Instituto Superior de Economia e Gestão
Universidade de Lisboa
Mestrado em Econometria Aplicada e Previsão
(MSc in Applied Econometrics and Forecasting)
Econometrics
Paulo M.D.C. Parente
Notes on Probability and Distribution Theory

Some definitions

Experiment: is any procedure that can be infinitely repeated and has a well-defined set of outcomes.
Example: We flip a coin 2 times and count the number of times the coin turns up heads.
Random variable: is a variable that takes on numerical values and has an outcome that is determined by an experiment.
Example: $X =$"number of Heads".

Discrete random variable

Discrete random variable: is a random variable that takes only a discrete set of values.
Example: We flip a balanced coin 2 times and define a random variable $X =$"number of Heads". We obtain:

X	Possible Outcomes	Probability
0	(Tails, Tails)	1/4
1	(Tails, Heads)(Heads, Tails)	2/4
2	(Heads, Heads)	1/4

Remark: There are 4 possible outcomes and all are equally likely.

In the example X takes 3 possible values $(0, 1, 2)$ and the associated probabilities are $(1/4, 1/2, 1/4)$ respectively. In general a discrete random variable takes k possible values $(x_1, x_2, ..., x_k)$ with associated probabilities $(p_1, p_2, ..., p_k)$ respectively. The probabilities are defined by

$$p_j = \mathcal{P}(X = x_j), \; j = 1, 2, ..., k,$$

where

$$0 \leq p_j \leq 1$$

and

$$p_1 + p_2 + ... + p_k = 1.$$

- The *probability distribution* of a discrete random variable is the list of all the possible values of the variable and the probability that each value occur.
- The *cumulative probability distribution* is the probability that the random variable is less than or equal to a particular value: $\mathcal{P}(X \leq x)$.

Example (cont):

X	Probability distribution	Cumulative probability distribution
0	1/4	1/4
1	2/4	3/4
2	1/4	1

Some well known discrete random variables

Bernoulli Random variable

We flip a coin and define a random variable

$$X = \begin{cases} 1 & \text{if Heads} \\ 0 & \text{if Tails} \end{cases}$$

Let us denote

$$\mathcal{P}(Heads) = \mathcal{P}(X = 1) = p$$

Then

$$\mathcal{P}(Tails) = \mathcal{P}(X = 0) = 1 - p.$$

This can be written as

$$\mathcal{P}(X = x) = p^x(1-p)^{1-x}, x = 0, 1.$$

If the coin is balanced $p = 0.5$.
A random variable that is defined as

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

is known as a *Bernoulli Random variable*, named after the Swiss mathematician Jacob Bernoulli (1654-1705).

One outcome is arbitrarily labeled a "success" (denoted $X = 1$) and the other a "failure" (denoted $X = 0$).

The Binomial random variable.

The Binomial random variable is defined as the number of successes in r trials, each of which has the probability of success p.
Remark: If $r = 1$ the Binomial random variable corresponds to the Bernoulli random variable.
2nd case: $r = 2$, for instance $X =$ number of boys in a family of 2 children.
Let us calculate the probability of $0, 1, 2$ boys in 2 births and define $\mathcal{P}(boy) = p$
We can have 4 possible cases:

$$(boy, boy)\,;(boy, girl)\,;(girl, boy)\,;(girl, girl)$$

Hence:

- $\mathcal{P}(X = 0) = \mathcal{P}\,(girl, girl) = (1-p)^2$
- $\mathcal{P}(X = 1) = \mathcal{P}((boy, girl) \text{ or } (girl, boy)) = \mathcal{P}\,(boy, girl) + \mathcal{P}\,(girl, boy) = 2p(1-p)$
- $\mathcal{P}(X = 2) = \mathcal{P}\,(girl, girl) = p^2$.

3th case: $r = 3$, for instance $X =$ number of boys in a family of 3 children.
Let us calculate the probability of $0, 1, 2, 3$ boys in 3 births.
We can have 8 cases:

$$(boy, boy, boy)\,;(boy, girl, boy)\,;(girl, boy, boy)\,;(girl, girl, boy)\,;$$
$$(boy, boy, girl)\,;(boy, girl, girl)\,;(girl, boy, girl)\,;(girl, girl, girl)\,.$$

Hence:

- $\mathcal{P}(X = 0) = \mathcal{P}\,(girl, girl, girl) = (1-p)^3$.
- $\mathcal{P}(X = 1) = \mathcal{P}((girl, girl, boy) \text{ or } (boy, girl, girl) \text{ or } (girl, boy, girl)) = 3(1-p)^2 p$.
- $\mathcal{P}(X = 2) = \mathcal{P}((boy, girl, boy) \text{ or } (girl, boy, boy) \text{ or } (boy, boy, girl)) = 3(1-p)p^2$.

- $\mathcal{P}(X = 3) = \mathcal{P}(boy, boy, boy) = p^3$.

General case: X = number of boys in a family of r children. One can show that

$$\mathcal{P}(X = x) = rCx \times p^x (1-p)^{r-x}$$

where

$$rCx = \frac{r!}{x!(r-x)!}$$

is the number of x combinations from a set with r elements.
and $n! = n \times (n-1) \times ... \times 2 \times 1$
$n!$ is read 'n factorial

The Poisson Distribution

The Poisson distribution, named after the French mathematician Simeon-Denis Poisson (1781-1840), is applicable in many situations where rare events occur. The Poisson distribution describes the number of occurrences within a randomly chosen unit of time or space. For example, within a minute, hour, day, square foot, or linear mile.
Examples:

- in the inspection and quality control of manufactured goods where the proportion of defective articles in a large lot can be expected to be small.
- number of customers arriving at a cash point in a given minute.
- number of file server virus infections at a data center during a 24-hour period.

Famous example: Bortkiewiz in 1898 used this distribution to study the number of soldiers killed by horse-kicks each year in each corps in the Prussian cavalry.

The Poisson model's only parameter is λ (Greek letter "lambda"): λ represents the mean number of events per unit of time or space.

The *Poisson probability function* is a discrete function defined for non-negative integers x. The Poisson distribution with parameter $\lambda > 0$, it is defined by

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Continuous random variable:

Continuous random variable: are random variables that take a continuum of possible values. Events are intervals and probabilities are areas underneath smooth curves. A single point has no probability

Associated with a continuous random variable there is usually a non-negative function known a *probability density function* $f(x)$ that provides information on the likely outcomes of the random variable. Associated with a continuous random variable there is usually a non-negative function known a *probability density function* $f(x)$ (PDF) that provides information on the likely outcomes of the random variable. The entire area under any PDF must be 1. Continuous probability functions are smooth curves. Unlike discrete distributions, the area at any single point = 0.

This function satisfies $\int_{-\infty}^{+\infty} f(x)dx = 1$.

When computing probabilities for continuous random variables it is easiest to work with *cumulative distribution functions:* (cdf) $F(x) = \mathcal{P}(X \leq x) = \displaystyle\int_{-\infty}^{x} f(x)\,dx$.

Example: *Uniform Continuous Distribution* If X is a random variable that is uniformly distributed between a and b, its PDF has constant height:

$$f(x) = \frac{1}{b-a}, a \leq x \leq b$$

The cumulative distribution function is

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Mean and Variance of a distribution

Discrete random variables.

Expected Value of a random variable: The expected value of a random variable, denoted as $E(X)$ or μ_X, also known as its population mean, is the weighted average of its possible values, the weights being the probabilities attached to the values

$$\mu_X = E(X) = x_1 p_1 + x_2 p_2 + ... + x_k p_k = \sum_{i=1}^{k} x_i p_i.$$

Expected value of a function of a random variable:

$$\begin{aligned} E(g(X)) &= g(x_1) p_1 + g(x_2) p_2 + ... + g(x_k) p_k \\ &= \sum_{i=1}^{k} g(x_i) p_i. \end{aligned}$$

The Population Variance: The expected value of the squared deviation from the population mean

$$\sigma_X^2 = Var(X) = E[(X - \mu_X)^2] = \sum_{i=1}^{k} (x_i - \mu_X)^2 p_i.$$

Standard deviation:

$$\sigma_X = \sqrt{Var(X)}.$$

Continuous random variable

Expected value

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

Expected value

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

The Population Variance: The expected value of the squared deviation from the population mean

$$\sigma_X^2 = Var(X) = E[(X - \mu_X)^2] = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f(x) dx.$$

Standard deviation:

$$\sigma_X = \sqrt{Var(X)}.$$

Properties of Expected values:

1. $E(X - \mu_X) = E(X) - \mu_X = 0.$
2. If c is a constant, $E(c) = c.$
3. If c is a constant, $E(cg(X)) = cE(g(X)).$
4. Given 2 functions $u(X)$ and $v(X)$, $E[u(X) + v(X)] = E[u(X)] + E[v(X)].$

Properties of the Variance:

1. $\sigma^2 = Var(X) = E(X^2) - \mu_X^2.$

2. If c is a constant, $Var\left(c\right) = 0$.
3. If a and b are constants, $Var\left(a + bX\right) = b^2 Var\left(X\right).$

Examples:

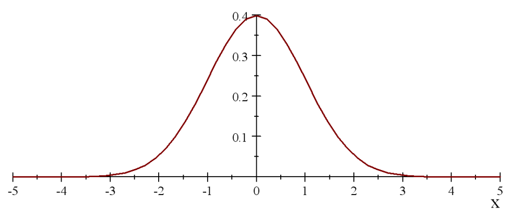
1. Bernoulli random variable with parameter p : $E(X) = p$, $Var(X) = p(1-p)$.
2. Binomial random variables with parameters p and r : $E(X) = rp$, $Var(X) = rp(1-p)$.
3. Poisson random variable with parameter λ. $E(X) = \lambda, Var(X) = \lambda$.
4. Uniform random variable $U(a,b)$: $E(X) = \frac{a+b}{2}, Var(X) = \frac{(b-a)^2}{12}$.

Other Moments

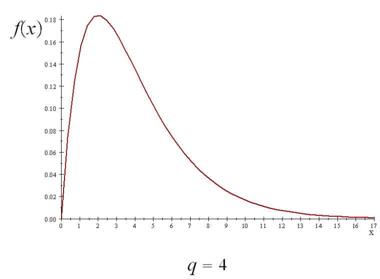
Skewness

$$skewness = \frac{E\left[(X - \mu_X)^3\right]}{\sigma_X^3}$$

It is a measure of asymmetry of a distribution.
Symmetric distribution:



$skewness = 0$
Asymmetric distribution:



$skewness \neq 0$

Kurtosis

$$kurtosis = \frac{E\left[(X - \mu_X)^4\right]}{\sigma_X^4}$$

It is a measure of mass in tails

Joint probability function:

Probability functions defined over a pair of random variables (X, Y) are denoted as joint probability functions: In the discrete case we have $p_{x,y} = \mathcal{P}(X = x, Y = y)$ where

$$\sum_x \sum_y p_{x,y} = 1.$$

In the continuous case we have the joint density function $f(x, y)$, where

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1.$$

Marginal distribution function:

The marginal distribution function is another name for the distribution function and it can be computed from the joint distribution. The marginal probability can be computed in the following way. If X can take k different values $x_1, ..., x_k$

$$\mathcal{P}(Y = y) = \sum_{i=1}^{k} \mathcal{P}(X = x_i, Y = y).$$

In the continuous case

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

Independence of random variables:

Two random variables X and Y are independent if and only if

$$\mathcal{P}(X = x, Y = y) = \mathcal{P}(X = x) \mathcal{P}(Y = y).$$

In the continuous case

$$f(x, y) = f_X(x) f_Y(y).$$

where $f_X(x)$ and $f_Y(y)$ are the density functions of X and Y respectively.

Expected values of functions of random variables

- $E[g(X, Y)] = \begin{cases} \sum_x \sum_y g(x, y) p_{x,y} \text{ in the discrete case} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy \text{ in the continuous case} \end{cases}$

Remark: If X and Y are independent, $E[XY] = E(X) E(Y)$.
Covariance:

$$Cov(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$

Properties:

- $Cov(X, Y) = E(XY) - E(X)E(Y)$.
- If X and Y are independent $Cov(X, Y) = 0$.
- If $Y = bZ$, where b is constant,
$$Cov(X, Y) = bCov(X, Z).$$
- If $Y = V + W$,
$$Cov(X, Y) = Cov(X, V) + Cov(X, W).$$
- If $Y = b$, where b is constant,
$$Cov(X, Y) = 0.$$

- If $Y = V + W$,

$$Var(Y) = Var(V) + Var(W) + 2Cov(V, W).$$

Correlation Coefficient:

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{Var(X)\,Var(Y)}}.$$

Properties:

- $-1 \le \rho_{X,Y} \le 1$.

If $Y = bX + a$, where b and a are constants

- $\rho_{X,Y} = 1$ if $b > 0$.
- $\rho_{X,Y} = -1$ if $b < 0$.
- $\rho_{X,Y} = 0$ if $b = 0$.

Conditional distribution:

A conditional distribution of a random variable Y given another variable X taking a specific value is denoted as

$$\mathcal{P}(Y = y|X = x) = \frac{\mathcal{P}(Y = y, X = x)}{\mathcal{P}(X = x)}.$$

in the continuous case

$$f(y|x) = \frac{f(y, x)}{f_X(x)}$$

Conditional Expectation: The conditional expectation of a random variable Y given another variable X taking a specific value is defined (if Y takes l possible values $y_1, ..., y_l$) as

$$E[Y|X = x] = \begin{cases} \sum_{i=1}^{l} y_i \mathcal{P}(Y = y_i|X = x) \\ \int_{-\infty}^{+\infty} y f(y|x) dx \end{cases}.$$

Law of iterated Expectations

$$E(Y) = E_X(E[Y|X])$$

The *redconditional variance* is defined as

$$Var[Y|X = x] = \begin{cases} \sum_{i=1}^{l} (y_i - E[Y|X = x])^2 \mathcal{P}(Y = y_i|X = x) \\ \int_{-\infty}^{+\infty} (y - E[Y|X = x])^2 f(y|x) dy \end{cases}$$

Review of some basic distributions

Normal distribution

A random variable X is said to have a normal distribution, i.e. $X \sim N(\mu, \sigma^2)$ if it has a density function given by:

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

where μ and σ^2 are the mean and variance of X, respectively.

Standard Normal Distribution

Any normal variable can be transformed to a *standard* normal Z defined as $Z = \frac{X-\mu}{\sigma}$, so that Z is $N(0,1)$ with density function

$$\phi(z) = (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}z^2\right). \tag{2}$$

95% of area underneath $\phi(z)$ lies between -2 and 2, almost all between -3 and 3. The distribution function for $\phi(z)$ is

$$\Phi(z) = \int_{-\infty}^{z} \phi(w)\, dw = \int_{-\infty}^{z} (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}w^2\right) dw.$$

Distributions derived from the normal

The χ^2 distribution

If $Z \sim N(0,1)$ then Z^2 is distributed as

$$Z^2 \sim \chi^2(1),$$

i.e. a *chi-squared* (χ^2) with 1 *degree of freedom* (d.o.f.). More generally, consider k *independent* random variables $Z_i, i = 1, \ldots, k$ with standard normal distribution, that is $Z_i \sim N(0,1), i = 1, \ldots, k$; then

$$X = \sum_{i=1}^{k} Z_i^2 \sim \chi^2(k),$$

i.e. a χ^2 distribution with k d.o.f.

Properties:

- Expected Value: $E(X) = k$
- Variance $Var(X) = 2k$.
- $Skewness = \sqrt{8/k}$
- $Kurtosis = 12/k$

The Student's t distribution

The Student's t distribution with k degrees of freedom is obtained as the ratio between a standard normal and the square root of a $\chi^2(k)$ divided by the number of degrees of freedom k. Formally if $X \sim N(\mu, \sigma^2)$, $\eta \sim \chi^2(k)$ and X and η are *independent* then

$$T = \frac{\frac{(X-\mu)}{\sigma}}{\sqrt{\frac{\eta}{k}}} = \frac{\sqrt{k}(X-\mu)}{\sigma\sqrt{\eta}} \sim t(k),$$

i.e. a Student t with k d.o.f.

Properties:

- Expected Value: $E(T) = 0$, if $k > 1$.
- Variance $Var(T) = k(k-2)$ if $k > 2$.
- $skewness = 0, k > 3$.
- $kurtosis = \frac{6}{k-4}, k > 4$.

The F distribution

The F distribution with k_1 and k_2 degrees of freedom (where k_1 denotes the degrees of freedom of the numerator and k_2 those of the denominator) is obtained as the distribution of the ratio of two independent random variables with χ^2 distribution, each divided by its number of degrees of freedom. Formally, if $\eta_1 \sim \chi^2(k_1)$ and $\eta_2 \sim \chi^2(k_2)$, and η_1 and η_1 are independent, then

$$F = \frac{\eta_1/k_1}{\eta_2/k_2} \sim F(k_1, k_2),$$

i.e. an F with k_1 and k_2 d.o.f. Notice that:

- The F distributed random variable can only take positive values (since it's the ratio of two random variables that can only take positive values).
- $t(k)^2 = F(1, k)$.
- If $F \sim F(k_1, k_2)$, then $k_1 F \overset{a}{\sim} \chi^2(k_1)$
- Expected Value: $E(F) = \dfrac{k_2}{k_2 - 2}$, provided $k_2 > 2$.
- Variance $Var(F) = \dfrac{2k_2^2(k_2 + k_1 - 2)}{k_1(k_2 - 2)^2(k_2 - 4)}$, provided $k_2 > 4$.