**Revision of some math details**

1.2 Distribution, Densities, and Moments

In econometrics, we are interested in ‘*random variables’* (e.g., r.v, or ), but we do not observe r.v themselves but their ‘realizations’ (e.g., as the realization). For a scalar r.v (e.g., a r.v which takes values on the (subset) of a real line, such as non-negative values), a realization of it is a single real value.

**Moments of r.v**

We are interested in the moments of r.v. (when we talk about their moments, we intend to mean their moments in the population – we may ‘estimate’ these moments using sampled data in later stage). The 1st moment of the r.v is referred as its expectation and written as (for the population).

For a discrete r.v, its expectation (e.g., 1st moment) is:

where is the possible value of the r.v and is the its associated possibility. Here represents the r.v, and , e.g., the 1st moment, is a computed value. For a continuous r.v, its expectation (e.g., 1st moment) is:

where is the PDF of the of the r.v. Here we assume that the PDF exists, otherwise the moment would not exist neither. For a continuous r.v, its moment is written as:

In practice, we are usually more interested in the “centred” moments of the r.v. e.g.,

we refer the 2nd centred moment of the r.v as its **variance**:

Where is called standard errors. For a continuous r.v., its centred moment is written as:

The moments are computed values.

**Multivariate distribution**

1. The PDF and CDF of multivariate r.v.

A **multivariate variable**, or a **random vector**, contains a vector of r.v which follow a single mathematic system, e.g., they follow a single joint distribution. e.g., a bivariate r.v contains two r.v in a vector (e.g., , ) and has a distribution function with the following CDF:

Where indicates set intersection. It can also be written as:

(1.10)

For a continuous r.v, its PDF, if exists, is a joint PDF as follows (e.g., derived from (1.10)):

(1.09)

The PDF has the following property:

2. To define “statistical independence”.

and are said to be **(statistically) independent** if the joint PDF of and is the product of the PDFs of and the PDFs of , e.g.,

(1.11)

Where the function is called the marginal CDF of and it is the CDF of considered by itself (e.g., it puts no constraint at all on ). Similarly, is called the marginal CDF of .

(1.11) is equivalent to:

, given that their PDF exist.

The proof of the equivalence above is:

From (1.09), we know:

(1.09)

From (1.10), we know:

Thus, if (1.11) holds, we have:

(1.11b)

According to (1.11b), we can rewrite (1.09) as:

(1.11c)

Thus, and are said to be **(statistically) independent** if:

Or

Or

, given that their PDF exist.

**Conditional probability**

Suppose A and B are any two events. Thus, we have:

, given that

In this example, we have which is the probability of event A conditional on the event of B. e.g., the chance of A happens given that B has happened.

We can extend this concept to the probability of an event conditional on a continuous r.v. e.g., suppose we have two continuous r.v, e.g., and , and we are interested in the PDF (and/or CDF) of conditional on a specified value (e.g., a realization) of . Therefore, we can **define** the conditional PDF as:

is the PDF of given that has a realized value of . Intuitively, it can be written as:

With the conditional PDF, we can calculate the conditional probability of , e.g.,

where represents the probability of the r.v to be smaller than a specific value of , given that the other r.v, e.g., takes a specific value of (note that this is very different from the concept of the probability of taking the value of which is zero for a continuous r.v. – here we assume that the event has happened and we are not interested in how unlikely it would happen). In this case, is a non-random calculated value (e.g., a probability value).

Example, suppose we know:

Thus, by definition, we have the PDF of given the value of (e.g., say, let’s denote , i.e.,

Based on this conditional PDF, we can calculate the probability for a range of (say, to be between 0 and 0.5, just for an example, as is a continuous r.v) conditional on the fact that , e.g.,

It is obvious that this conditional probability is a non-random value as we denote as a realization of from the beginning.

While is a non-random value, is a function which depends on the value of . It takes the value of when . Back to the example above, if is yet determined, then:

Thus, the conditional probability is a function of .

**Mean independence:**

From previous, we know that and are said to be **(statistically) independent** if , and also,

Therefore, if and are independent, then:

If we take the 1st moment for both sides, we have:

We call this mean independent. Independence suggests mean independent, but not *vice versa*.

**Conditional expectation**

Based on the conditional PDF (or CDF) above, we can compute the conditional moment. We are especially interested in the conditional 1st moment, that is, the conditional expectation:

This conditional expectation is a determined, non-random quantity because is determined (e.g., takes a realized value of ). More specifically, we have:

While is a non-random value, is a function which depends on the value of . It takes the value of when . Thus, is a new variable.

**Law of Iterated Expectation**

Since is a r.v., we can compute its 1st moment, and it can be found that:

Another property: any deterministic function of a r.v conditional on the r.v is the deterministic function of a r.v itself. e.g.,

These two properties lead to an important property:

“when , for any deterministic function of h(.), ”

We use this property when we derive the MM estimator: when , we have .

Proof: , according to Law of Iterated Expectation,

According to the second property, we have:

**Variance and covariance of vectors**

Suppose that is a *k* x *k* symmetric matrix, the diagonal element of is the which is the variance of the parameter. The off-diagonal element of is, and we have:

The full matrix of can be written as:

where

(3.22)

a special case is, when ,

The correlation can be computed based on :

If are independent, then we have , and are non-correlated, but this is not necessarily true conversely. It is possible for two variables which are not independent to have zero covariance.

**Positive definite matrices**

A *k* x *k* symmetric matrix A is said to be **positive definite** if, for all non-zero k-vector , the matrix product , which is a scalar, e.g.,

is positive. If is non-negative (e.g., can be zero), then A is said to be **positive semidefinite**. Therefore, we have the following findings:

Any form of (e.g., is a matrix)is positive semidefinite. This is because:

The identify matrix is a positive definite matrix, because:

We also have the following findings:

The diagonal elements of a positive definite matrix must all be positive.

A positive definite matrix cannot be singular.

**4.3 Some common distributions**

The PDF of the normal distribution is denoted as , thus the PDF of the evaluated at is:

There is no closed form for the CDF of normal distribution , but it is easy to evaluate numerically. That is, it is straightforward to calculate the p-value for any r.v which is standard normal.

Any normally distributed r.v can be expressed as:

**Linear combination of normal variables**

It can be proven that any linear combination of independent normally distributed r.v is itself normally distributed.

**The multivariate normal distribution**

It can be proven that any linear combination of normally distributed r.v is itself normally distributed even if these variables are not independent. i.e., if is an *m*-vector of a multivariate normal distribution, e.g., , then, any scalar **,** whereis an *m*-vector of fixed coefficients, is normally distributed with meanand variance.

There is a special property for multivariate normal distribution: suppose is a vector whose elements are independently distributed and follow the standard normal distribution, then we have . This is true because independence between these elements suggests they have zero covariance. Normally, it is not true for the other way around (e.g., zero covariance usually does not imply independence). However, it is true for multivariate normal distribution as a special case.

**The Chi-squared Distribution**

suppose is a vector whose elements are independently distribution and follow the standard normal distribution, then we have . Thus, we can form a new r.v as:

is said to follow a chi-squared distribution with *m* degrees of freedom and can be written as: . In the case of a test statistic, *m* turns out to be the number of restrictions. We can derive the mean and variance of as:

It can be proven that, if and , and and are independent, then we have .

Another property: when *m* increases, distribution approaches the distribution.

**The student’s t distribution**

Suppose and , and are independent, we can form a r.v as:

is said to follow a student’s *t* distribution with *m* degrees of freedom. The student’s t distribution approaches standard normal distribution when *m* increase.

**The F-distribution**

Suppose we have and , and and are independently distributed, then we can form a new r.v as:

is said to follow an F-distribution with and degrees of freedom. It can be written in a compact way as .

It can be shown that the square of a r.v which follows a student’s *t* distribution of would follow an F distribution of .