1: Probability and Distribution Basics

Bertille Antoine
(adapted from notes by Brian Krauth and Simon Woodcock)

Random Variables

Econometrics is the application of economic models to economic data. Economic data are measurements of some aspect of the economy. We usually think of them as being the outcome of some random process, sometimes called the **data generating process** or DGP. We don't observe the DGP. As econometricians, our objective is to learn about the DGP.

If X is an outcome of the DGP, we call it a **random variable** because its value is uncertain until the data are observed. I will use capital letters for the name of a random variable, and lower case letters for the values it takes.

Definition 1 *A random variable X is* ***discrete*** *if the set of possible outcomes is finite or countably infinite. X is* ***continuous*** *if the set of possible outcomes is uncountable.*

Probability Distributions

We use probabilities to describe uncertainty about the specific outcome a random variable will take. Informally, we call the set of possible outcomes of a random variable X and associated probabilities the **distribution** of X. We summarize this information with a **probability distribution function** when X is discrete, or a **probability density function** when X is continuous. In either case, we abbreviate this function as pdf, and denote it $f_X(x)$.

Definition 2 *For a discrete random variable X, $f_X(x) = \Pr(X = x)$ and satisfies:*

$$\begin{aligned} 0 &\leq f_X(x) \leq 1 \\ \sum_x f_X(x) &= 1. \end{aligned}$$

Definition 3 *For a continuous random variable X, $f_X(x)$ satisfies:*

$$\begin{aligned} f_X(x) &\geq 0 \\ \Pr(a \leq x \leq b) &= \int_a^b f_X(x)\, dx \geq 0 \\ \int_X f_X(x)\, dx &= 1. \end{aligned}$$

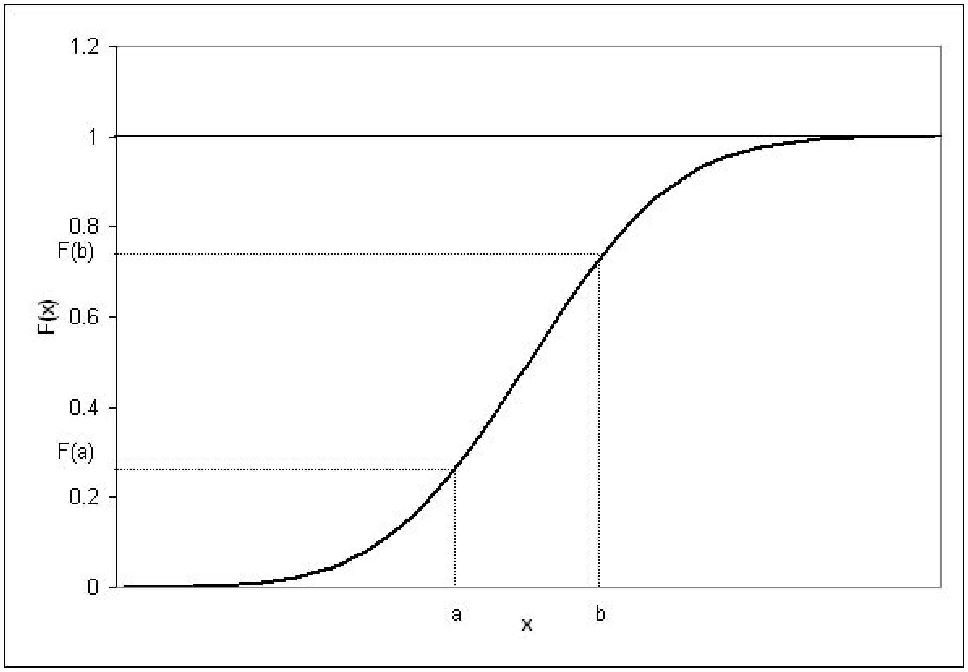

Figure 1: $\Pr(a < x \leq b) = F_X(b) - F_X(a)$

Cumulative Distribution Functions

For any random variable X, the **cumulative distribution function** (cdf), is

$$F_X(x) = \Pr(X \leq x).$$

Definition 4 *For a discrete random variable* X, $F_X(x) = \sum_{X \leq x} f_X(X)$. *For a continuous random variable* X, $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$. *In both cases,* $F_X(x)$ *satisfies:*

$$\begin{aligned} 0 &\leq F_X(x) \leq 1 \\ \textit{if } x &> y, \textit{ then } F_X(x) \geq F_X(y) \\ \lim_{x\to\infty} F_X(x) &= 1 \\ \lim_{x\to-\infty} F_X(x) &= 0 \end{aligned}$$

Notice that the definition of the cdf implies $\Pr(a < x \leq b) = F_X(b) - F_X(a)$. See Figure 1.

Transformations of Random Variables

Often we know the probability distribution of a random variable X, but our interest centers on some function of X. In this case, we need to know how to map the probability distribution of X into the distribution of a function of X. Let's define $\mathcal{X} = \{x : f_X(x) > 0\}$ and $\mathcal{Y} = \{y : y = g(x) \text{ for some } x \in \mathcal{X}\}$. The set $\mathcal{Y}$ is called the **image** of g. The sets $\mathcal{X}$ and $\mathcal{Y}$ describe the possible values that the random variables X and Y can take. We call this the **support** of X and Y.

Theorem 5 *Let X have cdf $F_X(x)$ and $Y = g(X)$. Then*
a. If g is an increasing function on $\mathcal{X}$ then $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.
b. If g is a decreasing function on $\mathcal{X}$ and X is a continuous random variable, $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.

Example 6 *Suppose $X \sim f_X(x) = 1$ for $0 < x < 1$ and 0 otherwise (this is called the uniform$(0,1)$ distribution). Here $\mathcal{X} = (0,1)$. Now define $Y = g(X) = -\ln X$. Clearly, $g(x)$ is decreasing, $\mathcal{Y} = (0,\infty)$, and $g^{-1}(y) = e^{-y}$. Therefore, for $y > 0$:*

$$F_Y(y) = 1 - F_X\left(g^{-1}(y)\right) = 1 - F_X\left(e^{-y}\right) = 1 - e^{-y}.$$

When the cdf of Y is differentiable, we can obtain the pdf of Y by differentiation. The result is summarized by the following theorem (its proof is a simple application of the chain rule and Theorem 5).

Theorem 7 *If X is a continuous random variable with pdf $f_X(x)$ and $Y = g(X)$ is continuous, monotone, and differentiable, then*

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{\partial}{\partial y} g^{-1}(y)\right|. \tag{1}$$

Example 8 *Let $f_X(x)$ be the* ***gamma pdf***

$$f_X(x) = \frac{1}{(n-1)!\beta^n} x^{n-1} e^{-x/\beta}, \quad 0 < x < \infty$$

where β is a positive constant and n is a positive integer. Suppose we want to know the pdf of $g(X) = 1/X$. Note that both X and $1/X$ have support $(0,\infty)$. If we let $y = g(x)$ then $g^{-1}(y) = 1/y$ and $\frac{\partial}{\partial y} g^{-1}(y) = -1/y^2$. Then for $y \in (0,\infty)$,

$$\begin{aligned} f_Y(y) &= f_X\left(g^{-1}(y)\right)\left|\frac{\partial}{\partial y} g^{-1}(y)\right| \\ &= \frac{1}{(n-1)!\beta^n}\left(\frac{1}{y}\right)^{n-1} e^{-1/(\beta y)} \frac{1}{y^2} \\ &= \frac{1}{(n-1)!\beta^n}\left(\frac{1}{y}\right)^{n+1} e^{-1/(\beta y)} \end{aligned}$$

which is a special case of the ***inverted gamma pdf****.*

The preceding is only applicable when g is monotone. If g is not monotone, then things are a little more complicated. Sometimes we can make some headway without complicated theorems, as the following example illustrates.

Example 9 *Suppose X is a continuous random variable. For $y > 0$ the cdf of $Y = X^2$ is*

$$\begin{aligned} F_Y(y) &= \Pr(Y \le y) \\ &= \Pr\left(X^2 \le y\right) \\ &= \Pr(-\sqrt{y} \le X \le \sqrt{y}) \\ &= \Pr(-\sqrt{y} < X \le \sqrt{y}) \qquad \textit{continuity of } X \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned} \tag{2}$$

Here's a theorem we can apply more generally to cases where the transformation g is not monotone.

Theorem 10 *Let X have pdf $f_X(x)$ and $Y = g(X)$. Suppose there exists a partition $A_0, A_1, ..., A_k$ of $\mathcal{X}$ such that $\Pr(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Suppose further there exist functions $g_1(x), ..., g_k(x)$ defined on $A_1, ..., A_k$ satisfying:*
a. $g(x) = g_i(x)$ for $x \in A_i$
b. $g_i(x)$ is monotone on A_i
c. the set $\mathcal{Y} = \{y : y = g_i(x)$ for some $x \in A_i\}$ is the same for each $i = 1, ..., k$
d. each $g_i^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$.
Then:

$$f_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X\left(g_i^{-1}(y)\right) \left|\frac{d}{dy} g_i^{-1}(y)\right| & \text{for } y \in \mathcal{Y} \\ 0 & \text{otherwise.} \end{cases}$$

The key to Theorem 10 is that we can partition the set of possible values of X into a collection of sets, such that g is monotone on each of the sets. We no longer need g to be monotone over all possible values of X. The set A_0 is typically ignorable – we just use it to handle technicalities, e.g., the endpoints of intervals.

Example 11 *Suppose X has a standard normal distribution, so that*

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right) \quad \text{for } -\infty < x < \infty$$

and let $Y = X^2$. Then $g(X) = X^2$ is monotone on $(-\infty, 0)$, and monotone on $(0, \infty)$, and $\mathcal{Y} = (0, \infty)$. Let $A_0 = \{0\}$, and

$$\begin{array}{ll} A_1 = (-\infty, 0), & g_1(x) = x^2, \quad g_1^{-1}(y) = -\sqrt{y} \\ A_2 = (0, \infty), & g_2(x) = x^2, \quad g_2^{-1}(y) = \sqrt{y}. \end{array}$$

Then from Theorem 10, we know:

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\left(-\sqrt{y}\right)^2/2\right) \left|-\frac{1}{2\sqrt{y}}\right| + \frac{1}{\sqrt{2\pi}} \exp\left(-\left(\sqrt{y}\right)^2/2\right) \left|\frac{1}{2\sqrt{y}}\right| \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} \exp(-y/2) \quad \text{for } 0 < y < \infty. \end{aligned}$$

We'll see this pdf again: it is the pdf of a chi-squared random variable with one degree of freedom.

Expectations of Random Variables

Definition 12 *The **mean** or **expected value** of a random variable is*

$$E[X] = \begin{cases} \sum_X x f_X(x) & \text{if } X \text{ is discrete} \\ \int_X x f_X(x)\, dx & \text{if } X \text{ is continuous} \end{cases}. \tag{3}$$

The expected value of a random variable X is a weighted average of the possible values taken by X, where the weights are their respective probabilities. We usually denote it by μ. The expectation operator is a linear operator, so that $E[a+bX] = a + bE[X]$ and $E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$. More generally, including the case of nonlinear functions, we have the following.

Proposition 13 *Let $g(X)$ be a function of X. The expected value of $g(X)$ is*

$$E[g(X)] = \begin{cases} \sum_X g(x) f_X(x) & \text{if } X \text{ is discrete} \\ \int_X g(x) f_X(x)\, dx & \text{if } X \text{ is continuous} \end{cases}.$$

Proposition 14 (Jensen's Inequality) *For any random variable X, if $g(x)$ is a convex function then $E[g(X)] \geq g(E[X])$.*

Definition 15 *The **variance** of a random variable X is*

$$\begin{aligned} Var[X] &= \begin{cases} \sum_X (x-\mu)^2 f_X(x) & \text{if } X \text{ is discrete} \\ \int_X (x-\mu)^2 f_X(x)\, dx & \text{if } X \text{ is continuous} \end{cases} \\ &= E\left[(X-\mu)^2\right] \\ &= E\left[X^2\right] - \mu^2. \end{aligned}$$

The variance of a random variable is a measure of dispersion in its distribution. We usually denote it by σ^2, and frequently work with σ (its square root), called the **standard deviation** of X.

By Jensen's inequality, the variance operator is not a linear operator. In fact, $Var[a+bX] = b^2 Var[X]$. More generally, for any function $g(X)$,

$$\begin{aligned} Var[g(X)] &= E\left[(g(x) - E[g(x)])^2\right] \\ &= \int_X (g(x) - E[g(x)])^2 f_X(x)\, dx. \end{aligned}$$

The expected value of a random variable is the first central **moment** of its distribution. The variance is the second central moment. The r^{th} central moment of the distribution of a random variable X is $E[(X-\mu)^r]$. We call the third central moment the skewness of a distribution. It is a measure of the symmetry of a distribution (or lack thereof). For symmetric distributions, skewness is zero. We call the fourth central moment kurtosis, and it measures the thickness of the tails of the distribution.

Some Important Probability Distributions

The Normal Distribution

Definition 16 *A random variable X has a normal distribution with mean μ and variance σ^2, denoted $X \sim N(\mu, \sigma^2)$, if and only if (iff)*

$$f_X\left(x|\mu,\sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty. \tag{4}$$

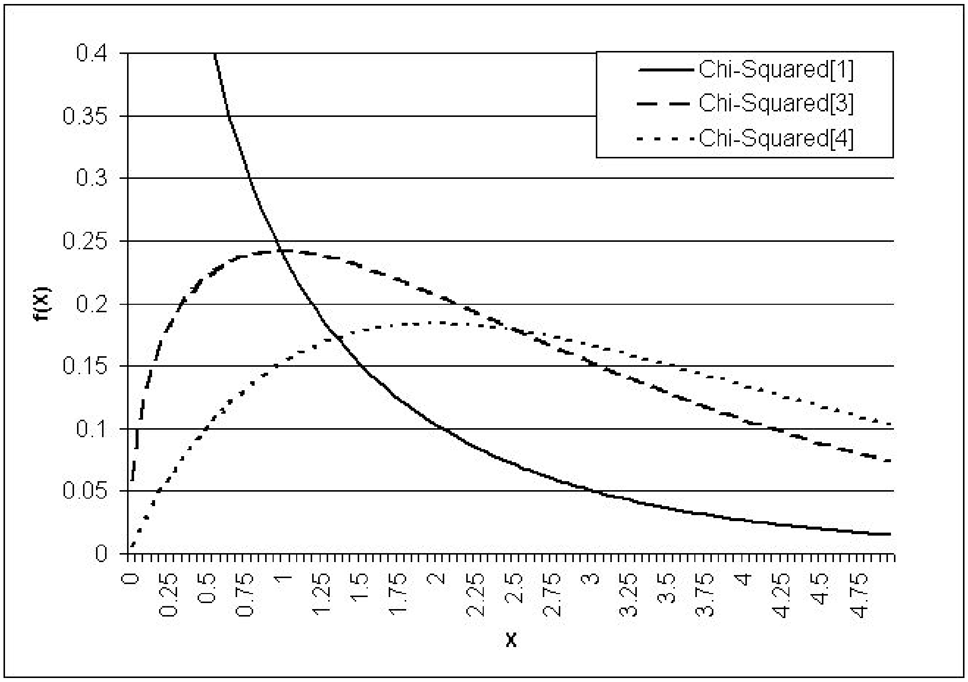

Figure 2: Some Chi-Squared Densities

The normal distribution plays a very important role in econometric theory, particularly for inference and asymptotic theory. It has many convenient properties. First, the normal distribution is completely characterized by two parameters (or by two moments): the mean and variance. Second, it is preserved under linear transformations. That is, if $X \sim N(\mu, \sigma^2)$ then $a + bX \sim N(a + b\mu, b^2\sigma^2)$. Because of this second property, we can always write a normally distributed random variable as a function of a random variable with a **standard normal distribution**, denoted $Z \sim N(0, 1)$. The usual notation for the standard normal pdf is $\phi(z)$; we usually use $\Phi(z)$ for the standard normal cdf. Notice that

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{5}$$

so that if $X \sim N(\mu, \sigma^2)$ we can write

$$f_X(x) = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right). \tag{6}$$

The Chi-Squared (χ^2) Distribution

The chi-squared distribution is another useful distribution for inference. Many test statistics have a χ^2 distribution. It is a special case of the gamma distribution, and is derived from the normal. It is defined by a single parameter: the degrees of freedom ν. It is a skewed distribution that takes positive values only. Figure 2 gives some sample plots of χ^2 densities with different degrees of freedom.

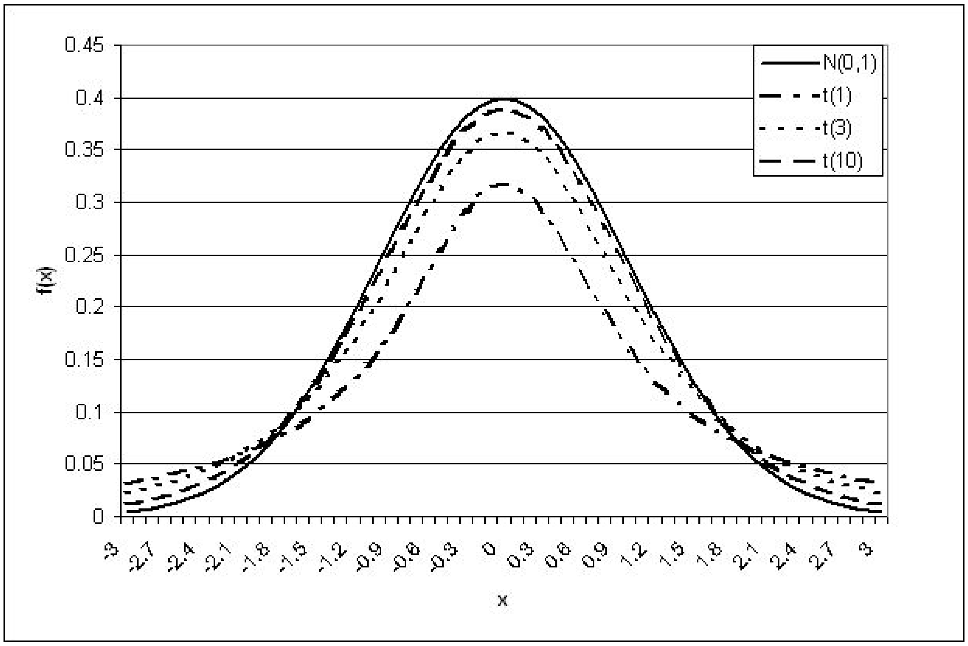

Figure 3: Some t Densities

Definition 17 *A random variable X has a χ^2 distribution with ν degrees of freedom, denoted $X \sim \chi^2_\nu$, iff*

$$f_X(x|\nu) = \frac{1}{\Gamma(\nu/2)\, 2^{\nu/2}} x^{(\nu/2)-1} e^{-x/2}, \quad 0 \leq x < \infty. \tag{7}$$

We call the function $\Gamma(n)$ the gamma function. It has no closed form unless n is an integer. In general, $\Gamma(n) = \int_0^\infty t^{n-1} e^{-t} dt$. When $n > 0$ is an integer, $\Gamma(n) = (n-1)!$.

Proposition 18 *If $Z \sim N(0,1)$, then $X = Z^2 \sim \chi^2_1$.*

Proposition 19 *If $X_1, ..., X_n$ are n independent χ^2_1 random variables, then $\sum_{i=1}^n x_i \sim \chi^2_n$.*

Proposition 20 *Suppose $X \sim \chi^2_n$. Then $E[X] = n$ and $Var[X] = 2n$.*

Proposition 21 *Suppose $X_1 \sim \chi^2_{n_1}$ and $X_2 \sim \chi^2_{n_2}$ are independent. Then $X_1 + X_2 \sim \chi^2_{n_1+n_2}$.*

Student's t Distribution

The t distribution has the same basic shape as a standard normal distribution, but with thicker tails. It is useful for inference: the usual "t test" statistic has a t distribution (we'll see this in an upcoming lecture). It is defined by a single parameter, the degrees of freedom ν. Figure 3 plots the standard normal density, and several t densities.

Definition 22 *A random variable X has a Student's t distribution (or simply a t distribution) with ν degrees of freedom, denoted $X \sim t_\nu$, iff*

$$f_X(x|t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)} \frac{1}{(\nu\pi)^{1/2}} \frac{1}{(1+x^2/\nu)^{(\nu+1)/2}}, \quad -\infty < x < \infty. \tag{8}$$

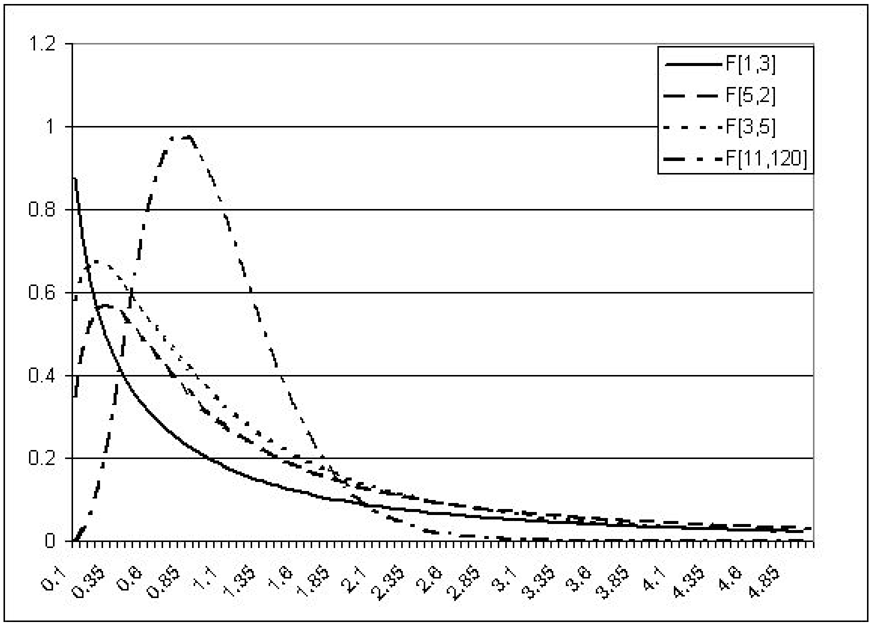

Figure 4: Some F Densities

Proposition 23 *If $Z \sim N(0,1)$, and $X \sim \chi^2_\nu$ is independent of Z, then*

$$t = \frac{Z}{\sqrt{X/\nu}} \sim t_\nu.$$

Proposition 24 *Suppose $X \sim t_\nu$. Then only $\nu - 1$ moments of the distribution exist (i.e., are finite), and $E[X] = 0$ if $\nu > 1$, and $Var[X] = \frac{\nu}{\nu-2}$ if $\nu > 2$.*

Snedecor's F Distribution

Like the t distribution, the F distribution is a "derived" distribution that is useful for inference. As you might guess, the usual "F test" statistic follows an F distribution (next lecture). It is defined by two degrees of freedom parameters, ν_1 and ν_2. Like the χ^2 distribution, it is skewed and takes positive values only. Figure 4 plots some F densities.

Proposition 25 *A random variable X has a Snedecor's F distribution (or simply an F distribution) with ν_1 and ν_2 degrees of freedom, denoted $X \sim F_{\nu_1,\nu_2}$, iff*

$$f_X(x|\nu_1,\nu_2) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2}\frac{x^{(\nu_1-2)/2}}{\left(1+\left(\frac{\nu_1}{\nu_2}\right)x\right)^{(\nu_1+\nu_2)/2}}, \quad 0 \le x < \infty. \tag{9}$$

Proposition 26 *If X_1 and X_2 are two* ***independent*** *chi-squared random variables with ν_1 and ν_2 degrees of freedom, respectively, then*

$$F = \frac{X_1/\nu_1}{X_2/\nu_2} \sim F_{\nu_1,\nu_2}.$$

Proposition 27 *If $X \sim t_\nu$ then $X^2 \sim F_{1,\nu}$.*