Chapter 3
Some Univariate Distributions

I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the 'law of error.' A savage, if he could understand it, would worship it as a god. It reigns with severity in complete self-effacement amidst the wildest confusion. The huger the mob and the greater the anarchy the more perfect is its sway. Francis Galton (1886) Presidential Address. *J. Anthrop. Institute*, **15**, 487-499.

Experimenters imagine that it [the normal distribution] is a theorem of mathematics, and mathematicians believe it to be an experimental fact. Gabriel Lippmann (attrib.)

3. Introduction

We now offer a catalog of some of the more commonly encountered probability distributions (density functions). When we first look at data, in what is sometimes called "exploratory data analysis," we often want to find a pdf that can act as a probability model for those data. If we can use one of the pdf's from this chapter, we are well launched on further studies, since there will be a lot of well-established results on how to work with data from that kind of pdf. Conversely, if we have to develop a whole new probability model, we are in for a lot more work. Seeing if data match a given pdf, and devising pdf's of our own from data, are both topics we will discuss later on; for now we simply offer a guide to some of the better-known, and most useful, probability density functions. We also provide, for most of these, a guide to numerical methods for generating a series of numbers that have one of these distributions; this is very important, because, as we will see, a number of useful statistical methods hinge on being able to do this.

This subject is usually made to look more complicated than it really is by the history and conventions of the field, which cause parameters with the same role to be named and symbolized differently for each distribution. And some of the names are not especially enlightening, with the prize for most obscure going to "number of degrees of freedom".

We can write nearly all pdf's for continuous random variables in the following form:

$$\frac{1}{cA_r(s)}\,\phi_s\!\left(\frac{x-l}{c}\right) \qquad L_b \le x \le L_c \tag{1}$$

where the L's give the range of applicability: usually from $-\infty$ to ∞, often from 0 to ∞, sometimes only over a finite range. The function ϕ_s is what gives the actual shape of the pdf; the constant $A_r(s)$ (the area under the function) is needed to normalize the integral of this function, over the limits given, to 1. We call s the **shape parameter**; not all pdf's have one. Almost all do have two others:

- l A **location parameter**, which has the effect of translating the pdf on the x-axis. This appears mostly for functions on $(-\infty, \infty)$, though we will see a few exceptions.

- c A **scale parameter**, which expands or contracts the scale of the x-axis. To keep the pdf normalized, it also has to scale the size of the pdf, and so multiplies $A_r(s)$.

For each pdf, we will give it in (1) a stripped-down form, usually with l set to 0 and c to 1; (2) in the full form with all the parameters symbolized through l, c, and s, and (3) with all the parameters given their conventional symbols. Even though the first two are not, strictly speaking, equal, we will use the equal sign to connect them. You should try to see

how equation (1) above can be applied to the stripped-down form to produce the conventional one.

3.1. Uniform Distribution

The simplest probability density function is the **uniform** or **rectangular** distribution, which has the probability density function (pdf)

$$\phi(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & x < 0 \ \text{ or } \ x > 1 \end{cases}$$

and cumulative distribution function (cdf)

$$\Phi(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 \leq x \leq 1 \\ 1 & x \geq 1 \end{cases}$$

In this particular case these formulas are the "conventional" form, and a random variable X with this distribution is said to satisfy $X \sim U(0,1)$. We can of course apply l and c to move the nonzero part to any location, and make it of any finite length, in which case we would have

$$\phi(x) = c^{-1} \quad \text{for} \quad l < x < l + c$$

This distribution is the basis for many others, and is used in various simulation techniques, some of which we will discuss in Chapter 5. A collection of uniformly distributed random numbers is what is supposed to be produced by repeated calls to a function named `ran` (or some similar name) in your favorite language or software package; we say "supposed to be" because the actual numbers produced depart from this in two ways. The first, harmless, one is that any such computer function has to actually return a deterministic set of numbers designed to "look random"; hence these are really **pseudorandom** numbers. The second and harmful departure is that many of the computer routines that have actually been used do not meet this requirement well. If you do not know the algorithm for the function you are using, you should use another one whose algorithm you do know. There are several good candidates, and an excellent discussion, in Press *et al* (1992)[1]—though much has been done since.
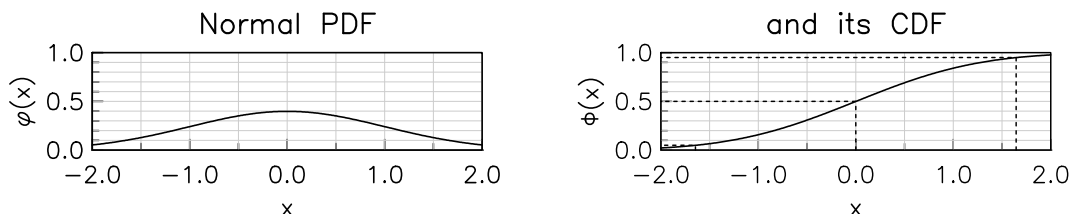

Figure 3.1

---

[1] Press, W. H., S. A. Teukolsky, W. T. Vettering, and B. P. Flannery (1992). *Numerical Recipes in Fortran: The Art of Scientific Computing, 2nd Ed.* (Cambridge: Cambridge University Press)

3.2. Normal (Gaussian) Distribution

We have already met this pdf, but present it again to show our different forms for a pdf, namely

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2} \;=\; \frac{1}{c\sqrt{2\pi}}\, e^{-(x-l)^2/2c^2} \;=\; \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}$$

where the location parameter (conventionally μ) is called the **mean** and the scale parameter (conventionally σ) is called the **standard deviation**. Figure 3.1 shows the pdf and cdf for this distribution, with the dashed lines showing where Φ attains the values 0.05, 0.50, and 0.95. The 0.5 and 0.95 lines bound 0.9 (90%) of the area under the pdf (often called, since this is a "density function," the mass). A summary table of mass against x-value (plus and minus) would include the following:

x	Mass fraction
±1.00	0.68
±1.65	0.90
±1.96	0.95
±2.58	0.99
±3.29	0.999
±3.90	0.9999

so we would, for example, expect that out of 1000 rv's with this distribution, no more than 1 would be more than 3.3σ away from the central location μ. We described in Chapter 2 why this pdf is special, namely that the central limit theorem says that (for example) sums of rv's approach this distribution somewhat irrespective of their own distribution. It very often simplifies the theory to assume Gaussian behavior (we shall see a number of examples of this), and this behavior is at least approximately satisfied by many datasets. But you should never casually assume it.

This is probably a good place to introduce another term of statistical jargon, namely what a **standardized** random variable is. We say that any normally distributed random variable $X \sim N(\mu, \sigma)$ may be transformed to the standard normal distribution by creating the new standardized rv $Z = (X - \mu)/\sigma$, which is distributed as $N(0,1)$. We will see other examples of such transformations.

3.2.1. Generating Normal Deviates

The title of this section contains one of those terms (like Love waves) liable to bring a smile until you get used to it; but "deviates" is the standard term for what we have called a collection of random numbers. There are quite a few ways of producing random numbers with a Gaussian pdf; with the advent of very large simulations it has become a matter of interest to use methods that are fast but that will also produce the relatively rare large values reliably.[2] In this section we describe two related methods that allow us to use the procedures developed in the last chapter for finding the pdf of functions of random variables.

Both methods start by getting two rv's with a uniform distribution; this pair of numbers specifies a point in a square with $0 < x_1 < 1$ and $0 < x_2 < 1$; clearly the distribution of these points is a uniform bivariate one, with $\phi(x_1, x_2) = 1$ over this square region. We term these two uniform rv's U_1 and U_2, though we use x_1 and x_2 for the actual numbers.

The first method, known as the **Box-Muller transform**, computes two values given by

---

[2] Thomas, D. B., W. Luk, P. H. W. Leong, and J. D. Villasenor (2007). Gaussian random number generators, *ACM Computing Surv.*, **39**, #11.

$$y_1 = \sqrt{-2\ln x_1}\,\cos 2\pi x_2 \qquad \text{and} \qquad y_2 = \sqrt{-2\ln x_1}\,\sin 2\pi x_2$$

and we assert that y_1 and y_2 are each distributed as a Gaussian.

To show this, we have to find the pdf's of

$$Y_1 = \sqrt{-2\ln U_1}\,\cos 2\pi U_2 \qquad \text{and} \qquad Y_2 = \sqrt{-2\ln U_1}\,\sin 2\pi U_2$$

The first point is to note that if we write the joint pdf in polar coordinates, as $\phi_Y(r,\theta)$, the distribution in angle θ is uniform because U_2 is. The pdf in radius is found from equation (22) of Chapter 2, with $W = g(R) = \sqrt{-2\ln R}$ and R uniformly distributed. Then $g^{-1}(w) = e^{-w^2/2}$, and the application of (22), and the uniformity of ϕ_R, gives

$$\phi_W = \phi_R(g^{-1}(w)) \left| \frac{d}{dw}\, g^{-1}(w) \right| = w e^{-w^2/2} \tag{2}$$

If Y_1 and Y_2 are iid variables with a Gaussian distribution, the joint pdf will be the product

$$\phi_Y = \frac{1}{2\pi}\, e^{-(x_1^2 + x_2^2)/2} = \frac{1}{2\pi}\, e^{-r^2/2}$$

Although this contains r only, it remains a joint pdf. If we integrate over all θ we get a pdf in r alone:

$$\phi_R' = \int_0^{2\pi} \phi_Y r\, d\theta = r e^{-r^2/2} \tag{3}$$

Equations (2) and (3) give the same result: the pdf found for the transformed variable $W = g(R) = \sqrt{-2\ln R}$ for R uniformly distributed matches that for the radius of the bivariate Gaussian. So we see that taking this function of a uniformly distributed rv, and then multiplying the result by the unit vector in a random direction (which is what $(\cos 2\pi U_2, \sin 2\pi U_2)$ is) will produce a pair of Gaussian rv's.

Another approach to this computation is to observe that if the points are uniformly distributed over the square, they also will be uniformly distributed over the circle inscribed within it, which they can be limited to if we remove all points for which $r^2 = x_1^2 + x_2^2 \geq 1$. Then we form

$$y_1 = \frac{x_1}{r}\,\sqrt{-2\ln r^2} = x_1\sqrt{\frac{-2\ln r^2}{r^2}} \qquad \text{and} \qquad y_2 = \frac{x_2}{r}\,\sqrt{-2\ln r^2} = x_2\sqrt{\frac{-2\ln r^2}{r^2}}$$

where the second form requires only one square root. The advantage of this method is that it avoids calls to trigonometric functions, and thus is usually faster even though it requires 27% more calls to the random number generator. It is equivalent to the Box-Muller transform because x_1/r and x_2/r are equivalent to a sine and cosine, and will be distributed uniformly in angle, while r^2 is also uniformly distributed, making it a valid replacement for x_1.

3.3. A Digression: Some Point Processes

To introduce the next probability distributions, we give a more formal description of the point process we introduced in Chapter 1: the Poisson process, which we used as a simple model for geomagnetic reversals. In this process we suppose that the probability of some event occurring is equal over all time, and is described by a **rate** (also called an **intensity**) which has the dimensions of probability over time. The conditions for a Poisson process with rate λ are that as a small time interval h approaches 0,

(A) the number of events in disjoint time intervals are independent: the number in $(0, t]$ is independent of (does not affect) the number in $(t, t + h)$.

(B) the probability of an event in an interval of length h is, as $h \to 0$, approximately proportional to h plus a remainder that is $o(h)$. (Remember that a function $g(h)$ is $o(h)$ if $\lim_{h\to 0} g(h)/h = 0$.)

(C)   the probability of more than one reversal in an interval of length h is $o(h)$.

The Poisson process is **memoryless**: there is no memory of what went before, so what happens at any time does not depend on the earlier history. This makes the Poisson process something of a "simplest case" point process; it is described by only one parameter and is fundamentally unpredictable. It therefore serves as a kind of reference model, for example in studies of earthquake statistics, as having the fewest complications and the least predictability.
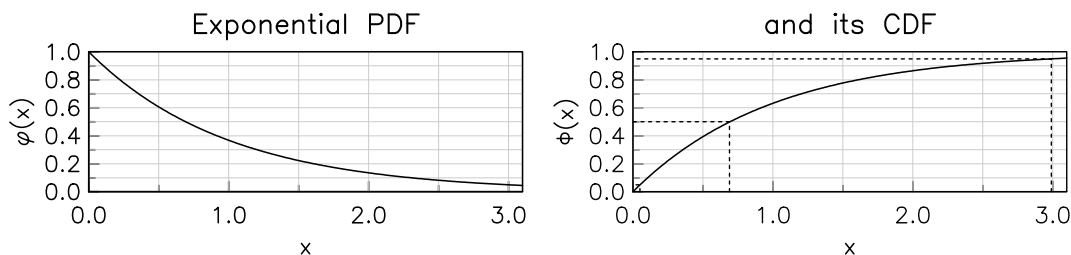
Figure 3.2

A more general type of point process is a **renewal process**, for which the probability of an event depends in some way on the time since the last event. The Poisson process might be called a renewal process that isn't, since in its case the probability is constant, and the time of the next event is uninfluenced by the time of the previous one. Continuing with the geomagnetic field example, we might suppose that immediately following a geomagnetic reversal there is some change in the probability of occurrence of a new event, e.g. , while the geodynamo recovers its stable polarity; so short intervals might be less probable.

3.4. Exponential Distribution

It can be shown that if the occurrences of events are given by a Poisson process, then the difference (time interval, or distance, or whatever) between successive occurrences will have an **exponential** distribution, with probability density function defined over $[0, \infty)$:

$$\phi(x) = e^{-x} = ce^{-x/c} = \lambda e^{-\lambda x}$$

and cumulative distribution function

$$\Phi(x) = 1 - e^{-\lambda x} \tag{4}$$

Since the pdf is peaked towards zero, short intervals have a much higher probability of occurring than long ones do. Figure 3.2 shows this distribution; again, the dotted lines show the values for Φ equal to 0.05, 0.50, and 0.95: much less symmetric than the same points for the Normal.

Producing random numbers for this distribution is very easy, and illustrates a method that can be applied to some other pdf's. We can think of taking uniformly distributed rv's and placing them on the y-axis of the cdf; then if we map these into the x-axis through the inverse cdf function, the result will have the distribution we want. This result is general; how usable it is depends on how easy it is to compute the inverse function for the cdf. In this case it is easy; from (4) the inverse cdf function is

$$\Phi^{-1}(y) = \frac{-\ln(1 - y)}{\lambda} = \frac{-\ln(y)}{\lambda}$$

where the last quality is gotten from the observation that if $1 - Y$ is uniformly distributed between 0 and 1, Y will be also. This method depends on having an easy way to compute the inverse cdf; because the Gaussian does not, more complicated methods such as the Box-Muller transform are needed.
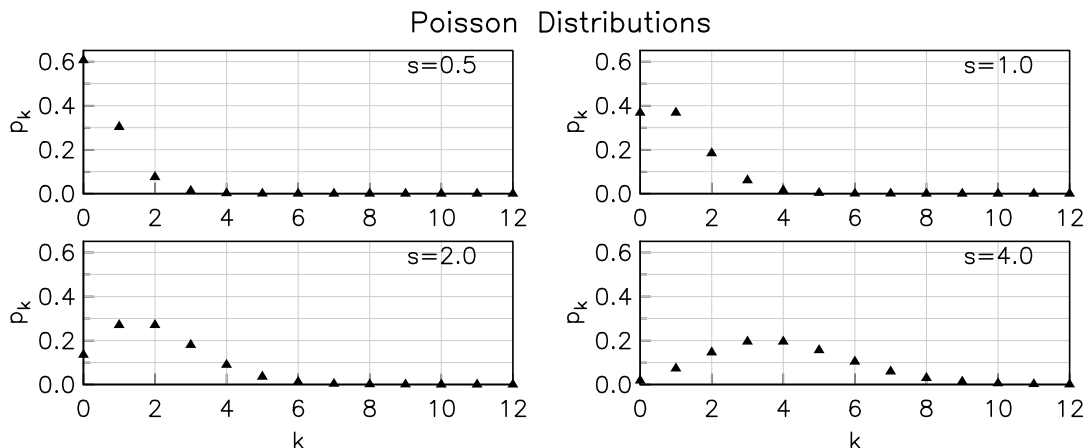

Figure 3

3.5. Poisson Distribution

Another distribution arises from the Poisson process if we consider a different random variable, namely the number of events k that occur in a time interval of length T. Then $k(T)$ follows a what is called a **Poisson distribution**. The probability of getting k events is given by

$$p_k = s^k \frac{e^{-s}}{k!}$$

where s is the shape parameter; This distribution does not have a scale or location parameter. For a Poisson process with rate λ, observed in an arbitrary interval of length T, $s = \lambda T$.

This distribution is discrete, so we write it as using delta functions as

$$\phi(x) = \sum_{k=0}^{\infty} p_k \delta(x - k)$$

As s becomes large, the distribution starts to look like a discrete version of the Normal. The expected value (first moment) of the distribution is just the shape parameter.

This distribution arises, not just for a time series, but whenever we have a small probability of something happening in each of a large number of instances, and want to know the distribution of the number of events. Our example so far has been reversals (with a small probability in each of very many years); earthquakes would be another point-process example. One that does not have a temporal component might be deaths from being struck by lightning; if we took a number of groups of (say) 100,000 people, we would expect the number of deaths in each group to be Poisson-distributed.[3]

---

[3] The classic example is, for each corps of the Prussian army, the number of soldiers killed each year by being kicked by a horse.
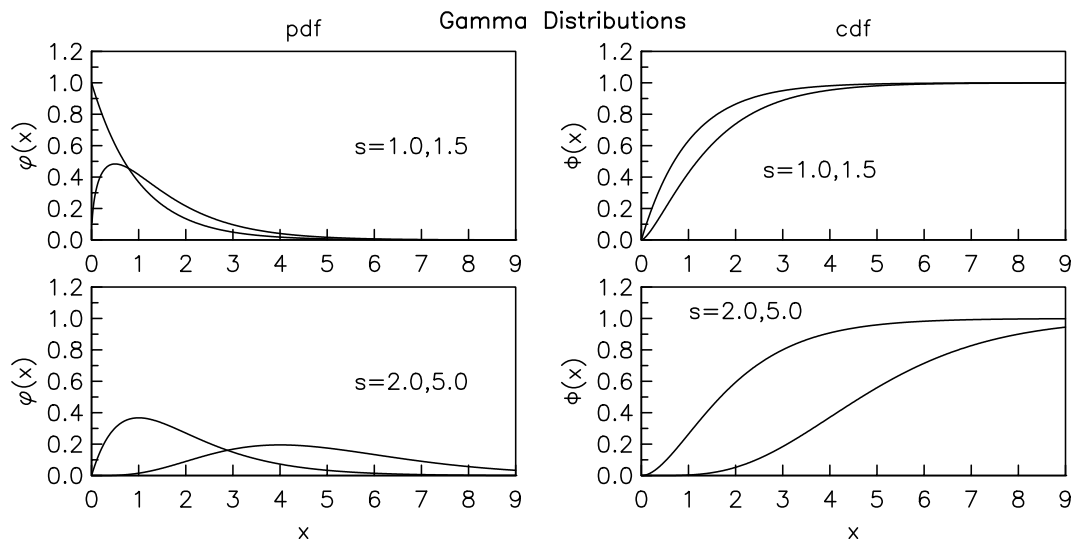
Figure 3.3

3.6. Gamma Distribution

For the more complicated case of a renewal process, the interval lengths often follow a **gamma distribution**, which has a pdf on $[0, \infty)$ that looks rather like a continuous version of the Poisson distribution:

$$\phi(x) = \frac{1}{\Gamma(s)} x^{s-1} e^{-x} = \frac{1}{c\Gamma(s)} \left(\frac{x}{c}\right)^{s-1} e^{-x/c} = \frac{\lambda^s}{\Gamma(s)} x^{s-1} e^{-\lambda x} \qquad (5)$$

where the gamma function (used to normalize the distribution) is defined by

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} \, du$$

As in the previous example we have followed the convention for dealing with point processes and written the scale parameter as $\lambda = c^{-1}$. We see from (5) that the exponential distribution is a special case of a gamma distribution, one for which $s = 1$.

Figure 3.3 shows the gamma density function for different values of s. For a renewal point process governed by this kind of probability density function, λ describes the rate of events well after each event, while s controls the shape of the probability function immediately following each event. Values of $s < 1$ correspond to an enhanced probability (relative to a Poisson process with the same value of λ) of another event immediately following one that has just occurred. Values of $s > 1$ indicate a diminished probability of another event immediately following any given one. For geomagnetic reversals, using the renewal model gives $s > 1$. The physical interpretation of this is controversial: it may be a consequence of the known inadequacies of the geological record of geomagnetic reversals, in which very short polarity intervals are missing; or it may reflect fundamental physical processes in the core.[4] Gamma distributions are also used in statistical studies of the times between

---

[4] McFadden, P.L. (1984). Statistical tools for the analysis of geomagnetic reversal sequences, *J. Geophys. Res.* **89** 3363-3372; McFadden, P.L. and R.T. Merrill (1984). Lower mantle convection and geomagnetism, *J. Geophys. Res.* **89** 3354-3362.

earthquakes: another example of a process that is not memoryless, since there is an enhanced probability of earthquake occurrence immediately following any given one (the aftershock sequence), compared with the probability later. Indeed, even after removing aftershocks, earthquakes are observed to cluster in time, making a gamma distribution with $s < 1$ a possible model.
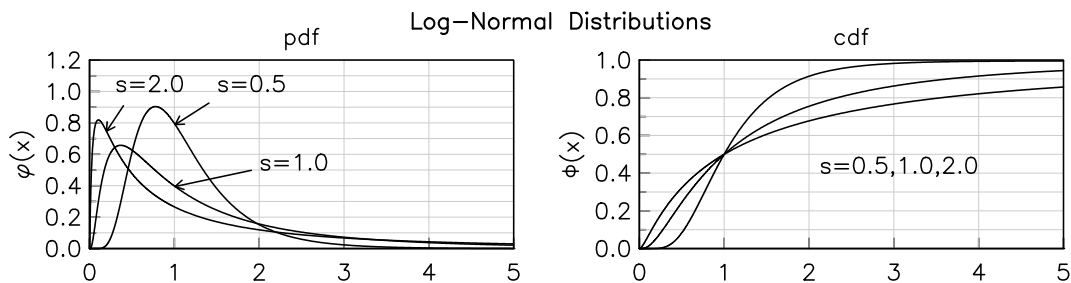
Log–Normal Distributions

Figure 5

3.7. Log-Normal Distribution

We next consider a variety of distributions derived in different ways from the normal distribution. One of the simplest comes if we exponentiate X: if X is normally distributed, then $Y = e^X$ is said to be **log-normally distributed**, since $X = \ln Y$. This is a mapping of the real line into the positive numbers, which may be useful for some quantities. The pdf is

$$\phi(x) = \frac{1}{s\sqrt{2\pi}} \frac{e^{-(\ln(x))^2/2s^2}}{x} = \frac{1}{s\sqrt{2\pi}} \frac{e^{-(\ln((x-l)/c))^2/2s^2}}{x - l} \quad \text{for} \quad x > l$$

One area where this distribution is relevant is grain size distributions in sediments; another is the observed distributions of magnetization of basaltic rocks. This distribution has also been used, with mixed success, to model times between large earthquakes: yet another kind of renewal process.
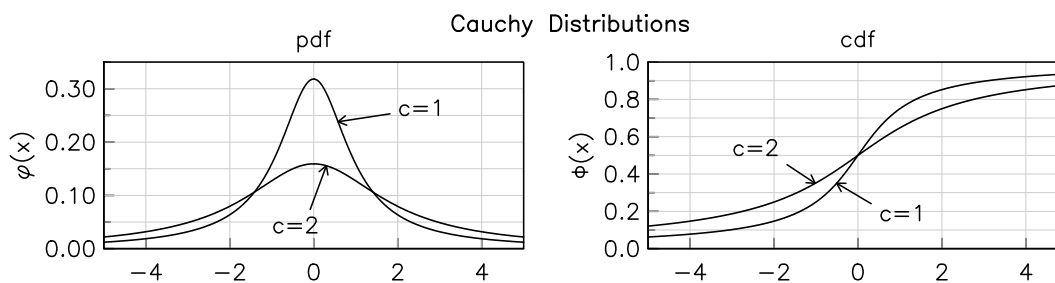
Cauchy Distributions

Figure 6

3.8. Cauchy Distribution

We have discussed in the previous chapter the distribution of the sum of two Normal random variables, which is again Normal; the difference will be the same. Things get more interesting if we consider the ratio of two such variables, so our new random variable would be $Y = X_1/X_2$, with $X_i \sim N(0, \sigma)$. It is easy to see that it would be relatively common for the denominator X_2 to be small, and hence for Y to be large; so we would expect the ratio to be

much more heavy-tailed than the Normal is.

We can use equation (26) of Chapter 2 to get the actual distribution, which is

$$\phi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |x| \, e^{-y^2/2} e^{-x^2 y^2/2} dx = \frac{1}{\pi} \int_{0}^{\infty} x e^{-y^2(x^2+1)/2} dx$$

A change of variables to $u = y^2$ makes this into the integral of an exponential in u:

$$\frac{1}{2\pi} \int_{0}^{\infty} e^{-u(x^2+1)/2} du = \frac{1}{\pi(1+x^2)} = \frac{c}{\pi(c^2 + (x-l)^2)}$$

which is known as the **Cauchy distribution**. This distribution is integrable (it has to be, to be a pdf), but the first and higher moments do not exist (that is to say, are infinite). The median does exist, however. This pdf is thus as heavy-tailed as a pdf can get, and can serve as a reminder that even such standard procedures as taking the mean may not necessarily be valid.

3.9. Chi-Squared Distribution

If, instead of taking the ratio of two Normal rv's, we take the product, we get the χ^2 distribution, one of several derived from the Normal that are used not less as models for data than for statistical tests—some of which we discuss in Chapter 5.

We start by squaring a random variable with a normal distribution; that is, if X is a random variable distributed with a normal pdf with mean 0 and standard deviation 1 (i.e., $X \sim N(0,1)$, then the distribution of the random variable $Z = X^2$ is called the **chi-square distribution with one degree of freedom**, denoted χ_1^2.

The pdf can be derived by the procedure described in Chapter 2:

$$\Phi_z(z) = \mathcal{P}(Z < z) = \mathcal{P}(-\sqrt{z} < X < \sqrt{z})$$

where the second expression comes from using the inverse function to x^2, namely the square-root. Rewriting the rightmost expression gives

$$\Phi_z(z) = \Phi_x(\sqrt{z}) - \Phi_x(-\sqrt{z})$$

and differentiating with respect to $z = x^2$ using the chain rule gives

$$\phi_z(z) = \frac{d\Phi(z)}{dz} = \frac{1}{\sqrt{z}} \left(\frac{d\Phi_x(\sqrt{z})}{dz} - \frac{d\Phi_x(-\sqrt{z})}{dz} \right) = \frac{1}{2\sqrt{z}} \left(\phi_x(\sqrt{z}) - \phi_x(-\sqrt{z}) \right)$$

Finally, substituting in the normal pdf for ϕ_x, we get

$$\phi_z(z) = \frac{1}{\sqrt{2\pi}} \frac{e^{-z/2}}{\sqrt{z}}$$

which is a special case of the gamma distribution, with $\lambda = s = \frac{1}{2}$.

Next, consider n random variables $Z_1, Z_2, \ldots, Z_n$ which are independent and each distributed as χ^2 with one degree of freedom. The distribution of $Y = Z_1 + Z_2 + \ldots + Z_Z$ is called the **chi-square distribution with n degrees of freedom**, denoted χ_n^2 (that is, $Y \sim \chi_n^2$). Each of the individual variables has a gamma distribution; and it is the case that the sum of independent random variables each with a gamma distribution (and all with the same value of λ) also follows a gamma distribution, so the χ^2 distribution with n degrees of freedom is a gamma distribution with $s = n/2$ and $\lambda = \frac{1}{2}$. The pdf is, like the gamma distribution, over $[0, \infty)$:

$$\phi(x) = \frac{1}{2^{n/2}\Gamma(n/2)}\, x^{(n/2)-1}e^{-x/2}$$

The expected value of the χ_n^2 distribution is just n, and the variance is $2n$. It is often useful (as we will discuss in later chapters) to look at the **reduced chi-square**, which is defined as

$$\chi_n^2/n$$

This has an expected value of 1, independent of n.

The χ^2 distribution finds widespread application in model fitting. If we have n observations o_i, with predicted values c_i and measurement errors σ_i, then we can form the **standardized residuals**

$$r_i^2 = \frac{(o_i - c_i)^2}{\sigma_i^2}$$

where the "standardized" part, as in the discussion in Section 3.2, comes from scaling the residuals $o - c$ by their errors. Then, if the residuals are distributed according to a Normal pdf, the **sum of squared residuals** (**ssr**, $\sum_i r_i^2$), will have a χ_n^2 distribution: so the value of the ssr is often used to see if the model fits the data adequately or not. The **reduced ssr** is the ssr divided by n, since this (for Gaussian errors) should be distributed as the reduced chi-square; so we would hope that the reduced ssr would be close to one. We will discuss all this in more detail in the chapter on hypothesis testing.
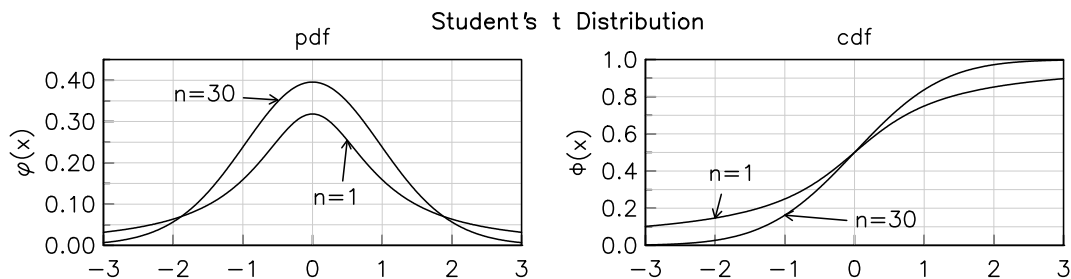

Figure 7

3.10. Student's t Distribution

If X is normally distributed ($X \sim N(0,1)$) and $Z \sim \chi_n^2$ and Z and X are independent, then the distribution of $X/\sqrt{Z/n}$ is called the **Student's t distribution** with n degrees of freedom.[5] This has the probability density function:

$$\phi(x) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\,\pi}\,\Gamma(n/2)}\left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

so that in this case the shape factor is, as with χ^2, an integer. The t distribution is symmetric about zero. As the number of degrees of freedom approaches infinity, the t distribution tends to the standard normal distribution. The t distribution is used in testing whether

---

[5] The name "Student" is the pseudonym that F. S. Gosset used for his statistical publications, to avoid problems with his employer, the Guinness brewery.

samples have statistically distinguishable means—we will, again, discuss this fully when we turn to hypothesis testing.

3.11. F Distribution

Next, suppose X is a random variable distributed as χ^2 with m degrees of freedom; then X/m is a similarly-distributed rv which has been scaled ("standardized") to make its expected value 1. Take Z to be another rv, independent of X, and distributed as χ^2 with n degrees of freedom. Now consider the random variable Y that is the ratio of these after both have been normalized by their degrees of freedom:

$$Y = \frac{X/m}{Z/n}$$

This variable is distributed according to what is called the F distribution with m and n degrees of freedom, denoted $F_{m,n}$. The probability density function of this is given by, for x over $[0, \infty)$:

$$\phi(x) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2} \cdot x^{m/2-1}\left(1 + \frac{m}{n}x\right)^{-(m+n)/2}$$

which we do not plot because of the complications of having to show two variables. Like the t and χ^2 distributions, the F distribution is not so much used as a pdf for modeling data, as in testing, in its case whether two sets of data (or, more often, residuals) have statistically distinguishable variances. This usage comes from this being a ratio of reduced χ^2; if the fit of two models is equally good, we would expect the reduced χ^2 for the residuals to be about the same, and Y to be near 1; if it is not, this may indicate than one model fits better than the other.
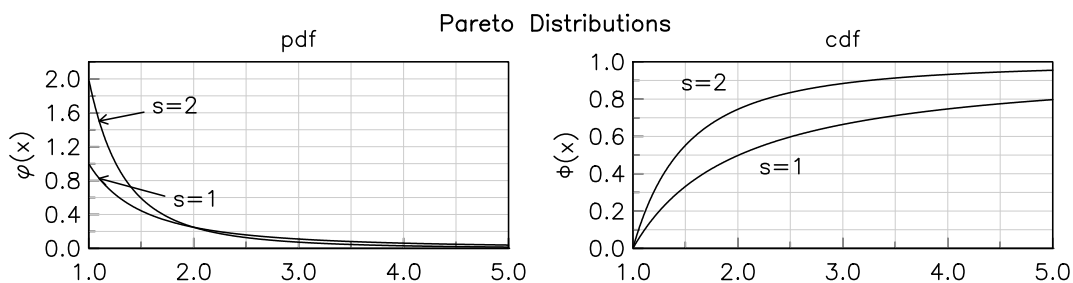

Figure 8

3.12. Pareto Distribution

This is a curiosity in geophysical data distributions, since everyone has heard of it but almost no one knows this name, even though it is the standard term in statistics, and particularly in economics, for which this distribution is a very common data model. It is yet another distribution on half of the real line (though on $[l, \infty)$ rather than on $[0, \infty)$); but it has a much larger tail than the other distributions we have seen for this interval, all of which have an exponential falloff. The Pareto distribution, by contrast, falls off as a power of x, so allowing a greater probability of very large values. The pdf is

$$\phi(x) = sl^s x^{-(s+1)}$$

where $s > 0$; note that for this distribution the location parameter determines the lower bound of the range. To see how this leads to a famous geophysical result, we first find the cdf

$$\Phi(x) = 1 - l^s x^{-s}$$

This probability distribution can be applied to data (such as a set of earthquakes, or the amount of money people have) by supposing that we have a total of n occurrences, with the probability of getting a value x (of earthquake magnitude or wealth) being given by a Pareto distribution. Then the number of occurrences less than some value of x, which we call $n(x)$, will have the pdf $n[1 - \Phi(x)] = nl^s x^{-s}$; taking the log of this we get

$$\ln[n(x)] = \ln(n\, l^s) - sx$$

which has exactly the form of the Gutenberg-Richter relationship for the number of earthquakes less than magnitude M, $\log(n) = a - M$. So the "b-value" much discussed in seismicity studies, is the shape factor for the Pareto distribution.
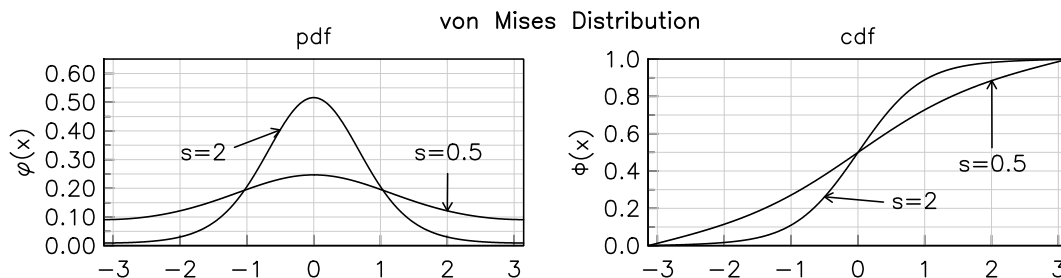

Figure 9

3.13. von Mises Distribution

Our next distribution is not usually discussed in elementary statistics, but is important for the geophysically common case of circular data, in which we are looking at directions in two dimensions, best described as an angle, and so as a variable in the range $[-\pi, \pi)$. With suitable scaling this distribution applies to any variable which is defined to be modulo some number: for example, we can use this distribution for the fractional part of a real-valued quantity, which would be limited to the range $[0, 1)$. The von Mises distribution serves as an analog to the Normal in this case; it has the pdf

$$\phi(x) = \frac{1}{2\pi I_0(s)} e^{s \cos(x-l)}$$

where I_0 is a modified Bessel function. Note that in this case the shape factor modifies the width of the function, as the scale factor does for the Normal. As this factor becomes very small, ϕ approaches a uniform distribution.

3.14. Rayleigh Distribution

We another special distribution arises from the following: suppose we add the squares of two Normal rv's together, and then take the square root? This would be, for example, the pdf of the amplitude of a two-dimensional vector, each of whose components was normally distributed—note that this is zero at the origin. We found the answer above in the process
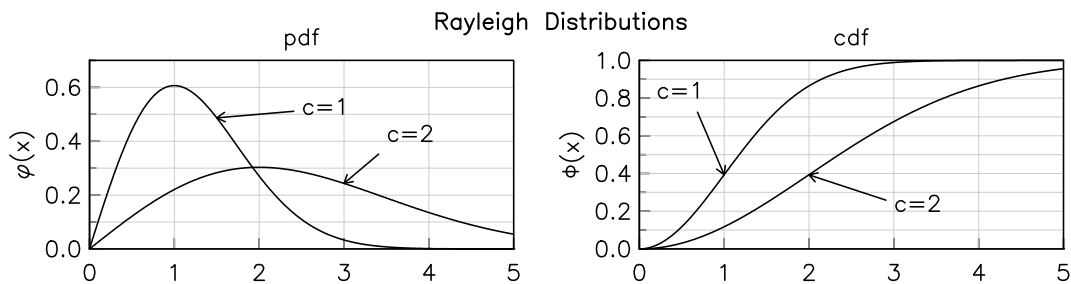
Figure 3.10

of deriving the Box-Muller transformation; another approach is to realize that the rv for the square of the amplitude (call this Z) is distributed as χ_2^2, with cdf $\Phi_0(x) = 1 - e^{-z/2}$. Then, we can find the cdf of $X = Z^{\frac{1}{2}}$ by the same kind of procedure as we used in Section 3.9: taking $\Phi(x)$ to be the cdf of X, we have

$$\Phi(x) = \mathcal{P}(X < x) = \mathcal{P}(Z < x^2) = \Phi_0(x^2) = 1 - e^{-x^2}$$

whence the pdf is

$$\phi(x) = 2xe^{-x^2} = \left(\frac{x}{c^2}\right)e^{-x^2/2c^2}$$

This is known as the **Rayleigh distribution**, and is used in the theory of random walks. Note that if we add two orthogonal vectors described by normal distributions, the expected value of the result is the origin; but the pdf of the amplitude is not zero: the most probable amplitude does not correspond to the most probable location.
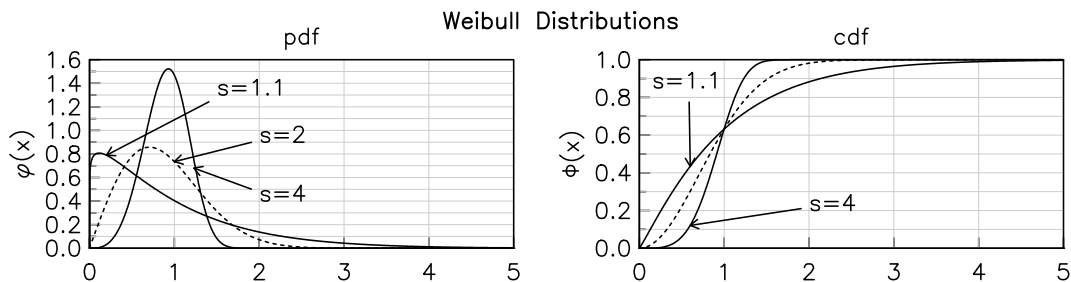

Figure 3.11

3.15. Weibull Distribution

This distribution was invented to describe failure rates, and as such it is popular in modeling renewal processes. The pdf is

$$\phi(x) = x^{s-1}e^{-x^s} = \left(\frac{s}{c}\right)\left(\frac{x-l}{c}\right)^{s-1}e^{-((x-l)/c)^s} \qquad \text{for} \quad x \geq l$$

which makes the cdf relatively simple:

$$\Phi(x) = 1 - e^{-(x)^s} = 1 - e^{-((x-l)/c)^s}$$

The shape and scale parameters are sufficient to provide a flexible distribution, so a nonzero location parameter is less often used. The exponential distribution is a special case, with $s = 1$, and the Rayleigh is another, with $s = 2$.