Solutions to Exercises of Chapters 3, 4 and 5

CHAPTER 3

3.1 The σ-algebra $\sigma(X)$ consists of all subsets of $X(\Omega)$. A set with n elements has precisely 2^n subsets. Hence $\sigma(X)$ consists of 2^n sets. Each of these sets can be obtained as follows: Let $x_1, \ldots, x_n$ be the distinct values of X. Enumerate the binary strings of 0's and 1's of length n and to each of them assign a set from $\sigma(X)$. The way to do this should be clear from the following example with $n = 3$:

000 : $\{X \neq x_1, X \neq x_2, X \neq x_3\} = \varnothing$
001 : $\{X \neq x_1, X \neq x_2, X = x_3\} = \{X = x_3\}$
010 : $\{X \neq x_1, X = x_2, X \neq x_3\} = \{X = x_2\}$
011 : $\{X \neq x_1, X = x_2, X = x_3\} = \{X = x_2 \text{ or } x_3\}$
100 : $\{X = x_1, X \neq x_2, X \neq x_3\} = \{X = x_1\}$
101 : $\{X = x_1, X \neq x_2, X = x_3\} = \{X = x_1 \text{ or } x_3\}$
110 : $\{X = x_1, X = x_2, X \neq x_3\} = \{X = x_1 \text{ or } x_2\}$
111 : $\{X = x_1, X = x_2, X = x_3\} = \Omega$

3.2 (i) If $x_1 < x_2$ then $\{X \leq x_1\} \subset \{X \leq x_2\}$.
(ii) $\lim_{x\to-\infty} F(x) = \lim_{n\to-\infty} F(n)$, where n ranges over the integers. But $F(n) = \mathbf{P}(X \leq n)$ decreases as n decreases, and so

$$\lim_{n\to-\infty} F(n) = \mathbf{P}\Big(\bigcap_{n\in\mathbb{Z}}\{X \leq n\}\Big) = \mathbf{P}(\varnothing) = 0.$$

(iii) Similarly,

$$\lim_{n\to+\infty} F(n) = \mathbf{P}\Big(\bigcup_{n\in\mathbb{Z}}\{X \leq n\}\Big) = \mathbf{P}(\Omega) = 1.$$

(iv)

$$\bigcap_{n\in\mathbb{N}}(-\infty, x + 1/n] = (-\infty, x].$$

3.3 For $a \leq b$: (i)

$$(a, b] = (-\infty, b] \setminus (-\infty, a]$$

and $(-\infty, a] \subset (-\infty, b]$.
(ii) Observe that

$$(a, b) = \bigcup_{n\in\mathbb{N}}(a, b - 1/n]$$

and use (i) to get

$$\mathbf{P}(X \in (a, b)) = \lim_{n\to\infty}(F(b - 1/n) - F(a)) = F(b-) - F(a).$$

(iii) Observe that

$$[a, b] = \bigcap_{n\in\mathbb{N}}(a - 1/n, b]$$

and use (i) to get

$$\mathbf{P}(X \in [a, b]) = \lim_{n\to\infty}(F(b) - F(a - 1/n)) = F(b) - F(a-).$$

(iv) Apply (iii) with $a = b$.

3.4 The way that F^{-1} is defined in the lecture notes is:

$$F^{-1}(t) := \inf A_t, \text{ where } A_t := \{x \in \mathbb{R} : F(x) \geq t\}.$$

By the definition of the infimum of a set of real numbers (the set A_t, in this case), $F^{-1}(t)$ is the largest of all numbers c which are lower bounds to the set (the existence of which is guaranteed by the completeness property of the set of real numbers–a consequence of its construction):

$$F^{-1}(t) \geq c \iff x \geq c \text{ for all } x \in A_t.$$

Equivalently,

$$F^{-1}(t) \geq c \iff \text{ for all } x < c, \quad x \notin A_t \iff \text{ for all } x < c, \quad F(x) < t.$$

Instead of saying "for all $x < c$" we can say "for all $x = c - \varepsilon$ with $\varepsilon > 0$", and so

$$F^{-1}(t) \geq c \iff F(c - \varepsilon) < t \text{ for all } \varepsilon > 0 \iff F(c-) < t.$$

Therefore,

$$F^{-1}(t) < c \iff t \leq F(c-),$$

This holds for all t and c and so, by setting $t = U$, we obtain

$$\mathbf{P}(F^{-1}(U) < c) = \mathbf{P}(U \leq F(c-)) = F(c-),$$

the latter due to the assumption that U is uniformly distributed on the interval $(0, 1)$. Hence

$$\mathbf{P}(F^{-1}(U) \leq x) = \lim_{n\to\infty} \mathbf{P}(F^{-1}(U) < x+1/n) = \lim_{n\to\infty} F((x+1/n)-) = \lim_{n\to\infty} F(x+1/n) = F(x).$$

3.5 The function $V : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ is defined by

$$V(\omega) = \sum_{k=1}^{\infty} 2\omega_k 3^{-k}, \quad \omega_1, \omega_2, \ldots \in \{0, 1\},$$

and $\mathbf{P}$ is a measure on the cylinder-σ-algebra $\mathscr{F}$ of $\{0,1\}^{\mathbb{N}}$ such that

$$\mathbf{P}(\omega_1 = \varepsilon_1, \ldots, \omega_n = \varepsilon_n) = 2^{-n},$$

for all $n \in \mathbb{N}$, and all $\varepsilon_1, \ldots, \varepsilon_n \in \{0, 1\}$. We consider the distribution function

$$F(x) = \mathbf{P}(V \leq x)$$

of V. Define $V'(\omega)$ by

$$V(\omega) = \frac{2\omega_1}{3} + \frac{1}{3}V'(\omega).$$

Observe that the $\mathbf{P}$-law of V' is the same as the $\mathbf{P}$-law of V (by considering finite-dimensional cylinder sets). Suppose $x < 1/3$. Then

$$F(x) = \mathbf{P}(V \leq x) = \mathbf{P}(\omega_1 = 0, \tfrac{1}{3}V' \leq x) = \frac{1}{2}F(3x).$$

Suppose $x > 2/3$. Then

$$F(x) = \mathbf{P}(V \leq x) = \mathbf{P}(\omega_1 = 1, \tfrac{2}{3} + \tfrac{1}{3}V' \leq x) = \frac{1}{2}F(3(x - \tfrac{2}{3})).$$

Observe that it is impossible for $V(\omega)$ to be strictly between $1/3$ and $2/3$. Indeed, if $\omega_1 = 0$, then the maximum value of $V(\omega)$ is $1/3$ (and this happens when $V'(\omega) = 1$; if $\omega_1 = 1$, the least value of $V(\omega)$ is $2/3$. So

$$\{1/3 < V < 2/3\} = \varnothing.$$

Since

$$\begin{aligned}
\mathbf{P}(V = 1/3) &= \mathbf{P}(\omega_1 = 0, V' = 1) = \mathbf{P}(\omega_1 = 0, \omega_n = 1 \text{ for all } n \geq 1) = 0, \\
\mathbf{P}(V = 2/3) &= \mathbf{P}(\omega_1 = 1, V' = 0) = \mathbf{P}(\omega_1 = 1, \omega_n = 0 \text{ for all } n \geq 1) = 0,
\end{aligned}$$

we have

$$\mathbf{P}(1/3 \leq V \leq 2/3) = 0$$

and so, for some constant c,

$$F(x) = \text{c}, \quad \text{if } 1/3 \leq x \leq 2/3.$$

Since V and $1 - V$ have the same law we have

$$c = 1/2.$$

In other words, if we define the operator $\mathcal{Q}$ on $[0,1]^{[0,1]}$ by

$$\mathcal{Q}f(x) = \begin{cases} \frac{1}{2} f(3x), & \text{if } 0 \leq x < 1/3 \\ 1/2, & \text{if } 1/3 \leq x \leq 2/3 \\ \frac{1}{2} f(3(x - 2/3)), & \text{if } 2/3 < x \leq 1, \end{cases}$$

we have just shown that F satisfies

$$F = \mathcal{Q}F.$$

Now restrict $\mathcal{Q}$ onto $C[0,1]$, the space of continuous functions on $[0,1]$, equipped with the usual norm

$$||f|| = \max_{0 \leq x \leq 1} |f(x)|.$$

Observe that

$$\begin{aligned}
||\mathcal{Q}f|| = \max_{0 \leq x \leq 1} |f(x)| &= \max_{0 \leq x < 1/3} |f(x)| \vee \max_{1/3 \leq x \leq 2/3} |f(x)| \vee \max_{2/3 < x \leq 1} |f(x)| \\
&= \max_{0 \leq x < 1/3} |\frac{1}{2} f(3x)| \vee (1/2) \vee \max_{2/3 < x \leq 1} |\frac{1}{2} f(3(x - 2/3))| = \frac{1}{2} ||f||.
\end{aligned}$$

So $\mathcal{Q}$ is a contraction, and so, by completeness of $C[0,1]$ (this is the analogous of the completeness as the completeness mentioned in Exercise 3.4), starting from any $F_0 \in C[0,1]$, the sequence defined recursively through

$$F_{n+1} = \mathcal{Q}F_n$$

converges uniformly to an $F^* \in C[0,1]$ which satisfies

$$\mathcal{Q}F^* = F^*.$$

Note that $\mathcal{Q}$ is an increasing operator, i.e. if f is increasing function, then so is $\mathcal{Q}f$. Therefore, the limit is also increasing. So F^* is an increasing continuous function. By observing that $\mathcal{Q}$ is a contraction also on the set $D[0,1]$ of functions on $[0,1]$ which have discontinuities of first kind, and using the fact that $F \in D[0,1]$, we have

$$F = F^*$$

and so F is continuous.

To show that F is not absolutely continuous, we show that

$$\mathbf{P}(V \in C) = 1$$

for some set $C \subset [0,1]$ with Lebesgue measure 0. Let H be a nonempty open subset of $[0,1]$ such that

$$\mathbf{P}(V \in H) = 0.$$

(We already saw that $H = (1/3, 2/3)$ is such a set.) Notice that

$$\frac{1}{3}H = \{x/3 : x \in H\} \subset (0, 1/3),$$
$$\frac{2}{3} + \frac{1}{3}H = \{(2/3) + x/3 : x \in H\} \subset (2/3, 1).$$

Hence

$$\mathbf{P}(V \in \tfrac{1}{3}H) = \mathbf{P}(\omega_1 = 0,\ \tfrac{1}{3}V' \in \tfrac{1}{3}H) = \frac{1}{2}\mathbf{P}(V \in H)$$
$$\mathbf{P}(V \in \tfrac{2}{3} + \tfrac{1}{3}H) = \mathbf{P}(\omega_1 = 1,\ \tfrac{2}{3} + \tfrac{1}{3}V' \in \tfrac{2}{3} + \tfrac{1}{3}H) = \frac{1}{2}\mathbf{P}(V \in H)$$

So, starting with a set $H_0 = H$ for which $\mathbf{P}(V \in H) = 0$, we can create a family of sets H_n for which $\mathbf{P}(V \in H_n) = 0$, recursively by

$$H_{n+1} = \frac{1}{3}H_n \cup \big(\frac{2}{3} + \frac{1}{3}H_n\big), \quad n = 1, 2, \ldots$$

Since the sets $\frac{1}{3}H_n$ and $\frac{2}{3} + \frac{1}{3}H_n$ are disjoint nonempty open sets, we have that their lengths add up. But both sets have length equal to a third of the length of H_n. So

$$|H_{n+1}| = \frac{2}{3}|H_n|,$$

i.e.

$$|H_n| = (2/3)^n |H_0|.$$

Let

$$D = \cup_{n=0}^{\infty} H_n,$$

where $H_0 = (1/3, 2/3)$. We have

$$\mathbf{P}(V \in D) \le \sum_{n=0}^{\infty} \mathbf{P}(V \in H_n) = \sum_{n=0}^{\infty} 0 = 0.$$

We have $|H_0| = 1/3$, $|H_n| = 2^n/3^{n+1}$. Observe that the H_n are here disjoint. So

$$|D| = \sum_{n=0}^{\infty} |H_n| = \sum_{n=0}^{\infty} 2^n/3^{n+1} = 1.$$

Finally, set

$$C = [0,1] - D.$$

We have $|C| = 1 - |D| = 0$, and $\mathbf{P}(V \in C) = 1 - \mathbf{P}(V \in D) = 1 - 0 = 1$.

3.6 $Y = 1$ if and only if X divides 3, and the only possibility is $X = 3$. So $\mathbf{P}(Y = 1) = 1/6$, $\mathbf{P}(Y = 2) = 5/6$.

3.7

$$\mathbf{P}(\varphi(X) \le t) = \mathbf{P}(X \ge \psi(t)) = \int_{\psi(t)}^{\infty} f(x)dx.$$

Taking derivative with respect to t, we find that the density of $\varphi(X)$ equals

$$\frac{d}{dt}\int_{\psi(t)}^{\infty} f(x)dx = -\psi'(t)f(\psi(t)).$$

But ψ is strictly decreasing, so $\psi'(t) < 0$, and so $-\psi'(t) = |\psi'(t)|$.

3.8 Let $a > 0$. The inverse of the function $\varphi(u) := u^{1/a}$ is $\psi(y) = y^a$. We have $\psi'(y) = ay^{a-1}$. So

$$f_Y(y) = f_U(\psi(y))\psi'(y) = ay^{a-1}.$$

Clearly, $0 < y < 1$ is the range over which f is $\neq 0$, because φ maps the interval $(0, 1)$ to itself. If $a = -b < 0$, then $\varphi(u) := u^{-1/b}$ maps $(0, 1)$ into $(1, \infty)$. So Y has nonzero density on $(1, \infty)$ and, by the exercise above, the density is given by

$$f_Y(y) = f_U(\psi(y))|\psi'(y)| = by^{-b-1}.$$

You should sketch the functions φ and ψ in both cases, and see why, qualitatively, the results are sound, i.e. that the mass gets transferred to the correct places. E.g., the function $Y = U^{30}$ has most of its density around 0, while $Y = U^{1/30}$ has most of its density around 1.

3.9 Let us write the solution in intuitive terms (using calculus), following, of course, a method which is totally equivalent to the theory. Let u be mapped to y via $y = e^u$. Then the mass assigned on an interval of tiny length $|du|$ sitting around the point u is transferred to an interval of tiny length $|dy|$ sitting around the point y:

$$f_Y(y)|dy| = f_U(u)|du|$$

and, since $\frac{dy}{du} = e^u$, we have

$$f_Y(y)e^u = f_U(u),$$

or

$$f_Y(y) = e^{-u} = e^{-\log y} = \frac{1}{y}.$$

The range of interest is the image of the interval $(0, 1)$ under the map $y = e^u$, i.e. y ranges over $(e^0, e^1) = (1, e)$.

Let us now consider the case $y = e^{au}$, where a is a real number. The function is one-to-one for all values of a, so we need not worry about multiple pre-images of a point. Again,

$$f_Y(y)|dy| = f_U(u)|du|$$

Here, $\frac{dy}{du} = ae^{au}$, and so

$$f_Y(y)|ae^{au}| = f_U(u),$$

i.e.

$$f_Y(y) = \frac{1}{|a|e^{au}}.$$

Let us not forget that we need to express this is a function of y, through $y = e^{au}$, i.e.

$$f_Y(y) = \frac{1}{|a|y}.$$

The range of y is the image of $(0,1)$ under the map $y = e^{au}$. If $a > 0$, then y ranges over $(1, e^a)$. If $a < 0$, then y ranges over $(e^{-|a|}, 1)$. If $a = 0$ we have a degenerate situation, because $Y = 1$ with probability 1 (no density).

3.10 (*Note there are some typos in the statement of the exercise.*) No sweat here: the method is as above:

$$f_Y(y)|dy| = f_X(x)|dx|$$

We have $y = e^{x/a}$, where $a > 0$, so $\frac{dy}{dx} = \frac{1}{a}e^{x/a}$:

$$f_Y(y)\frac{1}{a}e^{x/a} = f_X(x),$$

$$f_Y(y) = e^{-x}ae^{-x/a} = ay^{-a-1}.$$

Here y ranges between 1 and ∞ because $y = e^{x/a}$ maps $\{x > 0\}$ onto $\{1 < y < \infty\}$.

3.11 If the function $y = \varphi(x)$ is non-monotonic, then each y may have multiple pre-images. So an interval of tiny length $|dy|$ located around the point y may be the image of many tiny intervals located at the points $x_1, x_2, \ldots$, where the latter are all pre-images of y under the map $y = \varphi(x)$. Hence the "mass" $f_Y(y)|dy|$ is the sum of the masses $f_X(x_i)|dx|$, $i = 1, 2, \ldots$:

$$f_Y(y)|dy| = \sum_i f_X(x_i)|dx|.$$

Let us consider the case $Y = X^2$, where X is uniform on $[-1,1]$. Then $\frac{dy}{dx} = 2x$. There are 2 solutions of the equation $y = x^2$, namely, $x_1 = +\sqrt{y}$, $x_2 = -\sqrt{y}$. So

$$f_Y(y) = f_X(\sqrt{y})\left|\frac{dx}{dy}\right|_{x=\sqrt{y}} + f_X(-\sqrt{y})\left|\frac{dx}{dy}\right|_{x=-\sqrt{y}}.$$

But $f_X(x) = 1/2$, for $-1 \le x \le 1$ and $\left|\frac{dx}{dy}\right|_{x=\sqrt{y}} = \left|\frac{dx}{dy}\right|_{x=-\sqrt{y}} = 1/(2\sqrt{y})$, so

$$f_Y(y) = 1/(2\sqrt{y}), \quad 0 \le y \le 1.$$

Alternatively, we can argue directly as follows:

$$\mathbf{P}(X^2 \le y) = \mathbf{P}(-\sqrt{y} \le X \le \sqrt{y}).$$

So

$$f_Y(y) = \frac{d}{dy}\mathbf{P}(X^2 \le y) = \frac{d}{dy}\int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2}dy' = \frac{1}{2\sqrt{y}}.$$

3.11 In the second case,

$$Y = X^2, \qquad \text{where } X \text{ is uniform on [-1,2]}$$

every $y \in (0,1)$ has two pre-images, $x = \pm\sqrt{y}$, but every $y \in (1,4)$ has one pre-image, $x = +\sqrt{y}$. So

$$f_Y(y) = \begin{cases} f_X(\sqrt{y}) \left|\frac{dx}{dy}\right|_{x=\sqrt{y}} + f_X(-\sqrt{y}) \left|\frac{dx}{dy}\right|_{x=-\sqrt{y}}, & 0 < y < 1, \\ f_X(\sqrt{y}) \left|\frac{dx}{dy}\right|_{x=\sqrt{y}}, & 1 < y < 4. \end{cases}$$

Here $f_X(x) = 1/3$, $-1 \leq x \leq 2$ and $\left|\frac{dx}{dy}\right| = 1/(2\sqrt{y})$ in all cases, so

$$f_Y(y) = \begin{cases} 1/(3\sqrt{y}), & 0 < y < 1, \\ 1/(6\sqrt{y}), & 1 < y < 4. \end{cases}$$

The density has a jump at $y = 1$, and we don't bother to define its value there: it is irrelevant because the density is useful only as an object to be integrated. In other words, not defining the value of a density at finitely many points won't matter at all.

3.12 Same story here:

$$f_Y(y)|dy| = (f_X(x_1) + f_X(x_2))|dx|,$$

where x_1, x_2 are the two pre-images of $y = \cosh(x)$. Here y ranges over $[1, \infty)$, and $|\frac{dy}{dx}| = |\sinh(x)| = \sqrt{y^2 - 1}$. The latter follows from the identity $\cosh^2(x) - \sinh^2(x) = 1$. Notice that $|x_1| = |x_2| = \cosh^{-1}(y)$ and the density of X is symmetric around zero. Hence

$$f_Y(y) = \frac{2c}{1 + \cosh^{-1}(y)^2} \frac{1}{\sqrt{y^2 - 1}} = \frac{c}{\sqrt{y^2 - 1}y^2 + y^3 - y}$$

3.13 We have

$$\mathbf{P}(X = a^n) = 2^{-n}, \quad n \in \mathbb{N}.$$

Therefore,

$$\mathbf{E}\, X = \sum_{n=1}^{\infty} (a/2)^n$$

which is finite iff $a/2 < 1$. The sum equals $a/(2 - a)$.

3.14 We have

$$\mathbf{P}(X = n^k) = cn^{-2}, \quad n \in \mathbb{N}.$$

Therefore,

$$\mathbf{E}\, X = \sum_{n=1}^{\infty} cn^{-(2-k)},$$

and the sum is finite iff $2 - k > 1$. (Use, e.g., the ratio test.)

3.15

$$\mathbf{E}(X;A) = \sum_{\substack{n\in\mathbb{N}\\ n \text{ odd}}} (a/2)^n = \sum_{m=0}^{\infty} (a/2)^{2m+1} = \frac{2a}{a^2-4}.$$

Setting $a = 1$ in the above we obtain

$$\mathbf{P}(A) = \sum_{m=0}^{\infty} (1/2)^{2m+1} = \frac{2}{3},$$

so $\mathbf{E}(X|A) = \mathbf{E}(X;A)/\,\mathbf{P}(A) = 3a/(a^2-4)$. The reason that $\mathbf{E}(X|A^c) = a\,\mathbf{E}(X|A)$ is *obvious* is that $X = a^{\xi}$, where $\mathbf{P}(\xi = n) = 2^{-n}$, $n \in \mathbb{N}$, i.e. ξ is a geometric–and hence memoryless–random variable.

3.16 *The solution depends on the way that theory was presented.*

3.17

$$\mathbf{E}\,N = \sum_{n=1}^{\infty} ne^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda}\lambda\sum_{n=1}^{\infty}\frac{\lambda^{n-1}}{(n-1)!} = e^{-\lambda}\lambda e^{\lambda} = \lambda,$$

where we used the Taylor expansion of the exponential function.

$$\mathbf{E}(N^2-N) = \sum_{n=2}^{\infty}(n^2-n)e^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda}\lambda^2\sum_{n=2}^{\infty}\frac{\lambda^{n-2}}{(n-2)!} = e^{-\lambda}\lambda^2 e^{\lambda} = \lambda^2.$$

Hence $\mathbf{E}(N^2) = \mathbf{E}(N^2-N) + \mathbf{E}\,N = \lambda^2 + \lambda$.

CHAPTER 4

4.1 We use Chernoff's inequality as follows:

$$\mathbf{P}(X > na) \le \frac{\mathbf{E}\, e^{\theta X}}{e^{\theta na}},$$

where θ is a positive constant. Now,

$$\mathbf{E}\, e^{\theta X} = \sum_{k=0}^{n} e^{\theta k} \binom{n}{k} 2^{-k} = \sum_{k=0}^{n} \binom{n}{k} (e^{\theta} 2^{-1})^k = (1 + e^{\theta} 2^{-1})^n.$$

Hence,

$$\mathbf{P}(X > na) \le \left(\frac{1 + e^{\theta} 2^{-1}}{e^{a\theta}} \right)^n,$$

and the bound holds for any $\theta > 0$. We can get a value for the bound by choosing some specific θ, e.g. $\theta = \log 4$, which gives $(3 \times 4^{-a})^n$, or we can be smart and find the best bound by choosing the value of θ for which the bound is least. To find this θ, take the logarithm of the expression inside the parenthesis, differentiate with respect to θ, and set the derivative equal to zero:

$$0 = \frac{d}{d\theta} (\log(1 + e^{\theta} 2^{-1}) - a\theta) = \frac{e^{\theta} 2^{-1}}{1 + e^{\theta} 2^{-1}} - a,$$

$$\text{whence, } e^{\theta} = \frac{2a}{1-a}.$$

For this value, we have

$$\mathbf{P}(X > na) \le \left(\frac{1 + \frac{a}{1-a}}{(\frac{2a}{1-a})^a} \right)^n = 2^{-an} a^{-an} (1-a)^{-(1-a)n}$$

4.2 This is an exercise about a nonnegative random variable Z for which we know that $\mathbf{E}\, Z = 0$. If Z is simple, then it is immediate that $\mathbf{P}(Z > 0) = 0$. For general Z, let Z_n be a sequence of simple random variables such that $Z_n \uparrow Z$. Then $\{Z > 0\} = \cup_n \{Z_n > 0\}$ and since the latter events have $\mathbf{P}$ equal to zero, it follows that $\mathbf{P}(Z > 0) = 0$.

Now apply this to the random variable $Z = (X - \mathbf{E}\, X)^2$. We learnt that $\mathbf{P}(Z = 0) = 1$, i.e. $\mathbf{P}(X = \mathbf{E}\, X) = 1$.

4.3 Let $Z = |X - \mathbf{E}\, X|^2$. Markov's inequality says

$$\mathbf{P}(Z > t^2) \le \frac{\mathbf{E}\, Z}{t^2},$$

for all $t > 0$. But

$$\mathbf{P}(Z > t^2) = \mathbf{P}(|X - \mathbf{E}\, X| > t).$$

4.4 Just use the binomial theorem:

$$(a(X - \mathbf{E}\, X) + b(Y - \mathbf{E}\, Y))^2 = a^2 (X - \mathbf{E}\, X)^2 + 2ab(X - \mathbf{E}\, X)(Y - \mathbf{E}\, Y) + b^2 (Y - \mathbf{E}\, Y)^2$$

4.5 Again, use the binomial theorem:

$$(\lambda X + Y)^2 = \lambda^2 X^2 + 2\lambda XY + Y^2.$$

Hence

$$Q(\lambda) := \lambda^2 \,\mathbf{E}\, X^2 + 2\lambda \,\mathbf{E}\, XY + \mathbf{E}\, Y^2 \geq 0$$

for all values of λ. The polynomial $Q(\lambda)$ is minimised at the λ which solves

$$0 = Q'(\lambda) = 2\lambda \,\mathbf{E}\, X^2 + 2 \,\mathbf{E}\, XY,$$

i.e. for $\lambda = -\,\mathbf{E}\, XY/\,\mathbf{E}\, X^2$. Hence

$$0 \leq Q(-\,\mathbf{E}\, XY/\,\mathbf{E}\, X^2) = \frac{\mathbf{E}\, X^2 \,\mathbf{E}\, Y^2 - (\mathbf{E}\, XY)^2}{\mathbf{E}\, X^2},$$

which gives the inequality we need.

4.6 Apply the previous inequality to $X - \mathbf{E}\, X$, $Y - \mathbf{E}\, Y$ in lieu of X, Y, respectively.

4.7 Just use the property $e^{x+y} = e^x e^y$ of the exponential function.

$$e^{\theta(a+bX)} = e^{\theta a} e^{\theta b X}$$
$$\mathbf{E}\, e^{\theta(a+bX)} = e^{\theta a} \,\mathbf{E}\, e^{\theta b X} = e^{\theta a} M_X(\theta b).$$

4.8 If $\lambda > \theta$,

$$M(\theta) = \int_{-\infty}^{\infty} e^{\theta x} \lambda e^{-\lambda x} dx = \lambda \int_{-\infty}^{\infty} e^{-x(\lambda - \theta)} dx = \frac{\lambda}{\lambda - \theta}.$$

If $\lambda \leq \theta$ the integral diverges. Now,

$$M'(\theta) = \frac{\lambda}{(\lambda - \theta)^2}, \quad M''(\theta) = \frac{2\lambda}{(\lambda - \theta)^3},$$

So,

$$\mathbf{E}\, X = M'(0) = \frac{1}{\lambda}, \quad \mathbf{E}\, X^2 = M''(0) = \frac{2}{\lambda^2}, \quad \operatorname{var} X = \mathbf{E}\, X^2 - (\mathbf{E}\, X)^2 = \frac{1}{\lambda^2}.$$

4.9 A necessary condition for the integral defining $M(\theta)$, for $\theta > 0$, to converge, is that the right tail of the density decays exponentially. This is not so for the Cauchy distribution and so $M(\theta) = \infty$ for $\theta > 0$. By symmetry, $M(\theta) = M(-\theta)$ for $\theta < 0$, and so $M(\theta) = \infty$, for all values of $\theta \neq 0$.

4.10 Let Y_1, Y_2 be real random variables. The modulus inequality is equivalent to

$$\sqrt{\mathbf{E}\, Y_1^2 + \mathbf{E}\, Y_2^2} \leq \mathbf{E}\, \sqrt{Y_1^2 + Y_2^2}.$$

But, by Jensen's inequality, the square of the right hand side is larger than or equal to $\mathbf{E}(Y_1^2 + Y_2^2)$, which proves the inequality.

4.11 The exponential function satisfies $e^{z+w} = e^z e^w$ for any two complex numbers z, w.

$$e^{it(a+bX)} = e^{ita} e^{itbX}$$
$$\mathbf{E}\, e^{it(a+bX)} = e^{ita}\, \mathbf{E}\, e^{itbX} = e^{ita} \varphi_X(tb).$$

4.12

$$\int_0^1 e^{itu} du = \frac{e^{it} - 1}{it}.$$

If X is uniform on (a, b) then $X = (b - a)U + a$, where U is uniform on $(0, 1)$.

4.13 We need to show that

$$\int_0^\infty e^{i\theta x} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - i\theta}.$$

Note that the function $f(z) = e^z$ is analytic for all $z \in \mathbb{C}$, and has itself as primitive. So

$$\int_\gamma e^z dz = e^b - e^a,$$

for any simple curve γ with endpoints $a, b \in \mathbb{C}$. Let γ be the curve $z = -cx$, $0 \le x \le x_0$, where c is a fixed complex number. Then

$$\int_\gamma e^z dz = \int_0^{x_0} e^{-cx}(-c) dx.$$

Hence

$$\int_0^{x_0} e^{-cx} dx = \frac{1 - e^{-cx_0}}{c}.$$

If the real part of c is positive, $\lim_{x_0 \to \infty} e^{-cx_0} = 0$, and so

$$\int_0^\infty e^{-cx} dx = \frac{1}{c},$$

which gives what we need if we set $c = \lambda - i\theta$.

4.14 For $x \le y$, both ranging in $\{1, \ldots, 6\}$,

$$\mathbf{P}(X \ge x, Y \le y) = \mathbf{P}(x \le N_1 \le y,\ x \le N_2 \le y) = \left(\frac{y - x}{6}\right)^2.$$

The marginals are as follows:

$$\mathbf{P}(X \ge x) = \left(\frac{6 - x}{6}\right)^2$$
$$\mathbf{P}(Y \le y) = \left(\frac{y}{6}\right)^2$$

We can work out $\mathbf{P}(X = x, Y = y)$ by using the additivity of $\mathbf{P}$, or, simply, by

$$\mathbf{P}(X = x, Y = y) = 2\,\mathbf{P}(N_1 = x, N_2 = y) = 1/6, \quad x \le y.$$

To find the conditional probabilities, just use division.

4.15 Just do an integral using Fubini:

$$a^{-1} = \int_{\substack{0\le x\le 1\\0\le y\le 1}} (x+y^2)d(x,y) = \int_{0\le x\le 1} dx \int_{0\le y\le 1} dy(x+y^2) = \frac{1}{2}+\frac{1}{3}.$$

$$F_{X,Y}(x,y) = a\int_{\substack{0\le x'\le x\\0\le y'\le y}} (x'+y'^2)d(x',y') = a\frac{x^2}{2}y + ax\frac{y^3}{3}.$$

$$\begin{aligned}\mathbf{P}(X>Y) &= \int_{\substack{0\le x\le 1\\0\le y\le 1\\x>y}} a(x+y^2) = a\int_{0\le x\le 1} dx \int_{0\le y\le 1} dy \mathbf{1}(x>y)(x+y^2)\\ &= a\int_{0\le x\le 1} x^2dx + a\int_{0\le y\le 1}(1-y)y^2dy = a(\tfrac{1}{3}+\tfrac{1}{3}-\tfrac{1}{4}).\end{aligned}$$

$$\begin{aligned}\mathbf{P}(X^2>Y) &= \int_{\substack{0\le x\le 1\\0\le y\le 1\\x^2>y}} a(x+y^2) = a\int_{0\le x\le 1} dx \int_{0\le y\le 1} dy \mathbf{1}(x^2>y)(x+y^2)\\ &= a\int_{0\le x\le 1} x^3dx + a\int_{0\le x\le 1}\frac{(x^2)^3}{3}dx = a(\tfrac{1}{4}+\tfrac{1}{3\cdot 7}).\end{aligned}$$

4.16 We have that X, W are independent exponentials with $\mathbf{E}X = 2$, $\mathbf{E}W = 1$, and $Y = \frac{1}{2}X - W$. As noted, $Y \ge \frac{1}{2}X$, with probability 1. We compute the distribution function in the form

$$G(x,y) := \mathbf{P}(X>x, Y>y),$$

because it's easier to do the integrals. The variable y ranges from $-\infty$ to $+\infty$ because Y *can* take negative values.

$$\begin{gathered}G(x,y) = \mathbf{P}(X>x, X>W+2y)\\ = \int_0^\infty dw\ e^{-w}\ \mathbf{P}(X>x, X>2w+2y) = \int_0^\infty dw\ e^{-w}\ \mathbf{P}(X> x\vee(2w+2y))\\ = \int_0^\infty dw\ e^{-w}\ e^{-[(x/2)\vee(w+y)]}\end{gathered}$$

This is an easy integral in the variable w, as long as we split it into two integrals: one over $w<(x/2)-y$ and one over $w>(x/2)-y$. We here assume (I) that $x/2>y$ (otherwise the first integral is vacuus.)

$$\begin{aligned}G(x,y) &= \int_0^{(x/2)-y} dw\ e^{-w}\ e^{-x/2} + \int_{(x/2)-y}^\infty dw\ e^{-w}\ e^{-w-y}\\ &= e^{-x/2}\int_0^{(x/2)-y} dw\ e^{-w} + e^{-y}\int_{(x/2)-y}^\infty dw\ e^{-2w}\\ &= e^{-x/2}(1-e^{-\frac{x}{2}+y}) + \frac{1}{2}e^{-y}e^{-x+2y} = e^{-x/2} - \frac{1}{2}e^{-x+y}.\end{aligned}$$

If (II) $x/2<y$ then $(x/2)\vee(w+y) = w+y$ for all $w>0$, and so

$$G(x,y) = \int_0^\infty dw\ e^{-w}\ e^{-w-y} = \frac{1}{2}e^{-y}.$$

So the answer is:

$$\mathbf{P}(X > x, Y > y) = \begin{cases} e^{-x/2} - \frac{1}{2}e^{-x+y}, & \text{if } x \geq 2y \\ \frac{1}{2}e^{-y}, & \text{if } x \leq 2y. \end{cases}$$

We can also find $F_{X,Y}(x,y)$ using additivity:

$$1 - G(x,y) = \mathbf{P}(X \leq x \text{ or } Y \leq y) = 1 - \mathbf{P}(X > x) + 1 - \mathbf{P}(Y > y) - F_{X,Y}(x,y).$$

4.17

$$\begin{aligned} \mathbf{E}\, e^{\eta X' + \theta Y'} &= \mathbf{E}\, e^{\eta(X+Y)+\theta(X-Y)} = \mathbf{E}\, e^{(\eta+\theta)X + (\eta-\theta)Y} \\ &= \mathbf{E}\, e^{(\eta+\theta)X}\, \mathbf{E}\, e^{(\eta-\theta)Y} \\ &= e^{(\eta+\theta)^2} e^{(\eta-\theta)^2} \\ &= e^{2\eta^2 + 2\theta^2}. \end{aligned}$$

Since

$$e^{2\eta^2+2\theta^2} = e^{2\eta^2} e^{2\theta^2},$$

X', Y' are independent, both normal with zero mean and variance 2.

4.18

$$\sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} (e^{\theta} n) = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^{\theta})^n}{n!} = e^{-\lambda} e^{\lambda e^{\theta}} = e^{\lambda(e^{\theta}-1)}.$$

See Exercise 5.2 below.

CHAPTER 5

5.1 This is the same as Exercise 4.1, except that there p was equal to 1/2. The method is precisely the same.

5.2 We use generating functions:

$$\prod_{j=1}^{n} M_{X_j}(\theta) = \prod_{j=1}^{n} \exp\left(\lambda_j(e^\theta - 1)\right) = \exp\left(\sum_{j=1}^{n} \lambda_j(e^\theta - 1)\right),$$

and the latter is the generating function of a Poisson random variable with rate $\sum_j \lambda_j$.

5.3 A brute-force proof is as follows:

$$\mathbf{P}(X_1 = n_1, \dots, X_d = n_d \mid X_1 + \cdots + X_d = n) = \frac{\mathbf{P}(X_1 = n_1, \dots, X_d = n_d)}{\mathbf{P}(X_1 + \cdots + X_d = n)}$$

$$= \frac{\prod_{k=1}^{d} \frac{\lambda_k^{n_k}}{n_k!} e^{-\lambda_k}}{\frac{\lambda^n}{n!} e^{-\lambda}} = \frac{n!}{\prod_{k=1}^{d} n_k!} \prod_{k=1}^{d} (\lambda_k/\lambda)^{n_k} = \binom{n}{n_1, \dots, n_d} \prod_{k=1}^{d} (\lambda_k/\lambda)^{n_k}.$$

(A proof can also be devised by using Exercise 5.8 below.)

5.4 Since

$$\mathbf{P}(X > k | X > k-1) = q$$

we have

$$\mathbf{P}(X > k) = q\, \mathbf{P}(X > k-1) = q^2\, \mathbf{P}(X > k-2) = \cdots = q^k,$$

and so X is geometric. The values $\mathbf{P}(X = 1)$ is called parameter. So the parameter is $\mathbf{P}(X = 1) = \mathbf{P}(X > 0) - \mathbf{P}(X > 1) = q^0 - q^1 = 1 - q$.

5.5

$$M(\theta) = \mathbf{E}\, e^{\theta X} = \sum_{n=1}^{\infty} e^{\theta n}(1-p)^{n-1} p = pe^\theta \sum_{n=1}^{\infty} ((1-p)e^\theta)^{n-1} = \frac{pe^\theta}{1-(1-p)e^\theta}.$$

$$\mathbf{E}\, X = M'(0) = 1/p, \quad \mathbf{E}\, X^2 = M''(0) = (2-p)/p^2, \quad \operatorname{var} X = (1-p)/p^2.$$

5.6 We have

$$\begin{aligned}
\mathbf{P}(X > Y + n, X > Y) &= \sum_k \mathbf{P}(X > k+n, X > k, Y = k) \\
&= \sum_k \mathbf{P}(X > k+n | X > k)\, \mathbf{P}(X > k)\, \mathbf{P}(Y = k) \\
&= \mathbf{P}(X > n) \sum_k \mathbf{P}(X > k)\, \mathbf{P}(Y = k) \\
&= \mathbf{P}(X > n)\, \mathbf{P}(X > Y).
\end{aligned}$$

Dividing by $\mathbf{P}(X > Y)$ we obtain the result. The result can be interpreted as follows: If X represents the duration of my sleep (which is geometrically distributed) then: given that I have not waken up by the unknown time Y that an explosion will occur in Australia, my remaining sleeping time $X - Y$ will be distributed as X, i.e. as if the explosion occur ed when I went to bed.

5.7 From the formula of density transformation, it is obvious that $cX + d$ has constant density. Since the function $y = cx+d$ maps the interval $[a, b]$ onto the interval with endpoints $ca + d$ and $cb + d$, the result follows.

5.8 We can get the answer easily if we think combinatorially. We want to compute

$$\mathbf{P}(A) \equiv \mathbf{P}(S_n^1 = m_1, \ldots, S_n^d = m_n),$$

for all non-negative integers $m_1, \ldots, m_n$ adding up to n. The probability that the first m_1 of the U_j's fall in I_1 and the next m_2 of them in I_2, and so on, equals $p_1^{m_1} \cdots p_d^{m_d}$. The event whose probability we just computed is one of the many events comprising A. Each of these events has exactly the same probability and there are $\binom{n}{m_1,\ldots,m_d}$ such events. Therefore,

$$\mathbf{P}(A) = \binom{n}{m_1, \ldots, m_d} p_1^{m_1} \cdots p_d^{m_d}.$$

For the analytically minded, we can *verify* (and prove) that the result is correct by checking that the generating functions of both sides are equal. First, for the multinomial, we have

$$\sum \binom{n}{m_1, \ldots, m_d} p_1^{m_1} \cdots p_d^{m_d} \theta_1^{m_1} \cdots \theta_d^{m_d} = (p_1\theta_1 + \cdots + p_d\theta_d)^n,$$

where the sum extends over all non-negative integers $m_1, \ldots, m_n$ adding up to n, and where we have used the multinomial theorem. Second, for the probability we are seeking, we have

$$\sum \mathbf{P}(S_n^1 = m_1, \ldots, S_n^d = m_n) \theta_1^{m_1} \cdots \theta_d^{m_d} = \mathbf{E}\left[\theta_1^{S_n^1} \cdots \theta_d^{S_n^d}\right],$$

where the sum extends over the same region as before. The latter further equals

$$\mathbf{E} \prod_{r=1}^{d} \prod_{j=1}^{n} \theta_r^{\mathbf{1}(U_j \in I_r)} = \mathbf{E} \prod_{j=1}^{n} \prod_{r=1}^{d} \theta_r^{\mathbf{1}(U_j \in I_r)} = \prod_{j=1}^{n} \mathbf{E} \prod_{r=1}^{d} \theta_r^{\mathbf{1}(U_j \in I_r)}$$

The function inside the expectation is a function of U_j only and is a simple function: it takes value θ_r with probability p_r; therefore it has expectation $\sum_{r=1}^{d} p_r\theta_r$. Hence the two generating functions agree.

5.9 First note that, with probability 1, all random variables are distinct. There are $d!$ ways to order $U_1, \ldots, U_d$. Since the probabilities of each order are equal and since the probabilities add up to 1, it follows that the probability of a specific order is $1/d!$. So $\mathbf{P}(U_1 < \cdots < U_d) = 1/d!$.

5.10 If X, Y, Z are the lengths of the sticks then $X + Y + Z = 1$, and the sticks form a triangle[1] if (from Euclidean Geometry) each of the sticks has length smaller than the sum of the lengths of the other two:

$$X \le Y + Z$$
$$Y \le Z + X$$
$$Z \le X + Y.$$

The model for stick breaking is, undoubtedly, as follows: Let the stick be the interval $[0,1] = \{x \in \mathbb{R} : 0 \le x \le 1\}$. Pick two i.i.d. random variables B_1, B_2, uniformly distributed in this interval. These represent the break points. Let $U = B_1 \wedge B_2$, $V = B_1 \vee B_2$. The intervals $[0,U], [U,V], [V,1]$ represent the three smaller sticks, and have lengths $X = U$, $Y = V - U$, $Z = 1 - V$. Therefore the above inequalities are written as

$$U \le 1/2$$
$$V - U \le 1/2$$
$$V \ge 1/2.$$

We thus need to compute the probability

$$p = \mathbf{P}(U \le 1/2, V \ge 1/2, V - U \le 1/2).$$

Since X_1, X_2 are interchangeable,

$$p = 2\,\mathbf{P}(X_1 \le X_2, X_1 \le 1/2, X_2 \ge 1/2, X_2 - X_1 \le 1/2).$$

Since (X_1, X_2) is uniformly distributed in the square $[0,1]^2$, it's obvious that $p/2$ is the area of the set

$$\{(x_1, x_2) \in [0,1]^2 : \; x_1 \le x_2, x_1 \le 1/2, x_2 \ge 1/2, x_2 - x_1 \le 1/2\},$$

which is a right isosceles triangle two sides of which have length $1/2$ and thus its area is $1/8$. Hence $p = 1/4$.

5.11 We give the intuition behind this identity in law. Imagine you have d alarms in your bedroom, each set to ring at an exponential time with rate 1 (hour^{-1}, say). The alarms function independently of one another. (This ensures that you will, at some point, get out of bed.) The first alarm will ring at a time (the minimum of d i.i.d. exponential random variables) which is exponentially distributed with rate the sum of the rates, i.e. rate d. In other words, the first alarm rings at time X_1/d, where X_1 is an exponential random variable with rate 1. After the first alarm rings, there are $d-1$ alarms remaining. Note two things: (i) by the memoryless property, the remaining times are independent of the first ring and (ii) they are independent of the first ring. Therefore, the second alarm will ring at a remaining time which is exponential with rate $d-1$. In other words, the second alarm rings at a time $X_2/d - 1$ (where X_2 is an exponential random variable with rate 1) *after* the first ring. Thus, the second alarm rings at time distributed like

$$\frac{X_1}{d} + \frac{X_2}{d-1}$$

[1] Euclid (ca. −300): *Elements*, Alexandria.

where X_1, X_2 are i.i.d. Exp(1). Continuing in this manner, we see that the d-th alarm will ring at time

$$\frac{X_1}{d} + \frac{X_2}{d-1} + \cdots + \frac{X_{d-1}}{2} + X_d,$$

where $X_1, \ldots, X_d$ are i.i.d. Exp(1).

You *can* make the argument formal by first proving the analogue of Exercise 5.6, namely that if X, Y are independent positive random variables, with X being exponential, then

$$\mathbf{P}(X - Y > t | X > Y) = \mathbf{P}(X > t),$$

for all $t > 0$, and then by using induction.

5.12 This follows immediately from Exercise 3.4. But let's prove it directly.

$$\mathbf{P}(-\ln U/\lambda > t) = \mathbf{P}(U < e^{-\lambda t}) = e^{-\lambda},$$

because U is uniform.

5.13 This is called *regenerative property* of the Gamma function. We have

$$\Gamma(\beta) = \int_0^\infty y^{\beta-1}(-e^{-y})' dy = [y^{\beta-1}e^{-y}]_0^\infty - \int_0^\infty (-e^{-y})(y^{\beta-1})' dy$$
$$= \int_0^\infty e^{-y}(\beta-1)y^{\beta-2} dy = (\beta-1)\Gamma(\beta-1).$$

A few remarks on rigour: We take $\beta > 1$, so that is why the value of $y^{\beta-1}e^{-y}$ at $y = 0$ equals 0. Second, $y^{\beta-1}e^{-y} \to 0$, as $y \to \infty$, that is why $[y^{\beta-1}e^{-y}]_0^\infty = 0$. Third, all integrals in the derivation above converge.

5.14 We have

$$\Gamma(1/2) = \int_0^\infty y^{-1/2} e^{-y} dy.$$

Change variable by

$$y = x^2/2$$

so that

$$y^{-1/2} = \frac{\sqrt{2}}{x}, \quad e^{-y} = e^{-x^2/2}, \quad dy = x dx.$$

Then

$$\Gamma(1/2) = \sqrt{2}\int_0^\infty e^{-x^2/2} = \sqrt{2}\sqrt{2\pi}\frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-x^2/2} = 2\sqrt{\pi}\,\mathbf{P}(N > 0),$$

where N is a standard normal random variable. So $\mathbf{P}(N > 0) = 1/2$, and the result follows.

5.15 Here X has law $\Gamma(\beta,\lambda)$, where $\beta,\lambda>0$. Picking θ small enough so that the integral below converges (we will see later how small), we have

$$\begin{aligned}M_X(\theta) &= \int e^{\theta x}\frac{\lambda^\beta}{\Gamma(\beta)}x^{\beta-1}e^{-\lambda x}dx\\ &= \frac{\lambda^\beta}{\Gamma(\beta)}\int x^{\beta-1}e^{-(\lambda-\theta)x}dx\\ &= \frac{\lambda^\beta}{\Gamma(\beta)}\frac{1}{(\lambda-\theta)^\beta}\int y^{\beta-1}e^{-y}dy\\ &= \frac{\lambda^\beta}{\Gamma(\beta)}\frac{1}{(\lambda-\theta)^\beta}\Gamma(\beta)\\ &= \left(\frac{\lambda}{\lambda-\theta}\right)^\beta.\end{aligned}$$

Looking at the second line above, we see that the integral converges for all $\theta\in(-\infty,\lambda)$, and diverges if $\theta\ge\lambda$.

Compute a couple of derivatives:

$$M'(\theta)=\beta\left(\frac{\lambda}{\lambda-\theta}\right)^{\beta-1}\frac{\lambda}{(\lambda-\theta)^2}$$

$$M''(\theta)=\left(\frac{\lambda}{\lambda-\theta}\right)^{\beta}\beta^2(\lambda-\theta)^{-2}+\left(\frac{\lambda}{\lambda-\theta}\right)^{\beta}\beta(\lambda-\theta)^{-2}$$

Set $\theta=0$:

$$\mathbf{E}\,X=M'(0)=\frac{\beta}{\lambda},\quad \mathbf{E}\,X^2=\frac{\beta^2}{\lambda^2}+\frac{\beta}{\lambda}.$$

5.16 Let (see Exercise 5.15)

$$M_{\beta,\lambda}(\theta)=\left(\frac{\lambda}{\lambda-\theta}\right)^{\beta}$$

be the generating function of a $\Gamma(\beta,\lambda)$ probability measure. Then

$$M_{\beta_1,\lambda}(\theta)M_{\beta_2,\lambda}(\theta)=\left(\frac{\lambda}{\lambda-\theta}\right)^{\beta_1+\beta_2}=M_{\beta_1+\beta_2,\lambda}(\theta),$$

and this proves the claim, because the generating function (being a Laplace transform) characterises the measure.

5.17 By changing variable $t=x/\sqrt{2}$, it is enough to prove that

$$I:=\int_{-\infty}^{\infty}e^{-t^2}dt=\sqrt{\pi},$$

But

$$I^2=\left(\int_{-\infty}^{\infty}e^{-x^2}dx\right)\left(\int_{-\infty}^{\infty}e^{-y^2}dy\right)$$

and, by Fubini's theorem,

$$I^2=\int_{\mathbb{R}^2}e^{-(x^2+y^2)}d(x,y).$$

Since the function to be integrated is rotationally invariant on the plane, we use polar coordinates, i.e.

$$x = r\cos\theta, \quad y = r\sin\theta, \quad r > 0, \quad 0 \le \theta < 2\pi.$$

Since $x^2 + y^2 = r^2$, and since $d(x, y) = rd(\rho, \theta)$, we have

$$I^2 = \int_{\mathbb{R}^2} e^{-r^2} r d(r, \theta).$$

Using Fubini's theorem once more,

$$I^2 = \int_0^\infty dr\ r\ e^{-r^2} \int_0^{2\pi} d\theta = 2\pi \int_0^\infty \tfrac{1}{2} s e^{-s} ds = \pi.$$

So, clearly,[2] $I = \sqrt{\pi}$,

5.18 First, complete the square in the exponent:

$$\tfrac{1}{2}x^2 - \theta x = \tfrac{1}{2}(x^2 - 2\theta x + \theta^2 - \theta^2) = \tfrac{1}{2}(x - \theta)^2 - \tfrac{1}{2}\theta^2.$$

Therefore,

$$\begin{aligned} M_X(\theta) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2 + \frac{1}{2}\theta^2} dx \\ &= e^{\frac{1}{2}\theta^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\theta)^2} dx \\ &= e^{\frac{1}{2}\theta^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy \\ &= e^{\frac{1}{2}}, \end{aligned}$$

because the last integral is the integral of a standard normal density, and hence it equals 1.

5.19 Recall the "stability property" of the Gamma distribution, namely, if $X_{\beta_j,\lambda}$, $j = 1, 2$ are *independent* random variables with laws $\Gamma(\beta_i, \lambda)$, respectively, then $X_{\beta_1,\lambda} + X_{\beta_2,\lambda}$ has law $\Gamma(\beta_1 + \beta_2, \lambda)$.
Here, Y_1^2, Y_2^2 are independent, both with law $\Gamma(1/2, 1/2)$. Hence $Y_1^2 + Y_2^2$ has law $\Gamma(1, 1/2) = Exp(1/2)$.

5.20 This requires reviewing the material you learnt and just summarising in an artsy manner.

[2] Lord Kelvin (1824-1907) was an admirer of Joseph Liouville. It is said that, one day, while Kelvin was lecturing in *Glasgow*, he asked his class: 'Do you know what a mathematician is?' He then wrote the following equation on the blackboard

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi},$$

and said: 'A mathematician is one to whom that is as obvious as that $2 \times 2 = 4$ is to you; Liouville was a mathematician.'

5.21 Let $U = X + 2Y$, $V = 3X - 4Y$. We want to compute $\mathbf{E}(U|V)$. We know that
1) $\mathbf{E}(U|V) = cV$, for some constant c,
2) $U - \mathbf{E}(U|V)$ is independent of V.
Therefore $U - cV$ and V are uncorrelated:

$$0 = \mathbf{E}((U - cV)V) = \mathbf{E}\, UV - c\,\mathbf{E}\, V^2,$$

whence

$$c = \frac{\mathbf{E}\, UV}{\mathbf{E}\, V^2}.$$

We have

$$\begin{aligned} \mathbf{E}\, UV &= \mathbf{E}(X + 2Y)(3X - 4Y) = 3\,\mathbf{E}\, X^2 - 8\,\mathbf{E}\, Y^2 + 0 = 3 - 8 = -5, \\ \mathbf{E}\, V^2 &= \mathbf{E}(3X - 4Y)^2 = 9\,\mathbf{E}\, X^2 + 16\,\mathbf{E}\, Y^2 + 0 = 25. \end{aligned}$$

So $c = -5/25 = -1/5$, and so $\mathbf{E}(U|V) = -V/5$.