9/22/2022

Continuous probability distributions

Continuous probability distributions obey the same properties as discrete probability distributions, but the random variable described by those distributions can take on any value in an interval (which may be a short interval, or all real numbers, or anything in between). The probability density function (pdf) is described as a function over real numbers. Sometimes we are able to find a compact expression for the cumulative distribution function (cdf) as well.

Let’s look at an example of a general continuous distribution.

Suppose that our probability density function is $f(x) = K(1 - x^2), 0 \leq x \leq 1$ (and 0 otherwise is assumed and not always stated). The probability is found by calculating the area under the curve. The area under the whole curve for a valid probability density function must equal 1 (the sum of all possible probabilities), so we must find the value of K that makes this area equal to 1.

$$\int_0^1 K(1 - x^2)dx = K\left[x - \frac{1}{3}x^3\right]_0^1 = K\left[1 - \frac{1}{3}\right] = \frac{2}{3}K = 1$$

If $\frac{2}{3}K = 1$, then the value of K that makes this a valid distribution is $\frac{3}{2}$. Thus, $f(x) = \frac{3}{2}(1 - x^2)$.

To find $P(a \leq X \leq b)$, we integrate over the range.

$$P(a \leq X \leq b) = \int_a^b \frac{3}{2}(1 - x^2)dx$$

If a and b are inside the range of 0 to 1, this is all you: calculate this integral. For any limits outside the range of [0,1], then the probability outside that interval is 0 (this where the otherwise condition comes in).

It’s also the case that no part of the density function can be negative, since no probability can be negative.

The cumulative distribution is similar, but the upper limit is x.

$$F(X) = P(X \geq 0) = \int_0^x \frac{3}{2}(1 - t^2)dt = \frac{3}{2}\left(x - \frac{1}{3}x^3\right)$$

This will be true for any value of x less than or equal to 1.

One important note about continuous distributions that differ from discrete ones. The $P(X = a)$ for a being any particular number is always 0, since the area under a single point is 0 (our estimating rectangle has zero width). That means that $P(X \leq a) = P(X < a)$. In other words, we don’t have to worry about whether we include the endpoints or not. The answer will be the same. This isn’t the case in discrete distributions where the inclusion of the equal sign can matter a great deal.

Let's look at some common continuous distributions.

Uniform distribution

The uniform distribution has the same probability for every value. Since the value of the area must be 1, the height of the distribution is everywhere equal to 1/range.

$$f(x) = \begin{cases} \frac{1}{360}, 0 \leq x \leq 360 \\ 0, \quad otherwise \end{cases}$$

Is an example of a uniform distribution. We can easily calculate the probability for a range of numbers inside this interval by finding $P(a \leq X \leq b) = \frac{1}{360}(b - a)$. This is what you would get from integrating over this same interval as we did before, assuming $a, b \in [0,360]$.

In general, the uniform distribution is given as $f(x) = \frac{1}{B-A}, A \leq x \leq B$ and 0 otherwise.

Normal distribution

The normal distribution is perhaps the most ubiquitous distribution in statistics. We will use it in many future statistics applications. The probability density function for a general normal distribution is given by

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

This distribution is valid on the entire set of real numbers.

The normal distribution is defined by two parameters: the mean μ and the standard deviation σ. Especially before we had technology to calculate the probability values for this distribution, we often used the standard normal distribution instead (and looked up probabilities in a table). That distribution sets the mean to 0 and the standard deviation to 1.

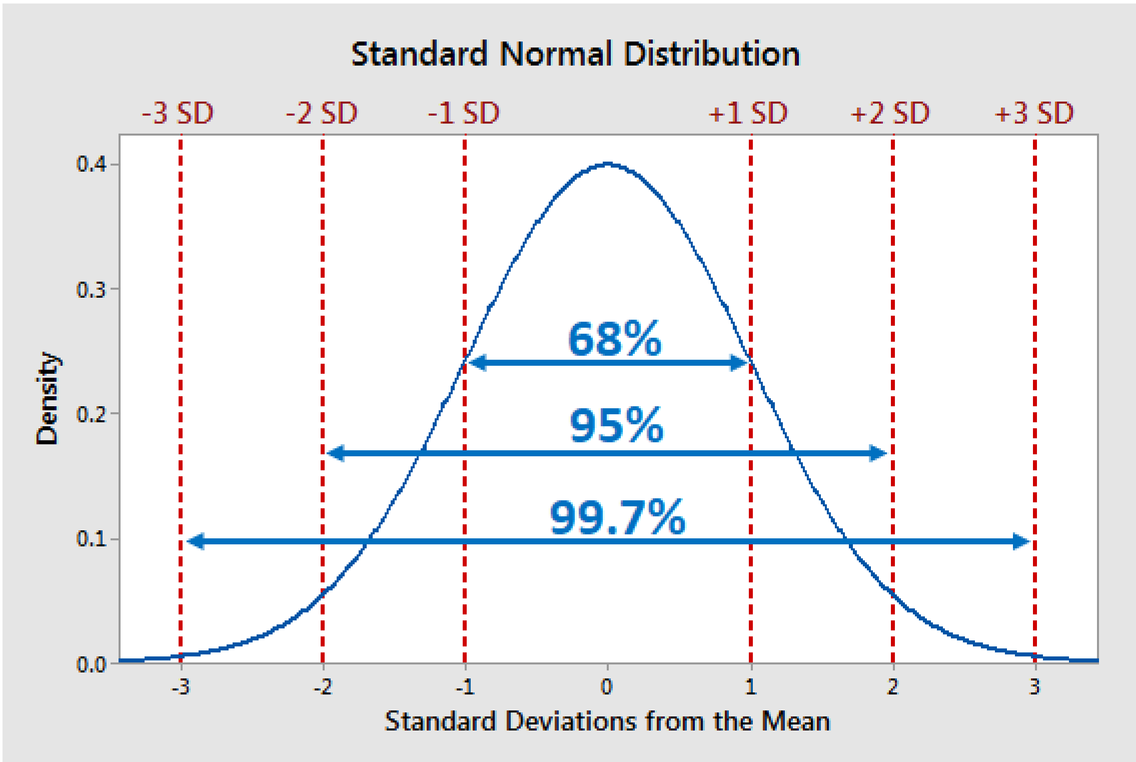
$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

It's common for statisticians to use z instead of x as the variable for the standard normal distribution. One can convert x to z using the formula for the standard score.

$$z = \frac{x - \mu}{\sigma}$$

Sometimes this formula is used to compare results from different normal distributions since this z value, the standard score, essentially converts an observation to units of the standard deviation. A standard score of 2 thus means that an observation is 2 standard deviations above the mean. A standard score of -3, would mean that the observation was three standard deviations below the mean.

The use of the normal distribution is so common that there are short-hand rules for estimating with the normal distribution. The **empirical rule** describes the probability between units of the standard deviation in a standard normal distribution.



We will tend to rely on precise calculations in this course, but it's a useful rule of thumb to keep in mind, and can be useful to checking the logic of calculations, to help you determine if something is reasonable. The empirical rule is also useful to see the relationship between "unusual" values and some of our definitions for outliers that correspond to being outside 2 standard deviations of the mean: they are basically based on the same general idea.

There is no compact expression for the cdf for the normal distribution since e^{-x^2} can't be easily integrated on its own. It can be done numerically, or one can use series methods to obtain an expression that can be worked with algebraically. These series methods are covered in Calculus II. In referring to the normal cdf in the abstract, statisticians will sometimes uses $\Phi(z)$. Recall that this means $\Phi(z) = P(X \leq z)$.

Exponential distribution

The exponential distribution was mentioned in the last lecture because of its relationship to the Poisson distribution. As we noted, it finds the probability that another Poisson process event will occur in a specific amount of time.

The pdf is given by

$$f(x;\lambda) = \lambda e^{-\lambda x}, x \geq 0$$

This distribution, we can integration over its interval to obtain a cdf.

$$F(x;\lambda) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

We will be able to prove a bit later that the mean is $E(X) = \frac{1}{\lambda}$ and the variance is $V(X) = \frac{1}{\lambda^2}$.

To find the **mode** of the distribution, you need the peak of the continuous distribution in most cases. You may be able to determine this graphically, particularly if the peak is at one endpoint of an interval of finite length, such as in our $f(x) = \frac{3}{2}(1 - x^2)$ distribution. If there is no peak inside the interval, then we need to check the endpoints (doing so, shows that the mode is at 0). When the mode is inside the interval, it will be at a critical point. So, we would need to take the derivative and find the critical point.

Consider the standard normal distribution $\phi(z) = e^{-z^2/2}$. The derivative is $\phi'(z) = -ze^{-z^2/2}$. If we set this equal to zero, the exponential part is never zero, so we are left with $z = 0$. And that is the location of the central peak in the standard normal distribution (it's the same as the mean in this case).

To find the median, we use the same technique we would need to find any other percentile (the median is the 50th percentile). Let's find the median of our example distribution $f(x) = \frac{3}{2}(1 - x^2)$.

Let η (eta) be the value at the given percentile p (expressed as a fraction or a decimal). Then η goes into the upper limit of the integral and we set the value of the integral equal to p. Or, set p equal to the cumulative distribution and set x equal to η. Then solve for η.

$$p = \int_0^{\eta} \frac{3}{2}(1 - x^2)dx$$

$$p = \frac{3}{2}\left(\eta - \frac{1}{3}\eta^3\right)$$

If we want to find the median, set $p = 0.50$ or $p = \frac{1}{2}$.

$$\frac{1}{2} = \frac{3}{2}\left(\eta - \frac{1}{3}\eta^3\right)$$

$$\frac{1}{3} = \eta - \frac{1}{3}\eta^3$$

$$\eta^3 - 3\eta + 1 = 0$$

We will need to solve this particular equation numerically and make sure the value we choose is inside the required interval [0,1].

This expression has three zeros, but only one is in the interval [0,1]. So $\eta = \tilde{\mu} \approx 0.34729$. This method is the same method used to apply the "inverse" distribution. For common distributions, the inverse function is precalculated. The inverse distribution takes a percentile and converts it to an observation in the distribution. We will likewise use this same method to find critical values for confidence intervals and rejection regions for a given confidence level from the normal distribution. These critical values (not the same thing as the calculus critical value) are usually notated with a subscript that represents their percentile: $z_{0.975}$ represents the value of the 97.5th percentile. We will discuss this more when we talk about confidence intervals and hypothesis tests after the first exam.

Mean and Variance

We can find the expected value a continuous distribution by multiplying the pdf by x and then integrating over the range the function is defined on.

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

Recall that if a probability distribution is non-zero on an interval smaller than this, then we only need to integrate over that smaller interval. Consider our example distribution we've been looking at.

$$f(x) = \frac{3}{2}(1 - x^2)$$

The expected value of this distribution is

$$E(X) = \int_0^1 x\left[\frac{3}{2}(1 - x^2)\right] dx = \frac{3}{2}\int_0^1 x - x^3 dx = \frac{3}{2}\left[\frac{1}{2}x^2 - \frac{1}{4}x^4\right]_0^1 = \frac{3}{2}\left(\frac{1}{4}\right) = \frac{3}{8} = \mu$$

We can perform this calculation for the normal distribution since the extra x makes the integral doable with u-substitution. We can show the mean is μ.

The variance is calculated similarly as we did in the discrete case. We multiply the pdf by $(x - \mu)^2$ and integrate to find the variance.

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

Looking at our sample distribution, we'll have to do a lot of algebra to find the variance.

$$V(X) = \sigma^2 = \frac{3}{2}\int_0^1 \left(x - \frac{3}{8}\right)^2 (1 - x^2)dx = \frac{3}{2}\int_0^1 -x^4 + \frac{3}{4}x^3 + \frac{55}{64}x^2 - \frac{3}{4}x + \frac{9}{64}dx =$$

$$\frac{3}{2}\left[-\frac{1}{5}x^5 + \frac{3}{16}x^4 + \frac{55}{192}x^3 - \frac{3}{8}x^2 + \frac{9}{64}x\right]_0^1 = \frac{3}{2}\left(-\frac{1}{5} + \frac{3}{16} + \frac{55}{192} - \frac{3}{8} + \frac{9}{64}\right) = \frac{3}{2}\left(\frac{19}{480}\right) = \frac{19}{320}$$

The standard deviation is then the square root of this.

This is a lot of algebra. As we saw with the discrete case, there is a bit of a shortcut we can take using moments. We can use the same formula for variance we did last time:

$$V(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$$

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

We can also use moments to find skew and kurtosis, two other statistics that can describe distributions.

Skew and kurtosis
There are a couple of other statistics we can calculate using moments. There is a skewness statistic based on a formula similar to the variance but with cubes.

$$skew(X) = E\left[\left(\frac{x_i - \mu}{\sigma}\right)^3\right] = \frac{E(X^3) - 3E(X)E(X^2) + 2[E(X)]^3}{[E(X^2) - [E(X)]^2]^{3/2}} = \frac{E(X^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3}$$

Right-skewed distributions produce a positive value for the skew. Left-skewed distributions produce a negative value for the skew parameter. If the distribution is symmetric, then the skewness parameter is 0. However, we can't say that if the skewness is 0 the distribution is symmetric.

Kurtosis is a measure of how much the distribution "bulges". Normally this value is compared to the normal distribution (which is why there is an extra -3 on the end of the formula).

$$\kappa(X) = kurt(X) = E\left[\left(\frac{x_i - \mu}{\sigma}\right)^4\right] - 3 = \frac{E(X^4) - 4E(X)E(X^3) + 6[E(X)]^2E(X^2) - 3[E(X)]^4}{(E(X^2) - [E(X)]^2)^2} - 3 =$$

$$\frac{E(X^4) - \mu E(X^3) + 6\mu^2\sigma^2 + 3\mu^4}{\sigma^4} - 3$$

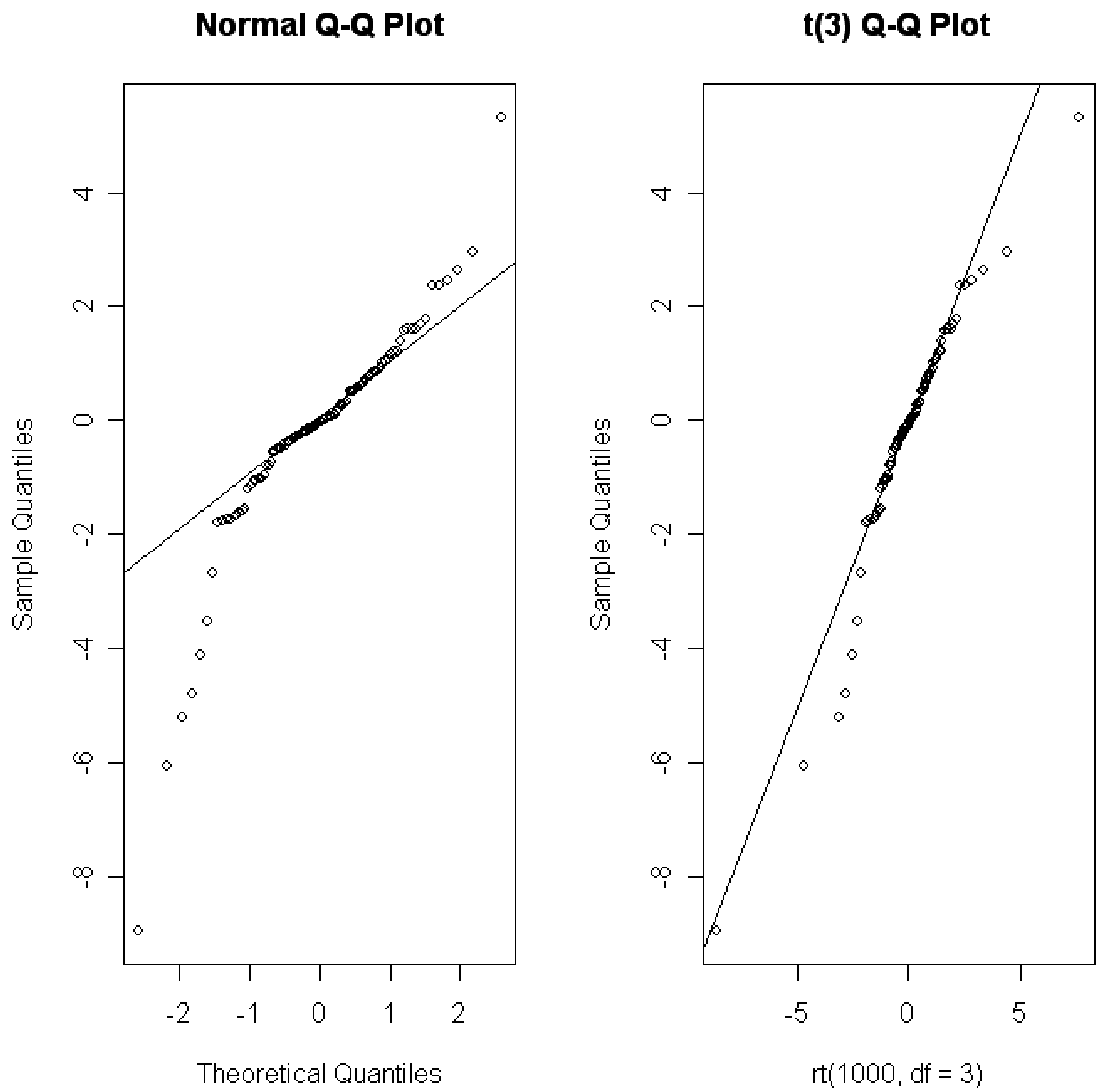
We won't calculate these on purpose very often in this course, but sometimes summary statistics will produce these values in the output and so we should be aware of them and what they mean.

The sample calculations for these values are somewhat similar. I've included a link in the references to the calculations both statistics using data rather than distributions. The differences in the formulas are similar to what you'd expect from the small sample size approximations we'll discuss below, and differences in the variance formula we used here compared to the variance formula for the population.

Before we talk briefly about a couple of other continuous probability distributions, we want to talk about a couple more things related to the normal distribution that we will make use of later in the course.

Normal probability plots
These graphs are also sometimes referred to as Q-Q plots. The idea is to compare the percentiles of observations in a data set with where is should be if it was normally distributed. The axes are sometimes given as percentiles, sometimes as standard scores, but the idea is the same in either case. And example is shown below.

If your data is normally distributed, then it should fall (more or less) on a straight line, which is also usually draw on the graph. The graph on the left is data that is not normally distributed (the data strongly deviations from the expected straight line on the left side). The second graph shows something that is normally distributed. A point or two that deviates from the line at the ends is okay, but the bulk of the data sits at or near the line everywhere else.



We'll discuss how to create this plot later in the course when we need to assess our assumptions about whether our data is normal enough.

In some cases we can approximate discrete distributions as continuous distributions. We can look at two cases: Approximating the binomial distribution, and approximating the Poisson distribution.

To approximate the binomial distribution with the normal distribution, we generally need to the sample size to be sufficiently large. By that, the standard is usually given in terms of the variance of the binomial. We want $npq \geq 5$ for a rough estimate or $npq \geq 10$ for a more accurate estimate. Some authors will make this a two-part test instead, asking for both $np > 10$ and $nq > 10$. The variance test covers both these cases since it includes both p and q. This rule is thumb is required for the mode of the distribution to be far enough away from 0 or 1 to avoid a strong skew. We use the mean from the binomial $\mu = np$ as the center of the normal distribution, and the standard deviation as $\sigma = \sqrt{npq}$.

The second thing we have to keep in mind is we have to account for any value that could be rounded to the required integer. So if we want to estimate $P(X = 45) = B(45; 100,0.5)$ we would need to include $44.5 \leq x < 45.5$ when we estimate with the normal distribution.

Let's compare. Our scenario here could be flipping a fair coin 100 times, finding the probability of getting exactly 45 heads.

$$B(45; 100,0.5) \approx \frac{1}{\sqrt{2\pi(100)(0.5)(0.5)}} \int_{44.5}^{45.5} e^{-\frac{(x-50)^2}{2(100)(0.5)(0.5)}} dx$$

The normal approximation (integrated numerically) gives 0.04839
The binomial distribution gives 0.04847.

The variance here is 25, which satisfies our rule of thumb and we can see these values are pretty close. This method is pretty useful if we need to find a range of values with the binomial distribution, which can be pretty tedious if we have to use a simple calculator or each one.

Similarly, there are cases where we can use the normal distribution to approximate the Poisson distribution, with somewhat similar restrictions. We want the distribution to be sufficiently normal (not skewed), and use the mean and variance of the original distribution in place of μ and σ. Different authors have slightly different standards for using this approximation. One such standard is $\lambda > 20$. If in doubt, check the shape of the distribution.

Other continuous distributions
There are many other continuous distributions, some of which we will encounter later in the course in specific contexts. We will go through a couple of them. All the ones discussed here are programmed in R, along with the discrete ones from last lecture.

Gamma distribution
The gamma function is given by

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

This function has the following properties:

- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1), \alpha > 1$
- If $\alpha = n$ (integer), then $\Gamma(\alpha) = (n - 1)!$
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

We use this in the standard gamma distribution

$$f(x; \alpha) = \begin{cases} \dfrac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)}, x \geq 0 \\ 0, \qquad otherwise \end{cases}$$

A slightly more general gamma function actually has two parameters (the second one is set to 1 for the standard gamma).

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{x^{\alpha-1}e^{-x}}{\beta^{\alpha}\Gamma(\alpha)}, x \geq 0 \\ 0, \qquad otherwise \end{cases}$$

$$E(X) = \mu = \alpha\beta, V(X) = \sigma^2 = \alpha\beta^2$$

You will sometimes encounter references to the incomplete gamma function (what this means may vary some, but often omits the gamma function from the denominator (and so if off by a scalar multiple). Look at the Devore reference (#1 below) for more on this.

Beta distribution
This distribution is similar to the gamma, but is not defined on the same interval. It's an important distribution in Bayesian statistics.

$$f(x;\alpha,\beta,A,B) = \frac{1}{B-A} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \left(\frac{x-A}{B-A}\right)^{\alpha-1} \left(\frac{B-x}{B-A}\right)^{\beta-1}, A \leq x \leq B$$

The standard beta distribution sets $A = 0, B = 1$.

$$E(X) = \mu = A + (B-A) \cdot \frac{\alpha}{\alpha+\beta}, V(X) = \sigma^2 = \frac{(B-A)^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

The Chi-Squared distribution (χ^2)
The χ^2 distribution is important for tests of independence which we'll encounter at the end of semester, and for sample standard deviation distributions.

$$f(x;\nu) = \frac{1}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-x/2}, x \geq 0$$

The Greek letter ν (nu) is the degrees of freedom parameter.

Weibull distribution

$$f(x;\alpha,\beta) = \frac{\alpha}{\beta^{\alpha}} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}, x \geq 0$$

Lognormal distribution

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma x} e^{[\ln(x)-\mu]^2/2\sigma^2}, x \geq 0$$

Pareto distribution

$$f(x;\alpha,\lambda) = \lambda\alpha^{\lambda}x^{-(\lambda+1)}, x \geq \alpha$$

Logistic distribution

$$f(x;\mu,\nu) = \frac{e^{(x-\mu)/\nu}}{\nu(1+e^{(x-\mu)/\nu})^2}$$

Some distributions are so complex, it's easier to describe them in terms of how they are derived rather than the formulas themselves. We'll encounter these later in the course, and we'll use technology to evaluate them. Two examples are the F and Student-t distribution.

The *Student-t distribution* is derived from the standard normal distribution and the χ^2 distribution (with n degrees of freedom. The result is that the t-distribution also depends on a degrees of freedom.

$$T_n = \frac{Z}{\sqrt{\frac{\chi_n^2}{n}}}$$

The *F-distribution* is derived from two different χ^2 distributions with degrees of freedom n and m respectively.

$$F_{n,m} = \frac{\frac{\chi_n^2}{n}}{\frac{\chi_m^2}{m}}$$

The t-distribution looks similar to the standard normal distribution, but has a larger kurtosis score (it's fatter in the tails). As the degrees of freedom increase, it looks more like the normal distribution.

The F-distribution, in most of the applications we will consider, is skewed right, like the χ^2 distribution it is based on.

1. https://faculty.ksu.edu.sa/sites/default/files/probability_and_statistics_for_engineering_and_the_sciences.pdf
2. https://assets.openstax.org/oscms-prodcms/media/documents/IntroductoryStatistics-OP_i6tAI7e.pdf
3. https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm
4. https://www.randomservices.org/random/expect/Skew.html
5. https://statisticsbyjim.com/probability/empirical-rule/
6. https://www.statmethods.net/advgraphs/probability.html
7. Introduction to Probability and Statistics for Engineers and Scientists, 5th ed., Sheldon M. Ross, 2014.
8. http://www.socr.ucla.edu/Applets.dir/NormalApprox2PoissonApplet.html