PLSC 502: “Statistical Methods for Political Research”

Continuous Probability Distributions
September 29, 2016

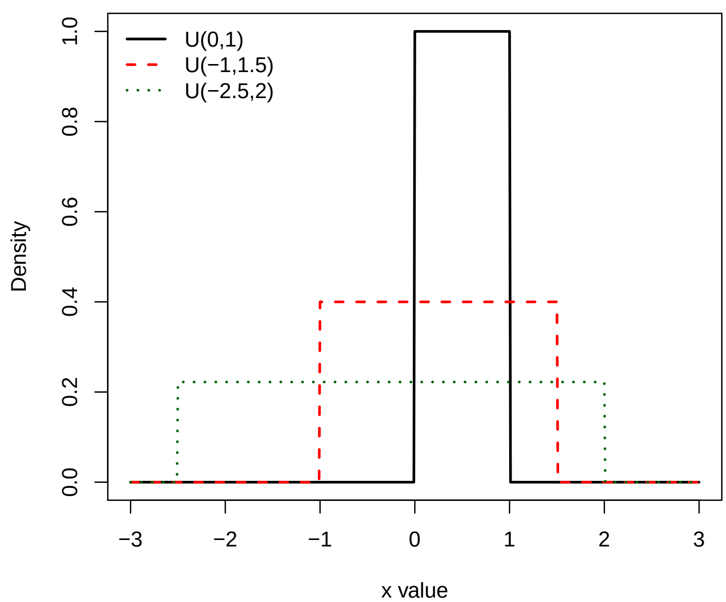
The Uniform Distribution

If a variate X is *uniformly* distributed on the range $[a, b]$, then it's PDF is:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{for } x < a \text{ or } x > b. \end{cases} \quad (1)$$

We write this as $X \sim U(a, b)$; this illustrates that the uniform distribution is a two-parameter distribution, where the parameters are the minimum and the maximum (the “bounds”), which may fall anywhere in $\mathcal{R}$. We can most easily think of this as a rectangular shape of probability, located between a and b in the real number line, with “length” $b-a$ and “height” equal to $\frac{1}{b-a}$. Some examples are illustrated in Figure 1.

Figure 1: Various Uniform PDFs

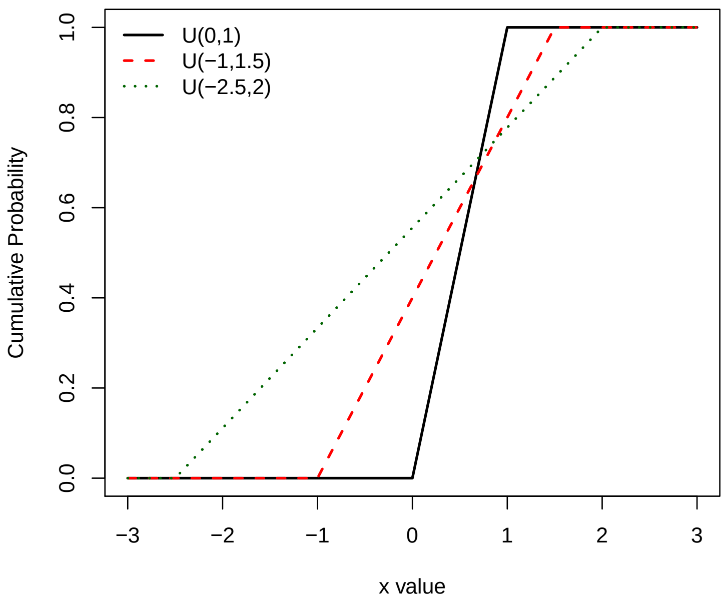


A draw from a uniform distribution has equal probability of falling anywhere on the real line between a and b. Not surprisingly, the CDF takes on an especially simple form:

$$F(x) = \int f(x)dx = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \leq x < b \\ 1 & \text{for } x \geq b \end{cases}$$

Because we are just integrating over a "constant" value of $f(x)$, the CDF looks like a sloped line extending from 0 to 1 over the range from a to b; this is illustrated in Figure 2:

Figure 2: Various Uniform CDFs



All of this means that uniform variates have especially simple characteristics. For example, for a uniform (a, b) variate X,

$$\text{E}(X) = \check{X} = \frac{a+b}{2}$$

and

$$\text{Var}(X) = \frac{(b-a)^2}{12}.$$

Note as well that the mode of a uniform variate is *any* value in $[a, b]$, and that its skewness is zero.

The Standard Uniform Distribution

A uniform variate with $a = 0$ and $b = 1$ is often referred to as a *standard uniform* variable. The standard uniform has the interesting (and, at times, useful) property that, for a standard uniform variate X,

$$X \sim 1 - X \sim U(0, 1).$$

That is, the variable X and its complement have the same distribution.

The standard uniform distribution turns out to be very useful in generating other random variates, since it is the range over which a probability varies. Thus, if we want to generate random (equiprobable) data from some distribution, we start with a standard uniform variate, and then transform that variate by the inverse of the relevant PDF. More on this below.

The Normal Distribution

We are all used to seeing normal distributions described, and to hearing that something is "normally distributed." We know that a normal distribution is "bell-shaped," and symmetrical, and probably that it has some mean and some standard deviation. We'll try to dig a bit deeper than that today.

Formally, if X is a *normally distributed* variate with mean μ and variance σ^2, then:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{2}$$

We denote this $X \sim N(\mu, \sigma^2)$, and say "X is distributed normally with mean mu and variance sigma squared." The symbol ϕ is often used as a shorthand to represent the normal density in (2):

$$X \sim \phi_{\mu,\sigma^2}.$$

The corresponding normal CDF – which is the probability of a normal random variate taking on a value less than or equal to some specified number – is (as always) the indefinite integral of (2). This has no simple closed-form solution[1] so we typically just write:

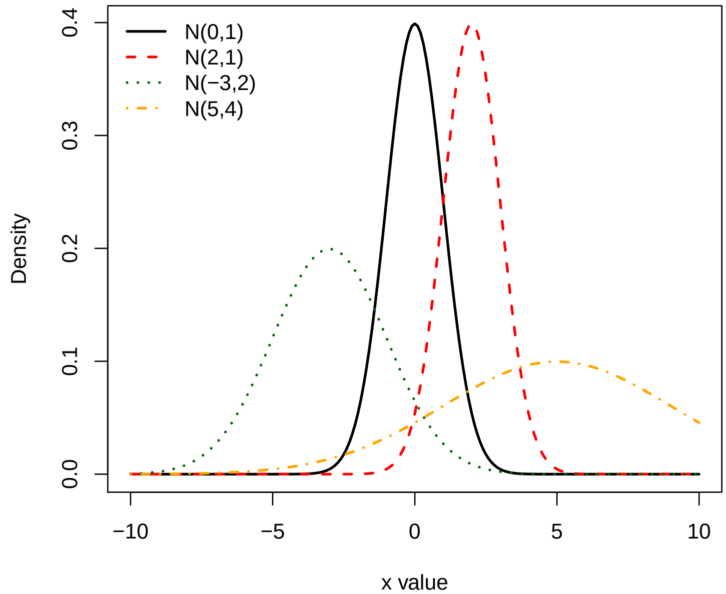
[1] If you really want to know, the CDF is

$$F(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right],$$

where $\text{erf}(\cdot)$ is the "error function"

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

Figure 3: Various Normal Densities



$$F(x) \equiv \Phi_{\mu,\sigma^2}(x) = \int \phi_{\mu,\sigma^2} f(x) dx. \quad (3)$$

Bases for the Normal Distribution

The most common justification for the normal distribution has its roots in the *central limit theorem*. Consider $i = \{1, 2, ...N\}$ independent, real-valued random variates X_i, each with finite mean μ_i and variance $\sigma_i^2 > 0$. If we consider a new variable X defined as the sum of these variables:

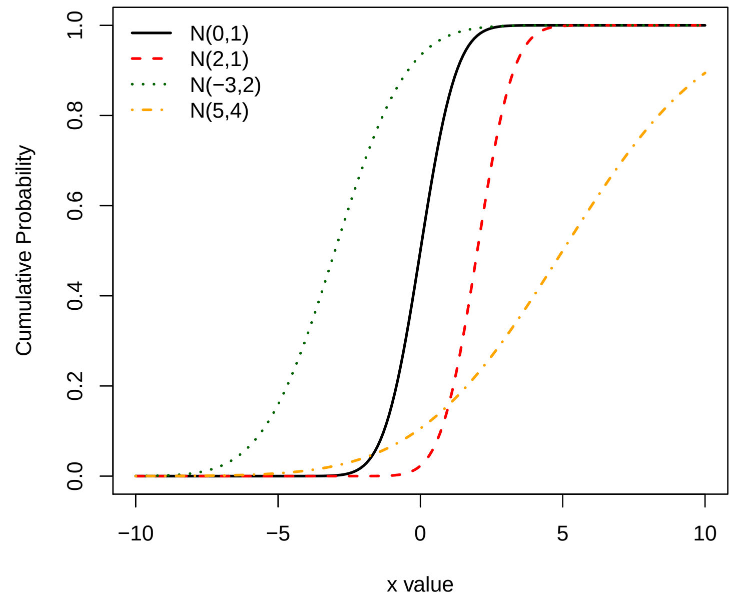
$$X = \sum_{i=1}^{N} X_i,$$

then we know that

$$\begin{aligned} \mathrm{E}(X) &= \sum_{i=1}^{N} \mu_i \\ &= \mu < \infty \end{aligned}$$

and

Figure 4: Various Normal CDFs (parameters as in Figure 3)



$$\begin{aligned} \mathrm{Var}(X) &= \sum_{i=1}^{N} \sigma_i^2 \\ &= \sigma^2 < \infty. \end{aligned}$$

The central limit theorem states that:

$$\lim_{N\to\infty} X = \lim_{N\to\infty} \sum_{i=1}^{N} X_i \xrightarrow{D} N(\cdot) \tag{4}$$

where the notation $\xrightarrow{D}$ indicates convergence in distribution. That is, as N gets sufficiently large, the distribution of the sum of N independent random variates with finite mean and variance will converge to a normal distribution. As such, we often think of a normal distribution as being appropriate when the observed variable X can take on a range of continuous values, and when the observed value of X can be thought of as the product of a large number of relatively small, independent "shocks" or perturbations.

Properties of the Normal Distribution

- A normal variate X has support in $\mathfrak{R}$.
- The normal is a two-parameter distribution, where $\mu \in (-\infty, \infty)$ and $\sigma^2 \in (0, \infty)$.
- The normal distribution is always symmetrical ($M_3 = 0$) and mesokurtic.
- The normal distribution is preserved under a linear transformation. That is, if $X \sim N(\mu, \sigma^2)$, then $aX + b \sim N(a\mu + b, a^2\sigma^2)$. (Why? Recall our earlier results on μ and σ^2).

The Standard Normal Distribution

One linear transformation is especially useful:

- $b = \frac{-\mu}{\sigma}$,
- $a = \frac{1}{\sigma}$.

This yields:

$$\begin{aligned} ax + b &\sim N(a\mu + b, a^2\sigma^2) \\ &\sim N(0, 1) \end{aligned}$$

This is the *standard normal density function*. We often denote this $\phi(\cdot)$, and say that "X is distributed as standard normal." We can also get this by transforming ("standardizing") the normal variate X...

- If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{(x-\mu)}{\sigma} \sim N(0, 1)$.
- The density function then reduces to:

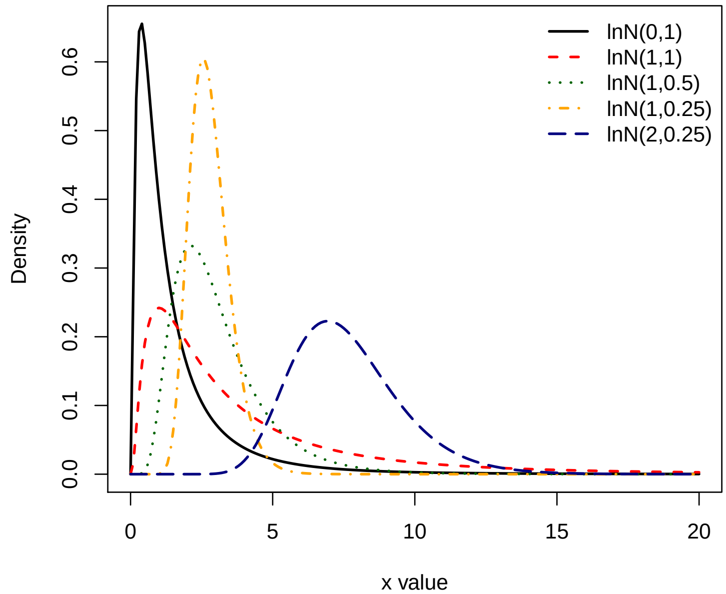
$$f(z) = \equiv \phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(z)^2}{2}\right]$$

Similarly, we often write the CDF for the standard normal as $\Phi(\cdot)$.

Why do we care about the normal distribution?

The normal distribution's importance lies in its relationship to the central limit theorem. As we'll discuss at more length later, the central limit theorem means that as one's sample size increases, the distribution of sample means (or other estimates) approaches a normal distribution.

Figure 5: Five Log-Normal Densities



The Log-Normal Distribution

Mathematically, the exponential function (e^x) can be useful, for a variety of reasons. Exponentiating a normally-distributed variable yields the *log-normal* distribution, so named because one can take the (natural) logarithm of such a variable to get a normal variate. Formally, for a normal variate X with mean μ and variance σ^2, the log-normal distribution is just the distribution of

$$Y = \exp(X) \sim \text{LogN}(m, s^2) \tag{5}$$

A lognormal variable has PDF equal to

$$f(y) = \frac{1}{ys\sqrt{2\pi}} \times \exp\left[\frac{-(\ln y - m)^2}{2s^2}\right]. \tag{6}$$

Such a variable has support on the non-negative real line: $y \in [0, \infty)$; as for the normal distribution, the CDF does not have a closed-form solution. Also like the normal distribution, the log-normal is a two-parameter distribution; m is the central tendency parameter and s the variance. The log-normal distribution is always positively skewed, though the precise degree of skewness depends on the precise value of s.[2] Illustrations of five log-normal densities with different values of m and s are presented in Figure 5

[2]Specifically, the coefficient of skewness of the log-normal distribution is equal to $[\exp(s^2) - 2]\sqrt{\exp(s^2) - 1}$, indicating that the skewness is increasing in s.

The importance of the log-normal lies in the fact that it is the distribution of the exponent of a normal variate. That means that, if we believe that a strongly-skewed variable should be logged, and that the subsequent (transformed) variate is normally distributed, then the original variate must be log-normal.

The χ^2 Distribution

The χ^2, t and F distributions can be derived from various products of normally-distributed variables. All three are used extensively in statistical inference and applied statistics, so it's useful to understand them in a bit of depth.

Gill discusses the χ^2 distribution as a special case of the gamma PDF. That's fine, but there's actually a much more intuitive way of thinking about it, and one that comports more closely with how it is (most commonly) used in statistics. Formally, a variable W that is distributed as χ^2 with k degrees of freedom has a density of:

$$\begin{aligned} f(w) &= \frac{1}{2^k\Gamma(k)} w^k \exp\left[\frac{-w}{2}\right] \\ &= \frac{w^{\frac{k-2}{2}} \exp(\frac{-w}{2})}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} \end{aligned} \tag{7}$$

where $\Gamma(k) = \int_0^\infty t^{k-1} \exp(-t)\, dt$ is the gamma integral (see, e.g., Gill, p. 222). As with the normal distribution, the need to write the distribution in this fashion reflects the fact that it has no closed-form solution. The corresponding CDF is

$$F(w) = \frac{\gamma(k/2, w/2)}{\Gamma(k/2)} \tag{8}$$

where $\Gamma(\cdot)$ is as before and $\gamma(\cdot)$ is the lower incomplete gamma function. We write this[3] as $W \sim \chi^2_k$, and say "W is distributed as chi-squared with k degrees of freedom."

The chi-square distribution is a one-parameter distribution defined only on the nonnegative real line, $W \in [0, \infty)$. It is positively skewed, with

$$\mathrm{E}(W) = k$$

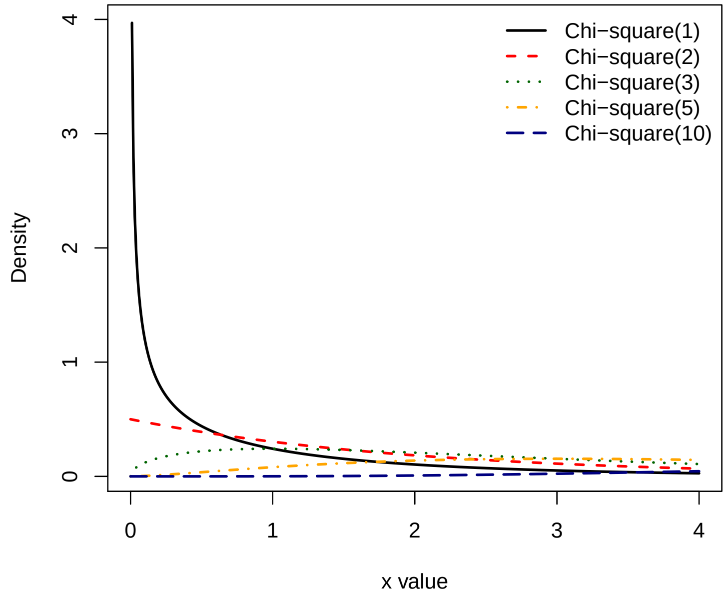
and

$$\mathrm{Var}(W) = 2k.$$

Figure 6 presents five χ^2 densities with different values of k.

[3]One also occasionally sees $W \sim \chi^2(k)$, with the degrees of freedom in parentheses.

Figure 6: Five χ^2 Densities



Characteristics of the χ^2 Distribution

More importantly, one needs to remember two key things about the chi-square distribution:

1. If $Z \sim N(0,1)$, then $Z^2 \sim \chi^2_1$. That is, *the square of a $N(0,1)$ variable is chi-squared with one degree of freedom.*

2. If W_j and W_k are independent χ^2_j and χ^2_k variables, respectively, then $W_j + W_k$ is $\sim \chi^2_{j+k}$; this result can be extended to any number of independent chi-squared variables.

The first of these is key, since it points out that the square of a standard normal variate is a one-degree-of-freedom chi-square variable. This explains why (e.g.) a chi-squared variate only has support on the nonnegative real numbers. The second point is also tremendously useful to know, in that it has a number of valuable corollaries. For example, it implies that

- if $W_1, W_2, ...W_k$ are all independent χ^2_1 variables, then $\sum_{i=1}^{k} W_i \sim \chi^2_k$. (The sum of k independent chi-squared variables is chi-squared with k degrees of freedom).

- By extension, the sum of the squares of k independent $N(0,1)$ variables are also $\sim \chi^2_k$.

All this means that, if we know a variable to be normally distributed, we can consider its squared, standardized values to be χ^2_1, and the sums of k of such variables to be χ^2_1.

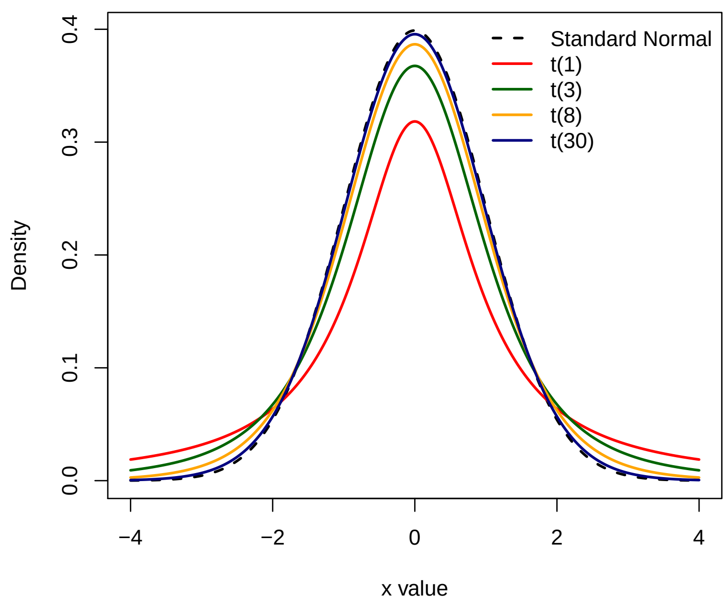
Student's t Distribution

(Insert stories about St. James' Gate here). For a variable X which is distributed as t with k degrees of freedom, the PDF function is:

$$f(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\,\Gamma(\frac{k}{2})}\left(1 + \frac{x^2}{k}\right)^{-(\frac{k+1}{2})} \tag{9}$$

where once again $\Gamma(\cdot)$ is the gamma integral. We write $X \sim t_k$, and say "X is distributed as Student's t with k degrees of freedom." The CDF is complicated, so I won't go into it here; Figure 7 presents t densities for five different values of k, along with a standard normal density for comparison.

Figure 7: Five t Distributions, with $N(0,1)$ for Comparison



Note a few things about t:

- The mean/mode/median of a t-distributed variate is zero, and its variance is $\frac{k}{k-2}$.
- t looks like a standard normal distribution (symmetrical, bell-shaped) but has thicker "tails" (read: higher probabilities of draws being relatively far from the mean/mode). However...
- ...as k gets larger, t converges to a standard normal distribution; at or above $k = 30$ or so, the two are effectively indistinguishable.

The importance of the t distribution lies in its relationship to the normal and chi-square distributions. In particular, if $Z \sim N(0,1)$ and $W \sim \chi^2_k$, and Z and W are independent, then

$$\frac{Z}{\sqrt{W/k}} \sim t_k$$

That is, the ratio of an $N(0,1)$ variable and a (properly transformed) chi-squared variable follows a t distribution, with d.f. equal to the number of d.f. of the chi-squared variable. Of course, this also means that

$$\frac{Z^2}{W/k} \sim t_k.$$

Since we know that $Z^2 \sim \chi^2_1$, this means that another derivation of the t distribution is as a ratio of a χ^2_1 variate and a χ^2_k variate. As we'll see in a week or so, that turns out to be quite important, and useful.

The F Distribution

An F distribution is best understood as the ratio of two chi-squared variates. Formally, if X is distributed as F with k and ℓ degrees of freedom, then the PDF of X is:

$$f(x) = \frac{\left(\frac{k\,x}{k\,x+\ell}\right)^{k/2}\left(1-\frac{k\,x}{k\,x+\ell}\right)^{\ell/2}}{x\;\mathrm{B}(k/2,\ell/2)} \tag{10}$$

where $\mathrm{B}(\cdot)$ is the "beta function.[4] The corresponding CDF is (once again) complicated, so we'll skip it. We write $X \sim F_{k,\ell}$, and say "X is distributed as F with k and ℓ degrees of freedom."

The F is a two-parameter distribution, with degrees of freedom parameters (say k and ℓ), both of which are limited to the positive integers. An F variate X has support on the non-negative real line; it has expected value equal to

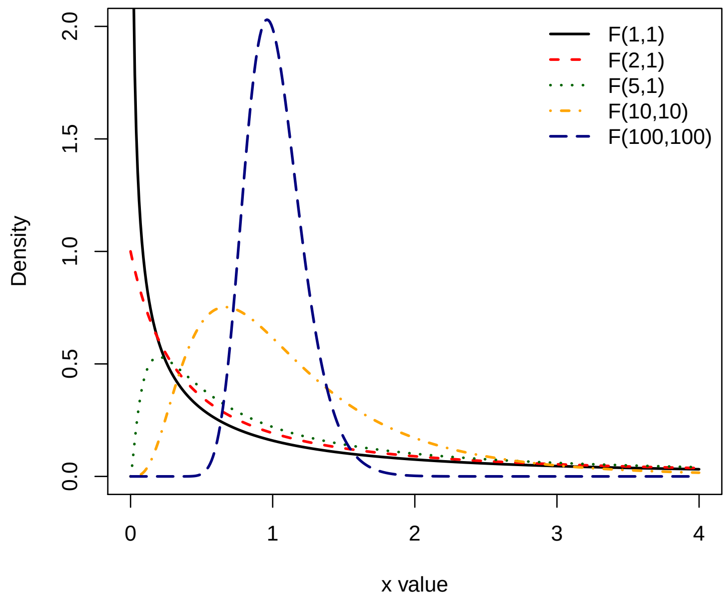
$$\mathrm{E}(X) = \frac{\ell}{\ell-2},$$

which implies that the mean of an F-distributed variable converges on 1.0 as $\ell \to \infty$. Likewise, it has variance

$$\mathrm{Var}(X) = \frac{2\,\ell^2\,(k+\ell-2)}{k(\ell-2)^2(\ell-4)},$$

which bears no simple relationship to either k or ℓ. It is (generally) positively skewed. Examples of some F densities with different values of k and ℓ are presented in Figure 8.

Figure 8: Five F Densities



As I noted a minute ago, if W_1 and W_2 are independent and $\sim \chi^2_k$ and χ^2_ℓ, respectively, then

$$\frac{W_1}{W_2} \sim F_{k,\ell}$$

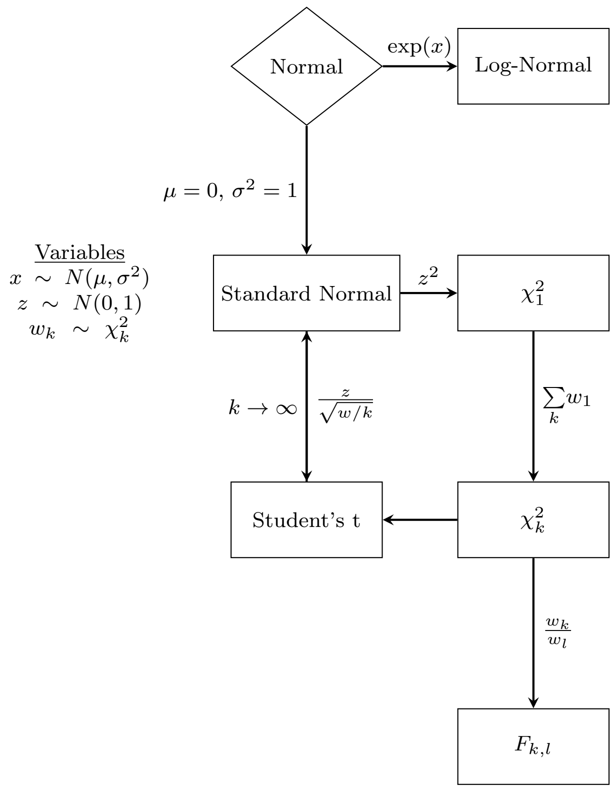
That is, the ratio of two chi-squared variables is distributed as F with d.f. equal to the number of d.f. in the numerator and denominator variables, respectively. This implies (at least) a couple of interesting things:

- If $X \sim F(k, \ell)$, then $\frac{1}{X} \sim F(\ell, k)$ (because $\frac{1}{X} = \frac{1}{(W_1/W_2)} = \frac{W_2}{W_1}$).
- The square of a t distributed variable is $\sim F(1, k)$ (*why*? – take the formula for t, and square it...)

The substantive importance of all these distributions will become apparent as we move on to sampling distributions, in our quest to (eventually) do statistical inference.

[4] That is, $\mathrm{B}(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\, dt$.

Summary: Relationships Among Continuous Distributions



For a much, much more complete (and interactive / clickable) "map" like this, check out http://www.math.wm.edu/ leemis/chart/UDR/UDR.html.

Generating Random Variates, etc.

The simplest way to draw from a probability distribution is to start with a uniform$(0, 1)$ random variable, then use the inverse of the PDF in question to map "back" from the probability space to the space of the variable.

As an example, suppose we wanted to draw 1000 times from a normal distribution with a mean of five and a variance of two (that's $X \sim N(5, 2)$). This means that the standard deviation of the distribution is $\sqrt{2}$, or about 1.414. If we start with a uniform variate (call it U) that is distributed uniform$(0, 1)$, we can simply do the following:

- Draw 1000 times from $U(0, 1)$,
- Calculate the inverse standard normal of each of those draws; this will give us a variate (call it Y) that is $Y \sim N(0, 1)$, and then

- "Rescale" Y to X by multiplying each value times 1.414 and then adding five, yielding a variate $X \sim N(5, 2)$.

Of course, most software packages have built-in routines for generating random variates of various types. So, for example, in the normal example we just discussed, we have this in R:

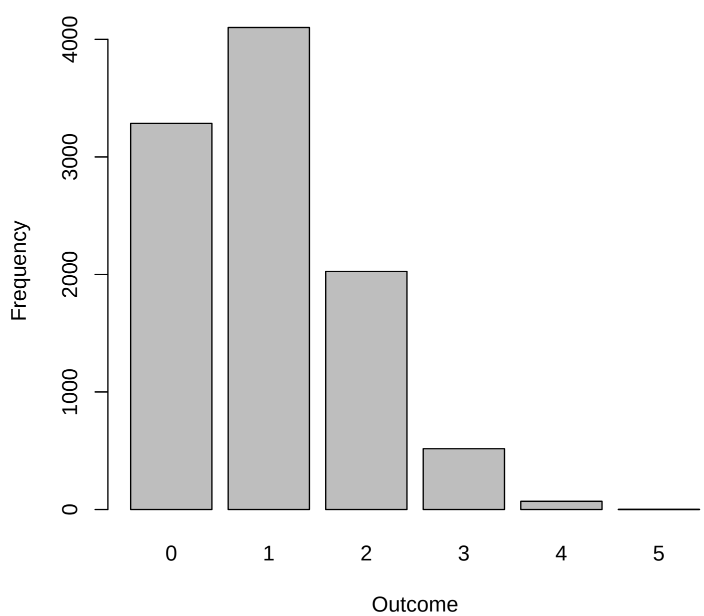
```
> Xnorm<-rnorm(1000, mean=5,sd=sqrt(2))
```

Similarly, if we want to draw 10000 observations from a binomial distribution with $n = 5$ and $\pi = 0.2$, we use

```
> Xbinom5point2<-rbinom(10000,5,0.2)
```

which gives you something like the following:

Figure 9: Ten Thousand Draws from a Binomial$(5, 0.2)$ Distribution



A handy table of R and Stata commands for generating random variates is:

Table 1: Commands for Generating Random Variates

Distribution	R	Stata
Binomial(n, π)	`rbinom()`	`rndbin`*
Geometric(π)	`rgeom()`	?
Negative Binomial(n, π)	`rnbinom()`	?
Poisson(λ)	`rpois()`	`rndpoi`*
Uniform$(0, 1)$	`runif()`	`uniform()`
Normal$(0, 1)$	`rnorm()`	`invnorm(uniform())`
Lognormal$(0, 1)$	`rlnorm()`	`xlgn`*
Student's $t(k)$	`rt()`	`rndt`*
Chi-Square(k)	`rchisq()`	`rndchi`*
$F(k, \ell)$	`rf`	`rndf`*

Note: Stata commands marked with an asterisk are from Hilbe's `rnd` group of commands. "?"s indicate that I'm not aware of any "canned" way of doing this, though one can always generate them "by hand" using the appropriate PDF function.

A Word About Generating "Random" Variates

When we generate random draws from a distribution, they're not really random in the truest sense. Instead, the values generated by random number generators are (usually) what we refer to as "psuedorandom," in that they start with some original number or set of numbers (called a *seed*) and then use functions of that number to generate random numbers from it.

I'm not going to go into any length on the details of random number generation, fascinating though they may be. The thing to bear in mind is that, given a particular seed, *all pseudo-random numbers generated from that seed will occur in exactly the same order*. That is, the seed determines the sequence of random numbers. The importance of this fact is manifold, but one thing that it allows one to do is to go back and replicate exactly what one has done before, even though the values generated are "random," *so long as we set the seed to some known number(s) at the outset* (and use the same pseudo-random number generation algorithm).

Consider this R example:

```
> seed<-3229 # calling "seed" some thing
> set.seed(seed) # setting the system seed
> rt(3,1) # three draws from a t distrib. w/1 d.f.
[1] -0.1113 -0.7306  1.9839
> seed<-1077
> set.seed(seed) # resetting the seed
> rt(3,1) # different values for the draws
```

```
[1]   -0.5211    7.9161 -155.3186
> seed<-3229 # original seed
> set.seed(seed)
> rt(3,1) # identical values of the draws
[1] -0.1113 -0.7306  1.9839
```

The moral of the story: Any time you are generating random draws from anything, always be sure to set the seed first, so that the results that you get are perfectly replicable. Type `?set.seed` in R or `help set seed` in Stata for more on these matters.