Chapter 5

Continuous Random Variables

5.1 Basic Results

5.1.1 Definitions: the c.d.f. and p.d.f.

Our previous definition of a discrete random variable was rather restrictive. A broader definition is as follows:

A *random variable* X defined on the probability space $(\mathcal{S}, \mathcal{F}, \mathrm{P})$ is a mapping of $\mathcal{S}$ into the set $\mathcal{R}$ of real numbers such that, if B_x denotes the subset of outcomes in $\mathcal{S}$ which are mapped onto the set $(-\infty, x]$, then

$$B_x \in \mathcal{F} \qquad \text{for all } x \in \mathcal{R}.$$

We write

$$\mathrm{P}(X \leq x) = \mathrm{P}(B_x).$$

We note that a discrete r.v. satisfies this definition.

Discrete r.v.s are generally studied through their probability functions; r.v.s in the broader sense are studied through their cumulative distribution functions. The *cumulative distribution function* (c.d.f.) F_X of a r.v. X is the function

$$F_X(x) = \mathrm{P}(X \leq x), \qquad -\infty < x < \infty. \tag{5.1}$$

As with the probability function, the suffix 'X' may be dropped when there is no ambiguity. The c.d.f. has the following properties:

(i) $F(x) \leq F(y)$ if $x \leq y$, i.e. $F(.)$ is monotonic non-decreasing;

(ii) $F(-\infty) = 0$, $F(+\infty) = 1$;

(iii) F is continuous from the right, i.e.

$$F(x+h) \rightarrow F(x) \quad \text{as } h \rightarrow 0^+;$$

(iv) $\mathrm{P}(a < X \leq b) = F(b) - F(a)$.

For a *discrete* r.v., the c.d.f. is a step-function, i.e. at certain points it is discontinuous (from the left). So it is natural to develop a theory for r.v.s with *continuous* c.d.f.s. However, it is sufficient for practical purposes to consider c.d.f.s which are also reasonably *smooth*.

A r.v. X defined on $(\mathcal{S}, \mathcal{F}, \mathrm{P})$ is said to be *continuous* if

(i) its c.d.f. $F(x), -\infty < x < \infty$, is a continuous function;

(ii) there exists a non-negative function $f(x)$ such that

$$F(x) = \int_{-\infty}^{x} f(t)dt, \quad -\infty < x < \infty. \tag{5.2}$$

An alternative form of the second condition is:

(ii*) whose derivative $\dfrac{dF(x)}{dx} = f(x)$ exists and is continuous except possibly at a finite number of points.

The function $f(x)$, $-\infty < x < \infty$, is called the *probability density function* (p.d.f.) of the r.v. X.

Theorem If X is a continuous r.v.,

$$\mathrm{P}(X = x) = 0 \qquad \text{for all x.} \tag{5.3}$$

Proof For any r.v. X with c.d.f. F,

$$\begin{aligned} \lim_{h\to 0^+} F(x+h) &= F(x) = \mathrm{P}(X \le x); \text{ and} \\ \lim_{h\to 0^+} F(x-h) &= \mathrm{P}(X < x), \qquad \text{for all } x. \end{aligned}$$

If X is continuous, $F(x)$ is a continuous function of x, i.e.

$$\lim_{h\to 0^+} F(x-h) = \lim_{h\to 0^+} F(x+h)$$

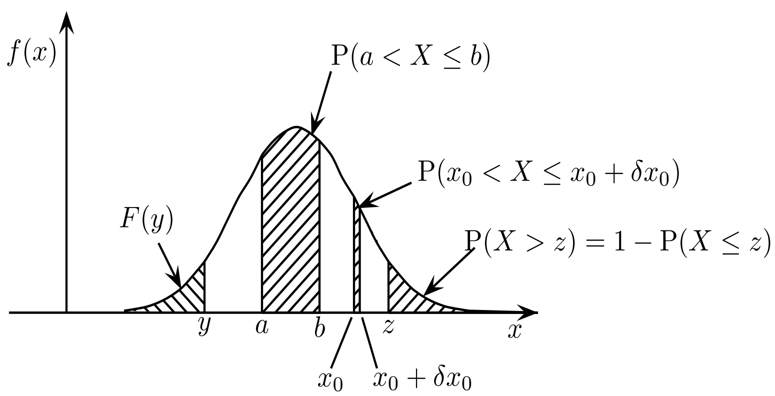
i.e.

$$\mathrm{P}(X < x) = \mathrm{P}(X \le x) = \mathrm{P}(X < x) + \mathrm{P}(X = x)$$

i.e.

$$\mathrm{P}(X = x) = 0 \qquad \text{for all } x. \qquad \square$$

For any meaningful statement about probabilities we must consider X lying in an interval or intervals. Probability is represented as an *area* under the curve of the p.d.f. $f(x)$.



$$P(x_0 \le X \le x_0 + \delta x_0) = \int_{x_0}^{x_0+\delta x_0} \approx f(x_0)\delta x_0 \text{ when } \delta x_0 \text{ is small.}$$

So $f(x)$ itself is not a probability, but in the above sense $f(x)$ is *proportional* to (or a measure of) probability.

To summarise: the p.g.f. $f(x)$ has the following properties:

(i) $f(x) \ge 0, \quad -\infty < x < \infty, \quad$ and $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1.$

(ii) $f(x)$ *is not a probability*: $f(x)\delta x \approx P(x < X \le x + \delta x)$.

(iii) Since $P(X = x) = 0$ for all x,

$$\begin{aligned} P(a < X < b) &= P(a \le X < b) = P(a < X \le b) = P(a \le X \le b) \\ &= F(b) - F(a) \\ &= \int_a^b f(t)dt. \end{aligned}$$

5.1.2 Restricted Range

Suppose that

$$f(x) = \begin{cases} > 0 & \text{for } A < x < B, \text{ a subset of } (-\infty, \infty) \\ = 0 & \text{otherwise.} \end{cases}$$

Then probability (and other) calculations may be based on the interval (A, B) rather than on $(-\infty, \infty)$: e.g.

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{A} f(x)dx + \int_{A}^{B} f(x)dx + \int_{B}^{\infty} f(x)dx = \int_{A}^{B} f(x)dx; \\ P(X \le y) &= \int_{-\infty}^{y} f(x)dx &= \int_{A}^{y} f(x)dx, \quad y \ge A; \\ P(X > z) &= \int_{z}^{\infty} f(x)dx &= \int_{z}^{B} f(x)dx, \quad z \le B. \end{aligned}$$

5.1.3 Expectation

The *expected value* or *expectation* of a continuous r.v. X with p.d.f. $f(x)$ is denoted by $E(X)$ and is defined as

$$E(X) = \int_{-\infty}^{\infty} x f(x)dx \tag{5.4}$$

provided that the integral is absolutely convergent (i.e. $\int_{-\infty}^{\infty} |x| f(x)dx$ is finite). As in the discrete case, $E(X)$ is often termed the *expected value* or *mean* of X.

The continuous analogue of the 'law of the unconscious statistician' states that

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \tag{5.5}$$

provided the integral is absolutely convergent: once again, this is a *result* and not a *definition*. An immediate application is to the *variance* of X, denoted by $\text{Var}(X)$ and defined as

$$\text{Var}(X) = E([X - E(X)]^2). \tag{5.6}$$

Writing $\mu = \mathrm{E}(X)$, we have

$$\begin{aligned} \mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \\ &= \int_{-\infty}^{\infty} x^2 f(x)dx - 2\mu \int_{-\infty}^{\infty} x f(x)dx + \mu^2 \int_{-\infty}^{\infty} f(x)dx \\ &= \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2, \end{aligned}$$

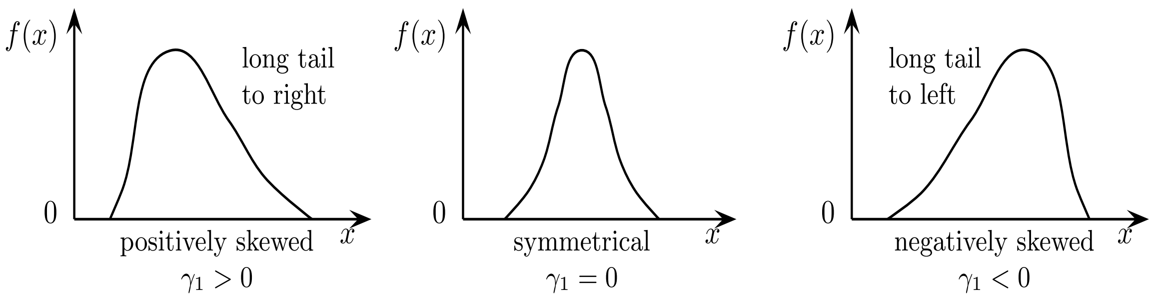
i.e.

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2, \tag{5.7}$$

just as in the discrete case. Other properties of the E and Var operators carry over in the same way – in the proofs we simply replace $\sum$ and probability function by $\int$ and p.d.f.

In modelling a set of data (see later), we are often interested in the *shape* of $f(x)$: this can be summarised by a measure of *asymmetry* called the *coefficient of skewness*, defined as

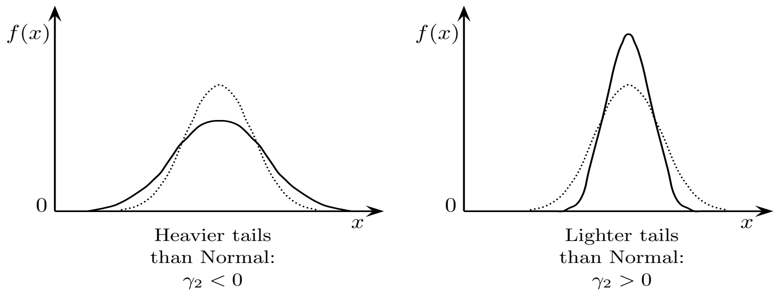
$$\gamma_1 = \frac{\mu_3}{\sigma^3}. \tag{5.8}$$



For symmetrical p.d.f.s, a measure of *peakedness* is the *coefficient of kurtosis*, defined as

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3. \tag{5.9}$$

Comparison is with the normal distribution, for which $\gamma_2 = 0$ (see §5.4.1).



Percentiles The $100q^{\text{th}}$ percentile point is the value $x_{[100q]}$ such that

$$\mathrm{P}(X \leq x_{[100q]}) = F(x_{[100q]}) = q, \quad 0 < q < 1. \tag{5.10}$$

In particular, the *median* m is $x_{[50]}$ and

$$F(m) = 0.5. \tag{5.11}$$

5.2 Transformations

5.2.1 Simpler Cases

Given a transformation $Y = g(X)$, where X is a continuous r.v. with known p.d.f $f_X(x)$, how can we determine the distribution of the r.v. Y? There are two straightforward cases:

(i) If $Y = g(X)$ where Y is a *discrete* r.v. and $Y = y_i$ corresponds to the interval $a_i < X < b_i$ (or a set of intervals), then

$$\mathrm{P}(Y = y_i) = \int_{a_i}^{b_i} f(x)dx.$$

(ii) Suppose $Y = g(X)$ where g is one-to-one and differentiable. Then Y is a continuous r.v. with p.d.f.

$$f_Y(y) = f_X\{g^{-1}(y)\}\left|\frac{dx}{dy}\right|, \quad g(-\infty) < y < g(+\infty). \tag{5.12}$$

(You may find it helpful to remind yourself of the proof of this result given in SOR101).

5.2.2 The Many-to-One Case

Now suppose that the transformation is no longer one-to-one, but *many-to-one*. There are two possible procedures, which we shall illustrate by considering the transformation

$$Y = X^2$$

where the range of X is $(-\infty, \infty)$ – a two-to-one transformation.

Method 1 Proceed through the c.d.f. (compare the proof of the result for a one-to-one transformation): thus

$$\begin{aligned} F_Y(y) &= \mathrm{P}(Y \le y) \\ &= \mathrm{P}(X^2 \le y) \\ &= \mathrm{P}(-\sqrt{y} \le X \le \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}), \quad y \ge 0. \end{aligned}$$

Then differentiating we get

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} \\ &= \frac{d}{dy}\{F_X(\sqrt{y}) - F_X(-\sqrt{y})\} \\ &= f_X(\sqrt{y})\left(\frac{1}{2\sqrt{y}}\right) - f_X(-\sqrt{y})\left(-\frac{1}{2\sqrt{y}}\right) \\ &= \frac{1}{2\sqrt{y}}\{f_X(\sqrt{y}) + f_X(-\sqrt{y})\}, \quad 0 \le y < \infty. \end{aligned}$$

Method 2 Express the transformation in terms of separate one-to-one transformations (so that the result for the one-to one case can be invoked). Here

$$X = \begin{cases} +\sqrt{Y}, & \text{for } 0 \le X < \infty \\ -\sqrt{Y}, & \text{for } -\infty < X < 0. \end{cases}$$

Let

$$f_X(x) = f_X^+(x) + f_X^-(x),$$

where

$$f_X^+(x) = \begin{cases} f_X(x), & 0 \le x < \infty \\ 0, & \text{otherwise} \end{cases}$$

and

$$f_X^-(x) = \begin{cases} f_X(x), & x < 0 \\ 0, & \text{otherwise} \end{cases}$$

We now use the formula for the separate one-to-one transformations:

$$\begin{aligned} f_Y^+(y) &= f_X^+(+\sqrt{y}) \left| \frac{1}{2\sqrt{y}} \right|, \quad 0 \le y < \infty \\ f_Y^-(y) &= f_X^-(-\sqrt{y}) \quad \left| -\frac{1}{2\sqrt{y}} \right|, \quad 0 \le y < \infty. \end{aligned}$$

Hence

$$\begin{aligned} f_Y(y) &= f_Y^+(y) + f_Y^-(y) \\ &= \frac{1}{2\sqrt{y}} \{f_X(\sqrt{y}) + f_X(-\sqrt{y})\}, \quad 0 \le y < \infty. \end{aligned}$$

- as obtained by Method 1.

5.2.3 Truncation

Suppose that the continuous r.v. X has p.d.f. $f_X(x), -\infty < x < \infty$ and the r.v. Y has similar properties to X in the interval (A, B) and is defined to be 0 elsewhere, i.e.

$$Y = \begin{cases} X, & A \le X \le B \\ 0, & \text{otherwise} \end{cases} .$$

Then

$$\begin{aligned} F_Y(y) &= \mathrm{P}(Y \le y) = \mathrm{P}(X \le y | A \le X \le B) \\ &= \frac{\mathrm{P}(X \le y \text{ and } A \le X \le B)}{\mathrm{P}(A \le X \le B)} \\ &= \frac{\mathrm{P}(A \le X \le y \le B)}{\mathrm{P}(A \le X \le B)} = \frac{F_X(y) - F_X(A)}{F_X(B) - F_X(A)}, \quad A \le y \le B. \end{aligned}$$

Hence

$$f_Y(y) = \begin{cases} \dfrac{f_X(y)}{\int_A^B f_X(x)dx}, & A \le y \le B \\ 0 & \text{otherwise} \end{cases} .$$

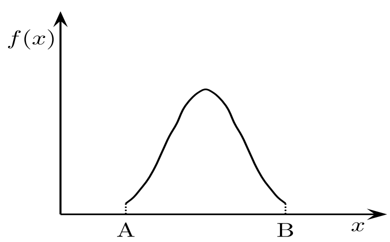
An alternative argument is as follows. Suppose that the p.d.f. of Y has a similar form to the p.d.f. of X in the interval (A, B) and is zero otherwise. Thus

$$f_Y(y) = \begin{cases} K f_X(y), & A \le y \le B \\ 0, & \text{otherwise} \end{cases} .$$

The normalisation requirement $\int_{-\infty}^{\infty} f_Y(y)dy = 1$ yields

$$K = \left[\int_A^B f_X(y)dy \right]^{-1}$$

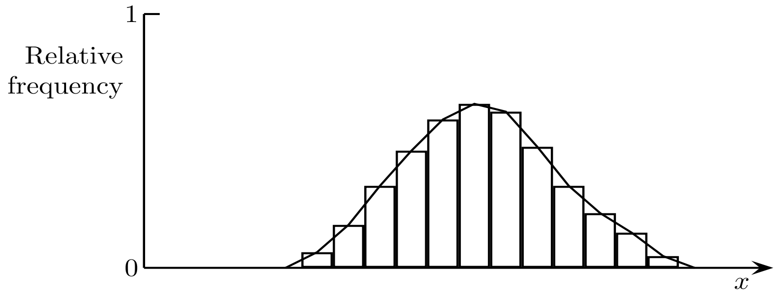
– the same result as above.



5.3 Modelling

Many *discrete* probability distributions are considered as suitable models for certain standard situations for which a probabilistic analysis is possible, e.g. binomial, geometric, Poisson, hypergeometic.

On the other hand, a *continuous* distribution is often chosen as a model because of its shape, particularly if a large sample of observations is available: if we construct a relative frequency histogram, the piecewise linear construction joining the mid-points of the tops of adjacent bars should approximate the curve of a suitable p.d.f.



Real data, which must lie within a finite interval, *may* be modelled by a p.d.f. defined over an infinite or semi-infinite range, since the infinite tail(s) of many p.d.f.s contain very little probability. For example, if $X \sim N(\mu, \sigma^2)$, $\mathrm{P}(X \geq \mu + 4\sigma) \approx 0.000032$. Thus, positive data may be modelled by a continuous r.v. which theoretically can take negative values.

5.4 Important Continuous Distributions

5.4.1 Normal distribution

Also known as the Gaussian distribution, this has two parameters (μ, σ^2): it is the most important distribution in statistics. If $X \sim N(\mu, \sigma^2)$, its p.d.f. is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad -\infty < x < \infty. \tag{5.13}$$

The r.v. $W = a+bX$ is distributed $N(a+b\mu, b^2\sigma^2)$ – an example of a one-to-one transformation. In particular, the r.v. $Z = (X-\mu)/\sigma$ is distributed $N(0,1)$ (the *standard normal* distribution), with p.d.f.

$$f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty. \tag{5.14}$$

Many properties of X (probabilities, moments, etc.) are readily derived from those of Z.

$$\begin{aligned}
\mathrm{E}(Z) &= \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\
&= \int_0^{\infty} \cdots + \int_{-\infty}^{0} \cdots \qquad \text{(set } y = -z \text{ in 2nd integral)} \\
&= \int_0^{\infty} \cdots + \int_{\infty}^{0} (-y) \frac{1}{\sqrt{\pi}} e^{-\frac{1}{2}y^2} (-dy) \\
&= \int_0^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz - \int_0^{\infty} y \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = 0.
\end{aligned}$$

Hence

$$\mathrm{E}\left[\frac{X-\mu}{\sigma}\right] = 0, \qquad \text{giving } \mathrm{E}(X) = \mu. \tag{5.15}$$

This result can also be deduced from the observation that $f_X(x)$ is symmetrical about $x = \mu$.

Since $\mathrm{P}(X \le \mu) = \frac{1}{2} = \mathrm{P}(X \ge \mu)$, μ is also the median (it is also the mode).

To find $\mathrm{Var}(X)$, we first consider $\mathrm{Var}(Z) = \mathrm{E}(Z^2)$.

$$\begin{aligned} \mathrm{E}(Z^2) &= \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\ &= \int_0^{\infty} \cdots + \int_{-\infty}^{0} \cdots \qquad \text{(set } y = -z \text{ in 2nd integral)} \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} z^2 e^{-\frac{1}{2}z^2} dz. \end{aligned}$$

For integrals over $(0, \infty)$ with an integrand consisting of a power term and an exponential term, one should try transforming into a *Gamma function*, defined as

$$\Gamma(p) = \int_0^{\infty} t^{p-1} e^{-t} dt, \qquad p > 0. \tag{5.16}$$

Given p, $\Gamma(p)$ can be found from tables or by means of a computer program. Some useful properties of the Gamma function are:

$$\begin{aligned} \Gamma(p+1) &= p\Gamma(p) \\ \Gamma(n+1) &= n! \qquad \text{for non-negative integer } n \\ \Gamma(\tfrac{1}{2}) &= \sqrt{\pi}. \end{aligned} \tag{5.17}$$

So, setting $t = \frac{1}{2}z^2, dt = zdz = \sqrt{2t}dz$, we get

$$\begin{aligned} \mathrm{E}(Z^2) &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} 2te^{-t} \frac{1}{\sqrt{2t}} dt \\ &= \frac{2}{\sqrt{\pi}} \int_0^{\infty} t^{\frac{1}{2}} e^{-t} dt \\ &= \frac{2}{\sqrt{\pi}} \Gamma(\tfrac{3}{2}) \\ &= \frac{2}{\sqrt{\pi}} . \frac{1}{2} \Gamma(\tfrac{1}{2}) = \frac{2}{\sqrt{\pi}} . \frac{1}{2} \sqrt{\pi} = 1. \end{aligned}$$

So $\mathrm{Var}(Z) = 1$ and

$$\mathrm{Var}(X) = \mathrm{Var}(\mu + \sigma Z) = \sigma^2 \mathrm{Var}(Z) = \sigma^2. \tag{5.18}$$

By a similar argument to that used for $\mathrm{E}(Z)$, we find that

$$\mathrm{E}(Z^3) = 0,$$

so

$$\mu_3 = \mathrm{E}[(X - \mu)^3] = 0$$

and the coefficient of skewness is

$$\gamma_1 = 0. \tag{5.19}$$

More generally,

$$\mathrm{E}(Z^{2r+1}) = 0 \text{ and } \mu_{2r+1} = 0, \quad r \ge 1.$$

Also, by a similar analysis to that for $\mathrm{E}(Z^2)$, it can be shown that $\mathrm{E}(Z^4) = 3$, so

$$\mu_4 = \mathrm{E}[(X - \mu)^4] = 3\sigma^4$$

and the coefficient of kurtosis is

$$\gamma_2 = 0. \tag{5.20}$$

5.4.2 (Negative) Exponential distribution

This important distribution has 1 parameter (λ): its p.d.f. and c.d.f. are

$$\begin{aligned} f(x) &= \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \lambda > 0, \\ 0, & x < 0. \end{cases} \\ F(y) &= \begin{cases} 1 - e^{-\lambda y}, & y \geq 0, \\ 0, & y < 0. \end{cases} \end{aligned} \qquad (5.21)$$

while the mean and variance are

$$\mathrm{E}(X) = 1/\lambda, \quad \mathrm{Var}(X) = 1/\lambda^2. \qquad (5.22)$$

(See fig. on p.71 for shapes). We write $X \sim \mathrm{Exp}(\lambda)$. This is the only continuous distribution with the 'no memory' property (see HW Examples 6).

In a Poisson process, with parameter (rate) λ, the time to the first event (and the time between successive events) is distributed $\mathrm{Exp}(\lambda)$ (see final chapter).

5.4.3 Gamma distribution

This distribution has 2 parameters (α, λ), and its p.d.f. is

$$f(x) = \begin{cases} \dfrac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, & x \geq 0; \quad \alpha, \lambda > 0 \\ 0, & x < 0, \end{cases} \qquad (5.23)$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt, (\alpha > 0)$ is the *Gamma function* already introduced in (5.16), with properties (5.17). Here α is an index or *shape* parameter, λ a *scale* parameter. We write $X \sim \mathrm{Gamma}(\alpha, \lambda)$.

For *integer* α this distribution is often termed the **Erlang** distribution: this case is of considerable importance because X can then be written as the sum of α i.i.d. exponential r.v.s. Note that in particular

$$\mathrm{Gamma}(\alpha = 1, \lambda) \equiv \mathrm{Exp}(\lambda). \qquad (5.24)$$

The Gamma distribution is very useful for modelling data over the range $(0, \infty)$: by selecting various values of α, quite a range of different shapes of p.d.f. can be obtained (see fig.).

The mean is

$$\begin{aligned} \mathrm{E}(X) &= \int_0^\infty x.\frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} dx \qquad (\text{set } t = \lambda x, dt = \lambda dx) \\ &= \frac{1}{\Gamma(\alpha)} \int_0^\infty t^\alpha e^{-t} \frac{1}{\lambda} dt \\ &= \frac{1}{\Gamma(\alpha)\lambda} \Gamma(\alpha + 1) = \frac{\alpha}{\lambda}. \end{aligned} \qquad (5.25)$$

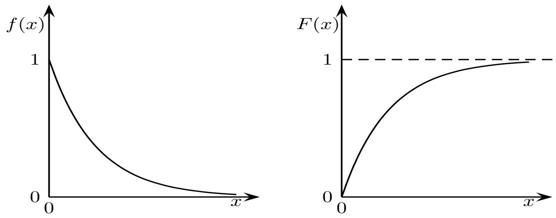
Similarly

$$\mathrm{E}(X^2) = \frac{\alpha(\alpha+1)}{\lambda^2}, \qquad \text{so} \quad \mathrm{Var}(X) = \frac{\alpha}{\lambda^2}. \qquad (5.26)$$

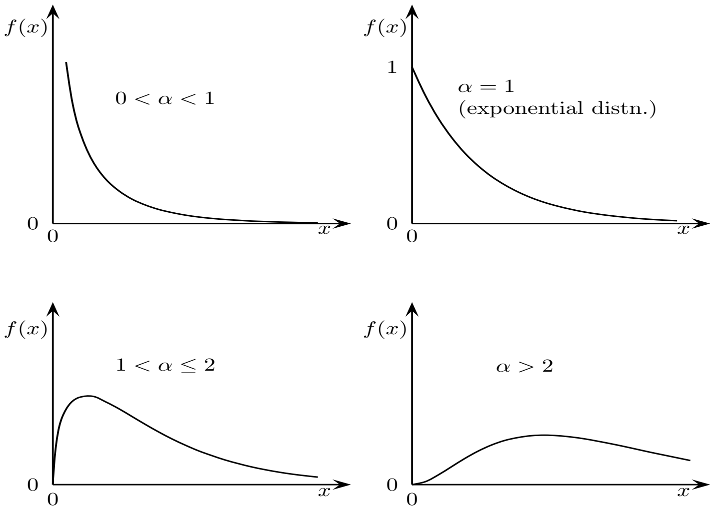
In a Poisson process with parameter (rate) λ, the time to the r^{th} event (and the time between the m^{th} and $(m+r)^{\text{th}}$ events) is distributed $\mathrm{Gamma}(r, \lambda)$ (see final chapter).

The shapes of some common distributions

Exponential

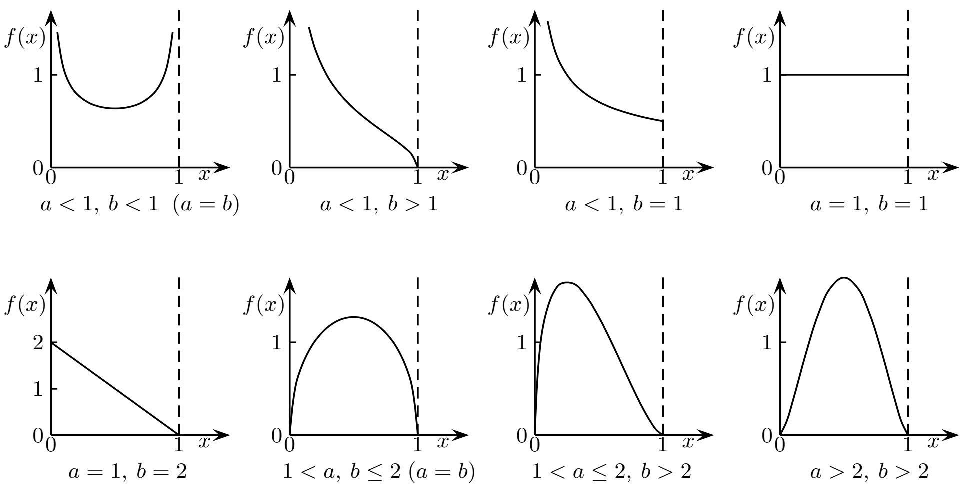


Gamma

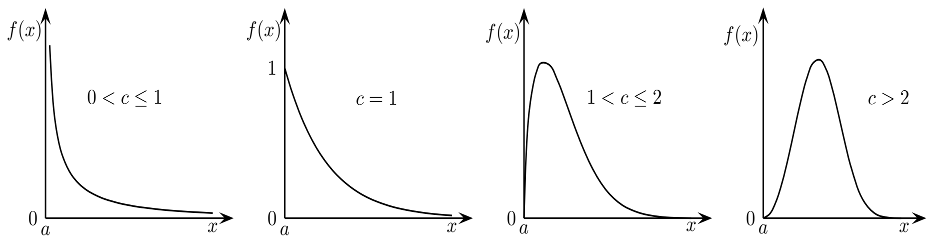


(the mode moves to the right as α increases)

Beta



Weibull



5.4.4 Beta distribution

This distribution has 2 parameters (a, b), and the p.d.f is

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, & 0 \le x \le 1; \quad a, b > 0 \\ 0, & \text{otherwise,} \end{cases} \tag{5.27}$$

where

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad (a, b > 0) \tag{5.28}$$

is the *Beta function.* Also

$$\begin{aligned} \mathrm{E}(X) &= \int_0^1 x . \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} dx \\ &= \frac{1}{B(a,b)} \int_0^1 x^a (1-x)^{b-1} dx \\ &= \frac{B(a+1,b)}{B(a,b)} = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} . \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \end{aligned}$$

i.e.

$$\mathrm{E}(X) = \frac{a}{a+b}. \tag{5.29}$$

Similarly we can show that

$$\mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}. \tag{5.30}$$

Again selection of values of a and b gives different shapes for the p.d.f. (see figures on p.71). Note that these shapes may be reversed by interchanging the values of a and b, since, if $X \sim \mathrm{Beta}(a, b)$, then $1 - X \sim \mathrm{Beta}(b, a)$.

This family of distributions is useful for modelling data over a *finite* range: the standard p.d.f. (given above) is defined over $[0, 1]$, but it may also be defined over $[A, B]$ where A and B are both finite. Thus, if we write $Z = A + (B - A)X$, i.e. $A \le Z \le B$, then

$$f_Z(z) = \frac{1}{B(a,b)} . \frac{(z-A)^{a-1}(B-z)^{b-1}}{(B-A)^{a+b-1}}, \qquad A \le z \le B. \tag{5.31}$$

5.4.5 Uniform (or Rectangular) distribution

This simple distribution has 2 parameters (a, b): the p.d.f. and c.d.f. are

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b, \\ 0, & \text{otherwise.} \end{cases} \tag{5.32}$$

and

$$F(x) = \begin{cases} 0, & \text{if } x \le a, \\ (x-a)/(b-a), & \text{if } a < x \le b, \\ 1, & \text{if } x > b. \end{cases} \tag{5.33}$$

Also,

$$\mathrm{E}(X) = \frac{1}{2}(a+b), \qquad \mathrm{Var}(X) = \frac{1}{12}(b-a)^2. \tag{5.33}$$

The case $a = 0, b = 1$ is particularly important (e.g. for random number generation in simulation).

5.4.6 Weibull distribution

This distribution (particularly associated with lifetime and reliability studies) has, in its most general form, 3 parameters (a, b, c): the p.d.f., c.d.f. and mean are

$$\begin{aligned} f(x) &= \begin{cases} \dfrac{c(x-a)^{c-1}}{b^c} \exp\left\{-\left(\dfrac{x-a}{b}\right)^c\right\}, & x \geq a, \\ 0, & x < a. \end{cases} \\ F(y) &= \begin{cases} 1 - \exp\left\{-\left(\dfrac{y-a}{b}\right)^c\right\}, & y \geq 0, \\ 0, & y < 0. \end{cases} \\ \mathrm{E}(X) &= a + b\Gamma(1 + 1/c). \end{aligned} \quad (5.35)$$

Selection of c determines the *shape* of the p.d.f. (see fig. for examples); b is a *scale* parameter and a a *location* parameter. This distribution has properties similar to the Gamma distribution. Note that $\quad$ $\text{Weibull}(a = 0, b, c = 1) \equiv \text{Exp}(1/b)$.

5.4.7 Chi-squared distribution

Several other distributions arise frequently in statistical inference. Here we mention only the *chi-squared* (χ^2) *distribution with* n *degrees of freedom*, sometimes written χ^2_n or $\chi^2(n)$, which has p.d.f.

$$f(x) = \begin{cases} \dfrac{1}{2\Gamma(\frac{1}{2}n)} (\tfrac{1}{2}x)^{\frac{1}{2}n-1} e^{-\frac{1}{2}x}, & x > 0, \\ 0, & x < 0 \end{cases} \quad (5.36)$$

and mean

$$\mathrm{E}(X) = n. \quad (5.37)$$

We observe that in fact

$$\chi^2_n \equiv \text{Gamma}(\alpha = n/2, n \text{ a positive integer}, \lambda = \tfrac{1}{2}). \quad (5.38)$$

5.5 Reliability

Let the continuous r.v. X, with c.d.f. $F(x)$ and p.d.f. $f(x)$, $x \geq 0$ denote the *lifetime* of some device or component: the device is said to *fail* at time X. There are a number of functions used in reliability studies:

Survival function	$\overline{F}(x) = 1 - F(x) = \mathrm{P}(X > x), \quad x \geq 0$
Hazard function	$H(x) = -\log(1 - F(x)), \quad x \geq 0$ (5.39)
Hazard rate function	$r(x) = \dfrac{f(x)}{\overline{F}(x)} = \dfrac{dH(x)}{dx}, \quad x \geq 0.$

The significance of $r(x)$ may be derived as follows. The probability that the device fails during $(x, x+h)$ given that it has not failed by time x is

$$\mathrm{P}(x \leq X \leq x+h | X > x) = \{F(x+h) - F(x)\}/\overline{F}(x).$$

Then

$$\begin{aligned}
\lim_{h \to 0} \left\{ \frac{\text{above prob.}}{h} \right\} &= \lim_{h \to 0} \frac{F(x+h) - F(x)}{h\overline{F}(x)} \\
&= \frac{1}{\overline{F}(x)} \frac{dF(x)}{dx} = \frac{f(x)}{\overline{F}(x)} = r(x)
\end{aligned}$$

i.e., $r(x)$ may be regarded as an instantaneous failure rate or intensity of the probability that a device aged x will fail. If $r(x)$ is an increasing function of x, this implies that the device is ‘wearing out’, while if it is a decreasing function of x, this implies that the device is ‘bedding in’, i.e. improving with age.

If $X \sim \mathrm{Exp}(\lambda)$, then $r(x) = \lambda$, $x \geq 0$: this constant hazard rate is consistent with the ‘lack-of memory’ property of this distribution - the device cannot ‘remember’ how old it is.