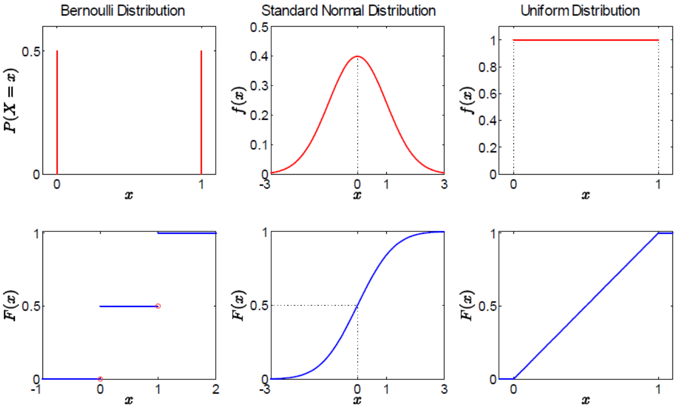
Econometricks: Short guides to econometrics

Trick 02: Specific Distributions

D. Rostam-Afschar

Content

Specific Distributions



Thanks to Ping Yu

The **Bernoulli distribution** for a single binomial outcome (trial) is

$$\begin{aligned} Prob(x = 1) &= p, \\ Prob(x = 0) &= 1 - p, \end{aligned}$$

where $0 \leq p \leq 1$ is the probability of success.

- $E[x] = p$ and
- $V[x] = E[x^2] - E[x]^2 = p - p^2 = p(1 - p)$.

Discrete distributions

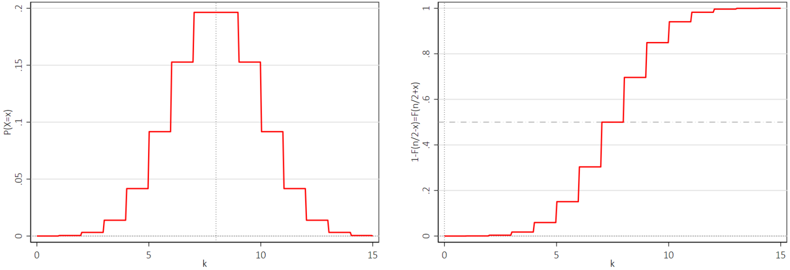
The distribution for x successes in n trials is the **binomial distribution**,

$$Prob(X = x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \quad x = 0, 1, \ldots, n.$$

The mean and variance of x are

- $E[x] = np$ and
- $V[x] = np(1-p)$.

Example of a binomial $[n = 15, p = 0.5]$ distribution:



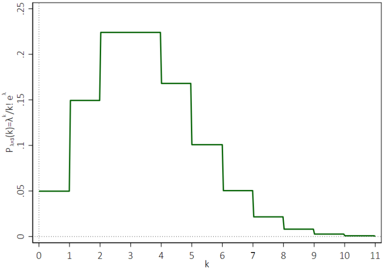
The limiting form of the binomial distribution, $n \to \infty$, is the **Poisson distribution**,

$$Prob(X = x) = \frac{e^{\lambda}\lambda^{x}}{x!}.$$

The mean and variance of x are

- $E[x] = \lambda$ and
- $V[x] = \lambda$.

Example of a Poisson [3] distribution:

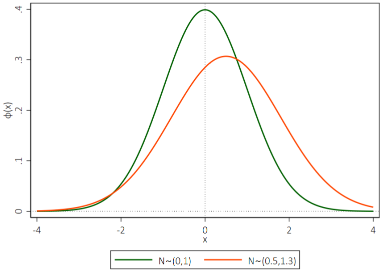


The normal distribution

Random variable $x \sim N[\mu, \sigma^2]$ is distributed according to the **normal distribution** with mean μ and standard deviation σ obtained as

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

The density is denoted $\phi(x)$ and the cumulative distribution function is denoted $\Phi(x)$ for the standard normal. Example of a standard normal, ($x \sim N[0,1]$), and a normal with mean 0.5 and standard deviation 1.3:



Continuous variable x may be transformed to a discrete variable y. Calculate the mean of variable x in the respective interval:

$$\begin{aligned}
Prob(Y = \mu_1) &= P(-\infty < X \leq a), \\
Prob(Y = \mu_2) &= P(a < X \leq b), \\
Prob(Y = \mu_3) &= P(b < X \leq \infty).
\end{aligned}$$

Method of transformations

If x is a continuous random variable with pdf $f_x(x)$ and if $y = g(x)$ is a continuous monotonic function of x, then the density of y is obtained by

$$Prob(y \leq b) = \int_{-\infty}^{b} f_x(g^{-1}(y))|g^{-1\prime}(y)|dy.$$

With $f_y(y) = f_x(g^{-1}(y))|g^{-1\prime}](y)|dy$, this equation can be written as

$$Prob(y \leq b) = \int_{-\infty}^{b} f_y(y)dy.$$

Example

If $x \sim N[\mu, \sigma^2]$, then the distribution of $y = g(x) = \frac{x-\mu}{\sigma}$ is found as follows:

$$g^{-1}(y) = x = \sigma y + \mu$$

$$g^{-1\prime}(y) = \frac{dx}{dy} = \sigma$$

Therefore with $f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}[(g^{-1}(y)-\mu)^2/\sigma^2]}|g^{-1\prime}(y)|$

$$f_y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-[(\sigma y+\mu)-\mu]^2/2\sigma^2}|\sigma| = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$

Properties of the normal distribution

- Preservation under linear transformation:

  If $x \sim N[\mu, \sigma^2]$, then $(a + bx) \sim N[a + b\mu, b^2\sigma^2]$.

- Convenient transformation $a = -\mu/\sigma$ and $b = 1/\sigma$:

  The resulting variable $z = \frac{(x-\mu)}{\sigma}$ has the standard normal distribution with density

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

- If $x \sim N[\mu, \sigma^2]$, then $f(x) = \frac{1}{\sigma}\phi[\frac{x-\mu}{\sigma}]$
- $Prob(a \leq x \leq b) = Prob\left(\frac{a-\mu}{\sigma} \leq \frac{x-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right)$
- $\phi(-z) = 1 - \phi(z)$ and $\Phi(-x) = 1 - \Phi(x)$ because of symmetry

Method of transformations

If $z \sim N[0,1]$, then $z^2 \sim \chi^2[1]$ with pdf $\frac{1}{\sqrt{2\pi y}} e^{-y/2}$.

Example

$$f_x(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$y = g(x) = x^2$$

$$g^{-1}(y) = x = \pm\sqrt{y} \text{ there are two solutions to } g_1, g_2.$$

$$g^{-1\prime}(y) = \frac{dx}{dy} = \pm 1/2y^{-1/2}$$

$$f_y(y) = f_x(g_1^{-1}(y))|g_1^{-1\prime}(y)| + f_x(g_2^{-1}(y))|g_2^{-1\prime}(y)|$$

$$f_y(y) = f_x(\sqrt{y})|1/2y^{-1/2}| + f_x(-\sqrt{y})| - 1/2y^{-1/2}|$$

$$f_y(y) = \frac{1}{2\sqrt{2\pi y}} e^{-\frac{y}{2}} + \frac{1}{2\sqrt{2\pi y}} e^{-\frac{y}{2}} = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}$$

Distributions derived from the normal

- If $z \sim N[0,1]$, then $z^2 \sim \chi^2[1]$ with $E[z^2] = 1$ and $V[z^2] = 2$.
- If $x_1, \ldots, x_n$ are n independent $\chi^2[1]$ variables, then
$$\sum_{i=1}^{n} x_i \sim \chi^2[n].$$
- If z_i, $i = 1, \ldots, n$, are independent $N[0,1]$ variables, then
$$\sum_{i=1}^{n} z_i^2 \sim \chi^2[n].$$
- If z_i, $i = 1, \ldots, n$, are independent $N[0,\sigma^2]$ variables, then
$$\sum_{i=1}^{n} \left(\frac{z_i}{\sigma}\right)^2 \sim \chi^2[n].$$
- If x_1 and x_2 are independent χ^2 variables with n_1 and n_2 degrees of freedom, then
$$x_1 + x_2 \sim \chi^2[n_1 + n_2].$$

The χ^2 distribution

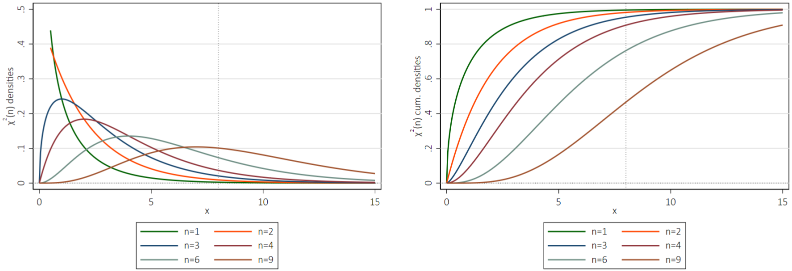
Random variable $x \sim \chi^2[n]$ is distributed according to the **chi-squared distribution** with n degrees of freedom

$$f(x|n) = \frac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}, \tag{2}$$

where Γ is the Gamma-distribution (more below).

- $E[x] = n$
- $V[x] = 2n$

Example of a $\chi^2[3]$ distribution:

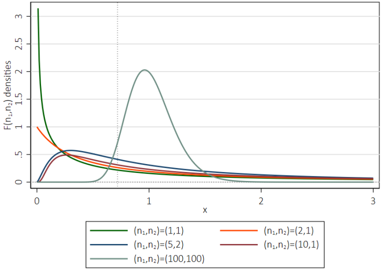


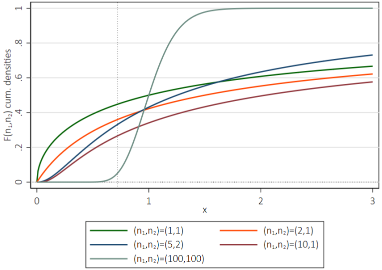
The F-distribution

If x_1 and x_2 are two independent chi-squared variables with degrees of freedom parameters n_1 and n_2, respectively, then the ratio

$$F[n_1, n_2] = \frac{x_1/n_1}{x_2/n_2} \tag{3}$$

has the **F distribution** with n_1 and n_2 degrees of freedom.



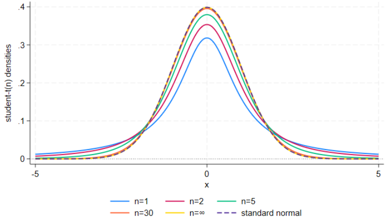


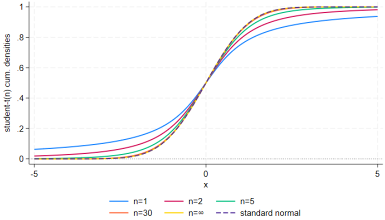
The student t-distribution

If z is an $N[0,1]$ variable and x is $\chi^2[n]$ and is independent of z, then the ratio

$$t[n] = \frac{z}{\sqrt{x/n}}. \tag{4}$$

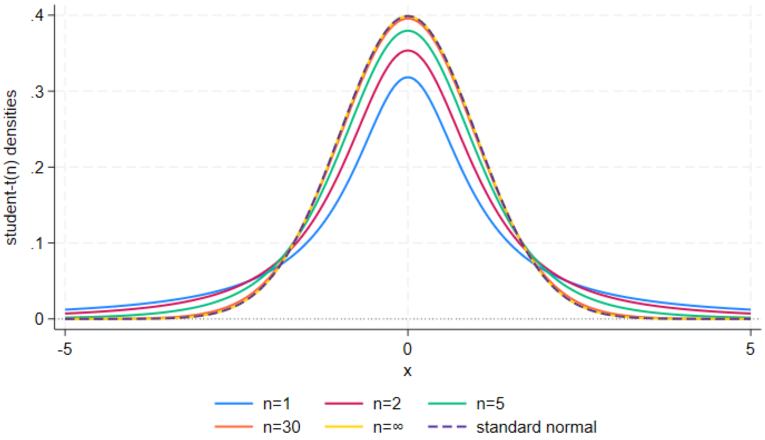
has the **t distribution** with n degrees of freedom.
Example for the t distributions with 3 and 10 degrees of freedom with the standard normal distribution.





Comparing (3) with $n_1 = 1$ and (4), if $t \sim t[n]$, then $t^2 \sim F[1, n]$.

The $t[30]$ approx. the standard normal



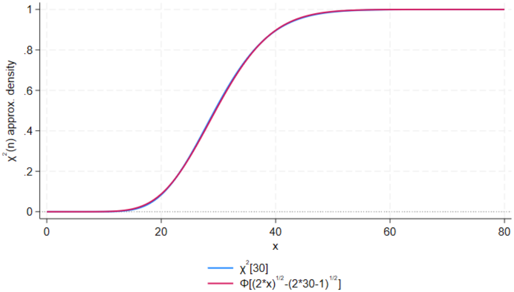
Approximating a χ^2

For degrees of freedom greater than 30 the distribution of the chi-squared variable x is approx.

$$z = (2x)^{1/2} - (2n-1)^{1/2}, \tag{5}$$

which is approximately standard normally distributed. Thus,

$$Prob(\chi^2[n] \leq a) \approx \Phi[(2a)^{1/2} - (2n-1)^{1/2}].$$



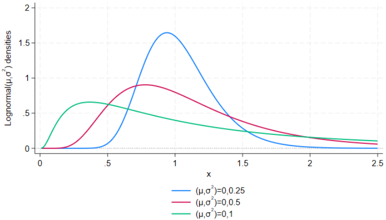
The **lognormal distribution**, denoted $LN[\mu, \sigma^2]$, has been particularly useful in modeling the size distributions.

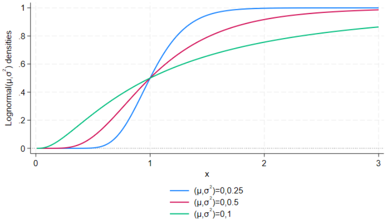
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2[(\ln x - \mu)/\sigma]^2}}, \qquad x > 0$$

A lognormal variable x has

- $E[x] = e^{\mu + \sigma^2/2}$, and
- $Var[x] = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$.

If $y \sim LN[\mu, \sigma^2]$, then $\ln y \sim N[\mu, \sigma^2]$.



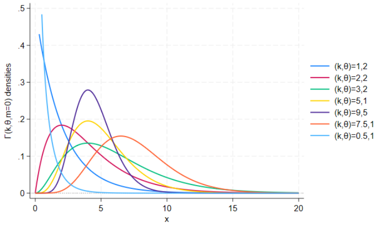


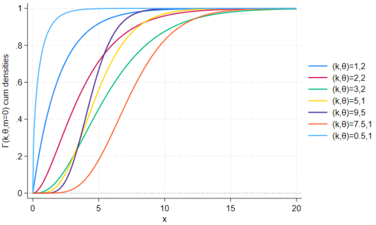
The gamma distribution

The general form of the **gamma distribution** is

$$f(x) = \frac{\lambda^P}{\Gamma(P)} e^{-\lambda x} x^{P-1}, \qquad x \geq 0, \lambda > 0, P > 0. \tag{6}$$

Many familiar distributions are special cases, including the **exponential distribution**$(P = 1)$ and **chi-squared**$(\lambda = 1/2, P = n/2)$. The **Erlang distribution** results if P is a positive integer. The mean is P/λ, and the variance is P/λ^2. The **inverse gamma distribution** is the distribution of $1/x$, where x has the gamma distribution.

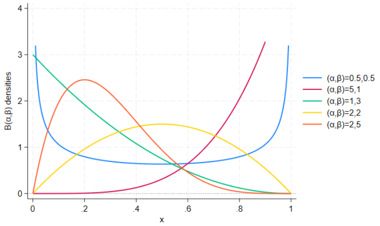


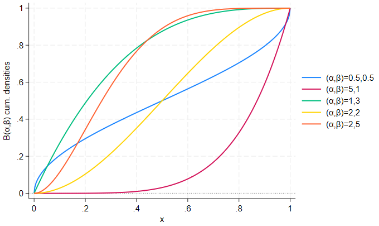


For a variable constrained between 0 and $c > 0$, the **beta distribution** has proved useful. Its density is

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{x}{c}\right)^{\alpha-1} \left(1-\frac{x}{c}\right)^{\beta-1} \frac{1}{c}, \qquad x \geq 0, \lambda > 0, P > 0.$$

It is symmetric if $\alpha = \beta$, asymmetric otherwise. The mean is $c\alpha/(\alpha+\beta)$, and the variance is $c^2\alpha\beta/[(\alpha+\beta+1)(\alpha+\beta)^2]$.



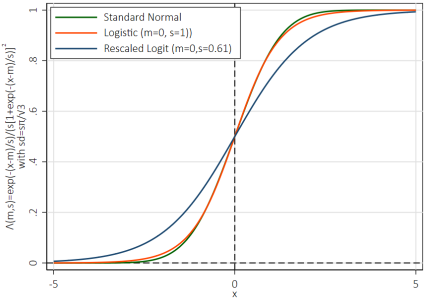


The logistic distribution

The **logistic distribution** is an alternative if the normal cannot model the mass in the tails; the cdf for a logistic random variable is

$$F(x) = \Lambda(x) = \frac{1}{1 + e^{-x}}.$$

The density is $f(x) = \Lambda(x)[1 - \Lambda(x)]$. The mean and variance of this random variable are zero and $\pi^2/3$.



The Wishart distribution

The **Wishart distribution** describes the distribution of a random matrix obtained as

$$f(\boldsymbol{W}) = \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)'.$$

where x_i is the ith of nK element random vectors from the multivariate normal distribution with mean vector, μ, and covariance matrix, Σ. The density of the Wishart random matrix is

$$f(\boldsymbol{W}) = \frac{\exp\left[-\frac{1}{2} trace(\Sigma^{-1}\boldsymbol{W})\right] |\boldsymbol{W}|^{-\frac{1}{2}(n-K-1)}}{2^{nK/2}|\Sigma|^{K/2}\pi^{K(K-1)/4}\prod_{j=1}^{K}\Gamma\left(\frac{n+1-j}{2}\right)}.$$

The mean matrix is $n\Sigma$. For the individual pairs of elements in $\boldsymbol{W}$,

$$Cov[w_{ij}, w_{rs}] = n(\sigma_{ir}\sigma_{js} + \sigma_{is}\sigma_{jr}).$$

The Wishart distribution is a multivariate extension of χ^2 distribution. If $\boldsymbol{W} \sim W(n, \sigma^2)$, then $\boldsymbol{W}/\sigma^2 \sim \chi^2[n]$.

References I

Greene, W. H. (2003): *Econometric Analysis*. Prentice Hall, 5 edn.