4 Continuous Distributions

The basic ideas of previous sections were the notions of a random variable, its probability distribution, expectation, and standard deviation. These ideas will now be extended from discrete distributions to continuous distributions on a line, in a plane, or in higher dimensions. This chapter concerns continuous probability distributions over an interval of real numbers. One example is the normal distribution, seen already as an approximation to various discrete distributions. A simpler example is the uniform distribution on an interval, defined by relative lengths. Another example, the exponential distribution, treated in Section 4.2, is the continuous analog of the geometric distribution. Each of these distributions is defined by a ***probability density function***, like the familiar normal curve associated with the normal distribution. The way a continuous distribution can be specified by such a density function is the subject of Section 4.1. Change of variable for distributions defined by densities is the subject of Section 4.4.

The concept of a continuously distributed random variable is an idealization which allows probabilities to be computed by calculus. This gives models for chance phenomena involving continuous variables. Such models arise both:

(i) as limits from discrete models (e.g., the normal distribution as an approximation to the binomial, or the exponential approximation to the geometric discussed in Section 4.2), and

(ii) directly from physical phenomena most naturally modeled by continuous variables (e.g., the normal distribution as a model for measurement error, or the exponential distribution as a model for the lifetime of an atom).

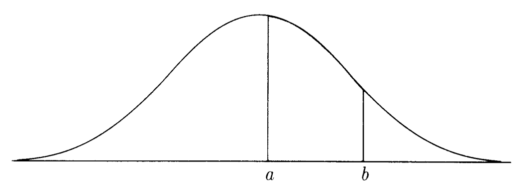
4.1 Probability Densities

In Chapters 2 and 3 the normal distribution was used as an approximation to the distribution of a sum or average of a large number of independent random variables. The idea there was to approximate a discrete distribution of many small individual probabilities by scaling the histogram to make it follow a continuous curve. The function defining such a curve is called a *probability density*, denoted $f(x)$ here. This function determines probabilities over an infinite continuous range of possible values.

The basic idea is that probabilities are defined by areas under the graph of $f(x)$. That is, a random variable X has density $f(x)$ if for all $a \le b$

$$P(a \le X \le b) = \int_a^b f(x)dx,$$

which is the area shaded in the following diagram:



The boxes on pages 262 and 263 show the analogy between a discrete distribution of a random variable X defined by the probabilities $P(x) = P(X = x)$ of individual values x, and a continuous distribution defined by a probability density $f(x)$. In the density case, it is of no use to consider $P(X = x)$. This probability is zero for every x for a distribution with a density, so it gives no information about the distribution. Rather, everything is determined by the density $f(x)$, which gives the probability per unit length for values near x. The individual probability $P(x)$ of the event $(X = x)$ is replaced everywhere by the infinitesimal probability $f(x)dx$ of the event $(X \in dx)$, and sums are replaced by integrals. Here $(X \in dx)$ stands for the event that X falls in an infinitesimal interval of length dx near x, for example, $(x \le X \le x + dx)$, or $(x - dx \le X \le x)$.

Assuming f is continuous at x, the area representing $P(X \in dx)$ is essentially a rectangle of sides $f(x)$ and dx, hence area $f(x)dx$. Note well that it is $f(x)dx$, not just $f(x)$, which is the analog of $P(x)$. It may well be that $f(x) > 1$ for some values of x. Thus $f(x)$ is *not* a probability, but a probability *density*. When multiplied by small lengths, $f(x)$ gives approximate probabilities of small intervals near x. If you cut the interval $[a, b]$ into lots of tiny intervals between a and b, add the probabilities of all the tiny intervals, and pass to the limit as the interval widths tend to zero, you get the integral formula for $P(a \le X \le b)$. So when integrated over an interval, $f(x)$ gives the exact probability of the interval. A probability density $f(x)$ thus describes a continuous distribution of probability over a number line.

Mean, variance and standard deviation. These are defined just as before in terms of expectations.

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

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$$

If the second integral is finite, then so is the first, and then $E(X^2)$ and $E(X)$ can be used to calculate $Var(X)$ and $SD(X)$ in the usual way

$$Var(X) = E(X^2) - [E(X)]^2 \qquad SD(X) = \sqrt{Var(X)}$$

The basic properties of expectation, variance, and standard deviation are the same as in the discrete case. For example, Chebychev's inequality holds just as well for X with a density as for a discrete random variable X. Proofs of such things parallel the discrete case, using properties of integrals instead of properties of sums.

Independence. Numerical random variables X and Y are called *independent* if the events $(X \in A)$ and $(Y \in B)$ are independent for any choice of two intervals A and B, or more generally any choice of subsets A and B of the line for which the probabilities of these events are defined. That is to say

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

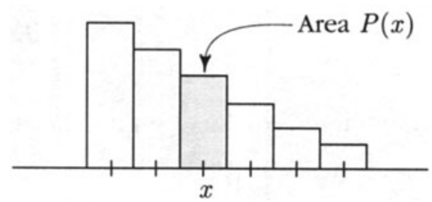
Only for discrete random variables can this definition be reduced to the case $A = [x, x]$ and $B = [y, y]$, when the rule becomes simply

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

If X has a distribution with a density, then $P(X = x) = 0$ for every x, which implies $P(X = x, Y = y) = 0 = P(X = x)P(Y = y)$ for all x and y for any random variable Y whatever. See Section 5.2 for a more careful treatment of independence of X and Y with densities in terms of their joint distribution. Independence of several variables is defined by a similar product rule. The basic properties of independent random variables are the same in the density case as in the discrete case. In particular, if X and Y are independent and both $E(X)$ and $E(Y)$ are defined and finite, then $E(XY) = E(X)E(Y)$. The addition rule for the variance of a sum of independent random variables follows from this.

Discrete Distributions

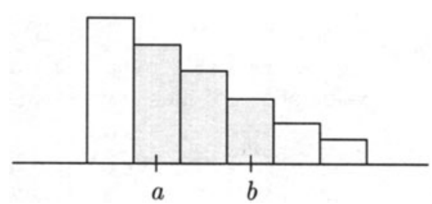
Point Probability:



$$P(X = x) = P(x)$$

So $P(x)$ is the probability that X has integer value x.

Interval Probability:



$$P(a \leq X \leq b) = \sum_{a \leq x \leq b} P(x)$$

the relative area under a histogram between $a - 1/2$ and $b + 1/2$.

Constraints: Non-negative with Total Sum 1

$$P(x) \geq 0 \quad \text{for all } x \qquad \text{and} \qquad \sum_{\text{all } x} P(x) = 1$$

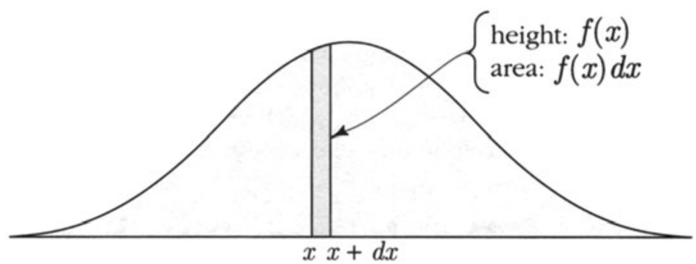
Expectation of a Function g of X, e.g., X, X^2:

$$E\left(g(X)\right) = \sum_{\text{all } x} g(x)P(x)$$

provided the sum converges absolutely.

Distributions Defined by a Density

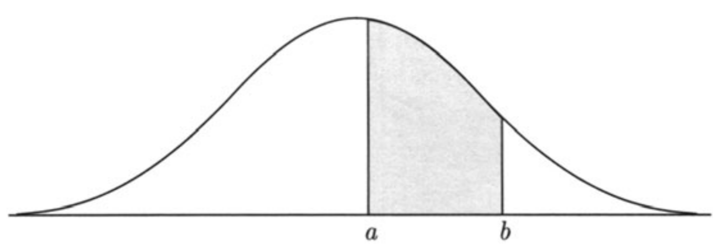
Infinitesimal Probability:



$$P(X \in dx) = f(x)dx$$

The *density* $f(x)$ gives the probability per unit length for values near x.

Interval Probability:



$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

the area under the graph of $f(x)$ between a and b.

Constraints: Non-negative with Total Integral 1

$$f(x) \geq 0 \quad \text{for all } x \qquad \text{and} \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

Expectation of a Function g of x, e.g., X, X^2:

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

provided the integral converges absolutely.

Special densities. There are a few particularly important probability densities which appear over and over again, both in theory and applications. Most notable are the uniform, normal, exponential, gamma, and beta densities. Why these few should be so important is not at first obvious, but emerges gradually after study of their properties and relationships, both with each other and with other discrete distributions. This section introduces only the uniform and normal densities. Further developments and examples involving other densities follow in subsequent sections. Also, summaries of these distributions are given in an Appendix. These include formulae for means, variances, etc., which are used routinely in calculations and which you are expected to look up as necessary.

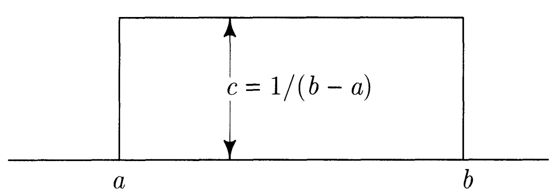
The Uniform Distribution

A random variable X has *uniform distribution* on the interval (a, b), if X has density $f(x)$ which is constant on (a, b), and 0 elsewhere. The uniform (a, b) density is

$$f(x) = \begin{cases} 1/(b-a) & \text{if } \; a < x < b \\ 0 & \text{otherwise} \end{cases}$$

The constant value c of the density on (a, b) is $1/(b-a)$, because the total area of the rectangle under the density function must be 1:

$$(b-a)c = 1 \quad \Longrightarrow \quad c = 1/(b-a)$$

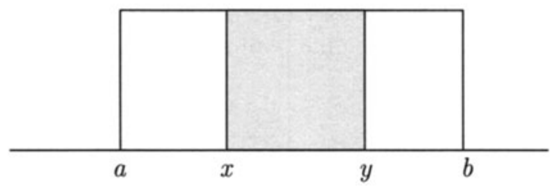


As suggested by the verticals at $x = a$ and $x = b$, the values of $f(x)$ at these endpoints do not affect the probabilities defined by areas under the graph. The area of a line is zero, and so is the probability that any continuously distributed random variable X takes any particular real value. This is an idealization based on the idea that a real number is specified with infinite precision. In practice, it would only ever be possible to know that X was equal to x to some finite number of decimal places. For X distributed uniformly on (a, b), and $a < x < b$, this event would always have strictly positive probability.

For a uniform distribution, probabilities reduce to relative lengths. So if X has uniform (a, b) distribution, then for $a < x < y < b$,

$$P(x < X < y) = \frac{\text{length } (x, y)}{\text{length } (a, b)} = \frac{y - x}{b - a}$$

as is obvious from the diagram.



For example, if X has uniform $(0, 2)$ distribution, the probability that X is 1.23 correct to two decimal places is

$$P(1.225 < X < 1.235) = \frac{1.235 - 1.225}{2} = 0.01/2 = 0.5\%$$

A simple rescaling transforms the interval (a, b) into $(0, 1)$. The uniform (a, b) distribution then transforms into the uniform $(0, 1)$ distribution, whose density is simply 1 on $(0, 1)$, and 0 elsewhere. In terms of random variables, any problem involving a uniform (a, b) random variable X reduces easily to one involving a uniform $(0, 1)$ random variable U defined by

$$U = (X - a)/(b - a) \qquad \text{so} \quad X = a + (b - a)U$$

This kind of *scaling* or *linear change of variable*, is a basic technique for reducing problems to the simplest case to avoid unnecessary calculation. To illustrate, the expected value of X is

$$\begin{aligned} E(X) &= E(a + (b - a)U) \\ &= a + (b - a)E(U) \\ &= a + (b - a)\frac{1}{2} = (a + b)/2 \end{aligned}$$

This is obvious anyway by symmetry, since $(a + b)/2$ is the midpoint of (a, b). The variance of X is

$$\begin{aligned} Var(X) &= Var(a + (b - a)U) \\ &= (b - a)^2 Var(U) \\ &= (b - a)^2[E(U^2) - (E(U))^2] = (b - a)^2[1/3 - (1/2)^2] = (b - a)^2/12 \end{aligned}$$

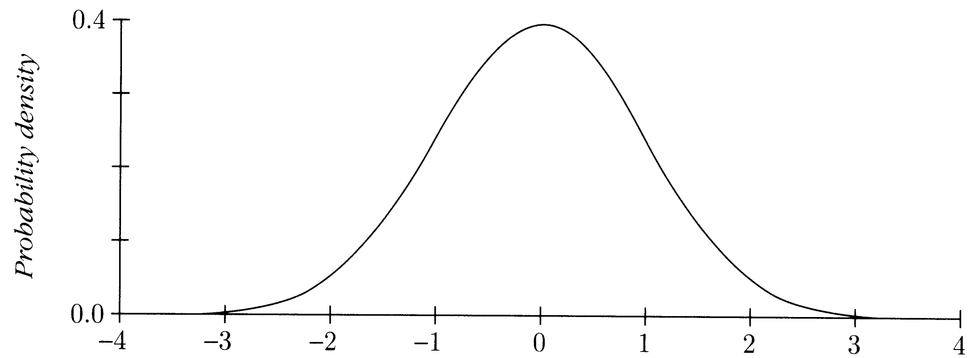
Here $E(U) = 1/2$ without calculation, but $E(U^2)$ requires an integral:

$$\begin{aligned} E(U^2) &= \int_{-\infty}^{\infty} u^2 f(u)du \\ &= \int_0^1 u^2 du \qquad \text{since } U \text{ has density } f(u) = 1 \text{ for } 0 < u < 1, \text{ 0 otherwise} \\ &= \frac{1}{3}u^3\Big|_0^1 = \frac{1}{3} \end{aligned}$$

The Normal Distribution

A random variable Z has *standard normal distribution* if Z has as its probability density the *standard normal density*

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad (-\infty < z < \infty)$$



The constant $1/\sqrt{2\pi}$ is put in the definition of the standard normal density so the total area under the standard normal curve $y = \phi(z)$ is 1. This is the first integral in the following box:

Standard Normal Integrals

$$\int_{-\infty}^{\infty} \phi(z)\,dz = 1; \qquad \int_{-\infty}^{\infty} z\phi(z)\,dz = 0; \qquad \int_{-\infty}^{\infty} z^2\phi(z)\,dz = 1.$$

The first and third of these integrals are evaluated in Section 5.3. The second and third integrals show that the standard normal distribution has mean 0 and second moment 1, hence variance 1. The mean of this distribution is zero, because of the symmetry about zero of the standard normal curve. The third integral in the box can be reduced to the first integral by integration by parts.

There is no simple formula for the standard normal probability of an interval

$$\Phi(a, b) = P(a < Z < b) = \int_a^b \phi(z)dz$$

Instead, this probability is found, as in Section 2.2, using a table of the standard normal c.d.f.

$$\Phi(b) = \Phi(-\infty, b) = P(Z \leq b) = \int_{-\infty}^{b} \phi(z)dz$$

Normal (μ, σ^2) Distribution

If Z has standard normal distribution and μ and σ are constants with $\sigma \geq 0$, then

$$X = \mu + \sigma Z$$

has mean μ, standard deviation σ, and variance σ^2. The distribution of X is called the *normal distribution with mean μ and variance σ^2*, abbreviated normal (μ, σ^2). So X has normal (μ, σ^2) distribution if and only if the standardized variable

$$Z = (X - \mu)/\sigma$$

has normal $(0, 1)$ or ***standard normal*** distribution. To find $P(c < X < d)$, change to standard units and use the standard normal table

$$P(c < X < d) = P(a < Z < b) = \Phi(b) - \Phi(a)$$

$$\text{where} \qquad a = (c - \mu)/\sigma \qquad Z = (X - \mu)/\sigma \qquad b = (d - \mu)/\sigma$$

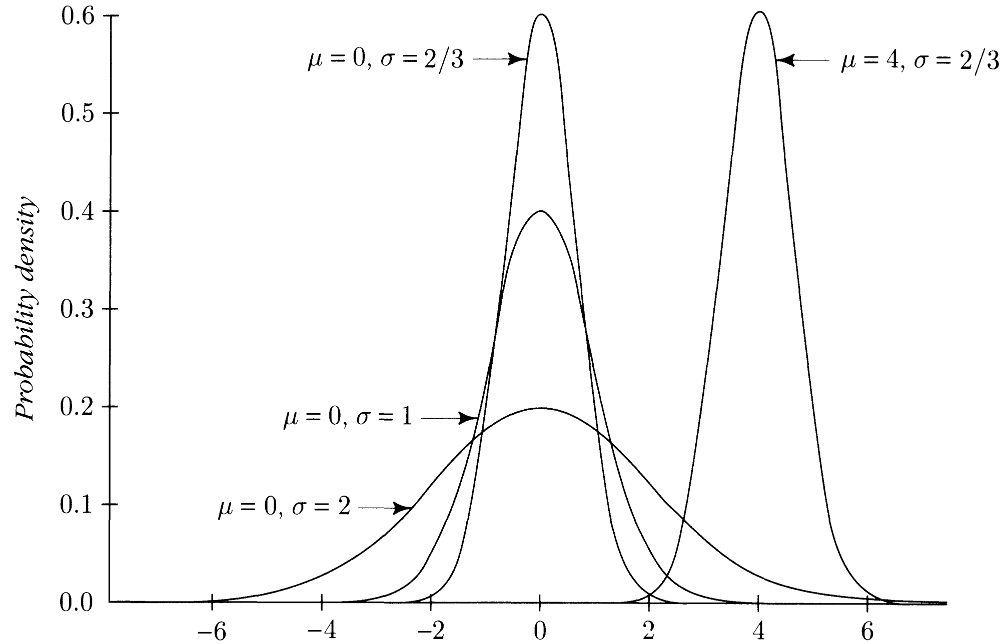
Formula for the normal (μ, σ^2) density. For $\sigma > 0$, the formula is

$$\frac{1}{\sigma}\phi((x - \mu)/\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} \qquad (-\infty < x < \infty).$$

This is the transformation of the standard normal density $\phi(z)$ corresponding to the linear change of variable from Z to $X = \mu + \sigma Z$. See Section 4.4 for details of this kind of transformation. This formula is rarely used in calculations. It is always simpler to transform to standard units as in Example 1 below. If $\sigma^2 = 0$ the normal (μ, σ^2) distribution is just the distribution of the constant random variable with value μ, with probability one at μ. For $\sigma^2 > 0$, the normal (μ, σ^2) distribution piles up around μ for small values of σ^2, and become more and more spread out as σ^2 increases. See Figure 1 on the next page.

Normal approximation to an empirical distribution. The normal distribution is often fitted to an empirical distribution of observations. The parameters μ and σ are usually estimated by the mean and standard deviation of the list of observations. This is justified by the integral approximation for averages discussed later in this section. How well such an approximation works depends on the source of the data and the measurement technique. Examples of the kinds of observations where the normal approximation has been found to be good are weighings on a chemical balance, and measurements of the angular position of a star.

FIGURE 1. Some normal (μ, σ^2) densities.



The central limit theorem. The appearance of the normal distribution in many contexts is explained by the central limit theorem, stated in Section 3.3. According to this result, for independent random variables with the same distribution and finite variance, as $n \to \infty$, the distribution of the standardized sum (or average) of n variables approaches the standard normal distribution. It can be shown that this happens no matter what the common distribution of the random variables summed or averaged, discrete or continuous, provided the distribution has finite variance. In particular, the central limit theorem implies that the distribution of the sum or average of a large number of independent measurements will typically tend to follow the normal curve, even if the distribution of the individual measurements does not. This mathematical fact is the basis for most statistical applications of the normal distribution.

History. The normal distribution is also known as the Gaussian distribution, and in France as Laplace's distribution. Gauss (1777–1855) and Laplace (1749–1827) brought out the central role of the normal distribution in the theory of errors of observation. Quetelet (1796–1874) and Galton (1822–1911) fitted the normal distribution to empirical data such as heights and weights in human and animal populations. But the normal distribution was actually first discovered around 1720 by Abraham De Moivre (1667–1754), as the approximation to the binomial (n, p) distribution for large n described in Section 2.2.

Example 1. Repeated measurements.

Suppose a long series of repeated measurements of the weight of a standard kilogram yield results that are normally distributed with a mean of one kilogram and an SD of 20 micrograms.

Problem 1. About what proportion of measurements are correct to within 10 micrograms?

Solution. By converting to standard units, this is $P(-0.5 \le Z \le 0.5) = 2\Phi(0.5) - 1 = 38.29\%$.

Problem 2. In 100 measurements, what is the probability that more than 45 measurements will be correct to within 10 micrograms?

Solution. It seems reasonable to assume that each measurement is correct to within 10 micrograms with chance 38.29%, independently of all others. Out of 100 measurements, the number correct to within 10 micrograms has the binomial (100, 0.3829) distribution. This is approximately normal, with

$$\mu = 38.29 \qquad \sigma = \sqrt{100 \times 0.3829 \times (1 - 0.3829)} = 4.86$$

The probability that more than 45 measurements are correct to within 10 micrograms is approximately

$$1 - \Phi\left(\frac{45.5 - 38.29}{4.86}\right) = 1 - \Phi(1.48) = 6.94\%$$

Problem 3. In the long series of measurements, some errors are positive and some are negative. What is the approximate average absolute size of these errors?

Solution. Here X = observed weight − 1 kilogram, in micrograms, and has normal $(0, 20^2)$ distribution. We want $E|X|$. In terms of a standard normal variable Z, $X = 20Z$, so

$$\begin{aligned} E|X| = 20E|Z| &= 20 \int_{-\infty}^{\infty} |z|\phi(z)dz \\ &= 40 \int_0^{\infty} z\phi(z)\,dz \quad \text{by symmetry} \\ &= 40 \int_0^{\infty} z\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz \\ &= -\frac{40}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\Big|_0^{\infty} = \frac{40}{\sqrt{2\pi}} = 15.96 \text{ micrograms.} \end{aligned}$$

Further Examples

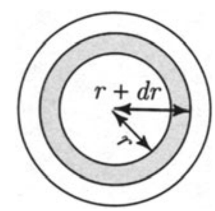
Example 2. Radial distance.

Suppose a bacterial colony appears at a point uniformly distributed at random on a circular plate of radius 1. Let R be the distance of the point from the center of the plate.

Problem 1. Find the probability density of R.

Solution. The basic assumption is that the probability of the colony appearing in any particular region of the plate is proportional to the area of the region. From the diagram, for $0 < r < 1$,

$$P(R \in dr) = \frac{\text{Area of annulus from } r \text{ to } r + dr}{\text{Total area}} = \frac{\pi(r+dr)^2 - \pi r^2}{\pi} = 2r\,dr$$



by ignoring the term involving $(dr)^2$. So R has density

$$f(r) = \begin{cases} 2r & 0 < r < 1 \\ 0 & \text{otherwise} \end{cases}$$

Problem 2. Find $P(a \le R \le b)$ for $0 < a < b < 1$.

Solution. $P(a \le R \le b) = \int_a^b 2r dr = r^2 \Big|_a^b = b^2 - a^2$

(This can also be done using areas in the plane.)

Problem 3. Find the mean and variance of R.

Solution.

$$E(R) = \int_{-\infty}^{\infty} r f(r) dr = \int_0^1 2r^2 dr = \frac{2}{3} r^3 \Big|_0^1 = \frac{2}{3}$$

$$E(R^2) = \int_{-\infty}^{\infty} r^2 f(r) dr = \int_0^1 2r^3 dr = \frac{2}{4} r^4 \Big|_0^1 = \frac{1}{2}$$

$$Var(R) = E(R^2) - (E(R))^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

Problem 4. Suppose 100 bacterial colonies are distributed independently and uniformly at random on a circular plate of radius 1. What is the probability that the mean distance of the colonies from the center of the plate is at least 0.7?

Solution. The problem is to find $P(A_{100} > 0.7)$ where

$$A_{100} = (R_1 + R_2 + \cdots + R_{100})/100$$

and the R_i are independent random variables with the same distribution as that of R calculated in Problem 1. Basic formulae for means and SDs derived in Chapter 3 still apply to give $E(A_{100}) = E(R) = 0.667$

$$SD(A_{100}) = SD(R)/\sqrt{100} = \sqrt{\frac{1}{18} \cdot \frac{1}{100}} \approx 0.0236$$

Using the normal approximation, the required probability is approximately

$$1 - \Phi\left(\frac{0.7 - 0.667}{0.0236}\right) = 1 - \Phi(1.40) = 8.7\%$$

Example 3. **A distribution with infinite mean.**

Suppose that X has probability density

$$f(x) = \begin{cases} 1/(1+x)^2 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Problem 1. Find $P(X > 3)$.

Solution. $P(X > 3) = \int_3^{\infty} \frac{1}{(1+x)^2}\, dx = -\frac{1}{1+x}\Big|_3^{\infty} = \frac{1}{4}.$

Problem 2. Let X_1, X_2, X_3, X_4 be independent random variables with the same distribution as X. Find the chance that exactly two of these variables are greater than 3.

Solution. Since $P(X_i > 3) = P(X > 3) = 1/4$, and the random variables X_i are independent, the events $(X_i > 3)$, $i = 1, 2, 3, 4$, are four independent events, each with probability $1/4$. The number of these events which occur is therefore a binomial $(4, 1/4)$ random variable. Call this random variable N. The required probability is then

$$P(N = 2) = \binom{4}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^2 = \frac{27}{128}$$

Problem 3. Find $E(X)$.

Solution.

$$E(X) = \int_0^{\infty} \frac{x}{(1+x)^2}\, dx = \int_0^{\infty} \left(\frac{1}{1+x} - \frac{1}{(1+x)^2}\right) dx$$

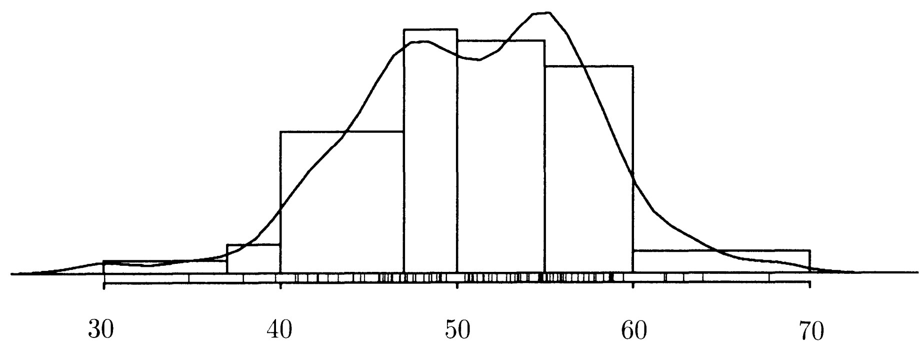
$$= \int_0^{\infty} \frac{1}{1+x}\, dx - 1 = \log(1+x)\Big|_0^{\infty} - 1 = \infty$$

Remark. The long-run interpretation is that the average $(X_1 + \cdots + X_n)/n$ of independent random variables chosen according to this distribution will, with overwhelming probability, tend to increase beyond all finite bounds as $n \to \infty$.

Fitting a Curve to an Empirical Distribution

The empirical distribution of a data list $(x_1, \ldots, x_n)$ can be displayed in a histogram, as in Figure 4 at the end of Section 1.3. This histogram smoothes out the data to display the general shape of the empirical distribution. Such a histogram often follows a smooth curve, say $y = f(x)$, as shown in Figure 2. Since histograms are non-negative it is natural to assume that $f(x) \geq 0$ for every x.

FIGURE 2. A smooth curve fitted to a data histogram.



The basic idea is that if (a, b) is a bin interval, then the area of the bar over (a, b) should approximately equal the area under the curve from a to b. Summing such approximations over bins, and interpolating between the cut points, suggests a more general approximation: for any interval (a, b) the proportion of data in the interval should be approximately the area under the curve from a to b. Since the area under the curve can be evaluated as an integral, this amounts to the following:

Integral Approximation for Empirical Proportions

If a histogram of an empirical distribution follows the curve $y = f(x)$, then the proportion $P_n(a, b)$ of observations between a and b is approximated by

$$P_n(a, b) \approx \int_a^b f(x)\, dx$$

Since $P_n(-\infty, \infty) = 1$, whatever the empirical distribution, any reasonable approximation $f(x)$ to a data histogram must satisfy

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \qquad (1)$$

Then $f(x)$ is a probability density function, and the empirical distribution of the data is approximated by the theoretical probability distribution with density $f(x)$.

Averages and Integrals

Given a data list $(x_1, \ldots, x_n)$ and an interval (a, b), the method of indicators provides a useful way to express the proportion of values in (a, b) as an average. Define the *indicator function* of (a, b) by

$$I_{(a,b)}(x) = \begin{cases} 1 & \text{if } x \in (a, b) \\ 0 & \text{otherwise} \end{cases}$$

Given a list $(x_1, \ldots, x_n)$, the number of i such that $x_i \in (a, b)$ can be calculated by going through the list and for each i adding 1 if $x_i \in (a, b)$ and adding 0 otherwise. The term added for the ith element of the list is $I_{(a,b)}(x_i)$. The empirical proportion of values in (a, b) is therefore

$$P_n(a, b) = \frac{1}{n} \sum_{i=1}^{n} I_{(a,b)}(x_i)$$

In words: the proportion of x-values in (a, b) is the average of $I_{(a,b)}(x)$ as x ranges over the n values in the list. Suppose now that the empirical distribution is well approximated by a theoretical distribution with density $f(x)$. The integral approximation for empirical proportions becomes an integral approximation for an empirical average:

$$\frac{1}{n} \sum_{i=1}^{n} I_{(a,b)}(x_i) = P_n(a, b) \approx \int_a^b f(x)dx = \int_{-\infty}^{\infty} I_{(a,b)}(x) f(x)\, dx$$

where the last equality holds because $I_{(a,b)}(x) = 0$ for x outside (a, b). The point of writing the integral approximation this way is that it suggests a very useful generalization for other functions $g(x)$ besides $g(x) = I_{(a,b)}(x)$.

Integral Approximation for Averages

If the empirical distribution of a list $(x_1, \ldots, x_n)$ is well approximated by the theoretical distribution with density $f(x)$, then the average of a function $g(x)$ over the n values in the list is approximated by the integral of $g(x)$ times the density $f(x)$ over all values of x:

$$\frac{1}{n} \sum_{i=1}^{n} g(x_i) \approx \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

Notice that the left-hand average is $E[g(X)]$ for X picked at random from the list of n values $(x_1, \ldots, x_n)$. The right-hand integral is $E[g(X)]$ for a random variable X with density $f(x)$.

Apart from indicator functions $g(x)$, the integral approximation is most commonly applied to the powers $g(x) = x^k$:

$$\frac{1}{n}\sum_{i=1}^{n} x_i^k \approx \int_{-\infty}^{\infty} x^k f(x)\, dx$$

The left side is the average value of x^k as x ranges over values in the data list, and is called the *kth moment of the empirical distribution*. The right side is called the *kth moment of the theoretical distribution* with density $f(x)$. The cases $k = 1$ and $k = 2$ together imply that the mean and variance of the empirical distribution are close to the mean and variance of the theoretical distribution. Thus if a data histogram looks like a normal curve, then the mean and variance of the data can be used to estimate the parameters of the normal curve.

Heuristic derivation of the integral approximation for averages. For $g(x)$ the indicator of an interval, this is just the integral approximation for proportions. A step function $g(x)$ that has a finite number of different values on a finite number of disjoint intervals can be written as a finite linear combination

$$g(x) = c_1 I_{(a_1,b_1)}(x) + \cdots + c_m I_{(a_m,b_m)}(x)$$

of indicator functions of intervals. So for a step function $g(x)$ the integral approximation for the average follows by combining the integral approximation for the proportions $P_n(a_i, b_i)$, using the linearity properties of sums and integrals. The approximation for a more general function $g(x)$ is obtained by approximating $g(x)$ by a step function, much as in the usual approximation of integrals by Riemann sums. □

How good is the integral approximation for an average? This depends both on how closely the empirical distribution conforms to the theoretical density $f(x)$, and on how rapidly $g(x)$ varies as a function of x. (If $g(x)$ grows too rapidly for large absolute values of x the integral $\int_{-\infty}^{\infty} g(x)f(x)\, dx$ might not even be defined.) Provided a data histogram follows the density curve closely, and $g(x)$ is a fairly smooth function of x that does not grow too rapidly for large $|x|$, the data average $\frac{1}{n}\sum_1^n g(x_i)$ will be well approximated by $\int_{-\infty}^{\infty} g(x)f(x)\, dx$.

The law of averages. This is a probabilistic way to make the statement of the previous paragraph more precise. If the data list $(x_1, \ldots, x_n)$ is obtained by a process of repeated measurements of some kind, it may be reasonable to assume that $(x_1, \ldots, x_n)$ is the result of independent random sampling of points from the theoretical distribution with density $f(x)$. More formally, $(x_1, \ldots, x_n)$ is regarded as the observed result of $(X_1, \ldots, X_n)$ for a sequence of independent random variables X_i, each distributed like X with density $f(x)$. According to the law of averages of

Section 3.3, which holds just as well for X with a density as for discrete X, provided the integral that defines $E[g(X)]$ is absolutely convergent, for large n it is highly probable that

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \approx E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Assuming $Var[g(X)] < \infty$, Chebychev's inequality gives for any $\epsilon > 0$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \int_{-\infty}^{\infty} g(x)f(x)dx\right| > \epsilon\right) \leq \frac{Var[g(X)]}{n\epsilon^2}$$

Provided n is large enough that $Var[g(X)]/n\epsilon^2$ is small, the integral approximation for the average of n values of $g(X_i)$ will probably be correct to within ϵ. Note that the variance of $g(X)$ will tend to be small provided $g(x)$ does not vary too rapidly over the typical range of values x of X, and provided $g(x)$ does not grow too rapidly for less typical values x in the tails of the distribution of X. So the factor $Var(g[X])$ in the above probability estimate captures nicely the idea of the previous paragraph that the integral approximation for averages will tend to work better for smoother functions $g(x)$. The estimate given by Chebychev's inequality is very conservative. More realistic approximations to the probability of errors of various sizes in the integral approximation for averages are provided by the normal approximation.

The Monte-Carlo method. It may be that the integral $\int_{-\infty}^{\infty} g(x)f(x)dx$ is difficult to evaluate by calculus or numerical integration, but it is easy to generate pseudo-random numbers X_i distributed according to density $f(x)$. The value of the integral can then be estimated by the average value of $g(X_i)$ for a large number of such X_i. For instance, the value of $\int_0^1 g(x)dx$ can be estimated this way using X_i with uniform $(0,1)$ distribution. Assuming that some bound on $Var[g(X)]$ is available (e.g., if $g(x)$ is a bounded function of x), error probabilities can be estimated using Chebychev's inequality or a normal approximation. The same method can be applied in higher dimensions to approximate multiple integrals.

Exercises 4.1

1. What is the probability that a standard normal random variable has value

a) between 0 and 0.001? b) between 1 and 1.001?

2. Suppose X has density $f(x) = c/x^4$ for $x > 1$, and $f(x) = 0$ otherwise, where c is a constant. Find a) c; b) $E(X)$; c) $Var(X)$.

3. Suppose X is a random variable whose density is $f(x) = cx(1-x)$ for $0 < x < 1$, and $f(x) = 0$ otherwise. Find:

a) the value of c; b) $P(X \leq 1/2)$; c) $P(X \leq 1/3)$;

d) $P(1/3 < X \leq 1/2)$; e) the mean and variance of X.

4. Suppose X with values in $(0, 1)$ has density $f(x) = cx^2(1-x)^2$ for $0 < x < 1$. Find:

a) the constant c; b) $E(X)$; c) $Var(X)$.

5. Suppose that X is a random variable whose density is

$$f(x) = \frac{1}{2(1+|x|)^2} \qquad (-\infty < x < \infty)$$

a) Draw the graph of $f(x)$. b) Find $P(-1 < X < 2)$.

c) Find $P(|X| > 1)$. d) Is $E(X)$ defined?

6. Suppose X has normal (μ, σ^2) distribution, and $P(X \leq 0) = 1/3$, $P(X \leq 1) = 2/3$.

a) What are the values of μ and σ? b) What if instead $P(X \leq 1) = 3/4$?

7. Suppose the distribution of height over a large population of individuals is approximately normal. Ten percent of individuals in the population are over 6 feet tall, while the average height is 5 feet 10 inches. What, approximately, is the probability that in a group of 100 people picked at random from this population there will be two or more individuals over 6 feet 2 inches tall?

8. Measurements on the weight of a lump of metal are believed to be independent and identically distributed; each measurement has mean 12 grams and SD 1.1 gram.

a) Find the chance that a single measurement is between 11.8 and 12.2 grams, assuming that individual measurements are normally distributed.

b) Estimate the chance that the average of 100 measurements is between 11.8 and 12.2 grams. For this calculation, is it necessary to assume that individual measurements are normally distributed? Explain.

9. Suppose X_1, X_2, X_3, X_4 are independent uniform $(0, 1)$ random variables, and let $S_4 = X_1 + X_2 + X_3 + X_4$. Use the normal approximation to calculate $P(S_4 \geq 3)$ approximately.

10. The distribution of repeated measurements of the weight of an object is approximately normal with a mean of 9.7800 gm and a standard deviation of 0.0031 gm. Calculate:

a) the chance that the next measurement will be between 9.7840 and 9.8000 gm;

b) the proportion of measurements smaller than 9.7794 gm;

c) the weight that the next measurement has a 10% chance of exceeding.

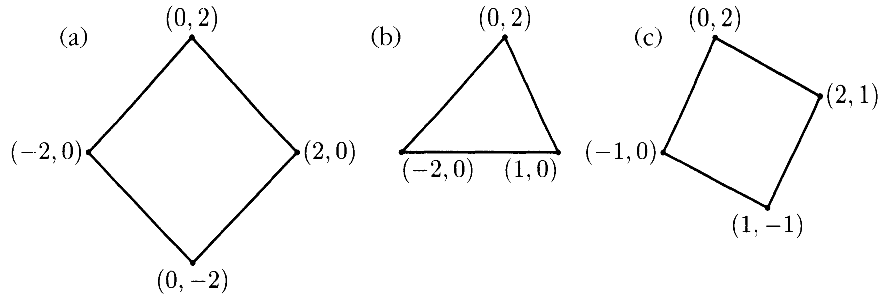
11. A large lot of marbles have diameters which are approximately normally distributed with a mean of 1 cm. One third have diameters greater than 1.1 cm. Find:

a) the standard deviation of the distribution;

b) the proportion whose diameters are within 0.2 cm of the mean;

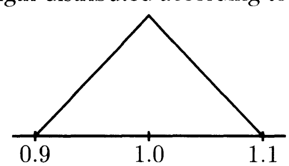
c) the diameter that is exceeded by 75% of the marbles.

12. Consider a point picked uniformly at random from the area inside one of the following shapes:



In each case find the density function of the x coordinate.

13. Suppose a manufacturing process designed to produce rods of length 1 inch exactly, in fact produces rods with length distributed according to the density graphed below.



For quality control, the manufacturer scraps all rods except those with length between 0.925 and 1.075 inches before he offers them to buyers.

a) What proportion of output is scrapped?

b) A particular customer wants 100 rods with length between 0.95 and 1.05 inches. Assuming lengths of successive rods produced by the process are independent, how many rods must this customer buy to be 95% sure of getting at least 100 of the prescribed quality?

14. Another manufacturer produces similar rods by a process that produces lengths with the same mean and standard deviation as in Exercise 13, but with a distribution following the normal curve. This manufacturer uses the same quality control procedure of scrapping rods not within 0.075 inches of 1 inch in length.

a) What proportion of output is scrapped by this manufacturer?

b) If you were the customer with requirements as in part b) in Exercise 13, which manufacturer would you prefer? Explain.

15. Standard normal c.d.f. in terms of the error function. Many calculators and computer languages have built in the *error function* $\text{erf}(x) = (2/\sqrt{\pi}) \int_0^x e^{-t^2} dt$.

a) Find μ and σ^2 so that $P(|X| \leq x) = \text{erf}(x)$ if X has normal (μ, σ^2) distribution.

b) Express $\text{erf}(x)$ in terms of the standard normal c.d.f. $\Phi(z)$.

c) Express $\Phi(z)$ in terms of $\text{erf}(x)$.

4.2 Exponential and Gamma Distributions

One of the things most commonly described by a distribution with a density is a *random time* of some kind. Some examples are:

(i) the lifetime of an individual picked at random from some biological population;

(ii) the time until decay of a radioactive atom;

(iii) the length of time a patient survives after an operation of some kind;

(iv) the time it takes a computer to process a job of some kind.

Such random times will be regarded as random variables with range the interval $[0, \infty)$. Assume the distribution of a random time T is defined by a probability density $f(t)$ for $0 \leq t < \infty$, so for $0 \leq a < b < \infty$

$$P(a < T \leq b) = \int_a^b f(t)\, dt$$

If T is interpreted as the lifetime of something, the probability of the thing surviving past time s is

$$P(T > s) = \int_s^\infty f(t)\, dt$$

This is a decreasing function of s, called the *survival function*. By the difference rule for probabilities

$$P(a < T \leq b) = P(T > a) - P(T > b)$$

So the probability of the random time falling in any interval can be found from the survival function.

The simplest model for a random time with no upper bound on its range is the *exponential distribution*. This distribution fits the lifetimes of a variety of inanimate objects that experience no aging effect. More importantly, many models for systems that evolve randomly over time, called *stochastic processes*, are built up from some combination of independent exponential random times. A case in point is the *Poisson process* on a time line, which models the times of successive arrivals of some kind, such as the times customers arrive at a store. In this model, the successive interarrival times are independent exponential random variables. And the time of the rth arrival has a *gamma* distribution. These exponential and gamma distributions, studied in this section, are the continuous analogs of the geometric and negative binomial distributions of Section 3.4.

The following section introduces the concept of a *death* or *hazard rate* associated with a random time. For the exponential distribution this is constant over time, but for more general distributions the death rate varies over time, indicating an aging effect.

Exponential Distribution

A random time T has *exponential distribution with rate* λ, denoted exponential (λ), where λ is a positive parameter, if T has probability density

$$f(t) = \lambda e^{-\lambda t} \qquad (t \geq 0)$$

Equivalently, for $0 \leq a < b < \infty$

$$P(a < T \leq b) = \int_a^b \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_a^b = e^{-\lambda a} - e^{-\lambda b}$$

To see that $f(t)$ is a probability density on $[0, \infty)$, let $a = 0$, and let $b \to \infty$ to find the total probability of 1 on $[0, \infty)$. Set $a = t$ and let $b \to \infty$ to get the next formula for the survival function. Calculation of the mean and SD are left as an exercise.

Exponential Survival Function

A random time T has exponential distribution with rate λ if and only if T has survival function

$$P(T > t) = e^{-\lambda t} \qquad (t \geq 0)$$

Mean and SD: $E(T) = SD(T) = \frac{1}{\lambda}$

Note that the rate λ is the inverse of the mean, so an exponential random time with a large rate is likely to be small, and one with a small rate is likely to be large. A better interpretation of λ as a hazard rate will be given shortly.

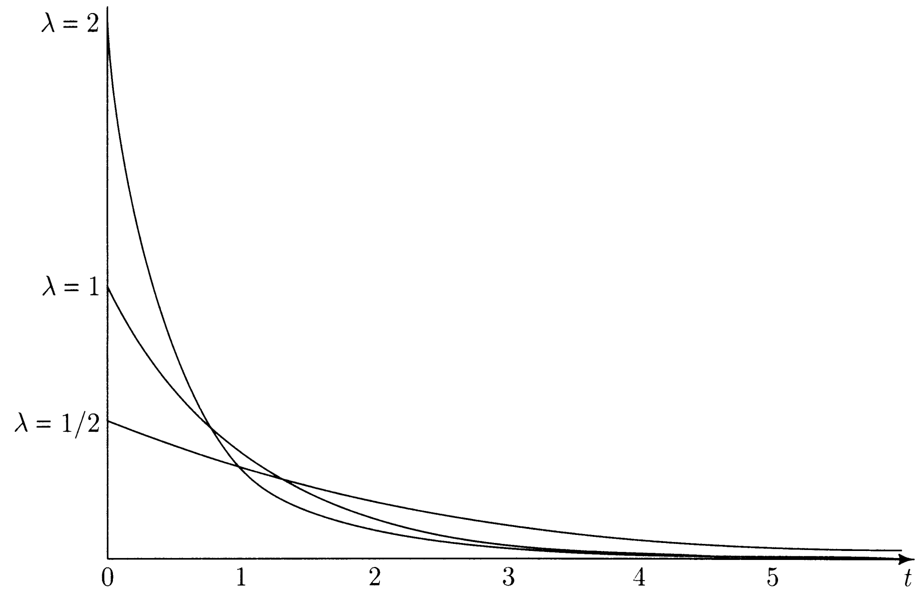
Memoryless Property of the Exponential Distribution

A positive random variable T has exponential (λ) distribution for some $\lambda > 0$ if and only if T has the *memoryless property*

$$P(T > t + s | T > t) = P(T > s) \qquad (s \geq 0, \quad t \geq 0)$$

In words: Given survival to time t, the chance of surviving a further time s is the same as the chance of surviving to time s in the first place.

FIGURE 1. Exponential densities for $\lambda = 0.5, 1, 2$.



The memoryless property follows immediately from the formula for the survival function, as you should check. The converse hinges on the fact that if T has the memoryless property then the survival function $G(t) = P(T > t)$ must be a solution of the functional equation

$$G(t+s) = G(t)G(s) \qquad (t > 0, \quad s > 0)$$

with $G(t)$ decreasing and bounded between 0 and 1. It can be shown that every such function $G(t)$ is of the form $e^{-\lambda t}$ for some λ.

Thinking of T as the lifetime of something, the memoryless property is this: Whatever the current age of the thing, the distribution of the remaining lifetime is the same as the original lifetime distribution. Some things, such as atoms or electrical components, have this property, hence exponential lifetime distribution. But most forms of life do not have exponential lifetime distribution because they experience an aging process.

Interpretation of the rate λ. For something with an exponentially distributed lifetime, λ is the constant value of the instantaneous *death rate* or *hazard rate*. That is to say, λ measures the probability of death per unit time just after time t, given survival up to time t. To see why, for a time t and a further length of time Δ, calculate

$$\begin{aligned} P(T \le t + \Delta \,|\, T > t) &= 1 - P(T > t + \Delta \,|\, T > t) \\ &= 1 - P(T > \Delta) \quad \text{by the memoryless property} \\ &= 1 - e^{-\lambda\Delta} \end{aligned}$$

$$= 1 - [1 - \lambda\Delta + \frac{1}{2}\lambda^2\Delta^2 - \cdots]$$
$$\approx \lambda\Delta \qquad \text{for small } \Delta$$

where $\approx$ is an approximation with error negligible in comparison to Δ as $\Delta \to 0$. Less formally, for an infinitesimal time increment dt, the result of this calculation is that

$$P(T \le t + dt \,|\, T > t) = \lambda\, dt \qquad \text{or}$$
$$P(t < T \le t + dt)/dt = \lambda P(T > t)$$

Since the left side is the density of T at time t, this explains why the exponential (λ) density at t is the death rate λ times the probability $e^{-\lambda t}$ of survival to time t. The characteristic feature of exponentially distributed lifetimes is that the death rate is constant, not depending on t. Other continuous distributions on $(0, \infty)$ correspond to a time-dependent death rate $\lambda(t)$: see Section 4.3.

Example 1. **Reliability.**

Under suitably constant conditions of use, some kinds of electrical components, for example, fuses and transistors, have a lifetime distribution well fitted by an exponential distribution. Such a component does not wear out gradually. Rather, it stops functioning suddenly and unpredictably. No matter how long the component has been in use, the chance that it survives a further time interval of length Δ is always the same. This probability must then be $e^{-\lambda\Delta}$ for some rate λ, called the *failure rate* in this context. The lifetime distribution is then exponential with rate λ. Roughly speaking, so long as it is still functioning, such a component is as good as new.

Problem 1. Suppose the average lifetime of a particular kind of transistor is 100 working hours, and that the lifetime distribution is approximately exponential. Estimate the probability that the transistor will work for at least 50 hours.

Solution. Since the mean of the exponential distribution is $1/\lambda$, put

$$1/\lambda = 100 \qquad \text{so} \quad \lambda = 0.01$$

and calculate $\quad P(T > 50) = e^{-\lambda 50} = e^{-0.5} = 0.606\ldots$

Problem 2. Given that the transistor has functioned for 50 hours, what is the chance that it fails in the next minute of use?

Solution. From the interpretation of $\lambda = 0.01$, as the instantaneous rate of failure per hour given survival so far, the chance is about $0.01 \times 1/60 \approx 0.00017$.

Example 2. **Radioactive decay.**

Atoms of radioactive isotopes like Carbon 14, Uranium 235, or Strontium 90 remain intact up to a random instant of time when they suddenly decay, meaning that they split or turn into some other kind of atom, and emit a pulse of radiation or particles of some kind. This radioactive decay can be detected by a Geiger counter. Let T be the random lifetime, or time until decay, of such an atom, starting at some arbitrary time when the atom is intact. It is reasonable to assume that the distribution of T must have the memoryless property. Consequently, there is a rate $\lambda > 0$, the *rate of decay* for the isotope in question, such that T has exponential (λ) distribution: $P(T > t) = e^{-\lambda t}$.

Probabilities here have a clear interpretation due to the large numbers of atoms typically involved (for example, a few grams of a substance will consist of around 10^{23} atoms). Assume a large number N of such atoms decay independently of each other. Then, by the law of large numbers, the proportion of these N atoms that survives up to time t is bound to be close to $e^{-\lambda t}$, the survival probability for each individual atom. This exponential decay over time of the mass of radioactive substance has been experimentally verified, confirming the hypothesis that lifetimes of individual atoms are exponentially distributed. The decay rates λ for individual isotopes can be measured with great accuracy, using this exponential decay of mass. These rates λ show no apparent dependence on physical conditions such as temperature and pressure.

A common way to indicate the rate of decay of a radioactive isotope is by the *half life* h. This is the time it takes for half of a substantial amount of the isotope to disintegrate. So

$$e^{-\lambda h} = 1/2 \quad \text{or} \quad h = \log(2)/\lambda$$

In other words, the half life h is the *median* of the atomic lifetime distribution

$$P(T \leq h) = P(T > h) = 1/2$$

The median lifetime is smaller than the mean lifetime $1/\lambda$, by the factor of $\log(2) = 0.693147\ldots$. This is due to the very skewed shape of the exponential distribution.

Numerical illustration. Strontium 90 is a particularly dangerous component of fallout from nuclear explosions. The substance is toxic, easily absorbed into bones when eaten, and has a long half-life of about 28 years. Assuming this value for the half-life h, let us calculate:

a) *The decay rate* λ: From above, this is

$$\lambda = \frac{\log(2)}{h} = 0.693147.../28 = 0.0248 \text{ per year}$$

b) *The mean lifetime of a Strontium* 90 *atom*: This is

$$\frac{1}{\lambda} = \frac{h}{\log(2)} = \frac{28}{0.693147...} = 40.4 \text{ years}$$

c) *The probability that a Strontium* 90 *atom survives at least* 50 *years*: This is

$$P(T > 50) = e^{-\lambda 50} = e^{-0.0248 \times 50} = 0.29$$

d) *The proportion of one gram of Strontium* 90 *that remains after* 50 *years.* This proportion is the same as the above probability, by the law of large numbers.

e) *The number of years after a nuclear explosion before* 99% *of the Strontium* 90 *produced by the explosion has decayed.* Let y be the number of years. Then

$$e^{-0.0248y} = 1/100 \qquad \text{so} \qquad y = \log(100)/0.0248 \approx 186 \text{ years}$$

Relation to the geometric distribution. The exponential distribution on $(0, \infty)$ is the continuous analog of the geometric distribution on $\{1, 2, 3, \ldots\}$. For instance, in the formulation of the memoryless property it was assumed that s and t range over all non-negative real numbers. This property for integers s and t, and an integer-valued random variable T, is a characterization of the geometric distribution. An exponential distribution is the limit of rescaled geometric (p) distributions as the parameter p tends to 0. More precisely, if G has geometric (p) distribution, so that $P(G > n) = (1-p)^n$, and p is small so that $E(G) = 1/p$ is large, then the rescaled variable $G/E(G) = pG$ has approximately exponential distribution with rate $\lambda = 1$:

$$\begin{aligned} P(pG > t) = P(G > t/p) &\approx (1-p)^{t/p} \text{ (only} \approx \text{because } t/p \text{ may not be an integer)} \\ &\approx e^{-t} \end{aligned}$$

by the usual exponential approximation $(1-p) \approx e^{-p}$ for small p. This approximation has been used already in the gambler's rule example in Section 1.6. The factor of $\log(2)$ which appeared there was the median of the exponential distribution with rate 1.

Relation to a Poisson process. A sequence of independent Bernoulli trials, with probability p of success on each trial, can be characterized in two different ways as follows:

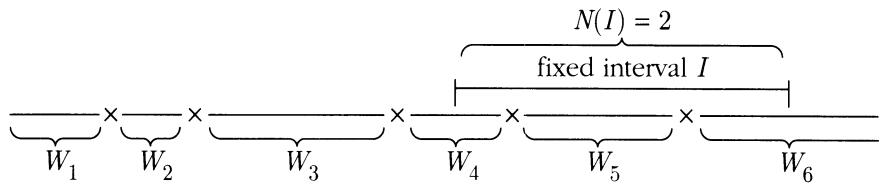
I. Counts of successes. The distribution of the number of successes in n trials is binomial (n, p), and numbers of successes in disjoint blocks of trials are independent.

II. Times between successes. The distribution of the waiting time until the first success is geometric (p), and the waiting times between each success and the next are independent with the same geometric distribution.

After a passage to the limit by discrete approximations, as in Section 3.5, these characterizations of Bernoulli trials lead to the two descriptions in the next box of a *Poisson arrival process with rate* λ. This means a Poisson random scatter of points, as in Section 3.5, for points now called *arrivals* on the interval $(0, \infty)$ interpreted as a time line, instead of hits on a region in the plane. In the diagram inside the box, arrivals are at times marked $\times$ on the time line. Think of arrivals representing something like calls coming into a telephone exchange, particles arriving at a counter, or customers entering a store.

Two Descriptions of a Poisson Arrival Process

I. Counts of arrivals. The distribution of the number of arrivals $N(I)$ in a fixed time interval I of length t is Poisson (λt), and numbers of arrivals in disjoint time intervals are independent.



II. Times between arrivals. The distribution of the waiting time W_1 until the first arrival is exponential (λ), and W_1 and the subsequent waiting times W_2, $W_3, \ldots$ between each arrival and the next are independent, all with the same exponential distribution.

These two descriptions of a random arrival process are equivalent.

Probabilities of events defined by a Poisson arrival process can be calculated from whichever of these two descriptions is more convenient.

Example 3. **Telephone calls.**

Suppose calls are coming into a telephone exchange at an average rate of 3 per minute, according to a Poisson arrival process. So, for instance, $N(2,4)$, the number of calls coming in between $t = 2$ and $t = 4$, has Poisson distribution with mean $\lambda(4-2) = 3 \times 2 = 6$; and W_3, the waiting time between the second and third calls, has exponential (3) distribution. Let us calculate:

a) *The probability that no calls arrive between* $t = 0$ *and* $t = 2$: Since $N(0,2]$, the number of calls arriving in this interval has Poisson (6) distribution, this is

$$P(N(0,2] = 0) = e^{-6} = 0.0025$$

b) *The probability that the first call after* $t = 0$ *takes more than 2 minutes to arrive.* From the exponential (3) distribution of W_1 this is

$$P(W_1 > 2) = e^{-3\times 2}$$

The answer is the same as in a) because the events are, in fact, identical.

c) *The probability that no calls arrive between $t = 0$ and $t = 2$ and at most four calls arrive between $t = 2$ and $t = 3$.* By independence of $N(0,2]$ and $N(2,3]$, this is

$$P(N(0,2] = 0) \cdot P(N(2,3] \leq 4) = e^{-6} \cdot e^{-3}(1 + 3 + \frac{3^2}{2!} + \frac{3^3}{3!} + \frac{3^4}{4!}) = 0.0020$$

d) *The probability that the fourth call arrives within 30 seconds of the third.* This is

$$P(W_4 \leq 0.5) = 1 - P(W_4 > 0.5) = 1 - e^{-3\times 0.5} = 0.7769$$

e) *The probability that the first call after $t = 0$ takes less than 20 seconds to arrive, and the waiting time between the first and second calls is more than 3 minutes.* By independence of W_1 and W_2, this is

$$P(W_1 < 1/3) \cdot P(W_2 > 3) = (1 - e^{-3\times 20/60})e^{-3\times 3}$$

f) *The probability that the fifth call takes more than 2 minutes to arrive.* Since the arrival time of the fifth call is the sum of the first five interarrival times, the problem is to find $P(W_1 + W_2 + W_3 + W_4 + W_5 > 2)$ where the W_i are independent, all with exponential (3) distribution. The general technique for finding the distribution of a sum of continuously distributed random variables is not discussed until Section 5.4. But this particular problem is solved easily by recoding it in terms of the Poisson distributed counts. The fifth call takes more than 2 minutes to arrive if and only if at most four calls arrive between $t = 0$ and $t = 2$. So the required probability is

$$\begin{aligned} P(W_1 + W_2 + W_3 + W_4 + W_5 > 2) &= P(N(0,2] \leq 4) \\ &= e^{-6}(1 + 6 + \frac{6^2}{2!} + \frac{6^3}{3!} + \frac{6^4}{4!}) = 0.2851 \end{aligned}$$

Gamma Distribution

As in the previous example, let $W_1, W_2, \ldots$ be independent exponential (λ) variables, and interpret the W_i as the waiting times between arrivals in a Poisson process with rate λ. The method used in the last part f) of the example can be used to find the distribution of the time T_r of the rth arrival, for any $r = 1, 2, \ldots$. Here is a general statement of the result:

Poisson Arrival Times (Gamma Distribution)

If T_r is the time of the rth arrival after time 0 in a Poisson process with rate λ, or if $T_r = W_1 + W_2 + \cdots + W_r$ where the W_i are independent with exponential (λ) distribution, then T_r has the *gamma* (r, λ) distribution defined by either (1) or (2) for all $t \geq 0$:

(1) **Density:** $$P(T_r \in dt)/dt = P(N_t = r-1)\lambda = e^{-\lambda t}\frac{(\lambda t)^{r-1}}{(r-1)!}\lambda$$

where N_t, the number of arrivals by time t in the Poisson process with rate λ, has Poisson (λt) distribution. In words, the probability per unit time that the rth arrival comes around time t is the probability of exactly $r-1$ arrivals by time t multiplied by the arrival rate.

(2) **Right tail probability:** $$P(T_r > t) = P(N_t \leq r-1) = \sum_{k=0}^{r-1} e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

because $T_r > t$ if and only if there are at most $r-1$ arrivals in the interval $(0, t]$.

(3) **Mean and SD:** $E(T_r) = r/\lambda \qquad SD(T_r) = \sqrt{r}/\lambda$

Formula (2) is the extension of the numerical example f) above from the case $r = 5, \lambda = 3, t = 2$ to general r, λ, and t. Formula (1) for the density can be derived from (2) by calculus. But here is a neater way. For the rth arrival to come in an infinitesimal interval of time of length dt just after time t, it must be that:

A: there is an arrival in the time dt,
where $P(A) = \lambda dt$, by the local interpretation of the arrival rate λ;

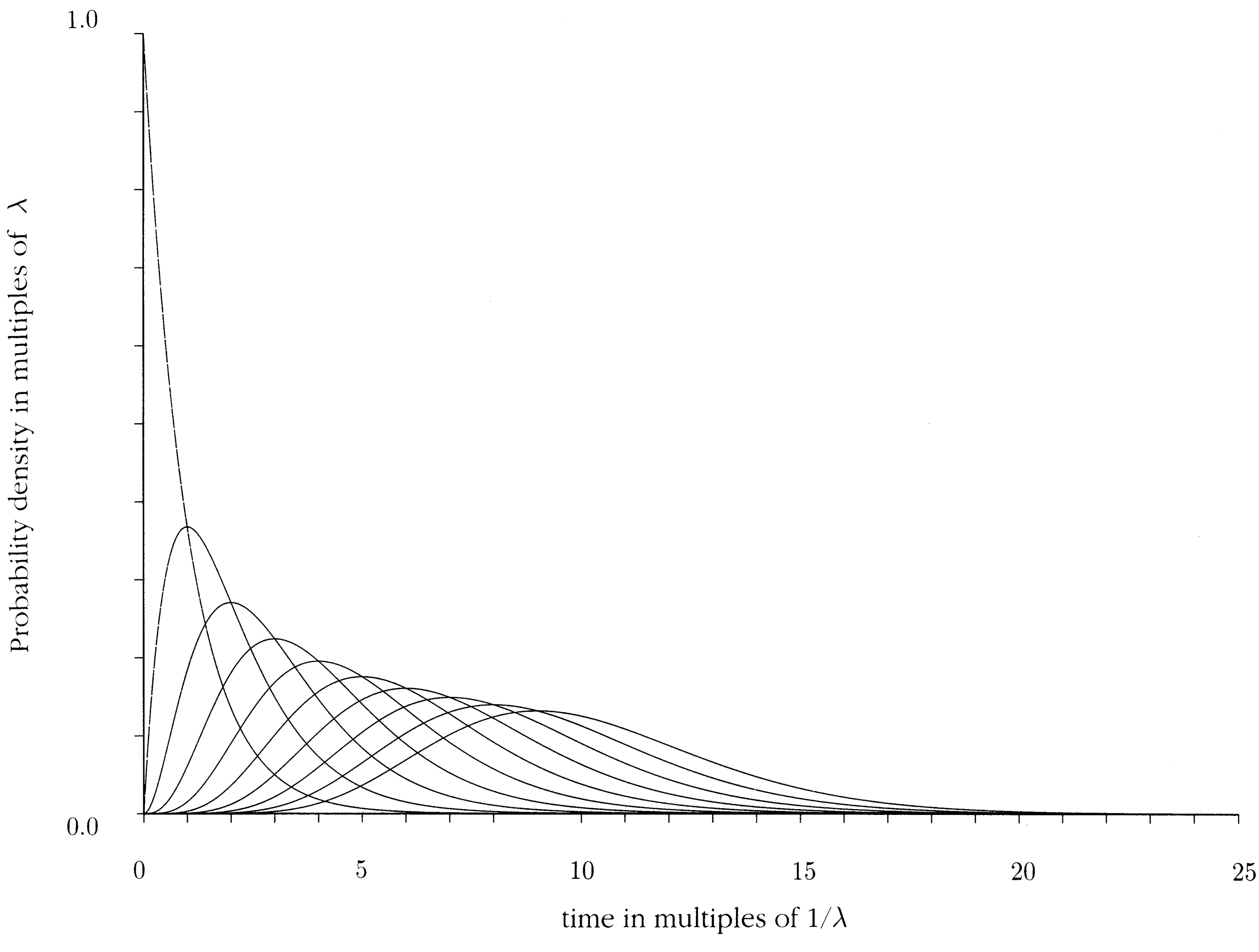
and (since the possibility of more than one arrival in the infinitesimal interval can be safely ignored), that:

B: there were exactly $r-1$ arrivals in the preceding time t,
where $P(B) = P(N_t = r-1) = e^{-\lambda t}(\lambda t)^{r-1}/(r-1)!$

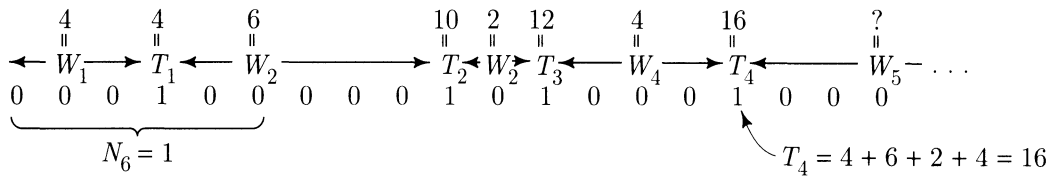
These events A and B are defined by arrivals in disjoint time intervals, so they are independent by the basic assumptions of a Poisson process. Multiplying their probabilities gives formula (1) for $P(AB) = P(T_r \in dt)$. The formulae (3) for the mean and SD are immediate from the representation of T_r as a sum of r independent exponential (λ) variables, and the formulae for the case $r = 1$, when the gamma $(1, \lambda)$ distribution is just exponential (λ).

The full extent of the analogy between Bernoulli trials and a Poisson process is brought out in the display on pages 288 and 289. In this analogy the continuous gamma (r, λ) distribution of the time until the rth arrival corresponds to the discrete negative binomial (r, p) distribution of the number of trials until the rth success, as derived in Section 3.4. As the display shows, the formulae relating the gamma to the Poisson distribution are like similar formulae relating the negative binomial to the binomial distribution.

FIGURE 2. Gamma density of the rth arrival for $r = 1$ to 10. Note how the distributions shift to the right and flatten out as r increases, in keeping with the formulae r/λ and $\sqrt{r}/\lambda$ for the mean and SD. Due to the central limit theorem, the gamma (r, λ) distribution becomes asymptotically normal as $r \to \infty$.



Summary of Properties of a Bernoulli (p) Trials Process



1. a) The probability of success per trial is p.
   b) The events of successes on different trials are independent.
   c) The long-run average success rate is p.

2. a) The number N_n of successes in n trials has binomial (n, p) distribution with

$$P(N_n = k) = \binom{n}{k} p^k q^{n-k} \qquad (k = 0, 1, \ldots, n, \quad n = 1, 2, \ldots)$$

$$E(N_n) = np \qquad \text{and} \qquad SD(N_n) = \sqrt{npq} \qquad \text{where } q = 1 - p$$

   b) As $n \to \infty$ the asymptotic distribution of $(N_n - E(N_n))/SD(N_n)$ is standard normal.

3. The waiting times $W_1, W_2, \ldots$ between successes are independent geometric (p) random variables with

$$\begin{aligned} P(W_k > n) &= P(\text{no successes in } n \text{ trials}) \\ &= P(N_n = 0) = q^n \qquad (n = 1, 2, \ldots) \\ P(W_k = n) &= P(\text{no successes in first } n-1 \text{ trials and trial } n \text{ is a success}) \\ &= P(N_{n-1} = 0)p = q^{n-1}p \qquad (n = 1, 2, \ldots) \end{aligned}$$

$$E(W_k) = \frac{1}{p} \qquad \text{and} \qquad SD(W_k) = \frac{\sqrt{q}}{p} \qquad (k = 1, 2, \ldots)$$

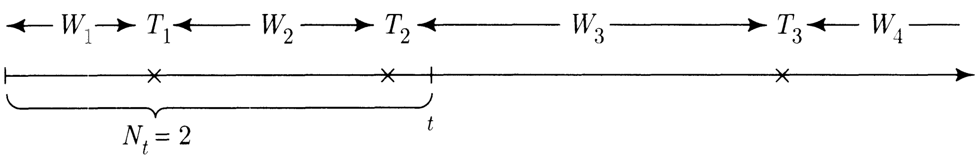
4. a) The waiting time $T_r = W_1 + \cdots + W_r$ until the rth success has negative binomial (r, p) distribution shifted to $\{r, r+1, \ldots\}$ with

$$\begin{aligned} P(T_r > n) &= P(N_n < r) \qquad (n = 1, 2, \ldots, \quad r = 1, 2, \ldots) \\ P(T_r = n) &= P(r-1 \text{ successes in first } n-1 \text{ trials and trial } n \text{ is a success}) \\ &= P(N_{n-1} = r-1)p \end{aligned}$$

$$E(T_r) = \frac{r}{p} \qquad \text{and} \qquad SD(T_r) = \frac{\sqrt{rq}}{p} \qquad (r = 1, 2, \ldots)$$

   b) As $r \to \infty$ the distribution of $(T_r - E(T_r))/SD(T_r)$ converges to standard normal.

Summary of Properties of a Poisson (λ) Arrival Process



1. a) $P(\text{arrival in interval } \Delta t) \approx \lambda \Delta t \quad$ as $\Delta t \to 0$.
   b) The events of arrivals in disjoint intervals are independent.
   c) The long-run average rate of arrivals per unit time is λ.

2. a) The number N_t of arrivals in time t has Poisson (λt) distribution with

$$P(N_t = k) = e^{-\lambda t}(\lambda t)^k / k! \qquad (k = 0, 1, \ldots, \quad t \geq 0)$$

$$E(N_t) = \lambda t \qquad \text{and} \qquad SD(N_t) = \sqrt{\lambda t}$$

   b) As $t \to \infty$ the asymptotic distribution of $(N_t - E(N_t))\,/SD(N_t)$ is standard normal.

3. The waiting times $W_1, W_2, \ldots$ between arrivals are independent exponential (λ) random variables with

$$\begin{aligned} P(W_k > t) &= P(\text{no arrivals in time } t) \\ &= P(N_t = 0) = e^{-\lambda t} \qquad (t \geq 0) \\ P(W_k \in dt) &= P(\text{no arrivals in time } t, \text{ arrival in time } dt) \\ &= P(N_t = 0) P(\text{arrival in time } dt) = e^{-\lambda t} \lambda dt \qquad (t \geq 0) \end{aligned}$$

$$E(W_k) = \frac{1}{\lambda} \qquad \text{and} \qquad SD(W_k) = \frac{1}{\lambda} \qquad (k = 1, 2, \ldots)$$

4. a) The waiting time $T_r = W_1 + \cdots + W_r$ until the rth arrival has gamma (r, λ) distribution with

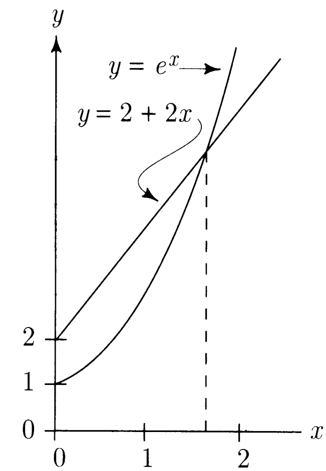
$$\begin{aligned} P(T_r > t) &= P(N_t < r) \qquad (t \geq 0, \quad r = 1, 2, \ldots) \\ P(T_r \in dt) &= P(r - 1 \text{ arrivals in time } t \text{ and arrival in time } dt) \\ &= P(N_t = r - 1) \lambda dt \end{aligned}$$

$$E(T_r) = \frac{r}{\lambda} \qquad \text{and} \qquad SD(T_r) = \frac{\sqrt{r}}{\lambda} \qquad (r = 1, 2, \ldots)$$

   b) As $r \to \infty$ the distribution of $(T_r - E(T_r))\,/SD(T_r)$ converges to standard normal.

Example 4. Sum of two lifetimes.

Problem. A component with lifetime that is exponentially distributed with failure rate 1 per 24 hours is put into service with a replacement component of the same kind which is substituted for the first one when it fails. What is the median of the total time to failure of both components?



Solution. The problem is to find m such that $P(T_2 \geq m) = 1/2$, where $T_2 = W_1 + W_2$ is the sum of two independent exponential lifetimes with rate $\lambda = 1/24$ per hour. But from formula (2) on page 286

$$P(T_2 \geq m) = P(N_m \leq 1) = e^{-\lambda m}(1+\lambda m)$$

where N_m has Poisson (λm) distribution. Put $x = \lambda m$. Then $m = x/\lambda$ where x solves

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

Some trial and error with a calculator gives $x \approx 1.675$. So the median is about $1.675/(1/24) \approx 40.3$ hours.

Discussion. Note how the Poisson formula for $P(T_2 \geq t)$ can be used here for the gamma $(2, \lambda)$ distribution of the sum $T_2 = W_1 + W_2$ of two independent exponential (λ) variables, even though these exponential random variables are not originally defined as interarrival times for a Poisson process. Technically, this is because the distribution of a sum of independent random variables is determined by the distributions of the individual variables. Section 5.4 goes into this in more detail. Intuitively, you may as well suppose the two lifetimes W_1 and W_2 are just the first two in an infinite sequence of independent exponentially distributed lifetimes of components replaced one after another. In that case the times of replacements would make a Poisson process, with N_t representing the total number of replacements by time t.

Gamma Distribution for Non-Integer Shape Parameter

A gamma distribution is defined for all positive values of the parameters r and λ by a variation of the density formula (1) on page 286 for integer r. A random variable T has *gamma distribution with parameters r and λ*, or *gamma (r, λ) distribution*, if

T has probability density

$$f_{r,\lambda}(t) = \begin{cases} [\Gamma(r)]^{-1}\lambda^r t^{r-1}e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \qquad \text{where} \quad \Gamma(r) = \int_0^\infty t^{r-1}e^{-t}dt$$

is a constant of integration, depending on r, called the *gamma function.* The parameter r is called the *index* or *shape* parameter. And $1/\lambda$ is a scale parameter. Comparison with formula (1) on page 286 shows that

$$\Gamma(r) = (r-1)! \qquad (r = 1, 2, \ldots)$$

You should think of the gamma function $\Gamma(r)$ as a continuous interpolation of the factorial function $(r-1)!$ for non-integer r. Integration by parts gives the following:

Recursion formula for the gamma function: $\Gamma(r+1) = r\Gamma(r) \quad (r > 0)$

Since it is easy to see that $\Gamma(1) = 1$, the recursion formula implies $\Gamma(r) = (r-1)!$ for integer r by mathematical induction.

But there is no explicit formula for $\Gamma(r)$ except in case r is a positive integer, or a positive half-integer, starting from $\Gamma(1/2) = \sqrt{\pi}$. See Exercise 5.3.15. Section 5.3 shows that for half integer r the gamma distributions arise from sums of squares of independent normal variables.

As will be shown in Section 5.4, several algebraic functions of gamma random variables have distributions which are easy to compute. See the gamma distribution summary for a survey. In applications, the distribution of a random variable may be unknown, but reasonably well approximated by some gamma distribution. Then results obtained assuming a gamma distribution might provide useful approximations.

For non-integer values of r the gamma (r, λ) distribution has a shape which varies continuously between the shapes for integers r, as illustrated by the following diagrams:

In Figures 3, 4, and 5, both horizontal and vertical scales change from one figure to the next. Figure 3 shows how the gamma (r, λ) density is unbounded near zero for $0 < r < 1$. As $r \to 0$ the distribution piles up more and more near zero, approaching the distribution of a constant random variable with value 0. This is a discrete distribution, which does not have a probability density, but assigns probability one to the point zero, and may be thought of as the gamma $(0, \lambda)$ distribution.

FIGURE 3. Gamma (r, λ) densities for $\lambda = 1$ and r a multiple of $1/4$, $0 < r \leq 1$.

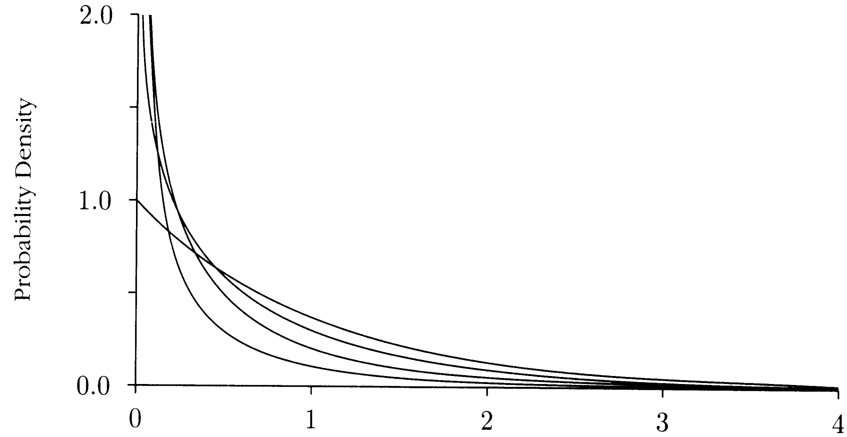


FIGURE 4. Gamma (r, λ) densities for $\lambda = 1$ and r a multiple of $1/4$, $1 \leq r \leq 2$.

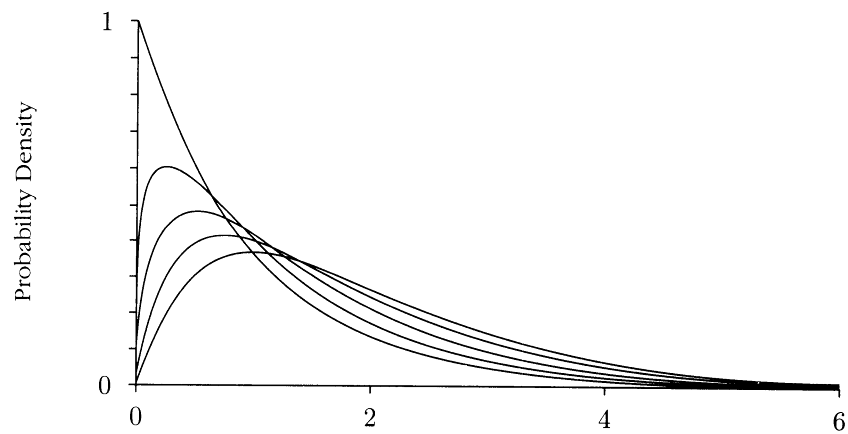
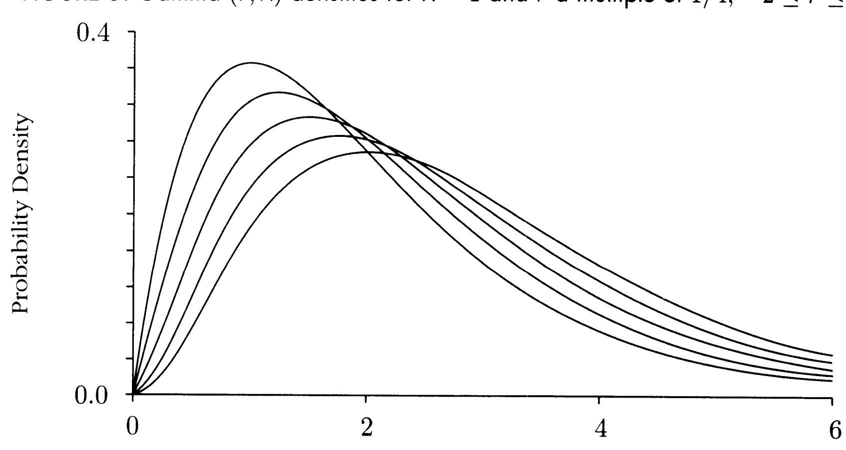


FIGURE 5. Gamma (r, λ) densities for $\lambda = 1$ and r a multiple of $1/4$, $2 \leq r \leq 3$.



Exercises 4.2

1. Suppose a particular kind of atom has a half-life of 1 year. Find:

a) the probability that an atom of this type survives at least 5 years;

b) the time at which the expected number of atoms is 10% of the original;

c) if there are 1024 atoms present initially, the time at which the expected number of atoms remaining is one;

d) the chance that in fact none of the 1024 original atoms remains after the time calculated in c).

2. A piece of rock contains 10^{20} atoms of a particular substance. Each atom has an exponentially distributed lifetime with a half-life of one century. How many centuries must pass before

a) it is most likely that about 100 atoms remain;

b) there is about a 50% chance that at least one atom remains. What assumptions are you making?

3. Suppose the time until the next earthquake in a particular place is exponentially distributed with rate 1 per year. Find the probability that the next earthquake happens within

a) one year; b) six months; c) two years; d) 10 years.

4. Suppose component lifetimes are exponentially distributed with mean 10 hours. Find:

a) the probability that a component survives 20 hours;

b) the median component lifetime;

c) the SD of component lifetime;

d) the probability that the average lifetime of 100 independent components exceeds 11 hours;

e) the probability that the average lifetime of 2 independent components exceeds 11 hours.

5. Suppose calls are arriving at a telephone exchange at an average rate of one per second, according to a Poisson arrival process. Find:

a) the probability that the fourth call after time $t = 0$ arrives within 2 seconds of the third call;

b) the probability that the fourth call arrives by time $t = 5$ seconds;

c) the expected time at which the fourth call arrives.

6. A Geiger counter is recording background radiation at an average rate of one hit per minute. Let T_3 be the time in minutes when the third hit occurs after the counter is switched on. Find $P(2 \leq T_3 \leq 4)$.

7. Let $0 < p < 1$. For the exponential distribution with rate λ, find a formula for the $100p$th percentile point t_p such that $P(T \leq t_p) = 100p\%$.

8. Transistors produced by one machine have a lifetime which is exponentially distributed with mean 100 hours. Those produced by a second machine have an exponentially distributed lifetime with mean 200 hours. A package of 12 transistors contains 4 produced by the first machine and 8 produced by the second. Let X be the lifetime of a transistor picked at random from this package. Find:

a) $P(X \geq 200 \text{ hours})$; b) $E(X)$; c) $Var(X)$.

9. Gamma function and moments of the exponential distribution. Consider the gamma function $\Gamma(r) = \int_0^\infty x^{r-1}e^{-x}dx \quad (r > 0)$

a) Use integration by parts to show that $\Gamma(r+1) = r\Gamma(r) \quad (r > 0)$

b) Deduce from a) that $\Gamma(r) = (r-1)! \quad (r = 1, 2, \ldots)$

c) If T has exponential distribution with rate 1, then

$$E(T^n) = n! \qquad (n = 0, 1, 2, \ldots) \qquad \text{and} \quad SD(T) = 1$$

d) If T has exponential distribution with rate λ, then show λT has exponential distribution with rate 1, hence

$$E(T^n) = n!/\lambda^n \qquad (n = 0, 1, 2, \ldots) \qquad \text{and} \quad SD(T) = 1/\lambda$$

10. Geometric from exponential.

a) Show that if T has exponential distribution with rate λ, then int(T), the greatest integer less than or equal to T, has a geometric (p) distribution on $\{0, 1, 2, \ldots\}$, and find p in terms of λ.

b) Let $T_m = \text{int}(mT)/m$, the greatest multiple of $1/m$ less than or equal to T. Show that T has exponential distribution on $(0, \infty)$ for some λ, if and only if for every m there is some p_m such that mT_m has geometric (p_m) distribution on $\{0, 1, 2, \ldots\}$. Find p_m in terms of λ.

c) Use b) and $T_m \leq T \leq T_m + 1/m$ to calculate $E(T)$ and $SD(T)$, from the formulae for the mean and standard deviation of a geometric random variable.

11. Suppose the probability that a given kind of atom disintegrates in any particular microsecond, given that it was alive at the beginning of the microsecond, is $\lambda \times 10^{-6}$ where $\lambda > 0$ is a constant. Let T be the random lifetime of the atom in seconds.

a) Show that the distribution of T is approximately exponential with parameter λ. [*Hint*: Consider $P(T \geq t)$ for t a multiple of 10^{-6}.]

b) What is the chance that the atom has a lifetime of between 1 and 2 seconds?

12. Gamma distribution. Derive the following features of the gamma (r, λ) distribution for all positive r:

a) For $r \geq 1$ the mode (i.e., the value that maximizes the density) is $(r-1)/\lambda$. What if $0 < r < 1$?

b) For $k > 0$, the kth moment of T with gamma (r, λ) distribution is

$$E(T^k) = \frac{1}{\lambda^k}\frac{\Gamma(r+k)}{\Gamma(r)}$$

In particular $E(T) = r/\lambda$.

c) $SD(T) = \sqrt{r}/\lambda$ and Skewness$(T) = 2/\sqrt{r}$.

13. Suppose that under normal operating conditions the operating time until failure of a certain type of component has exponential (λ) distribution for some $\lambda > 0$. And suppose that the random variables representing lifetimes of different components of this type may be regarded as independent.

a) The average lifetime of 10,000 components is found to be 20 days. Estimate the value of λ based on this information.

b) Assuming the exponential lifetime model with $\lambda = 5\%$ per day, let N_d be the number of components among 10,000 components which survive more than d days. Find $E(N_d)$ and $SD(N_d)$ for $d = 10, 20, 30$.

14. Interpretation of the rate. In Exercise 13, the exponential model with $\lambda = 5\%$ per day implies the probability of a component failing in the first day of its use is:
a) exactly 5%; b) approximately 5%, but slightly less;
c) approximately 5%, but slightly more. Without doing any numerical calculations, pick out which of a), b), or c) is true, and explain your choice. Confirm your choice by numerical calculation of the exact probability.

15. Satellite problem. Suppose that a system using one of the components described in Exercise 13, with failure rate 5% per day, is sent up in a satellite together with three spare components of the same type. Assume that as soon as the original component fails, it is replaced by one of the spares, and when that component fails it is replaced by a second spare, and so on. The total operating time of the component plus three spares is then $T_{\text{total}} = T_1 + T_2 + T_3 + T_4$ where T_1 is the operating time of the first component, T_2 is the operating time of the first spare, and so on. Assuming that the satellite launch is successful, and normal operating conditions obtain once the satellite is in orbit, calculate:

a) $E(T_{\text{total}}])$; b) $SD(T_{\text{total}})$; c) $P(T_{\text{total}} \geq 60 \text{ days})$.

16. In the satellite problem of Exercise 15, how many spares would have to be provided to make $P(T_{\text{total}} \geq 60 \text{ days})$ at least 90%?

17. Another type of component has lifetime distribution which is approximately gamma $(2, \lambda)$ with $\lambda = 10\%$ per day.

a) Redo Exercise 15 for this type of component, making similar independence assumptions. After calculating the answers to a) and b), guess without calculation whether the answer to c) should be larger or smaller than under the original assumptions of the satellite problem. Confirm your guess by calculation.

b) Redo Exercise 16 for this type of component.

4.3 Hazard Rates (Optional)

Let T be a positive random variable with probability density $f(t)$, where t ranges over $(0, \infty)$. Think of T as the lifetime of some kind of component. The *hazard rate* $\lambda(t)$ is the probability per unit time that the component will fail just after time t, *given* that the component has survived up to time t. Thus

$$P(T \in dt \,|\, T > t) = \lambda(t)dt$$

where $(T \in dt)$ stands for the event $(t < T \leq t + dt)$ that the component fails in an infinitesimal time interval of length dt just after time t. As usual, this is shorthand for a limit statement:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P\,(T \in (t, t + \Delta t)\,|\,T > t)}{\Delta t}$$

Depending on what lifetime T represents in an application, the hazard rate $\lambda(t)$ may also be called a *death rate* or *failure rate*. For example, T might represent the lifetime of some kind of component. Then $\lambda(t)$ would represent the failure rate for components that have been in use for time t, estimated, for example, by the number of failures per hour among similar components in use for time t.

In practice, failure rates can be estimated empirically as suggested above. Often it is found that empirically estimated hazard rates based on large amounts of data tend to follow a smooth curve. It is then reasonable to fit an ideal model in which $\lambda(t)$ would usually be a continuous function of t. The exponential distribution of the previous section is the simplest possible model corresponding to *constant* failure rate $\lambda(t) = \lambda$ for some $\lambda > 0$. Other distributions with densities on $(0, \infty)$ correspond to time-varying failure rates. The following box summarizes the basic terminology and analytic relationships between the probability density, survival function, and hazard rate.

Formulae (1), (2), and (3) in the box are simply definitions, and (4) is the usual integral for the probability of an interval. Formulae (4) and (5) are equivalent by the fundamental theorem of calculus. Informally, (5) results from

$$\begin{aligned} f(t)dt &= P(T \in dt) \qquad \text{by (1)} \\ &= P(T > t) - P(T > t + dt) \qquad \text{by the difference rule} \\ &= G(t) - G(t + dt) \qquad \text{by (2)} \\ &= -dG(t) \end{aligned}$$

Random Lifetimes

Probability density:	$P(T \in dt) = f(t)dt$	(1)
Survival function:	$P(T > t) = G(t)$	(2)
Hazard rate:	$P(T \in dt \mid T > t) = \lambda(t)dt$	(3)
Survival from density:	$G(t) = \int_t^\infty f(u)du$	(4)
Density from survival:	$f(t) = -\frac{dG(t)}{dt}$	(5)
Hazard from density and survival:	$\lambda(t) = \frac{f(t)}{G(t)}$	(6)
Survival from hazard:	$G(t) = \exp\left(-\int_0^t \lambda(u)du\right)$	(7)

To obtain (6), use $P(A|B) = P(AB)/P(B)$, with $A = (T \in dt)$, $B = (T > t)$. Since $A \subset B$, $AB = A$,

$$\lambda(t)dt = P(T \in dt \mid T > t) = \frac{P(T \in dt)}{P(T > t)} = \frac{f(t)dt}{G(t)}$$

by (1) and (2).

The most interesting formula is (7). To illustrate, in case $\lambda(t) = \lambda$ is constant,

$$\int_0^t \lambda(u)du = \lambda t$$

so (7) becomes the familiar exponential survival probability

$$P(T > t) = e^{-\lambda t} \qquad \text{if } T \text{ has exponential } (\lambda) \text{ distribution.}$$

In general, the exponential of the integral in (7) represents a kind of continuous product obtained as a limit of discrete products of conditional probabilities. This is explained at the end of the section. Formula (7) follows also from (5) and (6) by calculus as you can check as an exercise.

Example 1. Linear failure rate.

Problem. Suppose that a component has linear increasing failure rate, such that after 10 hours the failure rate is 5% per hour, and after 20 hours 10% per hour.

(a) Find the probability that the component survives 20 hours.

(b) Calculate the density of the lifetime distribution.

(c) Find the mean lifetime.

Solution. By assumption,

$$\lambda(t) = (t/2)\% = t/200$$

(a) The required probability is by (7)

$$P(\text{survive 20 hours}) = G(20) = \exp\left(-\int_0^{20} \lambda(u)du\right)$$

The integral inside the exponent is

$$\int_0^{20} \frac{u\,du}{200} = \frac{1}{400}u^2\Big|_0^{20} = 1$$

Thus $P(\text{survive 20 hours}) = e^{-1} \approx 0.368$

(b) Put t instead of 20 above to get

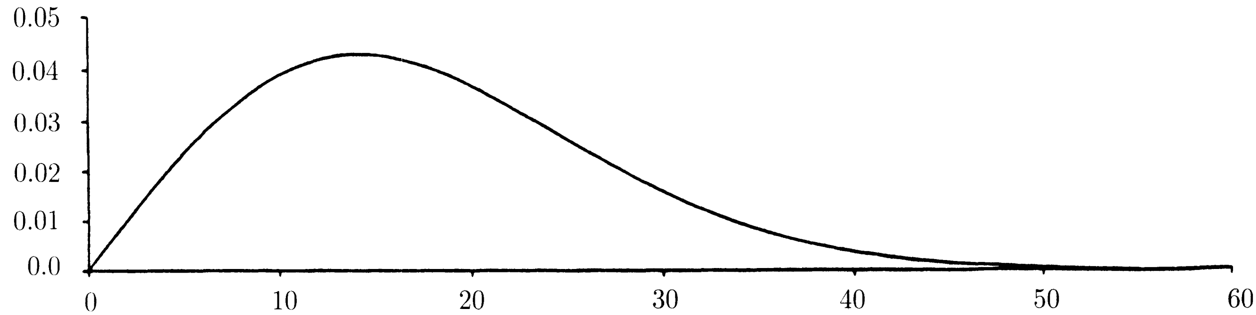
$$G(t) = \exp(-t^2/400)$$

Now by (5)

$$\begin{aligned} f(t) &= -\frac{d}{dt}G(t) \\ &= \frac{t}{200}\exp\left(-\frac{t^2}{400}\right) \end{aligned}$$

You can sketch the density by calculating a few points, as in the following table and graph:

t	0	5	10	15	20
$f(t)$	0	0.023	0.039	0.043	0.037



(c) The mean can be calculated from

$$E(T) = \int_0^\infty t f(t)\, dt$$

but there is a shortcut for examples like this where the survival function $G(t)$ is simpler than the density $f(t)$. This is to use the following formula:

Mean Lifetime from Survival Function

$$E(T) = \int_0^\infty G(t)\, dt \tag{8}$$

This follows by integration by parts from the previous formula for $E(T)$, using $\dfrac{dG(t)}{dt} = -f(t)$. It is the continuous analog of the formula

$$E(T) = \sum_{n=1}^{\infty} P(T \geq n)$$

valid for a random variable T with possible values $0, 1, 2, \ldots$. In the present example, (8) gives

$$E(T) = \int_0^\infty \exp(-t^2/400)\, dt \tag{9}$$

Now the problem is that you cannot integrate the function $\exp(-t^2/400)$ in closed form. But you should recognize this integral as similar to the standard Gaussian integral

$$\int_0^\infty e^{-z^2/2}\, dz = \frac{1}{2}\int_{-\infty}^\infty e^{-z^2/2}\, dz = \frac{1}{2}\sqrt{2\pi} = \sqrt{\frac{\pi}{2}}$$

Since $t^2/400 = \dfrac{1}{2}\left(\dfrac{t}{10\sqrt{2}}\right)^2$, make the change of variable $z = t/10\sqrt{2}$, $dz = dt/10\sqrt{2}$, $dt = 10\sqrt{2}dz$ in (9) to get

$$E(T) = 10\sqrt{2}\int_0^\infty e^{-z^2/2}\, dz = 10\sqrt{2}\sqrt{\frac{\pi}{2}} \approx 17.72$$

Derivation of the formula $G(t) = \exp\left(-\int_0^t \lambda(u)\,du\right)$. Recall that the exponential of a sum is the product of exponentials. An integral is a kind of continuous sum, so an exponential of an integral is a kind of continuous product. In this case, the continuous product is a limit of discrete products of conditional probabilities. To see how, divide the time interval $[0, t]$ into a very large number N of very small intervals of length say $\Delta = t/N$. Survival to time t means survival of each of the N successive intervals of length Δ between 0 and t

$$
\begin{aligned}
G(t) &= P(T > t) = P(T > N\Delta) \\
&= P(T > \Delta, T > 2\Delta, \cdots, T > N\Delta) \\
&= P(T > \Delta)P(T > 2\Delta \mid T > \Delta) \cdots P(T > N\Delta \mid T > (N-1)\Delta) \\
&= [1 - P(T \le \Delta)]\,[1 - P(\Delta \le T \le 2\Delta \mid T > \Delta)] \cdots \\
&\approx [1 - \Delta\lambda(0)]\,[1 - \Delta\lambda(\Delta)]\,[1 - \Delta\lambda(2\Delta)] \cdots [1 - \Delta\lambda\left((N-1)\Delta\right)] \\
&\qquad \text{for small } \Delta, \text{ by the definition of } \lambda(t) \\
&\approx e^{-\Delta\lambda(0)} e^{-\Delta\lambda(\Delta)} \cdots e^{-\Delta\lambda((N-1)\Delta)} \\
&\qquad \text{for small } \Delta, \text{ by the approximation } 1 - x \approx e^{-x} \text{ for small } x \\
&= \exp\left[-\Delta \sum_{i=0}^{N-1} \lambda(i\Delta)\right] \\
&\approx \exp\left[-\int_0^t \lambda(u)\,du\right] \\
&\qquad \text{for small } \Delta, \text{ by a Riemann sum approximation of the integral.}
\end{aligned}
$$

As $\Delta \to 0$, the errors in each of the three approximations $\approx$ above tend to zero. So the approximate equality between the first and last expressions not involving Δ must in fact be an exact equality. This is (7).

Note how the exponential appears here, as always, as the limit of a product of more and more factors all approaching 1 in the limit.

Exercises 4.3

1. For T with survival function $G(t) = P(T > t)$, find:
a) $P(T \le b)$; b) $P(a \le T \le b)$.

2. Use the formulae of this section to show that the hazard rate $\lambda(t)$ is constant if and only if the distribution is exponential (λ) for some λ.

3. Business enterprises have the feature that the longer an enterprise has been in business, the less likely it is to fail in the next month. This indicates a decreasing failure rate. One that has been successfully fitted to empirical data of lifetimes of businesses is $\lambda(t) = a/(b + t)$, where a, b, and t are greater than 0. For this $\lambda(t)$:
a) find a formula for $G(t)$; b) find a formula for $f(t)$.

4. Weibull distribution. Show that the following are equivalent:

(i) $\lambda(t) = \lambda\alpha t^{\alpha-1}$ for constants $\lambda > 0$ and $\alpha > 0$

(ii) $G(t) = e^{-\lambda t^{\alpha}}$

(iii) $f(t) = \lambda\alpha t^{\alpha-1}e^{-\lambda t^{\alpha}}$

This is called the *Weibull* distribution with parameters λ and α. This family of distributions is widely used in engineering practice. It can be verified both theoretically and practically that the distribution of the lifetime of a component which consists of many parts, and fails when the first of these parts fails, can be well approximated by a Weibull distribution.

5. Moments of the Weibull distribution. Let T have the Weibull distribution described in Exercise 4. a) Show that $E(T^k) = \Gamma(1 + \frac{k}{\alpha})\,\lambda^{-\frac{k}{\alpha}}$ b) Find $E(T)$ and $Var(T)$.

6. Suppose that a component is subject to failure at constant rate 5% per hour for the first 10 hours in use. After 10 hours the component is subject to additional stress producing a failure rate of 10% per hour.

a) Find the probability that the component survives 15 hours.

b) Calculate and sketch the survival probability function.

c) Calculate and sketch the probability density function.

d) Find the mean lifetime.

7. Second moment from survival function.

a) Show that $E(T^2) = 2\int_0^\infty tG(t)\,dt$

b) Use this formula to calculate the SD of the component in Example 1.

c) If 100 components of this type operate independently, what approximately is the probability that the average lifetime of these components exceeds 20 hours?

8. Suppose the failure rate is $\lambda(t) = at + b$ for $t \geq 0$.

a) For what parameter values a and b does this make sense?

b) Find the formula for $G(t)$. c) Find the formula for $f(t)$.

d) Find the mean lifetime. e) Find the SD of the lifetime.

9. Calculus derivation of $G(t) = \exp\{-\int_0^t \lambda(u)du\}$ (Formula (7)).

a) Use (5) and (6) to show $\lambda(t) = -\frac{d}{dt}\log G(t)$.

b) Now derive (7) by integration from 0 to t.

10. Suppose a component has failure rate $\lambda(t)$ which is an increasing function of t.

a) For $s, t > 0$, is $P(T > s + t | T > s)$ larger or smaller than $P(T > t)$?

b) Prove your answer.

c) Repeat a) and b) for $\lambda(t)$ which is decreasing.

4.4 Change of Variable

Many problems require finding the distribution of some function of X, say $Y = g(X)$, from the distribution of X. Suppose X has density $f_X(x)$, where a subscript is now used to distinguish densities of different random variables. Then provided the function $y = g(x)$ has a derivative dy/dx which does not equal zero on any interval in the range of X, the random variable $Y = g(X)$ has a density $f_Y(y)$ which can be calculated in terms of $f_X(x)$ and the derivative dy/dx. How to do this calculation is the subject of this section.

Linear Functions

To see why the derivative comes in, look first at what happens if you make a *linear* change of variable. For a linear function $y = ax + b$, the derivative is the constant $dy/dx = a$. The function stretches or shrinks the length of every interval by the same factor of $|a|$.

Example 1. **Uniform distributions.**

Suppose X has the uniform $(0, 1)$ distribution, with density

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Then for $a > 0$, you can see that $Y = aX + b$ has the uniform $(b, b + a)$ density

$$f_Y(y) = \begin{cases} 1/a, & b < y < b + a \\ 0 & \text{otherwise} \end{cases}$$

Similarly, if $a < 0$, then $Y = aX + b$ has the uniform $(b + a, b)$ distribution

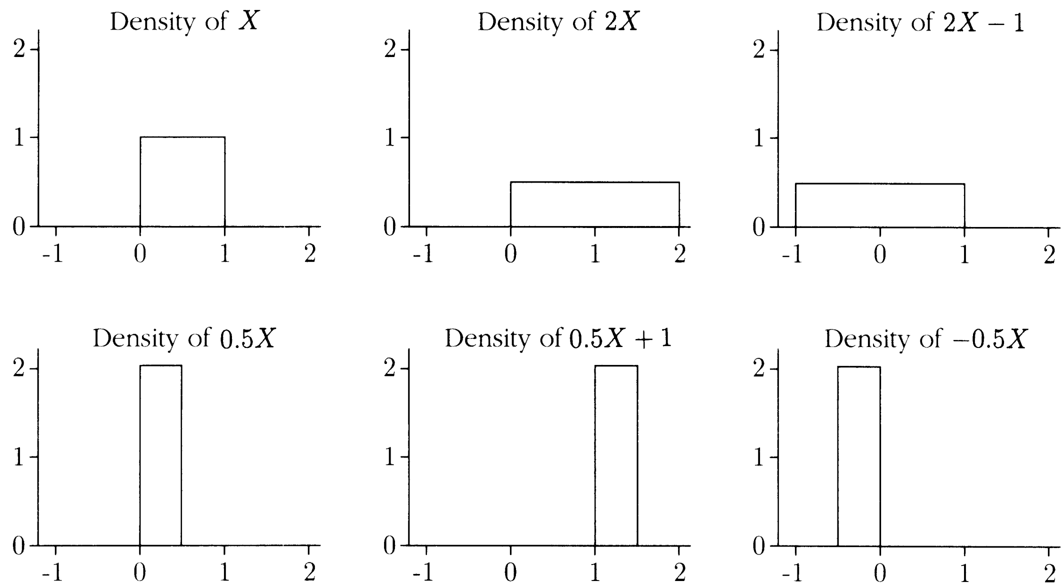
$$f_Y(y) = \begin{cases} 1/|a|, & b + a < y < b \\ 0 & \text{otherwise} \end{cases}$$

You might guess the density of $Y = aX + b$ at y was the density of X at the corresponding point $x = (y - b)/a$. But this must be divided by $|a|$, because the probability density gives probability per unit length, and the transformation from x to $ax + b$ multiplies lengths by a factor of $|a|$:

Linear Change of Variable for Densities

$$f_{aX+b}(y) = \frac{1}{|a|} f_X\left(\frac{y - b}{a}\right)$$

FIGURE 1. **Linear change of variable for uniform densities.** The graphs show the densities of $Y = aX + b$ for various a and b, where X has uniform $(0, 1)$ distribution. Notice how if $a > 1$ the range is spread out and the density decreased. And if $0 < a < 1$ the range is shrunk and the density increased. Adding $b > 0$ shifts to the right by b, and adding $b < 0$ shifts to the left by $-b$.



Example 2. Normal distributions.

Take X with standard normal density $\phi(x)$, $a = \sigma > 0$, and $b = \mu$. The linear change of variable formula then gives the density of the normal (μ, σ^2) distribution, displayed on page 267.

One-to-One Differentiable Functions

Let X be a random variable with density $f_X(x)$ on the range (a, b). Let $Y = g(X)$ where g is either strictly increasing or strictly decreasing on (a, b). For example, X might have an exponential distribution on $(0, \infty)$, and Y might be X^2, $\sqrt{X}$, or $1/X$. The range of Y is then an interval with endpoints $g(a)$ and $g(b)$.

The aim now is to calculate the probability density function $f_Y(y)$ for y in the range of Y. For an infinitesimal interval dy near y, the event $(Y \in dy)$ is identical to the event $(X \in dx)$, where dx is an infinitesimal interval near the unique x such that $y = g(x)$. See Figure 2, where each of the two shaded areas represents the probability of the same event

$$P(Y \in dy) = P(X \in dx) \quad \text{where} \quad y = g(x)$$

This identity $P(Y \in dy) = P(X \in dx)$, where $y = g(x)$, makes

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

and so
$$f_Y(y) = f_X(x)\frac{dx}{dy} = f_X(x)\Big/\frac{dy}{dx} \quad \text{where} \quad y = g(x)$$

The case of a decreasing function g is similar except that the calculus derivative dy/dx now has a negative sign. This sign must be ignored because it is only the *magnitude* of the ratio of lengths of small intervals which is relevant. To summarize:

One-to-One Change of Variable for Densities

Let X be a random variable with density $f_X(x)$ on the range (a, b).
Let $Y = g(X)$ where g is either strictly increasing or strictly decreasing on (a, b). The range of Y is then an interval with endpoints $g(a)$ and $g(b)$. And the density of Y on this interval is

$$f_Y(y) = f_X(x)\Big/\left|\frac{dy}{dx}\right| \qquad \text{where} \quad y = g(x)$$

The equation $y = g(x)$ must be solved for x in terms of y, and this value of x substituted into $f_X(x)$ and dy/dx. This will leave an expression for $f_Y(y)$ entirely in terms of y.

Example 3. Square root of an exponential variable (illustrated by Figure 2)

Problem. Let X have the exponential density, $f_X(x) = e^{-x} \qquad (x > 0)$
Find the density of $Y = \sqrt{X}$.

Solution. **Step 1.** Find the range of y: here $0 < x < \infty$, $y = \sqrt{x}$, so $0 < y < \infty$.

Step 2. Check the function is one-to-one by solving for x in terms of y: here $x = y^2$

Step 3. Calculate $\frac{dy}{dx}$: here $\frac{dy}{dx} = \frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}}$

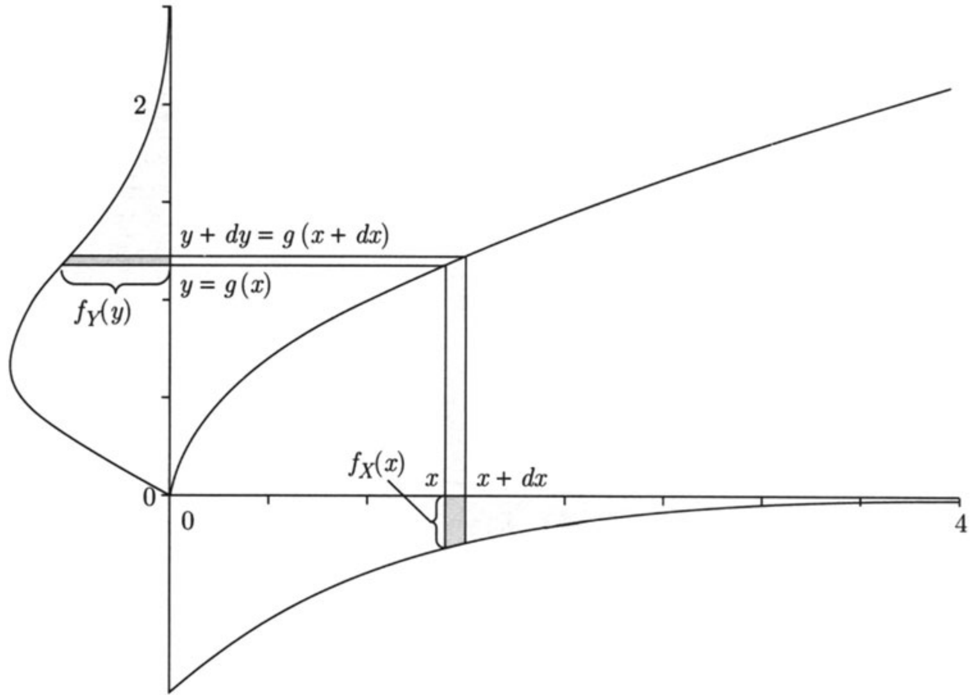
Step 4. Plug density of X and the result of Step 3 into $f_Y(y) = f_X(x)\Big/\left|\frac{dy}{dx}\right|$:

$$f_Y(y) = e^{-x}\Big/\frac{1}{2\sqrt{x}}$$

Step 5. Use result of Step 2 to eliminate x from the right side

$$f_Y(y) = e^{-y^2}\Big/\frac{1}{2\sqrt{y^2}} = 2ye^{-y^2} \qquad (y > 0)$$

FIGURE 2. Change of variable formula for densities. The diagram shows the graph of $y = g(x)$ for the increasing function $g(x) = \sqrt{x},\ x > 0$. Density $f_X(x)$ is graphed upside down below the x-axis. Density $f_Y(y)$ is graphed on the side of the y-axis. The densities are as in Example 3.



Example 4. **Log of uniform.**

Let X have uniform $(0, 1)$ distribution.

Problem 1. Find the distribution of $Y = -\lambda^{-1} \log(X)$, where $\lambda > 0$.

Solution. This follows the steps of the previous example in a slightly different order. Here $y = -\lambda^{-1} \log x$ has

$$\frac{dy}{dx} = -\frac{1}{\lambda x} < 0 \quad \text{for} \quad 0 < x < 1$$

so y decreases from ∞ to 0 as x increases from 0 to 1. The density of Y is then

$$f_Y(y) = f_X(x) \Big/ \left|\frac{dy}{dx}\right| = 1 \Big/ \frac{1}{\lambda x} = \lambda x$$

where $\quad -\lambda^{-1} \log x = y, \quad$ or $\quad x = e^{-\lambda y}, \quad$ so

$$f_Y(y) = \lambda e^{-\lambda y} \qquad (y > 0)$$

Conclusion: Y is exponentially distributed with rate λ.

Discussion. This way of obtaining an exponential variable as a function of a uniform $(0, 1)$ variable is a standard method of simulating exponential variables by computer. The next section shows how any distribution on the line can be obtained as the distribution of a function of a uniform variable.

Problem 2. Find the distribution of $-\lambda^{-1}\log(1-X)$, where $\lambda > 0$.

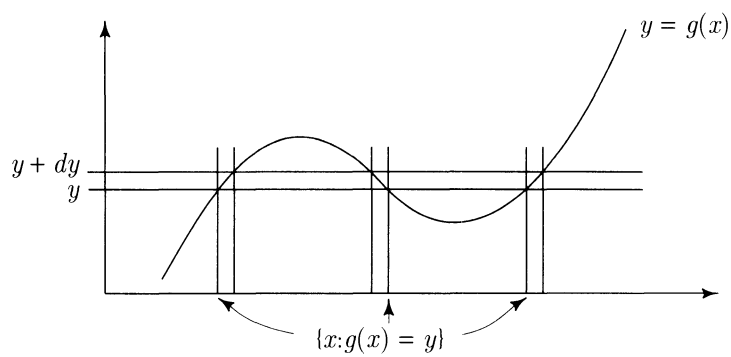
Solution. Clearly the technique used to solve Problem 1 could be repeated. But this is unnecessary. It is intuitively clear (and easy to check) that $X' = 1 - X$ is also a uniform $(0, 1)$ random variable, so $-\lambda^{-1}\log(1-X) = -\lambda^{-1}\log(X')$ has the same distribution as $-\lambda^{-1}\log(X)$. Therefore, $-\lambda^{-1}\log(1-X)$ also has exponential (λ) distribution.

Discussion. The justification of the short argument in the last solution is the change of variable principle. This principle, stated for discrete random variables in Section 3.1, is worth restating here. The principle can often be used as in the last example to eliminate calculations by reducing a change of variables problem to one whose solution is already known:

Change of Variable Principle

If X has the same distribution as Y, then $g(X)$ has the same distribution as $g(Y)$, for any function g.

Many-to-one functions. Suppose the function $y = g(x)$ has a derivative that is zero at only a finite number of points. Now some values of y may come from more than one value of x. Consider $Y = g(X)$ for a random variable X. As shown in the diagram, Y will be in an infinitesimal interval dy near y when X is in one of possibly several infinitesimal intervals dx near points x such that $g(x) = y$.



Now

$$P(Y \in dy) = \sum_{\{x: g(x)=y\}} P(X \in dx)$$

This gives

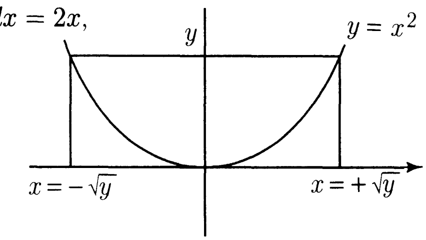
$$f_Y(y) = \sum_{\{x:\, g(x)=y\}} f_X(x) \Big/ \left|\frac{dy}{dx}\right|$$

Example 5. Density of the square of a random variable.

Problem. Suppose X has density $f_X(x)$. Find a formula for the density of $Y = X^2$.

Solution. Here, for $y > 0$, there are two values x such that $x^2 = y$, namely, $x = \sqrt{y}$ and $x = -\sqrt{y}$. Since $dy/dx = 2x$,

$$f_Y(y) = \sum_{\{x=\pm\sqrt{y}\}} f_X(x)/|2x|$$
$$= [f_X(\sqrt{y}) + f_X(-\sqrt{y})]/2\sqrt{y}.$$



Expectation of a function of *X*. If you just want to calculate the expectation of $Y = g(X)$, it is not necessary to calculate the density of Y, and usually simpler not to. For instance, there is no need to use the linear change of variable formula for densities to calculate $E(Y)$ or $SD(Y)$ for $Y = aX + b$. Instead use the simple scaling rules

$$E(aX+b) = aE(X) + b \qquad \text{and} \qquad SD(aX+b) = |a|SD(X)$$

whenever $E(X)$ or $SD(X)$ are defined. More generally, if $Y = g(X)$, where both X and Y have densities, then

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$$

Often the second integral is easier to evaluate than the first. The equality of the two integrals is the density analog of the basic discrete formula for the expectation of a function of X that was derived in Section 4.1. The equality of integrals can also be checked by the calculus technique of substitution

$$y = g(x), \qquad dy = g'(x)dx.$$

Further Examples

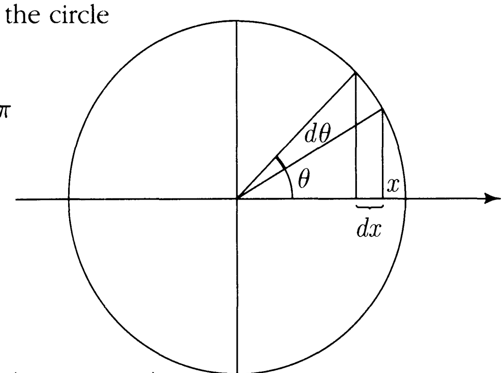
Here are some more geometric problems solved by the same basic technique of finding the probability in an infinitesimal interval by calculus.

Example 6. Projection of a uniform random variable on a circle.

A point is picked uniformly at random from the perimeter of a unit circle.

Problem 1. Find the probability density of X, the x-coordinate of the point.

Solution. From the diagram, since two places on the circle map to one x-value,



$$P(X \in dx) = 2|d\theta|/2\pi = |d\theta|/\pi$$

where $x = \cos\theta,\ \ 0 < \theta < \pi$. So

$$\frac{dx}{d\theta} = -\sin\theta = -\sqrt{1-x^2}$$

$$\frac{d\theta}{dx} = -\frac{1}{\sqrt{1-x^2}}$$

$$\frac{P(X \in dx)}{dx} = \frac{1}{\pi}\left|\frac{d\theta}{dx}\right| = \frac{1}{\pi\sqrt{1-x^2}} \qquad (-1 < x < 1)$$

Problem 2. Find $E(X)$.

Solution. Easily, $E(X) = 0$, since the density of X is symmetric about 0.

Problem 3. Find the probability density of $Y = |X|$, the absolute value of X.

Solution. Since two x values $+y$ and $-y$, with the same probability density, map to any given value of y with $0 < y < 1$, $P(Y \in dy) = 2 \times P(X \in dy)$, and so

$$f_Y(y) = \frac{2}{\pi\sqrt{1-y^2}} \qquad (0 < y < 1)$$

Problem 4. Find $E(Y)$.

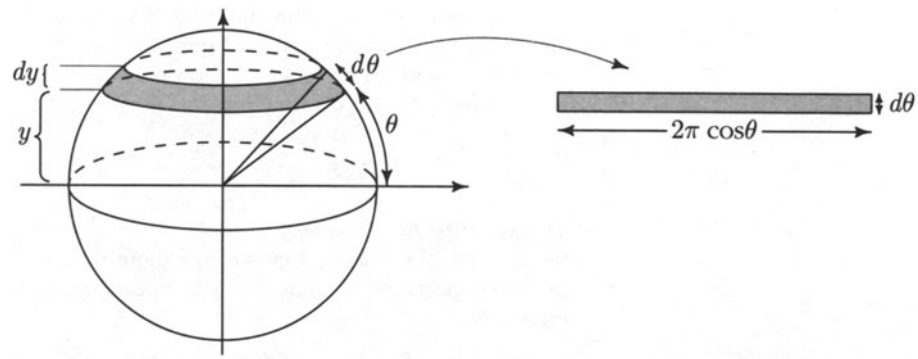
Solution. $$E(Y) = \frac{2}{\pi}\int_0^1 \frac{y}{\sqrt{1-y^2}}dy = -\frac{2}{\pi}\sqrt{1-y^2}\Big|_0^1 = \frac{2}{\pi}$$

Example 7. Projection of a uniform random variable on a sphere.

Let Θ be the latitude, between $-\pi/2$ and $\pi/2$, of a point chosen uniformly at random on the surface of a unit sphere.

Problem 1. Find the probability density of Θ.

Solution. From the diagram:



$$P(\Theta \in d\theta) = \frac{\text{Indicated Area}}{\text{Total Surface Area}} = \frac{2\pi \cos\theta d\theta}{4\pi}$$

$$f_\Theta(\theta) = \frac{\cos\theta}{2} \qquad (-\frac{\pi}{2} < \theta < \frac{\pi}{2})$$

Problem 2. Let Y be the vertical coordinate of the point on the sphere, between -1 and 1. Find the probability density of Y.

Solution. $P(Y \in dy) = P(\Theta \in d\theta)$ with $y = \sin\theta$, which implies that $dy = \cos\theta\, d\theta$ and

$$P(Y \in dy) = P(\Theta \in d\theta) = f_\Theta(\theta)d\theta = \frac{\cos\theta\, d\theta}{2} = \frac{dy}{2} \qquad (-1 < y < 1)$$

Conclusion: Y has uniform $(-1, 1)$ distribution.

Discussion. This calculation shows that the surface area of the sphere between two parallel planes cutting the sphere depends only on the distance between the planes, and not on exactly how they cut the sphere. This fact was discovered by Archimedes. The formula $4\pi r^2$ for the total surface area, used in Problem 1, is a consequence.

Exercises 4.4

1. Suppose X has an exponential (λ) distribution. What is the distribution of cX for a constant $c > 0$?

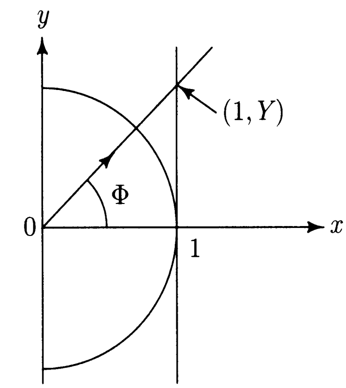
2. **Scaling of gamma distributions.** Show that a random variable T has gamma (r, λ) distribution, if and only if $T = T_1/\lambda$, where T_1 has gamma $(r, 1)$ distribution.

3. Suppose U has uniform $(0, 1)$ distribution. Find the density of U^2.

4. Suppose X has uniform distribution on $(-1, 1)$. Find the density of $Y = X^2$.

5. Suppose X has uniform $[-1, 2]$ distribution. Find the density of X^2.

6. **Cauchy distribution.** Suppose that a particle is fired from the origin in the (x, y)-plane in a straight line in a direction at random angle Φ to the x-axis, and let Y be the y-coordinate of the place where the particle hits the line $\{x = 1\}$. Show that if Φ has uniform $(-\pi/2, \pi/2)$ distribution, then



$$f_Y(y) = \frac{1}{\pi(1 + y^2)}$$

This is called the *Cauchy distribution*. Show that the Cauchy distribution is symmetric about 0, but that the expectation of a Cauchy random variable is undefined.

7. Show that if U has uniform $(0, 1)$ distribution, then $\tan(\pi U - \frac{\pi}{2})$ has the Cauchy distribution, as in Exercise 6.

8. **Arcsine distribution.** Suppose that Y has the Cauchy distribution as in Exercise 6. Let $Z = 1/(1 + Y^2)$.
   a) Show Z has density

$$f_Z(z) = \frac{1}{\pi\sqrt{z(1-z)}} \qquad (0 < z < 1)$$

   b) Show $P(Z \le x) = (2/\pi)\arcsin(\sqrt{x})$ $(0 < x < 1)$.
   c) Find $E(Z)$. d) Find $Var(Z)$.

   [This *arcsine distribution* of Z is the special case $r = s = 1/2$ of the beta(r, s) distribution. This distribution arises naturally in the context of random walks. If $S_n = X_1 + \cdots + X_n$ for X_i with values ± 1 determined by tosses of a fair coin, and L_n is the last time $k \le n$ such that $S_k = 0$, then the limit distribution of L_n/n as $n \to \infty$ is the arcsine distribution. See Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I.]

9. **Weibull distribution.**
   a) Show that if T has the Weibull (λ, α) distribution, with density

$$f(t) = \lambda\alpha t^{\alpha-1}e^{-\lambda t^\alpha} \qquad (t > 0)$$

   where $\lambda > 0$ and $\alpha > 0$, then T^α has an exponential (λ) distribution. (Note the special case when $\alpha = 1$.)
   b) Show that if U is a uniform $(0, 1)$ random variable, then $T = (-\lambda^{-1}\log U)^{\frac{1}{\alpha}}$ has a Weibull (λ, α) distribution.

10. Let Z be a standard normal random variable. Find formulae for the densities of each of the following random variables:
   a) $|Z|$; b) Z^2; c) $1/Z$; d) $1/Z^2$.

11. Explain how the calculations of Example 7 imply the formula $4\pi r^2$ for the surface area of a sphere of radius r.

4.5 Cumulative Distribution Functions

One way to specify a probability distribution on the line is to say how much probability is at or to the left of each point x. In terms of a random variable X with the given distribution, this probability is a function of x,

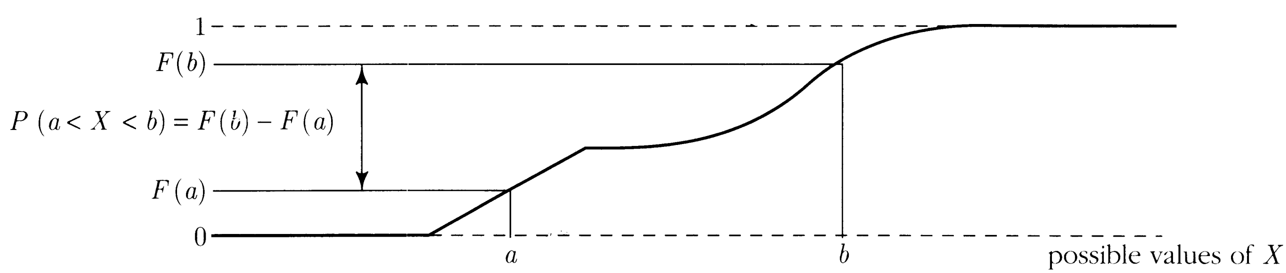
$$F(x) = P(X \leq x)$$

called the *cumulative distribution function (c.d.f.)* of X. For example, the standard normal c.d.f. is the function $F(x) = \Phi(x)$ used in calculations with the normal distribution. But the cumulative distribution function can be defined for any distribution of a random variable X over the line, whether continuous, discrete, or neither.

If you can define or calculate the c.d.f. of X then, by using the rules of probability, you can find the probability of any event determined by X, for example, the probability that X falls in an interval, or the probability that X is an even integer. To clarify terminology, the ***distribution*** of X refers broadly to the assignment of probabilities to all such events determined by X. Technically, this means probabilities defined for a collection of subsets of the line, satisfying the rules of probability, now including the infinite sum rule of Section 3.4. The c.d.f. just gives the probabilities of the intervals $(-\infty, x]$ as a function of the point x.

Interval probabilities. The formula $P(a < X \leq b) = F(b) - F(a)$, a consequence of the difference rule for probabilities, is familiar from the special case of the standard normal c.d.f. Because probabilities must be non-negative, this shows that a c.d.f. $F(x)$ must be a nondecreasing function of x

FIGURE 1. Graph of a continuous c.d.f



The distribution is called ***continuous*** if the c.d.f. is a continuous function. Then it can be shown that

$$P(X = x) = 0 \quad \text{for all } x$$

so it makes no difference in formulae involving the c.d.f. whether inequalities are strict or weak. For example, using the rule of complements,

$$P(X > x) = 1 - F(x) \qquad \text{whatever the distribution of } X$$

$$P(X \geq x) = 1 - F(x) \qquad \text{if the distribution of } X \text{ is continuous}$$

More generally, it can be shown that the c.d.f. determines the probability of every interval, and also the probability of more complicated sets by the addition rule. To summarize:

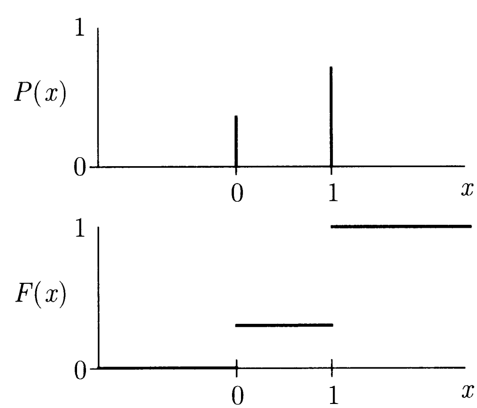
> A probability distribution over the line is completely determined by its c.d.f.

Most distributions of practical interest are either discrete or defined by densities. These two cases will now be discussed in more detail.

Discrete Case

Here is an illustration:

FIGURE 2. Individual probabilities and the c.d.f. for an indicator variable. Consider the c.d.f. of an indicator variable X which is 0 with probability 0.3 and 1 with probability 0.7. The value of $F(x)$ is 0 for $x < 0$ because there is no chance for $X \le x$ for a negative x. The value of $F(x)$ is 0.3 for $0 \le x < 1$, because for such an x the event $(X \le x)$ is the same as the event $(X = 0)$, which has probability 0.3. And the value of $F(x)$ is 1 for $1 \le x < \infty$, because for these x the event $(X \le x)$ is certain. Thus $F(x)$ jumps by $0.3 = P(0)$ at $x = 0$ and by $1 - 0.3 = 0.7 = P(1)$ at $x = 1$.



In general, the c.d.f. of a discrete random variable X looks like a staircase with a rise of $P(x) = P(X = x)$ at each possible value x of X:

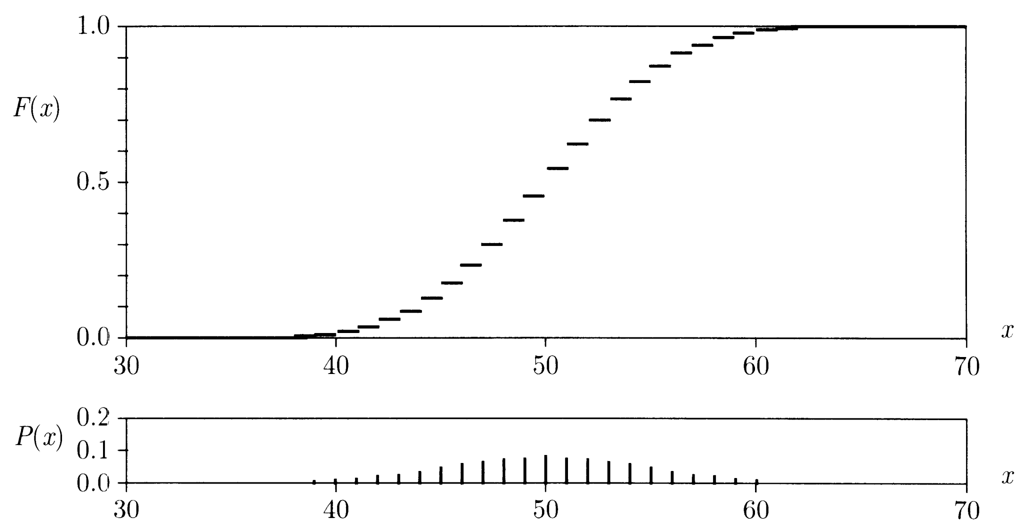
$$F(x) = \sum_{y \le x} P(y)$$

and $P(x)$ is the jump of the c.d.f. at x:

$$P(x) = F(x) - F(x-)$$

where $F(x-) = P(X < x)$ is the limit of values of F approaching x from the left. Figure 3 gives a more interesting example.

FIGURE 3. **The c.d.f. and individual probabilities for the binomial (100, 0.5) distribution.** Here $F(x)$ is the probability of getting x or less heads in 100 fair coin tosses, $P(x)$ is the probability of exactly x heads. The value of $F(x)$ is simply the sum of values $P(y)$ over all integers y less than or equal to x. Each integer x introduces a new term $P(x)$ into the sum. Thus the graph of F jumps by $P(x)$ at each integer x, and is flat between. Put another way, the probability $P(x)$ of an individual value x shows the *difference* between $F(x)$ and $F(x-)$, where $F(x-) = F(x-1)$ is the value of $F(y)$ for any y in the interval $[x-1, x)$.



Density Case

As usual in this case, sums become integrals. So if X has density $f(x)$, then $F(x)$ is the area under the density function to the left of x

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)dy$$

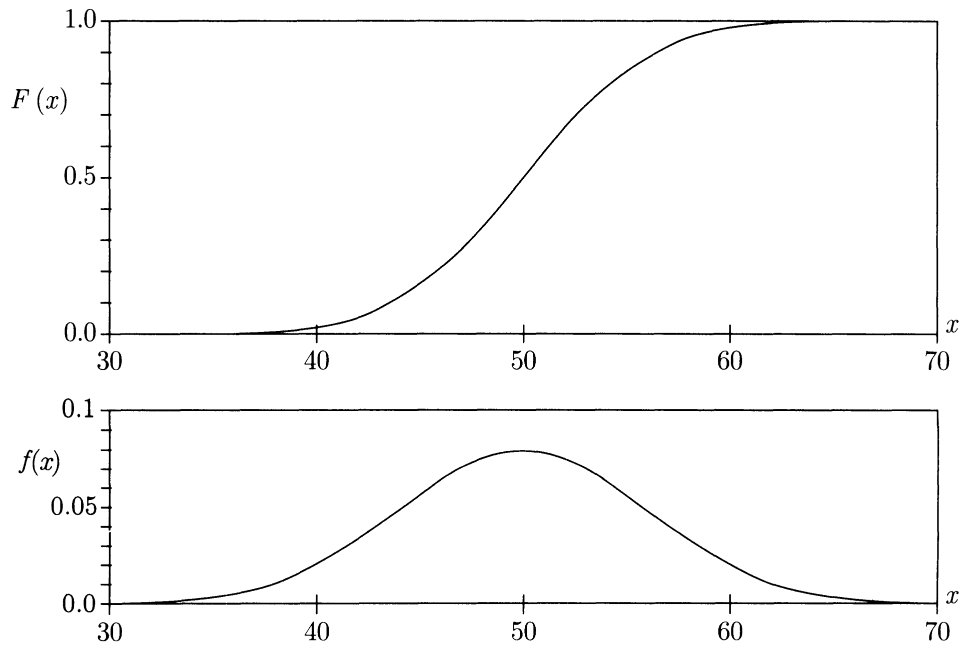
Similarly, discrete differences become derivatives,

$$dF(x) = F(x + dx) - F(x) = P(X \in dx) = f(x)dx$$

$$\text{so} \qquad f(x) = \frac{dF(x)}{dx} = F'(x)$$

That is to say, the density $f(x)$ is the slope at x of the c.d.f. This is an instance of the fundamental theorem of calculus. Conversely, it can be shown that if the c.d.f. is everywhere continuous, and differentiable at all except at perhaps a finite number of points, then the corresponding distribution has density $f(x) = F'(x)$. In this density case, $F(x)$ is a particular choice of an indefinite integral of $f(x)$, namely, the one which vanishes at $-\infty$.

FIGURE 4. **The c.d.f. and density for the normal $(50, 25)$ distribution.** This distribution, with mean 50 and variance 25, is the usual normal approximation to the preceding binomial distribution. Its c.d.f. and density are just scale changes of the standard normal ones plotted in Section 2.2.



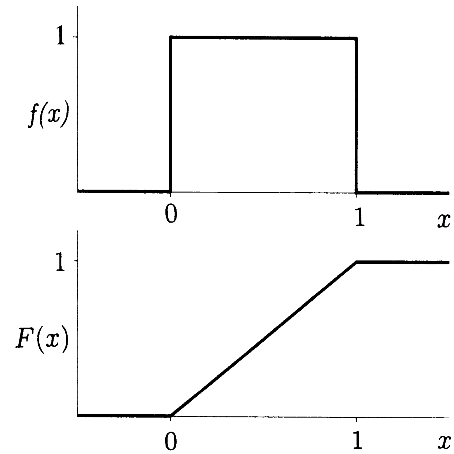
A distribution with a density can be specified by a formula for the density $f(x)$, or by a formula for the c.d.f. $F(x)$. Either of these functions can be obtained from the other by calculus.

You might think that every continuous distribution has a density, but this turns out not to be so. Still, you don't have to worry about continuous distributions without densities in this course. The famous mathematician Poincare thought such distributions "were invented by mathematicians to confound their ancestors". For a nice picture of one, see Mandelbrot's book, *The Fractal Geometry of Nature.*

Example 1. **The uniform (0, 1) distribution.**

The density is

$$f(x) = \begin{cases} 1 & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$



and the c.d.f. is

$$F(x) = \begin{cases} x & \text{for } 0 \le x \le 1 \\ 0 & \text{for } x < 0 \\ 1 & \text{for } x > 1 \end{cases}$$

Here is an application: If U is uniform $(0, 1)$, then so is $X = 2\left|U - \frac{1}{2}\right|$, because

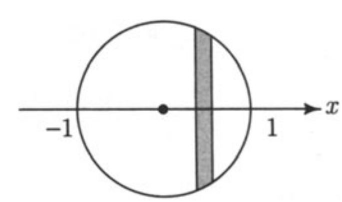
$$\begin{aligned} P(X \le x) &= P\left(2\left|U - \frac{1}{2}\right| \le x\right) \\ &= P\left(\frac{1}{2} - \frac{x}{2} \le U \le \frac{1}{2} + \frac{x}{2}\right) \\ &= F(x) \end{aligned}$$

as defined above. This technique is an alternative to the method of the previous section for calculating the distribution of a function of a random variable.

Example 2. **Uniform on a disc.**

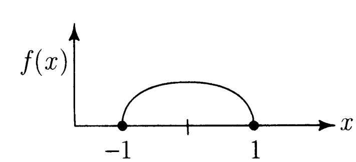
Let (X, Y) be a point chosen uniformly at random from the unit disc $\{(x, y) : x^2 + y^2 \le 1\}$. Calculate the c.d.f. and density function of X.

Solution. It is easiest to find the density function first. Suppose $|x| \le 1$. The event $(X \in dx)$ is shaded in the diagram. For small dx the event in question is approximately a rectangle with height $2\sqrt{1 - x^2}$ and width dx. Dividing by the total area π gives its probability, then dividing by dx gives the density



$$f(x) = \begin{cases} \frac{2}{\pi}\sqrt{1 - x^2} & |x| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

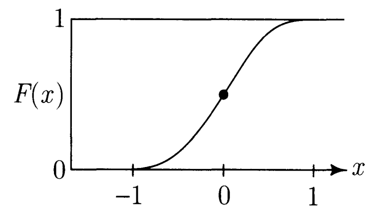
as graphed on the right. This is half an ellipse obtained by rescaling the upper semi-circle. The c.d.f. $F(x)$, which represents the relative area of the disc to the left of x, is now obtained by calculus



$$F(x) = \int_{-1}^{x} f(z)\,dz = \frac{1}{\pi}\int_{-1}^{x} 2\sqrt{1 - z^2}\,dz$$

This is not a very easy integral. Still, because $F(x)$ has derivative $f(x)$ which you know, and $F(x)$ is 0 for $x \leq -1$ and 1 for $x \geq 1$, you should be able to sketch the graph of $F(x)$ and see it must have the shape shown below. Some more calculus (or consulting a table of integrals) gives

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\left[x\sqrt{(1-x^2)} + \arcsin x\right] \qquad (|x| \leq 1).$$



Maximum and Minimum of Independent Random Variables

Cumulative distribution functions make it easy to find the distribution of the maximum and minimum

$$X_{\text{max}} = \max(X_1, \ldots, X_n) \qquad \text{and} \qquad X_{\text{min}} = \min(X_1, \ldots, X_n)$$

of a collection of independent random variables $X_1, X_2, \ldots, X_n$. Let F_i denote the c.d.f. of X_i, $i = 1, \ldots, n$. The c.d.f. of either the maximum or the minimum of the X's can be written in terms of the individual distribution functions F_i, once you notice the following key facts:

For any number x:

(a) X_{max} is less than or equal to x if and only if all the X's are less than or equal to x;

(b) X_{min} is greater than x if and only if all the X's are greater than x.

The c.d.f. of the maximum is then

$$\begin{aligned}
F_{\text{max}}(x) &= P(X_{\text{max}} \leq x) \qquad (-\infty < x < \infty) \qquad \text{by definition} \\
&= P(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x) \qquad \text{by (a)} \\
&= P(X_1 \leq x)P(X_2 \leq x) \cdots P(X_n \leq x) \qquad \text{by independence} \\
&= F_1(x)F_2(x) \cdots F_n(x)
\end{aligned}$$

The c.d.f. of the minimum is

$$\begin{aligned}F_{\min}(x) &= P(X_{\min} \le x) \quad (-\infty < x < \infty)\\ &= 1 - P(X_{\min} > x)\\ &= 1 - P(X_1 > x, X_2 > x, \ldots, X_n > x) \qquad \text{by (b)}\\ &= 1 - (1 - F_1(x))(1 - F_2(x)) \cdots (1 - F_n(x)).\end{aligned}$$

It is best not to try and memorize these formulae. Just remember (a) and (b), and derive the formulae when you need them.

Example 3. **Minimum of independent exponential variables is exponential.**

Let $X_1, X_2, \ldots, X_n$ be independent random variables, and suppose X_i has exponential distribution with rate λ_i, $i = 1, \ldots, n$.

Problem. Find the distribution of $X_{\min}$ the minimum of $X_1, \ldots, X_n$.

Solution. For $i = 1, \ldots, n$, the c.d.f. of X_i is

$$F_i(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda_i x} & \text{if } x \ge 0 \end{cases}$$

Since the X's are non-negative, so is their minimum. So $X_{\min}$ has c.d.f.

$$F_{\min}(x) = 0 \qquad (x < 0)$$

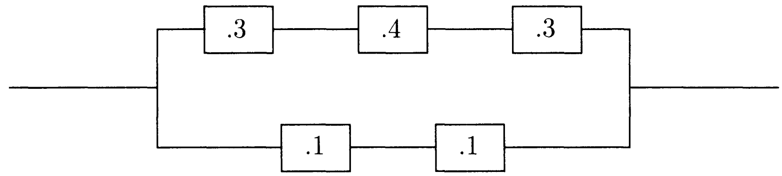
For $x \ge 0$,

$$\begin{aligned}F_{\min}(x) &= 1 - e^{-\lambda_1 x} e^{-\lambda_2 x} \cdots e^{-\lambda_n x}\\ &= 1 - e^{-(\lambda_1 + \lambda_2 + \cdots + \lambda_n)x}\end{aligned}$$

This is the c.d.f. of the exponential distribution with rate $\lambda_1 + \lambda_2 + \cdots + \lambda_n$. So the minimum of independent exponential variables with rates λ_i is simply a new exponential variable with rate the sum of the rates λ_i.

Example 4. **Expected lifetime of a circuit.**

An electrical circuit consists of five components, connected as in the following diagram. The lifetimes of the components, measured in days, have independent exponential distributions with rates indicated in the diagram.



Problem. What is the expected lifetime of the circuit?

Solution. We want $E(L)$, where L denotes the lifetime of the circuit. Let L_{top} and L_{bottom} denote the lifetimes of the top and bottom parts of the circuit. Then L_{top} and L_{bottom} are independent, and

$$L = \max(L_{\text{top}}, L_{\text{bottom}})$$

since the top and bottom parts are linked in parallel.

Now L_{top} is the minimum of three independent exponential lifetimes, since the top consists of three components linked in series. By Example 3, L_{top} has exponential distribution with rate $0.3 + 0.4 + 0.3 = 1$. So the top is expected to last about 1 day. By a similar argument, L_{bottom} has exponential (0.2) distribution, so the bottom is expected to last about $1/0.2 = 5$ days.

Since L is the maximum of L_{top} and L_{bottom}, its c.d.f. is

$$F_L(x) = \begin{cases} 0 & x < 0 \\ (1 - e^{-x})(1 - e^{-0.2x}) & x \geq 0 \end{cases}$$

Since L is a positive random variable

$$E(L) = \int_0^\infty (1 - F_L(x))dx$$

(See Exercise 9 .) For $x \geq 0$,

$$F_L(x) = 1 - e^{-x} - e^{-0.2x} + e^{-1.2x}$$
$$1 - F_L(x) = e^{-x} + e^{-0.2x} - e^{-1.2x}$$

so

$$E(L) = \int_0^\infty (e^{-x} + e^{-0.2x} - e^{-1.2x})dx = 1 + (1/0.2) - (1/1.2) = 5.17$$

So the circuit is expected to last about 5.17 days.

Note. Once you have the c.d.f. of L, you can, of course, compute its expectation by first differentiating to find the density, then using the density to find the expectation by integration. But that involves more work than the method used here.

Suppose now that in addition to being independent, the X's are continuous random variables with the same density. For example, the X's could be a sequence of random numbers produced by a uniform random number generator. Let f denote the common density function of the X's, and F the common c.d.f. The maximum $X_{\max}$ and minimum $X_{\min}$ are also continuous random variables, whose densities can be obtained by differentiating their c.d.f.'s

$$F_{\max}(x) = (F(x))^n \qquad (-\infty < x < \infty)$$

$$f_{\max}(x) = \frac{d}{dx}(F(x))^n = n\,(F(x))^{n-1} f(x) \qquad (-\infty < x < \infty)$$

by the chain rule of calculus. Similarly,

$$F_{\min}(x) = 1 - (1 - F(x))^n \qquad (-\infty < x < \infty)$$

$$f_{\min}(x) = n\,(1 - F(x))^{n-1} f(x) \qquad (-\infty < x < \infty)$$

These densities can also be found more directly by a differential calculation explained in the next section.

Percentiles and the Inverse Distribution Function

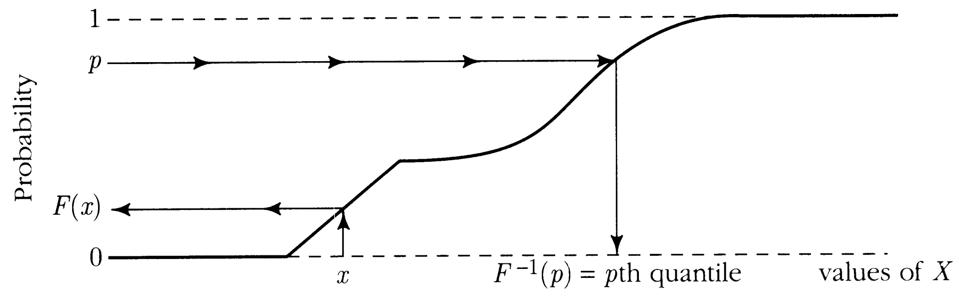
Given a distribution of X and a value x, the c.d.f. $F(x)$ gives the probability that X is less than or equal to x. Often the question gets turned around. For instance: For what value of x is there probability $1/2$ that X is less than or equal to x? Such an x is a *median* of the distribution. More generally, given a probability p, for what x is $P(X \leq x) = p$? By definition of the c.d.f. this x must solve the equation

$$F(x) = p$$

In the case of $F(x)$ given by a formula, the formula can usually be rearranged to express x in terms of p. In general, assuming this equation has a unique solution, as it does for most continuous distributions of interest and $0 < p < 1$, the solution of this equation defines the *inverse c.d.f.*

$$x = F^{-1}(p)$$

FIGURE 5. Relation between a c.d.f. and its inverse.



See Figure 5. This point x, such that $P(X \leq x) = p$, is called the pth ***quantile*** of the distribution of X. This term is a generalization of the more common ***quartile, decile,*** and ***percentile*** in case p is expressed as a multiple of 1/4, 1/10, or 1/100.

Example 5. Finding percentiles.

Problem 1. For the exponential (λ) distribution, find a formula for the pth quantile, $0 < p < 1$.

Solution. Since the c.d.f. is $F(x) = 1 - e^{-\lambda x}$ for $x > 0$, the required point x is found from

$$1 - e^{-\lambda x} = p \qquad \text{so} \qquad x = -\frac{1}{\lambda}\log(1-p)$$

Problem 2. Find the 75th percentile point of the standard normal distribution.

Solution. This is $\Phi^{-1}(0.75)$ where Φ is the standard normal c.d.f. Just as there is no simple formula for Φ, there is none for Φ^{-1}. But numerical values of Φ^{-1} are easily found by backwards lookup in the table of values of Φ. Inspection of the table gives $\Phi(0.67) = 0.7486$ and $\Phi(0.68) = 0.7517$, so $\Phi^{-1}(0.75) \approx 0.675$.

Simulation via Inverse Distribution Function

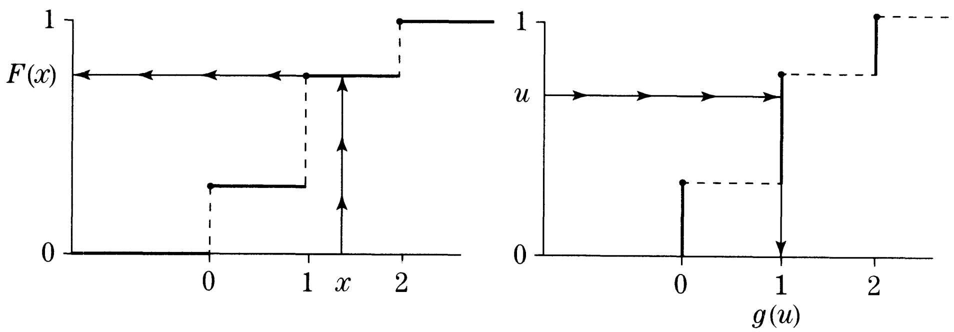
Given a distribution on the line, how can you create random variables with this distribution? This problem arises in computer simulation of random variables. The random number generator on a computer provides a sequence of numbers between 0 and 1, say $U_1, U_2, \ldots$, which behaves in most respects like a sequence of independent uniform $(0, 1)$ random variables. For example, the long-run proportion of values U_i in any subinterval of $[0, 1]$ will be very close to the length of the subinterval. How can these variables be transformed into a sequence simulating independent random variables with some other distribution? The problem is to find a function g such that if U has uniform $(0, 1)$ distribution, then $X = g(U)$ has a prescribed c.d.f., say $F(x)$:

$$P(g(U) \leq x) = F(x) \quad \text{for all } x$$

There are many ways to solve this problem by tricks depending on the desired distribution. Which method is best depends on considerations such as computational efficiency, not discussed here. One method will now be described which works no matter what the required distribution. Here is a simple example to illustrate the method.

Example 6. Simulating a binomial (2, 0.5) random variable.

The left graph shows the required c.d.f. The right graph shows a function g from $(0,1)$ to $\{0,1,2\}$. This graph should be read on its side as a kind of inversion of the graph of the c.d.f. The staircase is the same in both graphs. Imagine U picked at random from the vertical unit interval. Then $g(U) \in \{0,1,2\}$ has the required distribution.



In detail, as it would be programmed on a computer, the rule for getting from the uniform $(0,1)$ variable U to the binomial $(2,0.5)$ variable $g(U)$ is

$$\begin{array}{lll} \text{if } 0 \le U \le 0.25 & \text{then} & g(U) = 0 \\ \text{if } 0.25 < U \le 0.75 & \text{then} & g(U) = 1 \\ \text{if } 0.75 < U \le 1.0 & \text{then} & g(U) = 2 \end{array}$$

This g does the job because by construction the intervals on which g takes the values 0, 1, and 2 have lengths 0.25, 0.5, and 0.25, respectively, as required by the binomial $(2,0.5)$ distribution.

Simulation of a discrete distribution. The method of the previous example generalizes easily to any discrete distribution. For example, to get a random variable with discrete distribution on $1,2,\ldots$ defined by probabilities $p_1, p_2, \ldots$ define

$$g(u) = k \quad \text{if} \quad p_1 + \cdots + p_{k-1} < u \le p_1 + \cdots + p_{k-1} + p_k$$

Then if U has uniform $(0,1)$ distribution

$$P(g(U) = k) = P(p_1 + \cdots + p_{k-1} < U \le p_1 + \cdots + p_{k-1} + p_k) = p_k$$

since this is the length of the interval of U-values that make $g(U) = k$. This means $g(U)$ has the given discrete distribution.

The inverse distribution function. The function $g(u)$ defined in the discrete case above is always a kind of inverse of the c.d.f. $F(x)$, in the sense that

$$g(F(x)) = x \quad \text{for all possible values } x$$

Check this inverse relation in the example above for $x = 0, 1, 2$. Given any c.d.f. F, not necessarily continuous or strictly increasing, a function g satisfying the above inverse relation can be defined. Because of the inverse relation, $g(u)$ is usually denoted $F^{-1}(u)$, and called the *inverse c.d.f.* In general, the inverse c.d.f. $F^{-1}(u)$ can be defined as the least value x such that $F(x) \geq u$. This function has the following important property:

Inverse c.d.f. Applied to Standard Uniform

For any cumulative distribution function F, with inverse function F^{-1}, if U has uniform $(0, 1)$ distribution, then $F^{-1}(U)$ has c.d.f. F.

To restate this result more intuitively, if you pick a percentage uniformly at random on $(0, 100)$, then take that percentile point in a distribution, you get a random variable with that distribution.

Proof. The discrete case has already been treated. The continuous case is more interesting. Assume, for simplicity, that $F(x)$ is a continuous and strictly increasing function of x. Then $F^{-1}(u)$ is the usual inverse function of $F(x)$, as discussed earlier, and

$$w \leq x \iff F(w) \leq F(x)$$

The events $(F^{-1}(U) \leq x)$ and $(F(F^{-1}(U)) \leq F(x))$ are therefore identical. But since $F(F^{-1}(u)) = u$ for every u in $(0, 1)$, by definition of the inverse function, we can calculate

$$\begin{aligned} P(F^{-1}(U) \leq x) &= P(F(F^{-1}(U)) \leq F(x)) \\ &= P(U \leq F(x)) \\ &= F(x) \quad \text{from the c.d.f. of } U \end{aligned}$$

Thus the random variable $F^{-1}(U)$ has c.d.f. F. □

The method of generating random variables via F^{-1} is efficient computationally in simulations only if F^{-1} turns out to be a fairly simple function to compute, as it is for the uniform distribution on (c, d) for any $c < d$, or the exponential distribution. But F^{-1} is laborious to compute for the normal distribution. In this case it is quicker and nearly as accurate to approximate using the central limit theorem, using, for instance, a standardized sum of 12 independent uniform $(0, 1)$ variables. See also Exercise 5.3.13 for another method of generating normal variables from uniform ones.

Exercises 4.5

1. For the exponential (λ) distribution:

a) Show the c.d.f. is $F(x) = 1 - e^{-\lambda x}$ for $x \geq 0$. b) Sketch this c.d.f. for $\lambda = 1$.

2. Find and sketch the cumulative distribution functions of:

a) the binomial $(3, 1/2)$ distribution;

b) the geometric $(1/2)$ distribution on $\{1, 2, \ldots\}$.

3. Let (X, Y) be as in Example 2.

a) Find f_Y and F_Y. [*Hint*: No calculations required!]

b) Let $R = \sqrt{X^2 + Y^2}$. Sketch the event $\{R \leq r\}$ as a subset of the circle. Deduce a formula for the c.d.f. of R, and check by differentiating that you get the same density for R as in Example 4.1.2.

4. Let X be a random variable with c.d.f. $F(x)$. Find the c.d.f. of $aX + b$ first for $a > 0$, then for $a < 0$.

5. Find the c.d.f. of X with density function $f_X(x) = \frac{1}{2}e^{-|x|}$ $(-\infty < x < \infty)$.

6. Let X be a random variable with c.d.f. $F(x) = x^3$ for $0 \leq x \leq 1$. Find:

a) $P(X \geq \frac{1}{2})$; b) the density function $f(x)$; c) $E(X)$.

d) Let Y_1, Y_2, Y_3 be three points chosen independently and uniformly on the unit interval, and let X be the rightmost point. Show that X has the distribution described above.

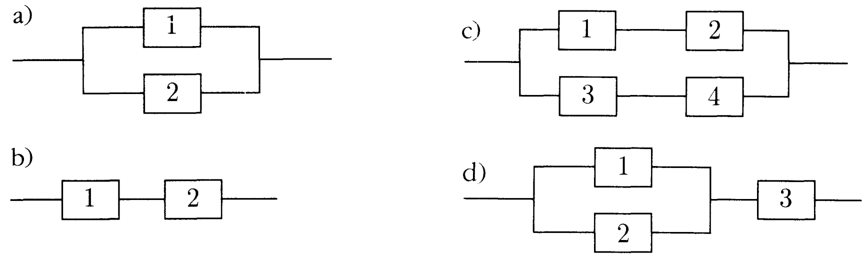
7. Let T have the exponential distribution with parameter λ, and let $Y = \sqrt{T}$.

a) Find the density of Y.

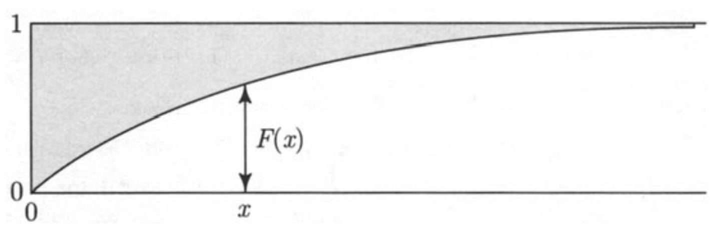
b) Find the expectation of Y, correct to two decimal places, for $\lambda = 3$.

c) A random number generator produces uniform [0, 1] random numbers. How could you use these to generate random numbers which have the distribution of Y?

8. Components in the following series-parallel systems have independent exponentially distributed lifetimes. Component i has mean lifetime μ_i. In each case, find a formula for the probability that the system operates for at least t units of time, and sketch the graph of this function of t in case $\mu_i = i$ for each i.



9. **Expectation from c.d.f.** Let X be a positive random variable, with c.d.f. F, as in the following diagram for example:



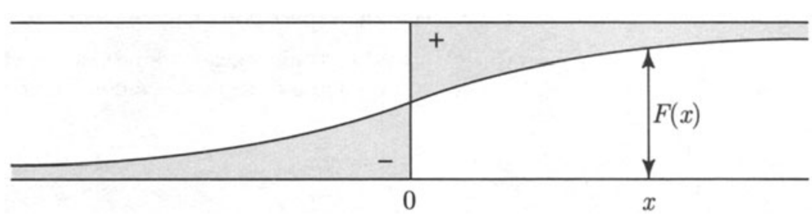
a) Show, using the representation $X = F^{-1}(U)$ for a uniform $[0, 1]$ random variable U, that $E(X)$ can be interpreted as the shaded area above the c.d.f. of X, both for X with a density, and for discrete X. Deduce that

$$E(X) = \int_0^\infty [1 - F(x)]\, dx = \int_0^\infty P(X > x)\, dx$$

b) Deduce that if X has possible values $0, 1, 2, \ldots$, then $E(X) = \sum_{n=1}^\infty P(X \geq n)$.

c) Use these formulae to rederive the means of the exponential and geometric distributions.

d) Show that for a random variable X with both positive and negative values (either discrete or with a density), $E(X) = E(X_+) - E(X_-)$ where $X_+ = XI(X > 0)$, and $X_- = (-X)I(X < 0)$, so $E(X)$ is area $(+)$ minus area $(-)$ defined in terms of the c.d.f. as indicated below:

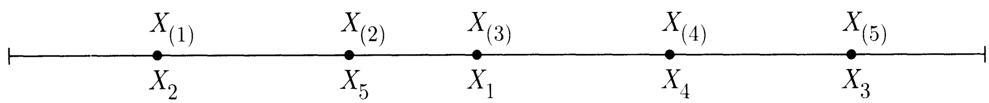


4.6 Order Statistics (Optional)

Let $X_1, X_2, \ldots, X_n$ be random variables. Let $X_{(1)}$ denote the smallest of the X's, $X_{(2)}$ the next smallest, and so on, so that

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$$

This relabeling of the X's corresponds to arranging them in increasing order, as shown below, for one particular ordering of five values $X_1, \ldots, X_5$.



Notice that

$$X_{(1)} = \min(X_1, \ldots, X_n)$$

$$X_{(n)} = \max(X_1, \ldots, X_n)$$

In general, $X_{(k)}$ is called the kth *order statistic* of $X_1, \ldots, X_n$.

This section deals with properties of order statistics of independent and identically distributed random variables. Beta distributions appear as the distributions of order statistics of independent uniform $(0, 1)$ random variables.

Let $X_1, X_2, \ldots, X_n$ be independent random variables, all with the same density function f and cumulative distribution function F. For example, the X's could be a sequence of random numbers produced by a uniform random number generator. The object is to find a formula for the density of the kth order statistic $X_{(k)}$. This has been done already in Section 4.5 in the case of the maximum $X_{(n)}$ and minimum $X_{(1)}$ by first finding the c.d.f., then differentiating. But here is another argument in these special cases which generalizes more easily. First of all, it can be shown that in a sequence $X_1, \ldots, X_n$ of independent continuous random variables, all n values are distinct with probability 1. Taking this for granted, here is a calculation of the density of the maximum $X_{(n)}$

$$\begin{aligned}
f_{(n)}(x)dx &= P(X_{(n)} \in dx) \\
&= P(\text{one of the } X\text{'s} \in dx, \text{ all others} < x) \\
&= P(X_1 \in dx, \text{ all others} < x) + P(X_2 \in dx, \text{ all others} < x) \\
&\qquad\qquad + \cdots + P(X_n \in dx, \text{ all others} < x) \\
&= nP(X_1 \in dx, \text{ all others} < x) \qquad \text{by symmetry} \\
&= nP(X_1 \in dx)P(\text{all others} < x) \qquad \text{by independence} \\
&= nf(x)dx\,(F(x))^{n-1}
\end{aligned}$$

in agreement with the previous calculation in Section 4.5. Similarly,

$$\begin{aligned} f_{(1)}(x)\,dx &= P(X_{(1)} \in dx) \\ &= P(\text{one of the } X\text{'s} \in dx, \text{ all others} > x) \\ &= nf(x)\,dx\,(1 - F(x))^{n-1} \end{aligned}$$

The same method can be used to derive a formula for the density of the kth order statistic of $X_1, \ldots, X_n$. Recall that $X_{(k)}$ is the kth smallest of $X_1, \ldots, X_n$. The density $f_{(k)}(x)$ of $X_{(k)}$ is found as follows. For $-\infty < x < \infty$

$$\begin{aligned} f_{(k)}(x)dx &= P(X_{(k)} \in dx) \\ &= P(\text{one of the } X\text{'s} \in dx, \text{ exactly } k-1 \text{ of the others} < x) \\ &= nP(X_1 \in dx, \text{ exactly } k-1 \text{ of the others} < x) \\ &= nP(X_1 \in dx)P(\text{exactly } k-1 \text{ of the others} < x) \\ &= nf(x)dx\binom{n-1}{k-1}(F(x))^{k-1}(1-F(x))^{n-k} \end{aligned}$$

using the binomial formula. To summarize:

Density of the kth Order Statistic

Let $X_{(k)}$ denote the kth order statistic of $X_1, X_2, \ldots, X_n$, where $X_1, \ldots, X_n$ are independent, identically distributed random variables with common density f and c.d.f. F. The density of $X_{(k)}$ is given by

$$f_{(k)}(x) = nf(x)\binom{n-1}{k-1}(F(x))^{k-1}(1-F(x))^{n-k} \qquad (-\infty < x < \infty)$$

It is best not to memorize the formula, but to remember how it is derived.

Order Statistics of Uniform Random Variables

Let $X_1, \ldots, X_n$ be independent random variables each with uniform distribution on $(0, 1)$. The common density of the X's is

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Their common c.d.f. is

$$F(x) = \begin{cases} 0 & x < 0 \\ x & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

By the boxed formula above, the density of the kth order statistic of the n uniform random variables is

$$f_{(k)}(x) = \begin{cases} n\binom{n-1}{k-1}x^{k-1}(1-x)^{n-k} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Some of these densities are graphed in Figure 1 on the next page.

Notice how as n increases, the density for the minimum gets more concentrated near 0, the density for the maximum gets more concentrated near 1, and the density for the middle value of the X's gets more concentrated near 1/2. This is what you would expect intuitively.

Notice also the functional form of the density: a constant, times x raised to a power, times $1-x$ raised to a power. This simple form for a density on $(0,1)$ appears in many settings. Here is a general definition:

Beta (*r*,*s*) Distribution

For $r, s > 0$, the *beta* (r, s) distribution on $(0,1)$ is defined by the density

$$\frac{1}{B(r,s)}x^{r-1}(1-x)^{s-1} \qquad (0 < x < 1)$$

where

$$B(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}dx$$

is the normalizing constant which makes the density integrate to 1.

Viewed as a function of r and s, $B(r,s)$ is called the *beta function*.

A comparison of the last two boxes shows the following:

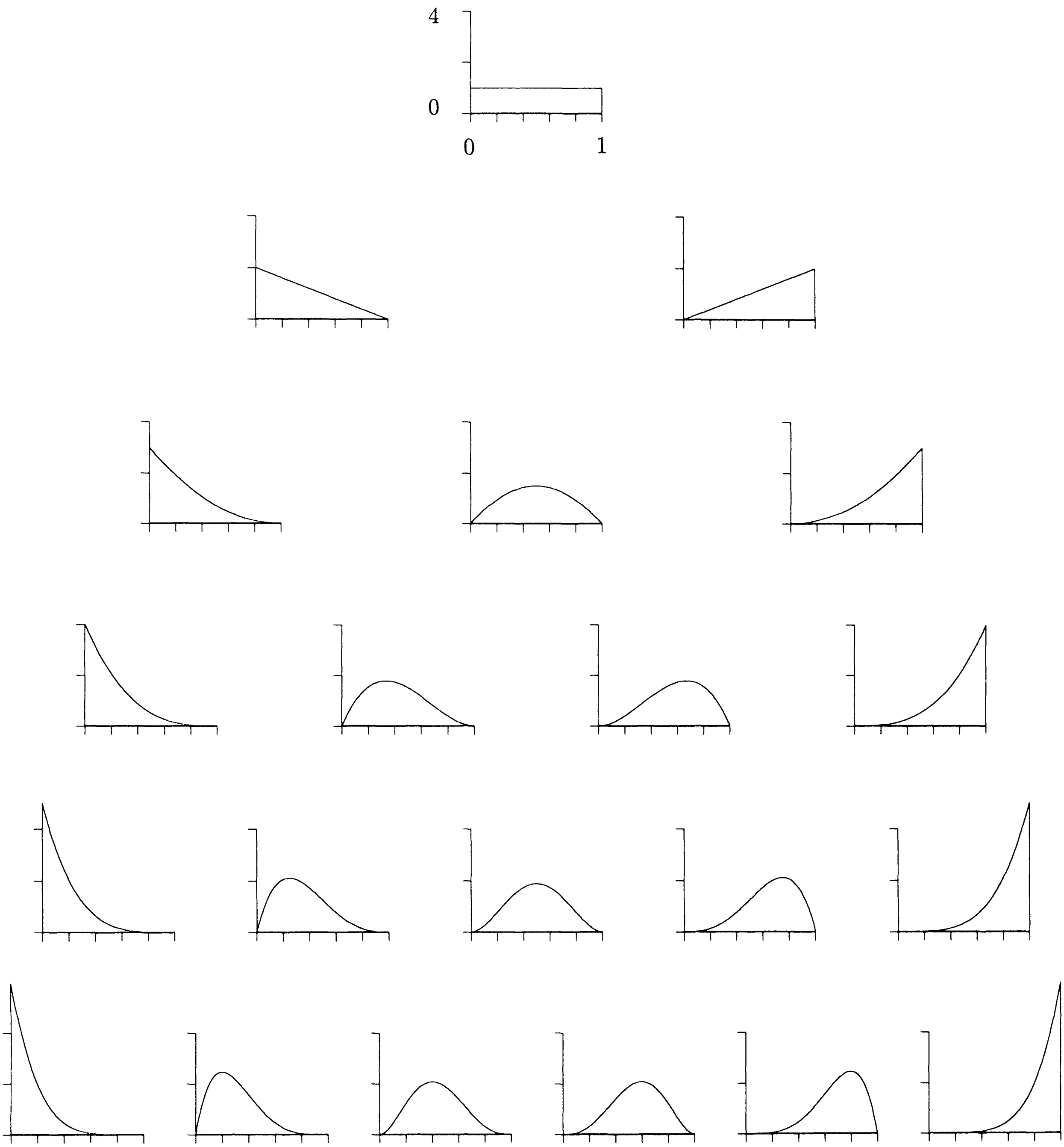
Beta Distribution of Uniform Order Statistics

The kth order statistic of n independent uniform $(0,1)$ random variables has beta $(k, n-k+1)$ distribution.

A nice corollary of the formula for the density of $X_{(k)}$ derived above is that for integers r and s, the beta function $B(r,s)$ is evaluated. Since $f_{(k)}$ is a density it must integrate to 1 over $[0,1]$. So

$$\int_0^1 x^{k-1}(1-x)^{n-k}dx = \frac{1}{n\binom{n-1}{k-1}} = \frac{(k-1)!(n-k)!}{n!}$$

FIGURE 1. **Densities of order statistics of independent uniform variables.** For $n = 1, 2, \ldots, 6$ and $k = 1, 2, \ldots, n$, the density of the kth order statistic of n independent uniform $(0, 1)$ random variables, which is the beta density with parameters k and $n - k + 1$, is plotted as the kth graph in the nth row of the diagram.



Substitute $r = k$ and $s = n - k + 1$ and recall that $\Gamma(r) = (r-1)!$ for positive integers r to get the following result for integers r and s:

Evaluation of the Beta Integral

For positive r and s

$$B(r, s) = \int_0^1 x^{r-1}(1-x)^{s-1}dx = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$$

The beta (r, s) distribution is defined, and the above evaluation of the beta integral is valid, for all positive r and s, not necessarily integers. See Section 5.4, especially Exercise 5.4.19 for a proof of this and explanation of the connection between the beta and gamma distributions.

Moments of the beta distribution. The expectation and variance of a beta random variable with integer parameters are now easy to calculate. If X has beta distribution with positive integer parameters r and s,

$$\begin{aligned}
E(X) &= \int_0^1 x \cdot \frac{1}{B(r,s)} x^{r-1}(1-x)^{s-1}dx \\
&= \frac{1}{B(r,s)} \int_0^1 x^{(r+1)-1}(1-x)^{s-1}dx \\
&= \frac{B(r+1,s)}{B(r,s)} \\
&= \frac{r!(s-1)!}{(r+s)!} \cdot \frac{(r+s-1)!}{(r-1)!(s-1)!} \\
&= \frac{r}{r+s}
\end{aligned}$$

$E(X^2)$ can be calculated in the same way, and used to find a formula for the variance of X. This is left as an exercise.

The kth order statistic of n independent uniform $(0,1)$ random variables has beta distribution with parameters k and $n - k + 1$, so

$$E(X_{(k)}) = \frac{k}{n+1}$$

Thus the smallest of four uniform random numbers is expected to be around $1/5$, the next smallest around $2/5$, the third smallest around $3/5$, and the largest around

4/5. In other words, if you think of picking four points at random from $[0, 1]$ as cutting the interval into five pieces

all the pieces are expected to have the same length. In fact, more is true: It can be shown that when an interval is split at random like this by any number of independent uniform random points, the length of each piece has the same beta distribution as the length of the first piece. See Chapter 6 Review Exercise 32.

Exercises 4.6

1. Four people agree to meet at a cafe at noon. Suppose each person arrives at a time normally distributed with mean 12 noon and SD 5 minutes, independently of all the others.

a) What is the chance that the first person to arrive at the cafe gets there before 11:50?

b) What is the chance that some of the four have still not arrived at 12:15?

c) Approximately what is the chance that the second person to arrive gets there within ten seconds of noon?

2. Let X have beta (r, s) distribution.

a) Find $E(X^2)$, and use the formula for $E(X)$ given in this section to find $Var(X)$.

b) Find a formula for $E(X^k)$, for integers $k \geq 1$.

3. Let $U_{(1)}, \ldots, U_{(n)}$ be the values of n independent uniform $(0, 1)$ variables arranged in increasing order. Let $0 \leq x < y \leq 1$. Find simple formulae for:

a) $P(U_{(1)} > x \text{ and } U_{(n)} < y)$; b) $P(U_{(1)} > x \text{ and } U_{(n)} > y)$;

c) $P(U_{(1)} < x \text{ and } U_{(n)} < y)$; d) $P(U_{(1)} < x \text{ and } U_{(n)} > y)$;

e) $P(U_{(k)} < x \text{ and } U_{(k+1)} > y)$ for $1 \leq k \leq n - 1$;

f) $P(U_{(k)} < x \text{ and } U_{(k+2)} > y)$ for $1 \leq k \leq n - 2$.

4. Let $X = \min(S, T)$ and $Y = \max(S, T)$ for independent random variables S and T with a common density f. Let Z denote the indicator of the event $S < T$.

a) What is the distribution of Z?

b) Are X and Z independent? Are Y and Z independent? Are (X, Y) and Z independent?

c) How can these conclusions be extended to the order statistics of three or more independent random variables with the same distribution?

5. C.d.f. of the beta distribution for integer parameters.

a) Let $X_1, X_2, \ldots, X_n$ be independent uniform $(0, 1)$ random variables, and let $X_{(k)}$ be the kth order statistic of the X's. Find the c.d.f. of $X_{(k)}$ by expressing the event $X_{(k)} \leq x$ in terms of the number of X_i that are $\leq x$.

b) Use a) to show that for positive integers r and s, the c.d.f. of the beta (r, s) distribution is given by

$$\sum_{i=r}^{r+s-1} \binom{r+s-1}{i} x^i (1-x)^{r+s-i-1} \qquad (0 \leq x \leq 1)$$

c) Expand the power of $(1-x)$ in the beta density using the binomial theorem, and then integrate, to obtain the following alternative formula for the c.d.f. of the beta (r, s) distribution:

$$\frac{x^r}{B(r,s)} \sum_{i=0}^{s-1} \binom{s-1}{i} (-1)^i x^i/(r+i) \qquad (0 \leq x \leq 1)$$

[Equating the results of these two calculations yields an algebraic identity that is not easy to prove directly.]

Continuous Distributions: Summary

For a random variable X with probability density $f(x)$:

Differential formula: $P(X \in dx) = f(x)dx$.

Integral formula: $P(a \le X \le b) = \int_a^b f(x)dx$.

Interpretation: $f(x)$ is the chance per unit length for values of X near x.

Properties of $f(x)$: Non-negative, total integral 1.

Expectation of a function g of X

$$E\left((g(X)\right) = \int_{-\infty}^{\infty} g(x)f(x)dx \quad \text{provided} \quad \int_{-\infty}^{\infty} |g(x)|f(x)dx < \infty$$

Uniform, exponential, normal distributions: See Distribution Summaries.

Hazard rates

Let T be a positive random variable with probability density f. Think of T as the lifetime of a component. The *hazard rate* (or *failure rate*, or *death rate*) function $\lambda(t)$ is the probability per unit time that the component will fail just after time t, given that it has survived up to time t

$$P(T \in dt | T > t) = \lambda(t)\, dt$$

For relations between λ and the density, survival function, etc., of T, see the table "Random Lifetimes" on page 297.

Expectation from the survival function: For a non-negative random variable T,

$$E(T) = \int_0^{\infty} G(t)dt$$

where $G(t) = P(T > t)$ is the survival function of T.

One-to-one change of variable for densities

Let X be a random variable with density $f_X(x)$ in the range (a, b).

Let $Y = g(X)$ where g is either strictly increasing or strictly decreasing on (a, b). The range of Y is then an interval with endpoints $g(a)$ and $g(b)$. And the density of Y on this interval is

$$f_Y(y) = f_X(x) \Big/ \left|\frac{dy}{dx}\right| \qquad \text{at} \quad x = g^{-1}(y)$$

where dy/dx is the derivative of $y = g(x)$, and g^{-1} is the inverse function of g.

Linear change of variable for densities:

$$f_{aX+b}(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Change of variable principle: If X has the same distribution as Y, then $g(X)$ has the same distribution as $g(Y)$, for any function g.

Cumulative distribution function of X: $F(x) = P(X \leq x)$

If the distribution has a *density* $f(x)$, then

$$F(x) = \int_{-\infty}^{x} f(y)\, dy$$

and the density function at x is the derivative of the c.d.f. at x

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

provided $F'(x)$ is continuous at x.

Percentiles

The kth percentile point of a distribution is the value x such that $F(x) = k/100$, written $x = F^{-1}(k/100)$, where F^{-1} is the *inverse c.d.f.*

Transformation by the inverse c.d.f.

If U has uniform $(0,1)$ distribution, then $F^{-1}(U)$ has c.d.f. F.

Order statistics

If $X_1, \ldots, X_n$ are independent with common density f and c.d.f. F, then the kth *order statistic* $X_{(k)}$, that is, the kth smallest value among the $X_1, \ldots, X_n$, has density

$$f_{X_{(k)}}(x) = nf(x)\binom{n-1}{k-1}(F(x))^{k-1}(1-F(x))^{n-k}$$

If the X_i have uniform $(0,1)$ distribution, then $X_{(k)}$ has beta $(k, n-k+1)$ distribution.

Review Exercises

1. Suppose atoms of a given kind have an exponentially distributed lifetime with rate λ. Let X_t be the number of atoms still present at time $t \geq 0$, starting from $X_0 = n$. Find formulae in terms of n, t, and λ for a) $E(X_t)$; b) $Var(X_t)$.

2. Find the constant c which makes the function $f(x) = c(x+x^2)$ for $0 < x < 1$ the density of a probability distribution on $(0, 1)$. Find the corresponding c.d.f. $F(x)$. Sketch the graphs of $f(x)$ and $F(x)$. Find the expectation μ and standard deviation σ of a random variable X with this distribution. Mark the points μ, $\mu + \sigma$ on your graphs.

3. Let Y_1, Y_2, and Y_3 be three points chosen independently and uniformly from $(0, 1)$, and let X be the rightmost (largest) point. Find the c.d.f., density function, and expectation of X.

4. Let X be a random variable with density $f(x) = 0.5e^{-|x|}$ $(-\infty < x < \infty)$. Find:

a) $P(X < 1)$; b) $E(X)$ and $SD(X)$; c) the c.d.f. of X^2.

5. An ambulance station, 30 miles from one end of a 100-mile road, services accidents along the whole road. Suppose accidents occur with uniform distribution along the road, and the ambulance can travel at 60 miles an hour. Let T minutes be the response time (between when accident occurs and when ambulance arrives).

a) Find $P(T > 30)$.

b) Find $P(T > t)$ as a function of t. Sketch its graph.

c) Calculate the density function of T.

d) Calculate the mean and standard deviation of T.

e) What would be a better place for the station? Explain.

6. Electrical components of a particular type have exponentially distributed lifetimes with mean 48 hours. In one application the component is replaced by a new one if it fails before 48 hours, and in case it survives 48 hours it is replaced by a new one anyway. Let T represent the potential lifetime of a component in continuous use, and U the time of such a component in use with the above replacement policy. Sketch the graphs of:

a) the c.d.f. of T; b) the c.d.f. of U. Is U discrete, continuous, or neither?

c) Find $E(U)$. [*Hint*: Express U as a function of T.]

d) Does the replacement policy serve any good purpose? Explain.

7. **Two-sided exponential distribution.** Suppose X with range $(-\infty, \infty)$ has density $f(x) = \alpha e^{-\beta|x|}$ where α and β are positive constants.

a) Express α in terms of β. b) Find $E(X)$ and $Var(X)$ in terms of β.

c) Find $P(|X| > y)$ in terms of y and β. d) Find $P(X \leq x)$ in terms of x and β.

8. **The principle of ignoring constants.** In calculating the density of a random variable X, a quick method is to ignore constant factors as you go along, to end up with an answer of the form $P(X \in dx)/dx = f(x)$ with $f(x) = c\,h(x)$ for a known function $h(x)$ and mystery constant c. The point is that provided your calculation has been consistent with the basic rules of probability, the density of X must integrate to 1, so

$$\int c\,h(x)\,dx = \int f(x)\,dx = 1$$

a) Use this identity to evaluate c in terms of $\int h(x)\,dx$.

b) You can often recognize at the end of a calculation that $h(x) = c_1 f_1(x)$ for some named density $f_1(x)$ (e.g. one of the densities displayed in the table on page 477 and some constant c_1. Deduce that then $c = 1/c_1$ and $f(x) = f_1(x)$.

Use this method to evaluate the constant factor c that makes $ch(x)$ a probability density for each of the following functions $h(x)$, assumed to be zero except for the indicated range of x, and find $E(X)$ and $Var(X)$ in each case from the table on page 477.

c) $e^{-\frac{1}{2}x^2}$ $(-\infty < x < \infty)$ d) x $(0 < x < 1)$

e) 1 $(0 < x < 10)$ f) e^{-5x} $(x > 0)$

9. Use the method of Exercise 8 to evaluate the constant factor c that makes $f(x) = ch(x)$ a probability density for each of the following functions $h(x)$, assumed to be zero except for the indicated range of x, where a and b are positive parameters. Also find $E(X)$ and $Var(X)$ in each case:

a) $e^{-(x-a)^2}$ $(-\infty < x < \infty)$; b) $e^{-(x-a)^2/b^2}$ $(-\infty < x < \infty)$;

c) $e^{-ax}x^5 (x > 0)$; d) $e^{-a|x|}$ $(-\infty < x < \infty)$;

e) $x^7(1-x)^9$ $(0 < x < 1)$; f) $x^7(b-x)^9$ $(0 < x < b)$.

10. Evaluate the following integrals:

a) $\displaystyle\int_0^\infty e^{-x^2}dx$; b) $\displaystyle\int_0^1 e^{-x^2}dx$; c) $\displaystyle\int_0^\infty x\,e^{-x^2}dx$; d) $\displaystyle\int_0^\infty x^2\,e^{-x^2}dx$.

11. Evaluate the following integrals:

a) $\displaystyle\int_0^\infty z^3 e^{-z^2}dz$; b) $\displaystyle\int_0^\infty x^7 e^{-2x}dx$; c) $\displaystyle\int_0^{100} x^2(100-x)^2 dx$.

12. A Geiger counter is recording background radiation at an average rate of 2 hits per minute; the hits may be modeled as a Poisson process. Let T be the time (in minutes) of the third hit after the machine is switched on. Find $P(1 < T < 3)$.

13. Local calls are coming into a telephone exchange according to a Poisson process with rate λ_{loc} calls per minute. Independently of this, long-distance calls are coming in at a rate of λ_{dis} calls per minute. Write down expressions for probabilities of the following events:

a) exactly 5 local calls and 3 long-distance calls come in a given minute;

b) exactly 50 calls (counting both local and long distance) come in a given three-minute period;

c) starting from a fixed time, the first ten calls to arrive are local.

14. Particles arrive at a Geiger counter according to a Poisson process with rate 3 per minute.

a) Find the chance that less than 4 particles arrive in the time interval 0 to 2 minutes.

b) Let T_n minutes denote the arrival time of the nth particle. Find

$$P(T_1 < 1,\ T_2 - T_1 < 1,\ T_3 - T_2 < 1)$$

c) Find the conditional distribution of the number of arrivals in 0 to 2 minutes, given that there were 10 arrivals in 0 to 4 minutes. Recognize this as a named distribution, and state the parameters.

15. Two Geiger counters record arrivals of radioactive particles. Particles arrive at Counter I according to a Poisson process, at an average rate of 3 per minute. Independently, particles arrive at Counter II at an average rate of 4 per minute, also according to a Poisson process. In a particular one-minute period, the counters recorded at total of 8 arrivals. Given this, what is the chance that each counter recorded four arrivals?

16. Cars arrive at a toll booth according to a Poisson process at a rate of 3 arrivals per minute.

a) What is the probability that the third car arrives within three minutes of the first car?

b) Of the cars arriving at the booth, it is known that over the long run 60% are Japanese imports. What is the probability that in a given ten-minute interval, 15 cars arrive at the booth, and 10 of these are Japanese imports? State your assumptions clearly.

17. Show that T has exponential distribution with rate λ if and only if

$$P(T \leq t) = 1 - e^{-\lambda t} \quad \text{for all} \quad 0 \leq t < \infty$$

18. Bus lines A, B, and C service a particular stop. Suppose the lines come as independent Poisson processes with rates λ_A, λ_B, and λ_C buses per hour respectively. Find expressions for the following probabilities:

a) exactly one A bus, two B buses, and one C bus come to the stop in a given hour;

b) a total of 7 buses come to the stop in a given two hour time period;

c) starting from a fixed time, the first A bus arrives after t hours.

19. A piece of rock contains 10^{20} atoms of a particular substance, each with a half-life of one century. How many centuries must pass before:

a) most likely about 100 atoms remain;

b) there is about a 50% chance that at least one atom remains.

20. **Hazard rates (refers to Section 4.3).** Suppose a component with constant failure rate λ is backed up by a second similar component. When the first component burns out the second is installed, and is thereafter subject to failure at the same rate λ, independently of when it was installed and how long it has been in use. Let T be the total time to failure of both components. Find for T:

a) the density function; b) the survival function; c) the hazard rate function.

d) Suppose $\lambda = 1$ per hour. Given $T \geq 2$ hours, what is the approximate probability of failure in the next minute?

21. Suppose R_1 and R_2 are two independent random variables with the same density function $f(x) = x \exp(-\frac{1}{2}x^2)$ for $x \geq 0$. Find

(a) the density of $Y = \min\{R_1, R_2\}$; b) the density of Y^2; c) $E(Y^2)$.

22. Let X be a random variable that has a uniform distribution on the interval $(0, a)$.

a) Find the c.d.f. of $Y = \min(X, a/2)$.

b) Is the distribution of Y continuous? Explain. c) Find $E(Y)$.

23. An earthquake of magnitude M releases energy X such that $M = \log X$. For earthquakes of magnitude greater that 3, suppose that $M - 3$ has an exponential distribution with mean 2.

a) Find $E(M)$ and $Var(M)$ for an earthquake of magnitude greater than 3.

b) For an earthquake as in part a), find the density of X.

c) Consider two earthquakes, both of magnitude greater than 3. What is the probability that the magnitude of the smaller earthquake is greater than 4? Assume that the magnitudes of the two earthquakes are independent of each other.

24. Suppose stop lights at an intersection alternately show green for one minute, red for one minute (ignore amber). Suppose a car arrives at the lights at a time distributed uniformly at random relative to this cycle. Let X be the delay of the car at the lights, neglecting any delay due to traffic congestion.

a) Find a formula for the c.d.f. of X, and sketch its graph.

b) Is X discrete, continuous, or neither? c) Find $E(X)$ and $Var(X)$.

d) Suppose that the car encounters a succession of ten such stop lights. Make an independence assumption and use the normal approximation to estimate the probability that the car will be delayed more than four minutes by the lights.

25. Suppose the random variable U is distributed uniformly on the interval $(0, 1)$. Find:

a) the density of the random variable $Y = \min\{U, 1 - U\}$ (indicate where the density is positive);

b) the density of $2Y$; c) $E(Y)$ and $Var(Y)$.

26. Suppose that the weight W_t of a tumor after time t is modeled by the formula $W_t = Xe^{tY}$ where X and Y are independent random variables, X distributed according to a gamma distribution with mean 2 and variance 1, and Y distributed uniformly on 1 to 1.5. Find formulae for: a) $E(W_t)$; b) $SD(W_t)$.

27. Suppose U_1, $U_2, \ldots$ are independent uniform $(0, 1)$ variables, and let N be the first $n \geq 2$ such that $U_n > U_{n-1}$. Show that for $0 \leq u \leq 1$:

a) $P(U_1 \leq u \text{ and } N = n) = \dfrac{u^{n-1}}{(n-1)!} - \dfrac{u^n}{n!} \quad n \geq 2;$

b) $P(U_1 \leq u \text{ and } N \text{ is even}) = 1 - e^{-u}$.

c) $E(N) = e$.

28. A point is chosen uniformly at random from the circumference of a circle of diameter 1. Let X be the length of the chord joining the random point to an arbitrary fixed point on the circumference. Find: a) the c.d.f. of X; b) $E(X)$; c) $Var(X)$.

29. A gambling game works as follows. A random variable X is produced; you win \$1 if $X > 0$ and you lose \$1 if $X < 0$. Suppose first that X has a normal $(0, 1)$ distribution. Then the game is clearly "fair". Now suppose the casino gives you the following option. You can make X have a normal $(b, 1)$ distribution, but to do so you have to pay $\$cb$ which is not returned to you even if you win. Here $c > 0$ is set by the casino, but you can choose any $b > 0$.

a) For what values of c is it advantageous for you to use this option?

b) For these values of c, what value of b should you choose?

30. A manufacturing process produces ball bearings with diameters which are independent and normally distributed with mean 0.250 inches and SD 0.001 inches. In a high-precision application, 16 bearings are arranged in a ring. The specifications are that:

(i) each bearing must be between 0.249 and 0.251 inches in diameter;

(ii) the sum of the diameters of the 16 bearings must be between 3.995 and 4.005 inches.

a) What is the expected number of bearings which must be produced by the process to obtain 16 satisfying specification (i)?

b) Given 16 bearings obtained like this, what is the chance that they meet specification (ii)?

[*Hint* for b): Write $x^2\phi(x) = x[x\phi(x)]$ and use integration by parts to show that

$$\int_{-z}^{z} x^2\phi(x)dx = 2\Phi(z) - 1 - 2z\phi(z).\ ]$$

31. The skew-normal pseudo-density. Referring to the end of Section 3.3, let

$$\Phi_\theta(z) = \Phi(z) - \frac{\theta}{6}(z^2 - 1)\phi(z)$$

This is the substitute for the normal c.d.f. $\Phi(z)$ which for $\frac{\theta}{6} \neq 0$ typically gives a better approximation than $\Phi(z)$ to the c.d.f. of a random variable with mean zero, variance 1 and third moment θ.

a) Let $\phi_\theta(z) = \frac{d}{dz}\Phi_\theta(z)$. Show $\phi_\theta(z) = [1 - \frac{\theta}{6}(3z - z^3)]\phi(z)$.

b) Show that for every θ

$$\int_{-\infty}^{\infty} \phi_\theta(z)dz = 1; \int_{-\infty}^{\infty} z\phi_\theta(z)dz = 0; \int_{-\infty}^{\infty} z^2\phi_\theta(z)dz = 1; \int_{-\infty}^{\infty} z^3\phi_\theta(z)dz = \theta$$

[So $\phi_\theta(z)$ is very like the probability density of a distribution with mean zero, variance 1 and third moment θ. This explains the choice $\theta = \text{Skewness}(X) = E(X_*^3)$ in the skew-normal approximation to the distribution of a standardized variable $X_* = (X - \mu)/\sigma$.]

c) Show that ϕ_θ is negative for large negative z if $\theta > 0$, and negative for large positive z if $\theta < 0$. So for $\theta \neq 0$, $\phi_\theta(z)$ is in fact *not* a probability density. It may be called instead a *pseudo-density.*

d) Find a probability in the Poisson(9) distribution whose normal approximation with continuity and skewness corrections is a negative number.

e) Explain carefully why, despite c) and d) the functions $\phi_{1/3}(z)$ and $\Phi_{1/3}(z)$ provide practically useful approximations to the Poisson(9) and other distributions which are roughly normal in shape but slightly skewed.