School of Oriental and African Studies **September 2013**
Department of Economics

Preliminary Statistics
Lecture 3: Probability Models and Distributions (Outline)

Gujarati D. *Basic Econometrics*, Appendix A.4 - A.6
Barrow M. *Statistics for Economics, Accounting and Business Studies* Ch. 3 & 6

1 Characteristics of Probability Distributions

Two of the most widely used characteristics of probability distributions are the expected value (mean) and the variance.

1.1 Expected Value

For a discrete random variable, the expected value (often denoted by the Greek letter μ) is the sum of each value it can take, x_i, multiplied by its corresponding probability

$$E(X) = \sum_{i=1}^{N} f(x_i)x_i = \mu$$

For continuous random variables the summation is replaced by an integral.

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

Notice that if all the values are equally likely, $f(x_i) = 1/N$, so the expected value is the arithmetic mean.

Properties of the Expected Value
Expected values behave like $N^{-1}\sum$.

- If a is a constant $E(a) = a$.
- If a and b are constants $E(a + bX) = a + bE(X)$.
  This can be generalised for N random variables.
- $E(X + Y) = E(X) + E(Y)$
- If X and Y are independent random variables, then

$$E(XY) = E(X)E(Y)$$

1.2 Variance

Let X be a random variable and let $E(X) = \mu$. The dispersion of the X values around the expected value can be measured by the variance,

$$\mathrm{Var}(X) = E(X - E(X))^2 = \sigma^2$$

The square root of the variance, σ, is defined as the standard deviation of X. The variance is computed as follows:

$$\mathrm{Var}(X) = \begin{cases} \sum_{i=1}^{N} f(x_i)(x_i - \mu)^2 & \text{if } X \text{ is a discrete random variable} \\ \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx & \text{if } X \text{ is a continuous random variable} \end{cases}$$

If $f(x_i) = 1/N$ this is just the same as the population variance we encountered in descriptive statistics.

Properties of Variance

- $\mathrm{Var}(X) = E(X - \mu)^2 = E(X^2) - \mu^2$

  Proof

$$\begin{aligned} \mathrm{Var}(X) &= E(X - \mu)^2 = E(X^2 + \mu^2 - 2X\mu) \\ &= E(X^2) + E(\mu^2) - E(2X\mu) = E(X^2) + \mu^2 - 2\mu E(X) \\ &= E(X^2) + \mu^2 - 2\mu^2 = E(X^2) - \mu^2 \end{aligned}$$

  □

- The variance of a constant is zero, $\mathrm{Var}(a) = 0$, since a fixed number does not have any dispersion.

- If a and b are constants, then

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$$

  Proof

$$\begin{aligned} \mathrm{Var}(aX + b) &= E(aX + b - E(aX + b))^2 = E(aX + b - aE(X) - b)^2 \\ &= E(a(X - E(X))^2 = E(a(X - \mu))^2 = E(a^2(X - \mu)^2) \\ &= a^2 E(X - \mu)^2 = a^2 \mathrm{Var}(X) \end{aligned}$$

  □

- If X and Y are two random variables and a, b constants, then

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$$

  and

$$\mathrm{Var}(aX + bY) = a^2 \mathrm{Var}(X) + b^2 \mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)$$

  where $\mathrm{Cov}(X, Y)$ is the covariance of the two variables introduced below.

Proof

$$\begin{aligned}
\mathrm{Var}(aX+bY) &= E[aX+bY-E(aX+bY)]^2 = E[aX+bY-aE(X)-bE(Y)]^2 \\
&= E[a(X-E(X))+b(Y-E(Y))]^2 \\
&= E[a^2(X-E(X))^2+b^2(Y-E(Y))^2+2ab(X-E(X))(Y-E(Y))] \\
&= a^2E(X-E(X))^2+b^2E(Y-E(Y))^2+2abE[(X-E(X))(Y-E(Y))] \\
&= a^2\mathrm{Var}(X)+b^2\mathrm{Var}(Y)+2ab\mathrm{Cov}(X,Y)
\end{aligned}$$

$\square$

- If X and Y are *independent* random variables, then

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{Var}(X)+\mathrm{Var}(Y) \\
\mathrm{Var}(X-Y) &= \mathrm{Var}(X)+\mathrm{Var}(Y) \\
\mathrm{Var}(aX+bY) &= a^2\mathrm{Var}(X)+b^2\mathrm{Var}(Y)
\end{aligned}$$

Covariance

The Covariance between two random variables is defined as

$$\mathrm{Cov}(X,Y) = E[(X-E(X))(Y-E(Y))] = E(XY)-E(X)E(Y).$$

The covariance is computed as follows

$$\mathrm{Cov}(X,Y) = \sum_y \sum_x (X-\mu_x)(Y-\mu_y)f(x,y)$$

for discrete variables, and

$$\mathrm{Cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(X-\mu_x)(Y-\mu_y)f(x,y)dxdy$$

for continuous random variables.

Properties of Covariance

- $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.
- For a constant, $\mathrm{Cov}(X,a) = 0$.
- If X and Y are independent random variables their covariance is zero, $\mathrm{Cov}(X,Y) = E(XY)-E(X)E(Y) = 0$, since for independent variables $E(XY) = E(X)E(Y)$. However, a covariance of zero does not imply that they are independent, independence is a stronger property.
- For a, b, c, d constants,

$$\mathrm{Cov}(a+bX, c+dY) = bd\mathrm{Cov}(X,Y)$$

Proof

$$\begin{aligned}
\mathrm{Cov}(a+bX,c+dY) &= E[((a+bX)-E(a+bX))((c+dY)-E(c+dY))] \\
&= E[(a+bX-E(a)-bE(X))(c+dY-E(c)-dE(Y))] \\
&= E[b(X-E(X))d(Y-E(Y))] = bdE[(X-E(X))(Y-E(Y))] \\
&= bd\mathrm{Cov}(X,Y)
\end{aligned}$$

$\square$

2 Probability Models and Distributions

2.1 The normal distribution

The most common distribution assumed for *continuous* random variables is the normal or Gaussian distribution. The normal distribution tends to arise when a random variable is the result of many independent, random influences added together, none of which dominates the others. Suppose that we have a random variable Y which is normally distributed with expected value μ and variance σ^2, then this is denoted as: $Y \sim \mathrm{N}(\mu, \sigma^2)$.

2.1.1 Shape and PDF of a Normal Distribution

The normal distribution is the exponential of a quadratic and it is

- bell-shaped
- symmetric, with mean, median and mode coinciding
- skewness and excess kurtosis are zero
- its PDF is
$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2}\left(\frac{y_i - \mu}{\sigma}\right)^2\}$$
We use the PDF to get probabilities of ranges, e.g. between 0 and 1.

2.1.2 Linear Transformations of Normal random variables

If one variable Y is normally distributed with mean μ and variance σ^2, then any linear function of Y is also normally distributed. That is,

$$\begin{aligned} Y &\sim N\left(\mu, \sigma^2\right) \\ W &= a + bY \sim N\left(a + b\mu, b^2\sigma^2\right) \end{aligned}$$

$$\begin{aligned} E(W) &= E(a + bY) = E(a) + E(bY) = a + bE(Y) = a + b\mu \\ \mathrm{Var}(W) &= E(W - E(W))^2 = E(a + bY - a - b\mu)^2 \\ &= E(b(Y - \mu))^2 = b^2 E(Y - \mu)^2 = b^2\sigma^2 \end{aligned}$$

So, if temperature over a year measured in centigrade is normally distributed, temperature in Farenheit (which is a linear transformation) is also normal.

Decomposing an observed random variable into its expected value and an error, u_i, is very convenient for many purposes. So, if $Y \sim \mathrm{N}(\mu, \sigma^2)$, then $Y_i = \mu + u_i$ and $u_i = Y_i - \mu$. Since, the error term is a linear function of the random variable Y it will also be normally distributed as, $u_i \sim \mathrm{N}(0, \sigma^2)$.

Linear transformations can apply to more than two random variables. So, if two (or more) random variables X and Y are normally and independently distributed, and a, b are constants, then the linear combination, $K = aX + bY$ is also normally distributed as $K \sim \mathrm{N}(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2)$. Another very important linear transformation of a normally distributed random variable Y is the standard normal.

2.2 Standard Normal Distribution

A random variable Y is standardised by subtracting its mean and dividing by its standard deviation,

$$z_i = \frac{Y_i - \mu}{\sigma}$$

Z is called the standard normal variable or the z-score and is a normally distributed variable with expected value zero and variance (and standard deviation) of one. Hence, $Z \sim \mathrm{N}(0,1)$, since

$$\begin{aligned} E(Z) &= E\left(\frac{Y-\mu}{\sigma}\right) = \frac{1}{\sigma}(E(Y)-\mu) = \frac{1}{\sigma}(\mu-\mu) = 0 \\ Var(Z) &= E\left(\frac{Y-\mu}{\sigma}\right)^2 = E\left(\frac{(Y-\mu)^2}{\sigma^2}\right) = \frac{1}{\sigma^2}E(Y-\mu)^2 = \frac{1}{\sigma^2}\sigma^2 = 1 \end{aligned}$$

The standard normal distribution is tabulated and hence used to calculate the areas under the normal distribution. Each normally distributed random variable is transformed to standard normal by

$$Z = \frac{Y-\mu}{\sigma}$$

in order to calculate areas under the distribution, and then converted back into the normal, using

$$Y = \sigma z + \mu.$$

The standard normal PDF is $\quad f(z_i) = (2\pi)^{-1/2} \exp\left(-\frac{z_i^2}{2}\right)$

The probability of being less than 0 for the standard normal, μ for the normal, is 0.5, the same as the probability of being greater than the mean, since they are symmetric. 95% of the normal distribution lies within 1.96 standard deviations from the mean. The probability of being 3 standard deviations from the mean is around 99.7%.

Random Sample and Sampling distribution

If we have an independent sample from the distribution of Y, Y_i; $i = 1, 2, ..., n$ we write this $Y_i \sim \mathrm{IN}(\mu, \sigma^2)$. This is said Y_i is independent normal with expected value μ and variance σ^2. A sample of size n, Y_i; $i = 1, 2, ..., n$ is random if all Xs are drawn independently from the same probability distribution. That is,

- Each X included in the sample has the same PDF
- Each X included in the sample is drawn independently from the others.

Each of the Xs is said to be independently and identically distributed random variable (*i.i.d.*).

An estimator for the population mean, such as the sample mean, can be treated as a random variable with its own PDF. We can define the sampling distribution of an estimator if the sample is an *i.i.d.* random sample.

2.3 Distributions related to the normal

2.3.1 Chi-squared

Suppose z_i is $IN(0,1)$, independently distributed, standard normal, then

$$A = \sum_{i=1}^{n} z_i^2 \sim \chi^2(n)$$

is said to have a Chi squared distribution with n degrees of freedom.

The χ^2 distribution

- is only defined over positive values
- its sole parameter that determines its shape is its degrees of freedom
- for small degrees of freedom is skewed to the right, though for large degrees of freedom it approaches the normal
- it arises naturally because we calculate estimates of the variance and $(n-1)s^2/\sigma^2$ has a χ^2 distribution with $n-1$ degrees of freedom, where

$$s^2 = \sum_{i=1}^{N} (x_i - \overline{x})^2/(n-1)$$

2.3.2 t distribution

A standard normal divided by the square root of a Chi-squared distribution, divided by its degrees of freedom, say n, is called the t distribution with n degrees of freedom

$$t(n) = z/\sqrt{\frac{\chi^2(n)}{n}}$$

The t-distibution

- is bell-shaped and symmetric
- has fatter tails than the normal, but as the sample size gets larger the distribution is indistinguishable from a normal
- we often divide an estimate of the mean or a regression coefficient (which are normally distributed) by their standard errors (which are the square root of a χ^2 divided by its degrees of freedom), and this is the formula for doing this

2.3.3 F distribution

Fisher's F distribution is the ratio of two independent Chi-squared divided by their degrees of freedom.

$$F(n_1, n_2) = \frac{\chi^2(n_1)/n_1}{\chi^2(n_2)/n_2}.$$

The F distribution

- is only defined over positive values

  for small degrees of freedom is skewed to the right, though for large degrees of freedom it approaches the normal
- it arises as the ratio of two variances