathrm{d}y && \text{(from (b))} \\
&= \int_0^\infty (1-\mathbb{P}(X\leq x))\,\mathrm{d}x - \int_0^\infty (1-\mathbb{P}(-X\leq y))\,\mathrm{d}y \\
& && \text{(for } a\geq 0,\ \max\{0,Y\}\leq a \iff Y\leq a) \\
&= \int_0^\infty (1-F_X(x))\,\mathrm{d}x - \int_0^\infty \mathbb{P}(X<-y)\,\mathrm{d}y \\
&= \int_0^\infty (1-F_X(x))\,\mathrm{d}x - \int_{-\infty}^0 F_X(x)\,\mathrm{d}x && \text{(substitution with } y=-x)
\end{aligned}$$

as desired. Note that $F_X(x) = \mathbb{P}(X\leq x) = \mathbb{P}(X<x)$ as X is continuous.

(d) Suppose (X, Y) has a joint density $f_{X,Y}(x, y)$ for $x, y \in \mathbb{R}^+$. Then the proof is nothing more than applying the previous tricks again:

$$
\begin{aligned}
\mathbb{E}[XY] &= \int_0^\infty \int_0^\infty st\, f_{X,Y}(s,t)\,\mathrm{d}s\,\mathrm{d}t \\
&= \int_0^\infty \int_0^\infty \left(\int_0^\infty \int_0^\infty f_{X,Y}(s,t)\mathbb{1}\{x < s, y < t\}\,\mathrm{d}x\,\mathrm{d}y \right) \mathrm{d}s\,\mathrm{d}t \\
&= \int_0^\infty \int_0^\infty \left(\int_0^\infty \int_0^\infty f_{X,Y}(s,t)\mathbb{1}\{s > x, t > y\}\,\mathrm{d}s\,\mathrm{d}t \right) \mathrm{d}x\,\mathrm{d}y \\
&= \int_0^\infty \int_0^\infty \mathbb{P}(X > x, Y > y)\,\mathrm{d}x\,\mathrm{d}y.
\end{aligned}
$$

Remark. Fubini's Theorem indeed grants us that $\int_0^\infty \int_0^\infty \mathbb{E}[\mathbb{1}\{X > x, Y > y\}]\,\mathrm{d}x\,\mathrm{d}y = \mathbb{E}[\int_0^\infty \int_0^\infty \mathbb{1}\{X > x, Y > y\}\,\mathrm{d}x\,\mathrm{d}y]$, so the proof can be generalized and the assumption of $f_{X,Y}(x, y)$ is not necessary, but have been made here to cater the scope of this course.

4. To show pairwise independence, it suffices to show the independence of the (X, Y) pair, as the remaining pairs could be shown via symmetry. Firstly, consider their joint density:

$$
f_{X,Y}(x,y) = \int_0^{2\pi} f(u,v,z)\,\mathrm{d}z = \frac{1}{4\pi^2}
$$

for $0 \le x, y \le 2\pi$. Then we can obtain the marginal density of X using

$$
f_X(x) = \int_0^{2\pi} f_{X,Y}(x,y)\,\mathrm{d}y = \frac{1}{2\pi}
$$

for $0 \le x \le 2\pi$, and likewise, $f_Y(y) = \frac{1}{2\pi}$ for $0 \le y \le 2\pi$. Note that all the densities above take value 0 at unspecified domains. Hence, it can be easily seen that $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for all $x, y \in \mathbb{R}$, which implies the independence of (X, Y). By symmetry, we conclude that X, Y and Z are pairwise independent.

But (X, Y, Z) are not independent, as the conditional distribution of (X, Z) given Y,

$$
f_{X,Z|Y}(x,z|y) = \frac{f(x,y,z)}{f_Y(y)} = \frac{1 - \sin x \sin y \sin z}{4\pi^2}
$$

for $0 \le x, z \le 2\pi$, is clearly not free of y.

5. Recall that if X follows exponential distribution with rate λ, then it has pdf $f_X(x) = \lambda e^{-\lambda x}$ and cdf $\mathbb{P}(X \le x) = 1 - e^{-\lambda x}$ for $x > 0$.

(a) We attempt to define the joint cdf of $F_{Z,W}(z, w)$ for $(z, w) \in \mathbb{R}^2$.

- If $w < 0$ or $z < 0$, then $F_{Z,W}(z, w) = \mathbb{P}(Z \le z, W \le w) = 0$.
- If $w \ge 1$ and $z \ge 0$, then

$$
\begin{aligned}
\mathbb{P}(Z \le z, W \le w) &= \mathbb{P}(Z \le z) = 1 - \mathbb{P}(\min\{X, Y\} > z) \\
&= 1 - \mathbb{P}(X > z)\mathbb{P}(Y > z) \\
&= 1 - e^{-\lambda z} e^{-\mu z} = 1 - e^{-(\mu+\lambda)z}.
\end{aligned}
$$

- If $0 \le w < 1$ and $z \ge 0$, then

$$\begin{aligned}
\mathbb{P}(Z \le z, W \le w) &= \mathbb{P}(Z \le z, W = 0) = \mathbb{P}(Y \le z, Y \le X) \\
&= \mathbb{P}(Y \le X \le z) + \mathbb{P}(Y \le z)\mathbb{P}(X > z) \\
&= \int_0^z \mu e^{-\mu y} \int_y^z \lambda e^{-\lambda x} \,\mathrm{d}x \,\mathrm{d}y + (1 - e^{-\mu z})e^{-\lambda z} \\
&= \frac{\mu}{\mu + \lambda} \left[1 - e^{-(\mu+\lambda)z}\right].
\end{aligned}$$

Combining the results gives us the desired cdf

$$F_{Z,W}(z, w) = \begin{cases} 0, & w < 0 \text{ or } z < 0 \\ \frac{\mu}{\mu+\lambda}\left[1 - e^{-(\mu+\lambda)z}\right], & 0 \le w < 1, z \ge 0 \\ 1 - e^{-(\mu+\lambda)z}, & w \ge 1, z \ge 0 \end{cases}.$$

(b) We have derived in (a) that $F_Z(z) = 1 - e^{-(\lambda+\mu)z}$ for $z \ge 0$, i.e. Z is also exponentially distributed, but with rate $\mu + \lambda$. On the other hand,

$$F_W(w) = F_{Z,W}(\infty, w) = \begin{cases} 0, & w < 0 \\ \frac{\mu}{\mu+\lambda}, & 0 \le w < 1 \\ 1, & w \ge 1 \end{cases},$$

i.e. W is a Bernoulli random variable with $\mathbb{P}(W = 0) = \frac{\mu}{\mu+\lambda}$ and $\mathbb{P}(W = 1) = \frac{\lambda}{\mu+\lambda}$.

(c) Using the results in (a) and (b), it is clear that $F_{Z,W}(z, w) = F_Z(z)F_W(w)$ for any $(z, w) \in \mathbb{R}^2$, which concludes the independence of Z and W.

6. (a) By the linearity and the law of total expectation, we have

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n \mathbb{E}[\mathbb{E}[X_i|P_i]] \\
&= \sum_{i=1}^n \mathbb{E}[P_i] = \sum_{i=1}^n \frac{\alpha}{\alpha + \beta} = \frac{n\alpha}{\alpha + \beta}.
\end{aligned}$$

(b) By the independence of X_i and the law of total variance, we have

$$\begin{aligned}
\mathrm{Var}(Y) &= \sum_{i=1}^n \mathrm{Var}(X_i) = \sum_{i=1}^n (\mathbb{E}[\mathrm{Var}(X_i|P_i)] + \mathrm{Var}(\mathbb{E}[X_i|P_i])) \\
&= \sum_{i=1}^n [\mathbb{E}[P_i(1 - P_i)] + \mathrm{Var}(P_i)] \\
&= \sum_{i=1}^n \left[\frac{\alpha}{\alpha + \beta} - \left(\frac{\alpha}{\alpha + \beta}\right)^2\right] = \sum_{i=1}^n \frac{\alpha\beta}{(\alpha + \beta)^2} = \frac{n\alpha\beta}{(\alpha + \beta)^2}.
\end{aligned}$$

Consider the distribution of each X_i. Since

$$\begin{aligned}
\mathbb{P}(X_i = 1) &= \int_0^1 \mathbb{P}(X_i = 1|P_i = x) f_{P_i}(x) \,\mathrm{d}x \\
&= \int_0^1 x f_{P_i}(x) \,\mathrm{d}x = \frac{\alpha}{\alpha + \beta}, \\
\mathbb{P}(X_i = 0) &= 1 - \mathbb{P}(X_i = 1)
\end{aligned}$$

for $i = 1, \cdots, n$, we conclude that each $X_i \sim \text{Bernoulli}\left(\frac{\alpha}{\alpha+\beta}\right)$ are iid. As a consequence, their sum $Y \sim \text{Binomial}\left(n, \frac{\alpha}{\alpha+\beta}\right)$, as expected.

(c) Apply the law of total expectation/variance with the properties of binomial distribution.

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{i=1}^{k} \mathbb{E}[X_i] = \sum_{i=1}^{k} \mathbb{E}[\mathbb{E}[X_i|P_i]] \\
&= \sum_{i=1}^{k} \mathbb{E}[n_i P_i] = \frac{\alpha}{\alpha+\beta} \sum_{i=1}^{k} n_i, \\
\mathrm{Var}(Y) &= \sum_{i=1}^{k} \mathrm{Var}(X_i) = \sum_{i=1}^{k} \left(\mathbb{E}[\mathrm{Var}(X_i|P_i)] + \mathrm{Var}(\mathbb{E}[X_i|P_i])\right) \\
&= \sum_{i=1}^{k} \left[\mathbb{E}[n_i P_i(1-P_i)] + \mathrm{Var}(n_i P_i)\right] \\
&= \sum_{i=1}^{k} \frac{n_i[\alpha(\alpha+\beta)(\alpha+\beta+1) - \alpha^2(\alpha+\beta+1) - \alpha\beta] + n_i^2 \alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
&= \sum_{i=1}^{k} \frac{n_i \alpha\beta(\alpha+\beta) + n_i^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\alpha\beta \sum_{i=1}^{k} n_i(\alpha+\beta+n_i)}{(\alpha+\beta)^2(\alpha+\beta+1)}.
\end{aligned}$$