27. The variance. More on improper integrals

In general, knowing only the mean of a distribution is not as useful as also knowing whether the distribution is clumped near the mean or spread more evenly across the sample space; in other words, whether the p.d.f. is sharply peaked or relatively flat. For that, we need an index of dispersion. In this lecture, we define one.

Accordingly, consider the dispersion density D defined by

$$D(x) \quad = \quad (x-\mu)^2 f(x), \qquad (27.1)$$

where f is the p.d.f. of a random variable X on $[0, \infty)$ and μ is its mean. The function D is never negative and has the property that its value is small *either* if x is very close to μ *or* if f is very small. Only if there is a significant probability of X being far from the mean can Area(D, $[0, \infty)$) be large. So a suitable index of dispersion is Int(D, $[0, \infty)$), called the **variance** of the distribution and denoted by σ^2. That is, the variance is

$$\sigma^2 \quad = \quad \int_0^\infty (x-\mu)^2 f(x)\,dx. \qquad (27.2)$$

The notation indicates that variance is measured in squared units; e.g., if μ is in mm, then σ^2 is in mm^2. The square root of the variance, σ, is called the **standard deviation**. It has the advantage of being in the same units as the mean. Indeed for many purposes a better index of dispersion than either σ or σ^2 is the **coefficient of variation**

$$\kappa \quad = \quad \frac{\sigma}{\mu}, \qquad (27.3)$$

which has the advantage of being a dimensionless ratio.

Variance is illustrated by Figure 1, where D is graphed next to the corresponding p.d.f. for Weibull distributions with shape parameters 10, 3 and 2, respectively. At each level, the unshaded area in the left-hand panel is 1, whereas the shaded area in the right-hand panel – i.e., the variance – increases as the p.d.f. flattens out.

Although (2) defines the variance, it is rarely used to calculate it (unless, as we will discuss in Lecture 28, the distribution is symmetric), because (12.25) implies

$$\begin{aligned}\int_0^\infty (x-\mu)^2 f(x)\,dx \quad &= \quad \int_0^\infty (x^2 - 2x\mu + \mu^2) f(x)\,dx \\ &= \quad \int_0^\infty x^2 f(x)\,dx \; - \; 2\mu\int_0^\infty x f(x)\,dx \; + \; \mu^2\int_0^\infty f(x)\,dx \qquad (27.4) \\ &= \quad \int_0^\infty x^2 f(x)\,dx \; - \; 2\mu\cdot\mu \; + \; \mu^2\cdot 1.\end{aligned}$$

Hence, from (2) and Int(f, $[0, \infty)$) = 1,

$$\sigma^2 \quad = \quad \int_0^\infty x^2 f(x)\,dx \; - \; \mu^2. \qquad (27.5)$$

Variance is a well defined measure of dispersion for all commonly used distributions. Nevertheless, there are well defined (and potentially useful) distributions for which σ^2 is not a finite quantity, even if μ is well defined.

To illustrate this point, and at the same time show how a variance calculation typically invokes the fundamental theorem of calculus, we consider the distribution defined on $[0, \infty)$ by

$$F(t) = \begin{cases} \dfrac{3\theta + A(\theta+2)}{2(\theta+1)}\left(\dfrac{t}{c}\right) - A\left(\dfrac{t}{c}\right)^2 + \dfrac{A(\theta+2)-\theta}{2(\theta+3)}\left(\dfrac{t}{c}\right)^3 & \text{if } 0 \le t \le c \\ 1 - \dfrac{3-A}{(\theta+1)(\theta+3)}\left(\dfrac{c}{t}\right)^\theta & \text{if } c \le t < \infty \end{cases} \quad (27.6)$$

or

$$f(t) = F'(t) = \begin{cases} \dfrac{3\theta + A(\theta+2)}{2(\theta+1)c} - \dfrac{2At}{c^2} + \dfrac{3\{A(\theta+2)-\theta\}t^2}{2(\theta+3)c^3} & \text{if } 0 \le t \le c \\ \dfrac{\theta(3-A)}{c(\theta+1)(\theta+3)}\left(\dfrac{c}{t}\right)^{\theta+1} & \text{if } c \le t < \infty \end{cases} \quad (27.7)$$

with

$$\frac{\theta}{\theta+2} < A < 3 \quad (27.8)$$

to ensure that f is a p.d.f. (see Exercise 1). In Figure 2 this distribution is fitted to the prairie-dog lifespan data from Table 19.1 with $\theta = 1.938$, $c = 2.805$ and $A = 1.476$ (so that $\theta/(\theta+2) = 0.492$, and (8) is satisfied). Minimum total error, 0.46×10^{-2}, is now more than twice as large as in Figure 19.1 and almost three times as large as in Figure 24.10. Nevertheless, the fit isn't wholly unreasonable. It at least suggests that the distribution is a potentially useful one, which is all that Figure 2 aims to achieve.

Before we can calculate the variance, we must first of all calculate the mean. On using (7), we have

$$t f(t) = \begin{cases} \dfrac{\{3\theta + A(\theta+2)\}t}{2(\theta+1)c} - \dfrac{2At^2}{c^2} + \dfrac{3\{A(\theta+2)-\theta\}t^3}{2(\theta+3)c^3} & \text{if } 0 \le t \le c \\ \dfrac{\theta(3-A)}{(\theta+1)(\theta+3)}\left(\dfrac{c}{t}\right)^{\theta} & \text{if } c \le t < \infty \end{cases} \quad (27.9)$$

So, from (26.10) and (16.20) in conjunction with Table 18.1,

$$\mu = \int_0^\infty t f(t)\,dt = \int_0^c t f(t)\,dt + \int_c^\infty t f(t)\,dt$$

$$= \int_0^c \left\{ \frac{\{3\theta + A(\theta+2)\}t}{2(\theta+1)c} - \frac{2At^2}{c^2} + \frac{3\{A(\theta+2)-\theta\}t^3}{2(\theta+3)c^3} \right\} dt + \frac{\theta(3-A)c^\theta}{(\theta+1)(\theta+3)} \int_c^\infty t^{-\theta}\,dt. \quad (27.10)$$

The first integral in (10) is quite straightforward: by the fundamental theorem (in Leibnitz notation), it reduces to

$$\int_0^c \frac{d}{dt}\left\{ \frac{\{3\theta + A(\theta+2)\}t^2}{4(\theta+1)c} - \frac{2At^3}{3c^2} + \frac{3\{A(\theta+2)-\theta\}t^4}{8(\theta+3)c^3} \right\} dt$$

$$= \left(\frac{\{3\theta + A(\theta+2)\}t^2}{4(\theta+1)c} - \frac{2At^3}{3c^2} + \frac{3\{A(\theta+2)-\theta\}t^4}{8(\theta+3)c^3} \right)\Bigg|_0^c$$

$$= \frac{\{6A + (45-7A)\theta + (9-A)\theta^2\}c}{24(\theta+1)(\theta+3)}. \quad (27.11)$$

The second integral, however, requires some care. In this regard, but also with a view to obtaining a more general result for improper integrals, we define u on $[c, \infty)$ by

$$u(t) \quad = \quad t^{-\alpha}. \tag{27.12}$$

Then, with $\alpha = \theta$, $\mathrm{Int}(u, [c, \infty))$ is the quantity to be evaluated in (10).

For $c = 2.805$, u is graphed in Figure 3 on successively larger subdomains $[c, K]$, for $\alpha = 2$ in the left-hand column, and for $\alpha = 1/2$ in the right-hand column. In the each case, the shaded area is $\mathrm{Int}(u, [c, K])$. In the left-hand column, u decreases rapidly enough that, as K increases, total shaded area remains finite. In the right-hand column, on the other hand, u decreases so slowly that, as K increases, total shaded area keeps on growing. In the first case, we say that the integral converges, and that the area is finite. In the second case, we say that the integral diverges, or that $\mathrm{Int}(u, [c, \infty)) = \infty$. We suspect that there is a critical value of α, between 1/2 and 2, at which the shaded area ceases to grow without bound and instead converges. To confirm this suspicion, note that, by the fundamental theorem and (22.31),

$$\int_c^K u(t)dt \quad = \quad \int_c^K t^{-\alpha}dt \quad = \quad \int_c^K \frac{d}{dt}\left\{-\frac{t^{-(\alpha-1)}}{\alpha-1}\right\}dt \quad = \quad -\frac{t^{-(\alpha-1)}}{\alpha-1}\bigg|_c^K \quad = \quad -\frac{K^{-(\alpha-1)}}{\alpha-1} \quad - \quad \left(-\frac{c^{-(\alpha-1)}}{\alpha-1}\right)$$
$$= \quad \frac{c^{-(\alpha-1)}}{\alpha-1} \quad - \quad \frac{1}{(\alpha-1)K^{(\alpha-1)}} \tag{27.13}$$

for any finite K. Now allow K to become infinitely large. If $\alpha > 1$, then $K^{(\alpha-1)}$ becomes infinitely large as well, so that $1/K^{(\alpha-1)}$ approaches zero and (13) reduces to

$$\int_c^\infty t^{-\alpha}dt \quad = \quad \frac{c^{-(\alpha-1)}}{\alpha-1}. \tag{27.14}$$

in the limit as $K \to \infty$. If $\alpha < 2$, however, then (13) becomes

$$\int_c^K t^{-\alpha}dt \quad = \quad \frac{c^{-(\alpha-1)}}{\alpha-1} \quad + \quad \frac{K^{(1-\alpha)}}{(1-\alpha)}, \tag{27.15}$$

which grows without bound. What happens if α is precisely equal to 1? Then

$$\int_c^K t^{-\alpha}dt \quad = \quad \int_c^K \frac{1}{t}dt \quad = \quad \int_c^K \frac{d}{dt}\{\ln(t)\}dt \quad = \quad \ln(t)\big|_c^K \quad = \quad \ln(K) \quad - \quad \ln(c); \tag{27.16}$$

and because the logarithm is a strictly increasing function, ln(K) becomes infinitely large as K approaches infinity. The upshot is that

$$\int_c^\infty t^{-\alpha}dt \quad = \quad \begin{cases} \infty & \text{if } \alpha \le 1 \\ \dfrac{c^{-(\alpha-1)}}{\alpha-1} & \text{if } \alpha > 1 \end{cases} \tag{27.17}$$

and hence, on setting $\alpha = \theta$, that μ is finite only if $\theta > 1$. Then, from (10), (11) and (17) with $\alpha = \theta$, we have

$$\mu \;=\; \frac{\{6A+(45-7A)\theta+(9-A)\theta^2\}c}{24(\theta+1)(\theta+3)} + \frac{\theta(3-A)c^{\theta}}{(\theta+1)(\theta+3)} \cdot \frac{c^{-(\theta-1)}}{\theta-1} \;=\; \frac{\{9\theta-A(\theta+2)\}c}{24(\theta-1)}, \tag{27.18}$$

after simplification. For example, according to the model of Figure 2, mean prairie-dog lifespan is 1.449 years.

Having calculated the mean, we now proceed to calculate the variance. From (7) or (9), we have

$$t^2 f(t) \quad = \quad \begin{cases} \dfrac{\{3\theta + A(\theta+2)\}t^2}{2(\theta+1)c} - \dfrac{2At^3}{c^2} + \dfrac{3\{A(\theta+2)-\theta\}t^4}{2(\theta+3)c^3} & \text{if } 0 \le t \le c \\ \dfrac{\theta(3-A)c}{(\theta+1)(\theta+3)}\left(\dfrac{c}{t}\right)^{\theta-1} & \text{if } c \le t < \infty \end{cases} \qquad (27.19)$$

Thus, on using (5),

$$\mu^2 + \sigma^2 = \int_0^\infty t^2 f(t)\,dt = \int_0^c t^2 f(t)\,dt + \int_c^\infty t^2 f(t)\,dt$$
$$= \int_0^c \left\{ \frac{\{3\theta + A(\theta+2)\}t^2}{2(\theta+1)c} - \frac{2At^3}{c^2} + \frac{3\{A(\theta+2)-\theta\}t^4}{2(\theta+3)c^3} \right\} dt$$
$$+ \frac{\theta(3-A)c^\theta}{(\theta+1)(\theta+3)} \int_c^\infty t^{-(\theta-1)}\,dt. \qquad (27.20)$$

On setting $\alpha = \theta - 1$ in (17), we find that σ^2 is finite only if $\theta - 1 > 1$, or $\theta > 2$. Then, from (17), (20) and the fundamental theorem,

$$\mu^2 + \sigma^2 = \int_0^c \frac{d}{dt}\left\{ \frac{\{3\theta + A(\theta+2)\}t^3}{6(\theta+1)c} - \frac{At^4}{2c^2} + \frac{3\{A(\theta+2)-\theta\}t^5}{10(\theta+3)c^3} \right\} dt + \frac{\theta(3-A)c^\theta}{(\theta+1)(\theta+3)} \frac{c^{-(\theta-2)}}{\theta-2}$$
$$= \left(\frac{\{3\theta + A(\theta+2)\}t^3}{6(\theta+1)c} - \frac{At^4}{2c^2} + \frac{3\{A(\theta+2)-\theta\}t^5}{10(\theta+3)c^3} \right)\Bigg|_0^c + \frac{\theta(3-A)c^2}{(\theta+1)(\theta+3)(\theta-2)}$$
$$= \frac{\{3A + 4(9-2A)\theta + (6-A)\theta^2\}c^2}{30(\theta+1)(\theta+3)} - 0 + \frac{\theta(3-A)c^2}{(\theta+1)(\theta+3)(\theta-2)}$$
$$= \frac{\{6\theta - A(\theta+2)\}c^2}{30(\theta-2)}, \qquad (27.21)$$

after simplification. Now, from (18), we obtain

$$\sigma^2 = \left(\frac{\{6\theta - A(\theta+2)\}}{30(\theta-2)} - \frac{\{9\theta - A(\theta+2)\}^2}{576(\theta-1)^2} \right) c^2. \qquad (27.22)$$

Note in particular that if $1 < \theta \le 2$, as in Figure 2, then the mean exists, but not the variance.

As we remarked above, however, mean and variance both exist for all of the distributions we commonly use. One example is the Weibull distribution. Another example is the **Gamma distribution**. The p.d.f. of the Gamma with shape parameter c and scale parameter s was defined in Exercise 26.9 by

$$f(x) = \frac{x^{c-1}e^{-x/s}}{s^c\Gamma(c)}. \qquad (27.23)$$

The Gamma distribution has numerous biological applications; for example, Troy and Robson (1992, p. 540) used it to model variation among interspike intervals (times between action potentials) for maintained discharges of cat retinal ganglion cells. General expressions for its mean and variance are therefore of interest. From (23) and (26.10), the mean is

$$\mu = \int_0^\infty x f(x)\,dx = \int_0^\infty \frac{x^c e^{-x/s}}{s^c\Gamma(c)}\,dx = \frac{1}{s^c\Gamma(c)} \int_0^\infty x^c e^{-x/s}\,dx. \qquad (27.24)$$

Because (23) defines a distribution for any c (≥ 1) and s (> 0), however, we must always have Int(f, $[0, \infty)$) = 1. Hence

$$1 \;=\; \int_0^\infty \frac{x^{c-1}e^{-x/s}}{s^c\Gamma(c)}\,dx \;=\; \frac{1}{s^c\Gamma(c)}\int_0^\infty x^{c-1}e^{-x/s}\,dx \tag{27.25}$$

for any c or s, implying

$$\int_0^\infty x^{c-1}e^{-x/s}\,dx \;=\; s^c\Gamma(c). \tag{27.26}$$

But if (26) is true for c, then it must also be true for c + 1. Therefore

$$\int_0^\infty x^{c}e^{-x/s}\,dx \;=\; s^{c+1}\Gamma(c+1). \tag{27.27}$$

Substituting in (24), we find that

$$\mu \;=\; \frac{s^{c+1}\Gamma(c+1)}{s^c\Gamma(c)} \;=\; sc, \tag{27.28}$$

on using the Gamma function's recursive property

$$\Gamma(r+1) \;=\; r\Gamma(r) \tag{27.29}$$

(with r = c). Similarly, because (26) must hold for c + 2 if it holds for c,

$$\int_0^\infty x^{c+1}e^{-x/s}\,dx \;=\; s^{c+2}\Gamma(c+2). \tag{27.30}$$

Thus

$$\int_0^\infty x^2 f(x)dx \;=\; \int_0^\infty \frac{x^{c+1}e^{-x/s}}{s^c\Gamma(c)}\,dx \;=\; \frac{1}{s^c\Gamma(c)}\int_0^\infty x^{c+1}e^{-x/s}\,dx \;=\; \frac{s^{c+2}\Gamma(c+2)}{s^c\Gamma(c)}. \tag{27.31}$$

But setting r = c + 1 in (29) yields $\Gamma(c + 2) = (c + 1)\Gamma(c + 1) = (c + 1)c\Gamma(c)$, because $\Gamma(c + 1) = c\Gamma(c)$. So (30) implies

$$\int_0^\infty x^2 f(x)dx \;=\; s^2c(c+1). \tag{27.32}$$

Now (5), (28) and (32) imply

$$\sigma^2 \;=\; s^2c(c+1) \;-\; s^2c^2 \;=\; s^2c. \tag{27.33}$$

Similarly, to calculate the variance of a Weibull distribution, for which

$$f(x) \;=\; \frac{c}{s}(x/s)^{c-1}e^{-(x/s)^c}, \tag{27.34}$$

we first obtain

$$\begin{aligned}\int_0^\infty x^2 f(x)dx \;&=\; c\int_0^\infty x(x/s)^c e^{-(x/s)^c}\,dx \\ &=\; s^2c\int_0^\infty u^{c+1}e^{-u^c}\,du \;=\; s^2c\int_0^\infty x^{c+1}e^{-x^c}\,dx,\end{aligned} \tag{27.35}$$

after using the substitution $u = \phi(x) = x/s$ as in Lecture 26; see Exercise 3. Substituting $u = x^c$ further reduces (35) to

$$\int_0^\infty x^2 f(x)dx \;=\; s^2\int_0^\infty u^{2/c}e^{-u}\,du \;=\; s^2\int_0^\infty u^{(2/c+1)-1}e^{-u}\,du \;=\; s^2\Gamma(1+2/c), \tag{27.36}$$

again as in Lecture 26, and again on using the definition of Γ; see Exercise 3. Moreover, $\mu = s\Gamma(1 + 1/c)$ from (26.34). So (5) and (36) imply

$$\sigma^2 \;=\; s^2\Gamma(1+2/c) \;-\; s^2\{\Gamma(1+1/c)\}^2. \qquad (27.37)$$

For example, if c = 10 then, from Figure 26.4, we have $\Gamma(1 + 1/c) = \Gamma(1.1) = 0.95135$ and $\Gamma(1 + 2/c) = \Gamma(1.2) = 0.91817$. So (37) implies $\sigma^2 = 0.91817\,s^2 - 0.90507s^2 = 0.0131\,s^2$. That is, the top shaded area in Figure 1 is $0.0131\,s^2$. Similarly, $\sigma^2 = 0.1053\,s^2$ if c = 3 because $\Gamma(4/3) = 0.89298$ and $\Gamma(5/3) = 0.90275$; and $\sigma^2 = 0.2146\,s^2$ if c = 2 because $\Gamma(1.5) = 0.88623$ and $\Gamma(2) = 1$, so that the middle and bottom shaded areas in Figure 1 are, respectively, $0.1053\,s^2$ and $0.2146\,s^2$. Note that the variance increases as the shape parameter decreases.

Collating our results, we find from (28) and (33) that

$$\mu \;=\; sc, \qquad \sigma \;=\; s\sqrt{c} \qquad (27.38)$$

for the Gamma distribution; whereas, from (26.34) and (37),

$$\mu \;=\; s\Gamma(1+1/c), \qquad \sigma \;=\; s\sqrt{\Gamma(1+2/c) \;-\; \{\Gamma(1+1/c)\}^2} \qquad (27.39)$$

for the Weibull. In both cases, the coefficient of variation depends only on the shape parameter: from (5) and (38)-(39),

$$\kappa \;=\; \frac{1}{\sqrt{c}} \qquad (27.40)$$

for the Gamma distribution but

$$\kappa \;=\; \sqrt{\frac{\Gamma(1+2/c)}{\{\Gamma(1+1/c)\}^2} \;-\; 1} \qquad (27.41)$$

for the Weibull. In either case, the coefficient of variation is a decreasing function of the shape parameter; e.g., (41) yields $\kappa = 0.5227$ for c = 2, $\kappa = 0.3634$ for c = 3 and $\kappa = 0.1203$ for c = 10 (Figure 1). In Figure 4, κ is plotted against c for both distributions.

Reference

Troy, J.B. & J.G.Robson (1992). Steady discharges of X and Y retinal ganglion cells of cat under photopic illuminance. Visual Neuroscience **9**, 535-553.

Exercises 27

27.1 Verify that (7) defines a p.d.f. if (8) is satisfied. Hint: Establish that Min(f, [0, c]) ≥ 0.

27.2 Show that both mean and variance of the distribution defined by (7) are positive if they exist.

27.3 Verify (35)-(36).

27.4 A probability density function is defined on [0, ∞) by

$$f(t) = \begin{cases} 2A(c-t) + \dfrac{\theta(1-Ac^2)}{(\theta+1)c} & \text{if } 0 \le t \le c \\ \dfrac{\theta(1-Ac^2)}{(\theta+1)c}\left(\dfrac{c}{t}\right)^{\theta+1} & \text{if } c \le t < \infty \end{cases}$$

where A, c and θ are positive numbers satisfying $Ac^2 < 1$.

(i) Does this distribution have a finite mean? If so, what is it?

(ii) Does this distribution have a finite variance? If so, what is it?

27.5 The probability density function of a distribution on [0, ∞) is f defined by

$$f(x) = \begin{cases} \frac{2}{3}x & \text{if } 0 \le x < 1 \\ \frac{1}{3}(3-x) & \text{if } 1 \le x < 3 \\ 0 & \text{if } 3 \le x < \infty \end{cases}$$

(i) Find the mean, μ

(ii) Find the variance, σ^2

(iii) Deduce that the coefficient of variation is

$$\kappa = \frac{1}{4}\sqrt{\frac{7}{2}} = 0.468.$$

27.6 Find both the variance and coefficient of variation of the distribution defined in Exercise 26.6.

27.7 Find both the variance and coefficient of variation of the distribution defined in Exercise 26.7.

27.8 The p.d.f. of a distribution on $[0, \infty)$ is f defined by

$$f(x) = \begin{cases} \frac{2}{5}x & \text{if } 0 \le x < 1 \\ \frac{1}{10}(5-x) & \text{if } 1 \le x < 5 \\ 0 & \text{if } 5 \le x < \infty \end{cases}$$

Find

(i) the mean, μ
(ii) the variance, σ^2
(iii) the median, M, and
(iv) the mode, m.
(v) Show that the coefficient of variation is

$$\kappa = \frac{1}{2}\sqrt{\frac{7}{6}}.$$

27.9 The p.d.f. of a distribution on $[0, \infty)$ is f defined by

$$f(x) = \frac{1}{15}\begin{cases} 2x & \text{if } 0 \le x < 3 \\ 15-3x & \text{if } 3 \le x < 5 \\ 0 & \text{if } 5 \le x < \infty \end{cases}$$

Find

(i) the mean, μ
(ii) the variance, σ^2
(iii) the median, M, and
(iv) the mode, m.
(v) Show that the coefficient of variation is

$$\kappa = \frac{1}{8}\sqrt{\frac{19}{2}}.$$

27.10 Find both the variance and coefficient of variation of the truncated exponential distribution defined in Exercise 26.10.

27.11* The p.d.f. of a distribution on $[0, \infty)$ is defined by

$$f(x) = \frac{1}{L} g(x)$$

where

$$g(x) = \begin{cases} x/2 & \text{if } 0 \le x < 2 \\ 3 - x & \text{if } 2 \le x < 3 \\ 0 & \text{if } 3 \le x < \infty \end{cases}$$

and L is a constant.

(i) Find L

(ii) Find μ, the mean

(iii) Find σ^2, the variance

(iv) Show that the coefficient of variation is

$$\kappa = \frac{1}{5}\sqrt{\frac{7}{2}}$$

(v) Find F, the cumulative distribution function, on [0, 3]

(vi) Show that the median is $M = \sqrt{3}$

27.12 The p.d.f. of a distribution on $[0, \infty)$ is defined by

$$f(x) = \begin{cases} x(4 - x)/L & \text{if } 0 \le x < 2 \\ (6 - x)/L & \text{if } 2 \le x < 6 \\ 0 & \text{if } 6 \le x < \infty \end{cases}$$

where L is a constant. Find

(i) L

(ii) The mean

(iii) The variance

(iv) The coefficient of variation

(v) The cumulative distribution function

(vi) The median

27.13 A smooth probability density function f is defined on $[0, \infty)$ by

$$f(t) = \begin{cases} At + 0.01t^3 & \text{if } 0 \le t < 2 \\ Bt^2 + Ct & \text{if } 2 \le t \le 4 \\ 0 & \text{if } 4 \le t < \infty \end{cases}$$

(i) Find the values of A, B and C

(ii) Find the mean of the distribution

(iii) Find the median of the distribution, at least approximately

(iv) Find the cumulative distribution function

(v) Find the variance

(vi) Find the coefficient of variation

27.14 The p.d.f. of a size distribution on $[0, \infty)$ is defined by $f(x) = \frac{1}{L}g(x)$ where

$$g(x) = \begin{cases} x & \text{if } 0 \le x < 2 \\ 2 & \text{if } 2 \le x < 4 \\ 2(5-x) & \text{if } 4 \le x < 5 \\ 0 & \text{if } 5 \le x < \infty \end{cases}$$

and L is a constant.

(i) What must be the value of L?
(ii) Find M, the median
(iii) Find μ, the mean
(iv) Find σ^2, the variance
(v) Show that the coefficient of variation is $\kappa = 1/\sqrt{6}$.
(vi) Find F, the cumulative distribution function, on $[0, \infty)$
(vii) What is the probability of a size between 1 and 3?

27.15 The p.d.f. of a distribution on $[0, \infty)$ is defined by $f(x) = \frac{1}{L}g(x)$ where

$$g(x) = \begin{cases} \frac{5}{4}x & \text{if } 0 \le x < 1 \\ 2 - \frac{3}{4}x & \text{if } 1 \le x < 2 \\ 1 - \frac{1}{4}x & \text{if } 2 \le x < 4 \\ 0 & \text{if } 4 \le x < \infty \end{cases}$$

and L is a constant.

(i) What must be the value of L?
(ii) Find M, the median
(iii) Find μ, the mean
(iv) Find σ^2, the variance
(v) Show that the coefficient of variation is $\kappa = (2/3)^{3/2}$
(vi) Find F, the cumulative distribution function, on $[0, \infty)$
(vii) What is the probability of a size between 1 and 3?

Answers and Hints for Selected Exercises

27.2 The mean exists if $\theta > 1$, in which case, (18) has the sign of $9\theta - A(\theta+2)$. Because $A < 3$, by (8), this quantity must exceed $9\theta - 3(\theta+2) = 6(\theta-1)$, which is positive. Similarly for the variance.

27.4 (i)
$$\mu = \int_0^\infty t f(t)\,dt = \int_0^c t f(t)\,dt + \int_c^\infty t f(t)\,dt$$
$$= \int_0^c \left\{A(2ct - 2t^2) + \frac{\theta(1-Ac^2)}{(\theta+1)c}t\right\}dt + \frac{\theta(1-Ac^2)c^\theta}{(\theta+1)}\int_c^\infty t^{-\theta}\,dt$$

So, setting $\alpha = \theta$ in (17), the mean is finite only if $\theta > 1$. Then
$$\mu = \int_0^c \frac{d}{dt}\left\{A\left(ct^2 - \tfrac{2}{3}t^3\right) + \frac{\theta(1-Ac^2)}{2(\theta+1)c}t^2\right\}dt + \frac{\theta(1-Ac^2)c^\theta}{(\theta+1)}\int_c^\infty t^{-\theta}\,dt$$
$$= \left(A\left(ct^2 - \tfrac{2}{3}t^3\right) + \frac{\theta(1-Ac^2)}{2(\theta+1)c}t^2\right)\Bigg|_0^c + \frac{\theta(1-Ac^2)c^\theta}{(\theta+1)}\cdot\frac{c^{-(\theta-1)}}{\theta-1}$$
$$= \frac{1}{3}Ac^3 + \frac{\theta(1-Ac^2)c}{2(\theta+1)} + \frac{\theta(1-Ac^2)c}{(\theta+1)(\theta-1)} = \frac{3\theta c - Ac^3(\theta+2)}{6(\theta-1)}.$$

(ii)
$$\mu^2 + \sigma^2 = \int_0^\infty t^2 f(t)\,dt = \int_0^c t^2 f(t)\,dt + \int_c^\infty t^2 f(t)\,dt$$
$$= \int_0^c \left\{A(2ct^2 - 2t^3) + \frac{\theta(1-Ac^2)}{(\theta+1)c}t^2\right\}dt + \frac{\theta(1-Ac^2)c^\theta}{(\theta+1)}\int_c^\infty t^{-(\theta-1)}\,dt$$

So, setting $\alpha = \theta-1$ in (17), the variance is finite only if $\theta-1 > 1$, or $\theta > 2$. Then
$$\mu^2 + \sigma^2 = \int_0^c \frac{d}{dt}\left\{A\left(\tfrac{2}{3}ct^3 - \tfrac{1}{2}t^4\right) + \frac{\theta(1-Ac^2)}{3(\theta+1)c}t^3\right\}dt + \frac{\theta(1-Ac^2)c^\theta}{(\theta+1)}\,\frac{c^{-(\theta-2)}}{\theta-2}$$
$$= \left(A\left(\tfrac{2}{3}ct^3 - \tfrac{1}{2}t^4\right) + \frac{\theta(1-Ac^2)}{3(\theta+1)c}t^3\right)\Bigg|_0^c + \frac{\theta(1-Ac^2)c^2}{(\theta+1)(\theta-2)}$$
$$= \frac{1}{6}Ac^4 + \frac{\theta(1-Ac^2)c^2}{3(\theta+1)} - 0 + \frac{\theta(1-Ac^2)c^2}{(\theta+1)(\theta-2)}$$
$$= \frac{1}{6}Ac^4 + \frac{\theta(1-Ac^2)c^2}{3(\theta-2)},$$

after simplification, implying
$$\sigma^2 = \frac{\{2\theta - (\theta+2)Ac^2\}c^2}{6(\theta-2)} - \frac{\{3\theta - (\theta+2)Ac^2\}^2c^2}{36(\theta-1)^2}.$$

27.5 (i) Define g on $[0, \infty)$ by

$$g(x) = \begin{cases} 2x & \text{if } 0 \le x < 1 \\ 3 - x & \text{if } 1 \le x < 3 \\ 0 & \text{if } 3 \le x < \infty \end{cases}$$

Then $f(x) = g(x)/3$, and so

$$\mu = \int_0^\infty x f(x)dx = \int_0^3 x f(x)dx + \int_3^\infty x f(x)dx = \int_0^3 x f(x)dx + 0 = \frac{1}{3}\int_0^3 x g(x)dx.$$

But

$$\int_0^3 x g(x)dx = \int_0^1 2x^2 dx + \int_1^3 (3x - x^2)dx$$

$$= \int_0^1 \frac{d}{dx}\{\tfrac{2}{3}x^3\}dx + \int_1^3 \frac{d}{dx}\{\tfrac{3}{2}x^2 - \tfrac{1}{3}x^3\}dx$$

$$= \tfrac{2}{3}x^3\Big|_0^1 + \{\tfrac{3}{2}x^2 - \tfrac{1}{3}x^3\}\Big|_1^3$$

$$= \tfrac{2}{3}\cdot 1^3 - 0 + \left(\tfrac{27}{2} - 9\right) - \left(\tfrac{3}{2} - \tfrac{1}{3}\right)$$

$$= \tfrac{2}{3} + \tfrac{9}{2} - \tfrac{7}{6} = 4.$$

So $\mu = 4/3$.

(iii) Similarly,

$$\int_0^3 x^2 g(x)dx = \int_0^1 2x^3 dx + \int_1^3 (3x^2 - x^3)dx$$

$$= \int_0^1 \frac{d}{dx}\{\tfrac{1}{2}x^4\}dx + \int_1^3 \frac{d}{dx}\{x^3 - \tfrac{1}{4}x^4\}dx$$

$$= \tfrac{1}{2}x^4\Big|_0^1 + \{x^3 - \tfrac{1}{4}x^4\}\Big|_1^3$$

$$= \tfrac{1}{2} - 0 + \left(27 - \tfrac{81}{4}\right) - \left(1 - \tfrac{1}{4}\right)$$

$$= \tfrac{1}{2} + 26 - \tfrac{80}{4} = \tfrac{1}{2} + 6 = \tfrac{13}{2}.$$

So

$$\sigma^2 = \int_0^6 x^2 f(x)dx - \mu^2 \Rightarrow$$

$$\sigma^2 + \mu^2 = \int_0^3 x^2\left\{\frac{1}{3}g(x)\right\}dx = \frac{1}{3}\int_0^6 x^2 g(x)dx = \frac{1}{3}\cdot\frac{13}{2} = \frac{13}{6}$$

$$\Rightarrow \sigma^2 = \frac{13}{6} - \left(\frac{4}{3}\right)^2 = \frac{39 - 32}{18} = \frac{7}{18}$$

(iv) $\sigma = \sqrt{\dfrac{7}{18}} = \dfrac{1}{3}\sqrt{\dfrac{7}{2}}$ $\kappa = \sigma/\mu = \dfrac{1}{4}\sqrt{\dfrac{7}{2}} = 0.468$

27.6 $\sigma^2 = \dfrac{4}{25}$, $\kappa = \dfrac{1}{3}$

27.7 $\sigma^2 = \frac{4}{25},\ \kappa = \frac{1}{2}$

27.8 (i) Define g on $[0, \infty]$ by

$$g(x) = \begin{cases} 4x & \text{if } 0 \le x < 1 \\ 5-x & \text{if } 1 \le x < 5 \\ 0 & \text{if } 5 \le x < \infty \end{cases}$$

Then $f(x) = g(x)/10$, implying $\mu = \int_0^\infty x f(x)dx = \frac{1}{10}\int_0^5 x g(x)dx$. But

$$\int_0^5 xg(x)dx = \int_0^1 4x^2\,dx + \int_1^5 (5x - x^2)dx$$
$$= \int_0^1 \frac{d}{dx}\{\tfrac{4}{3}x^3\}dx + \int_1^5 \frac{d}{dx}\{\tfrac{5}{2}x^2 - \tfrac{1}{3}x^3\}dx$$
$$= \tfrac{4}{3}x^3\Big|_0^1 + \{\tfrac{5}{2}x^2 - \tfrac{1}{3}x^3\}\Big|_1^5$$
$$= \tfrac{4}{3} - 0 + \tfrac{125}{6} - \tfrac{13}{6} = 20.$$

So $\mu = 20/10 = 2$.

(ii) Similarly,

$$\int_0^5 x^2g(x)dx = \int_0^1 4x^3\,dx + \int_1^5 (5x^2 - x^3)dx$$
$$= \int_0^1 \frac{d}{dx}\{x^4\}dx + \int_1^5 \frac{d}{dx}\{\tfrac{5}{3}x^3 - \tfrac{1}{4}x^4\}dx$$
$$= x^4\Big|_0^1 + \{\tfrac{5}{3}x^3 - \tfrac{1}{4}x^4\}\Big|_1^5$$
$$= 1 - 0 + \tfrac{625}{12} - \tfrac{17}{12} = \tfrac{155}{3}.$$

So

$$\sigma^2 = \frac{1}{10}\int_0^5 x^2 g(x)dx - \mu^2 = \frac{155}{30} - 4 = \frac{31-24}{6} = \frac{7}{6}$$

(iii) Because

$$\text{Int}(f,[1,\infty)) = \text{Area of triangle of height } f(1) \text{ with base } 5-1 = \frac{1}{2}\cdot 4\cdot\frac{2}{5} = \frac{4}{5}$$

exceeds $1/2$, implying $M > 1$, to find the median we solve $1/2 = \text{Int}(f, [M, \infty)) = \text{Int}(f, [M, 5]) = (5–M)f(M)/2 = (5-M)^2/20$. So $5 – M = \sqrt{10}$, or $M = 5 – \sqrt{10} \approx 1.84$.

(iv) From the triangular shape of the graph of f, it is clear that m = 1.

(v) $\sigma = \sqrt{\frac{7}{6}} \Rightarrow \kappa = \sigma/\mu = \frac{1}{2}\sqrt{\frac{7}{6}} = 0.54.$

27.9 (i) Define g on $[0, \infty]$ by

$$g(x) = \begin{cases} 2x & \text{if } 0 \le x < 3 \\ 15 - 3x & \text{if } 3 \le x < 5 \\ 0 & \text{if } 5 \le x < \infty \end{cases}$$

Then $f(x) = g(x)/15$, implying $\mu = \int_0^\infty x f(x)dx = \frac{1}{15}\int_0^5 x g(x)dx$. But

$$\begin{aligned}\int_0^5 xg(x)dx &= \int_0^3 2x^2\,dx + \int_3^5 (15x - 3x^2)dx \\ &= \int_0^3 \frac{d}{dx}\{\tfrac{2}{3}x^3\}dx + \int_3^5 \frac{d}{dx}\left\{\tfrac{15}{2}x^2 - x^3\right\}dx \\ &= \tfrac{2}{3}x^3\Big|_0^3 + \left\{\tfrac{15}{2}x^2 - x^3\right\}\Big|_3^5 \\ &= 18 - 0 + \tfrac{125}{2} - \tfrac{81}{2} = 40.\end{aligned}$$

So $\mu = 40/15 = 8/3$.

(ii) Similarly,

$$\begin{aligned}\int_0^5 x^2g(x)dx &= \int_0^3 2x^3\,dx + \int_3^5 (15x^2 - 3x^3)dx \\ &= \int_0^3 \frac{d}{dx}\{\tfrac{1}{2}x^4\}dx + \int_3^5 \frac{d}{dx}\left\{5x^3 - \tfrac{3}{4}x^4\right\}dx \\ &= \tfrac{1}{2}x^4\Big|_0^3 + \left\{5x^3 - \tfrac{3}{4}x^4\right\}\Big|_3^5 \\ &= \tfrac{81}{2} - 0 + \tfrac{625}{4} - \tfrac{297}{4} = \tfrac{245}{2}.\end{aligned}$$

So

$$\sigma^2 = \frac{1}{15}\int_0^5 x^2 g(x)dx - \mu^2 = \frac{245}{30} - \frac{64}{9} = \frac{147 - 128}{18} = \frac{19}{18}$$

(iii) Because

$$\text{Int}(f,[0,3)) = \text{Area of triangle of height } f(3) \text{ with base } 3 = \frac{1}{2}\cdot 3 \cdot \frac{2}{5} = \frac{3}{5}$$

exceeds $1/2$, implying $M < 1$, to find the median we solve $1/2 = \text{Int}(f, [0, M]) = Mf(M)/2$. So $Mf(M) = 1$ or $2M^2/15 = 1$, implying $M = \sqrt{15/2} \approx 2.74$.

(iv) From the triangular shape of the graph of f, it is clear that m = 3.

(v) $$\sigma = \sqrt{\frac{19}{18}} = \frac{1}{3}\sqrt{\frac{19}{2}} \Rightarrow \kappa = \sigma/\mu = \frac{1}{3}\sqrt{\frac{19}{2}} \div \frac{8}{3} = \frac{1}{8}\sqrt{\frac{19}{2}} = 0.385.$$

27.11 Go to http://www.math.fsu.edu/~mm-g/QuizBank/mac3311.f96.html (Problem #1)

27.12 (i) Define g on $[0, \infty]$ by

$$g(x) = \begin{cases} x(4-x) & \text{if } 0 \le x < 2 \\ (6-x) & \text{if } 2 \le x < 6 \\ 0 & \text{if } 6 \le x < \infty \end{cases}$$

Then $f(x) = g(x)/L$, and $\text{Int}(f, [0, \infty]) = 1$, so

$$L = \int_0^\infty g(x)dx = \int_0^6 g(x)dx = \int_0^2 \{4x - x^2\}dx + \int_2^6 (6-x)dx$$

$$= \int_0^2 \frac{d}{dx}\{2x^2 - \tfrac{1}{3}x^3\}dx + \int_2^6 \frac{d}{dx}\{-\tfrac{1}{2}(6-x)^2\}dx$$

$$= 2x^2 - \tfrac{1}{3}x^3\Big|_0^2 + \{-\tfrac{1}{2}(6-x)^2\}\Big|_2^6$$

$$= 2\cdot 2^2 - \tfrac{1}{3}\cdot 2^3 - \tfrac{1}{2}(6-6)^2 - \{-\tfrac{1}{2}(6-2)^2\}$$

$$= \tfrac{16}{3} - 0 + 8 = \tfrac{40}{3}$$

(ii) $$\mu = \int_0^6 x f(x)dx = \int_0^6 x\left\{\frac{g(x)}{L}\right\}dx = \frac{1}{L}\int_0^6 x g(x)dx = \frac{3}{40}\int_0^6 x g(x)dx.$$

But

$$\int_0^6 xg(x)dx = \int_0^2 \{4x^2 - x^3\}dx + \int_2^6 (6x - x^2)dx$$

$$= \int_0^2 \frac{d}{dx}\{\tfrac{4}{3}x^3 - \tfrac{1}{4}x^4\}dx + \int_2^6 \frac{d}{dx}\{3x^2 - \tfrac{1}{3}x^3\}dx$$

$$= \tfrac{4}{3}x^3 - \tfrac{1}{4}x^4\Big|_0^2 + \{3x^2 - \tfrac{1}{3}x^3\}\Big|_2^6$$

$$= \tfrac{4}{3}\cdot 2^3 - \tfrac{1}{4}\cdot 2^4 + 3\cdot 6^2 - \tfrac{1}{3}\cdot 6^3 - \{3\cdot 2^2 - \tfrac{1}{3}\cdot 2^3\}$$

$$= \tfrac{20}{3} + 36 - \tfrac{28}{3} = \tfrac{100}{3}.$$

So $\mu = \tfrac{3}{40}\tfrac{100}{3} = \tfrac{5}{2}$.

(iii) Similarly,

$$\int_0^6 x^2 g(x)dx = \int_0^2 \{4x^3 - x^4\}dx + \int_2^6 (6x^2 - x^3)dx$$

$$= \int_0^2 \frac{d}{dx}\{x^4 - \tfrac{1}{5}x^5\}dx + \int_2^6 \frac{d}{dx}\{2x^3 - \tfrac{1}{4}x^4\}dx$$

$$= x^4 - \tfrac{1}{5}x^5\Big|_0^2 + \{2x^3 - \tfrac{1}{4}x^4\}\Big|_2^6$$

$$= 2^4 - \tfrac{1}{5}\cdot 2^5 + 2\cdot 6^3 - \tfrac{1}{4}\cdot 6^4 - \{2\cdot 2^3 - \tfrac{1}{4}\cdot 2^4\}$$

$$= \tfrac{48}{5} + 108 - 12 = \tfrac{528}{5}.$$

So

$$\sigma^2 = \int_0^6 x^2 f(x)dx - \mu^2 \Rightarrow$$

$$\sigma^2 + \mu^2 = \int_0^6 x^2 \left\{\frac{g(x)}{L}\right\}dx = \frac{1}{L}\int_0^6 x^2 g(x)dx = \frac{3}{40}\frac{528}{5} = \frac{198}{25}$$

$$\Rightarrow \sigma^2 = \frac{198}{25} - \left(\frac{5}{2}\right)^2 = \frac{792-625}{100} = 1.67$$

(iv) $\sigma = \frac{1}{10}\sqrt{167} \Rightarrow$ $\kappa = \sigma/\mu = \sqrt{167}/25 = 0.517$

(v) Suppose $0 \le t \le 2$. Then

$$F(t) = \int_0^t f(x)dx = \frac{1}{L}\int_0^t g(x)dx = \frac{3}{40}\int_0^t \{4x - x^2\}dx$$

$$= \frac{3}{40}\int_0^t \frac{d}{dx}\{2x^2 - \tfrac{1}{3}x^3\}dx = \frac{3}{40}\left(2x^2 - \tfrac{1}{3}x^3\Big|_0^t\right)$$

$$= \frac{3}{40}\left(2\cdot t^2 - \tfrac{1}{3}\cdot t^3\right) = \frac{t^2}{40}(6-t)$$

Note in particular that $F(2) = 2/5$. Thus, for $2 \le t \le 6$,

$$F(t) = \int_0^t f(x)dx = \int_0^2 f(x)dx + \int_2^t f(x)dx = F(2) + \frac{1}{L}\int_2^t g(x)dx$$

$$= \frac{2}{5} + \frac{3}{40}\int_2^t (6-x)dx = \frac{2}{5} + \frac{3}{40}\int_2^t \frac{d}{dx}\{-\tfrac{1}{2}(6-x)^2\}dx$$

$$= \frac{2}{5} + \frac{3}{40}\{-\tfrac{1}{2}(6-x)^2\}\Big|_2^t = \frac{2}{5} + \frac{3}{40}\left\{-\tfrac{1}{2}(6-t)^2 - \{-\tfrac{1}{2}(6-2)^2\}\right\}$$

$$= 1 - \frac{3}{80}(6-t)^2.$$

(vi) $F(2) = 2/5$ is less than $1/2$, so $2 < M < 6$. Thus $1 - F(M) = 1/2 \Rightarrow$

$$1 - \frac{3}{80}(6-t)^2 = \frac{1}{2} \Rightarrow (6-M)^2 = \frac{40}{3} \Rightarrow M = 6 - \sqrt{\frac{40}{3}} = 2.35.$$

27.14 (i) L = Int(g, [0,5]) = Area under quadrilateral = 7

(ii) M = 11/4

(iii) $\mu = 19/7$

(iv) $\mu^2 + \sigma^2 = 361/42$, implying $\sigma^2 = 361/294$

(v) $\sigma = 19/\sqrt{294} = 19/7\sqrt{6}$, implying $\kappa = \sigma/\mu = 1/\sqrt{6}$

(vi) From

$$f(x) = \begin{cases} \frac{1}{7}x & \text{if } 0 \le x < 2 \\ \frac{2}{7} & \text{if } 2 \le x < 4 \\ \frac{2}{7}(5-x) & \text{if } 4 \le x < 5 \\ 0 & \text{if } 5 \le x < \infty \end{cases}$$

and F(x) = Int(f, [0, x]) we have

$$F(x) = \begin{cases} \frac{1}{14}x^2 & \text{if } 0 \le x < 2 \\ \frac{2}{7}(x-1) & \text{if } 2 \le x < 4 \\ 1-\frac{1}{7}(5-x)^2 & \text{if } 4 \le x < 5 \\ 1 & \text{if } 5 \le x < \infty \end{cases}$$

(vii) $F(3) - F(1) = 4/7 - 1/14 = 1/2$

27.15 L = Int(g, [0,4]) = Area under re-entrant quadrilateral = 2

(ii) M = 4/3

(iii) $\mu = 3/2$

(iv) $\mu^2 + \sigma^2 = 35/12$, implying $\sigma^2 = 2/3$

(v) $\sigma = \sqrt{2/3}$, implying $\kappa = \sigma/\mu = 1/\sqrt{6}$

(vi) From

$$f(x) = \begin{cases} \frac{5}{8}x & \text{if } 0 \le x < 1 \\ 1-\frac{3}{8}x & \text{if } 1 \le x < 2 \\ \frac{1}{8}(4-x) & \text{if } 2 \le x < 4 \\ 0 & \text{if } 4 \le x < \infty \end{cases}$$

and F(x) = Int(f, [0, x]) we have

$$F(x) = \begin{cases} \frac{5}{16}x^2 & \text{if } 0 \le x < 1 \\ x-\frac{3}{16}x^2-\frac{1}{2} & \text{if } 1 \le x < 2 \\ 1-\frac{1}{16}(4-x)^2 & \text{if } 2 \le x < 4 \\ 1 & \text{if } 4 \le x < \infty \end{cases}$$

(vii) $F(3) - F(1) = 13/16 - 5/16 = 1/2$