Master 2

Econometrics I

Nour Meddahi (Toulouse School of Economics)

Basics in Statistics

I. Moments of Random Variables

Moments of a random variable:

$$\text{Expectation or Mean} = E[X] = \mu$$

$$\text{Variance} = E[(X-\mu)^2] = E[X^2] - (E[X])^2 = \sigma^2$$

$$\text{Volatility} = \sigma$$

$$\text{Skewness coefficient} = \frac{E[(X-\mu)^3]}{(Var[X])^{3/2}}$$

$$\text{Kurtosis coefficient} = \frac{E[(X-\mu)^4]}{(Var[X])^2}$$

Dependence between two random variables:

$$\text{Covariance} = Cov[X,Y] = E[(X-\mu_x)(Y-\mu_y)] = E[XY] - \mu_X\mu_y$$

$$\text{Correlation} = Corr[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}}.$$

Sample counterparts. One observes $x_1, x_2, ..., x_n$:

$$\text{sample mean of X} : \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\text{sample variance of X} : \widehat{Var}[X] = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \ - \ (\bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$$

$$\text{sample skewness of X} : \widehat{Skew}[X] = \frac{1}{(\widehat{Var}[X])^{3/2}}\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^3$$

$$\text{sample kurtosis of X} : \widehat{Kurt}[X] = \frac{1}{(\widehat{Var}[X])^{2}}\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^4$$

$$\text{sample covariance of X and Y} : \widehat{Cov}[X,Y] = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})$$

$$\text{sample correlation of X and Y} : \widehat{Corr}[X,Y] = \frac{\widehat{Cov}[X,Y]}{\sqrt{\widehat{Var}[X]\widehat{Var}[Y]}}$$

II. Normal and Student distributions

• **Normal distribution.** If $X \sim \mathcal{N}(\mu,\sigma^2)$, with density function. $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. Then,

$$E[X] = \mu, \quad Var[X] = \sigma^2, \quad \text{Skew}[X] = 0, \quad \text{Kurt}[X] = 3.$$

• **Chi-squared distribution:** Let X_i for $i = 1, 2, \ldots, n$ be a set of independent standard normal random variables. Let $Y \equiv \sum_{i=1}^n X_i^2$. Then Y is Chi-squared distribution with n degrees of freedom.

• **Student distribution $\mathcal{T}(n)$.** Suppose that X is standard normal, Y is Chi-squared with n degrees of freedom, and X and Y are independent. Then $Z = X/\sqrt{Y/n}$ is T-distributed and has the density function:

$$f(x) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \text{ where } \Gamma(w) \equiv \int_0^\infty s^{w-1} \exp[-s] ds.$$

• The moments of order $m \geq n$ do not exist. When the moments exist, one has

$$E[\mathcal{T}(n)] = 0, \; Var[\mathcal{T}(n)] = \frac{n}{n-2}, \; Skew[\mathcal{T}(n)] = 0, \; Kurt[\mathcal{T}(n)] = 3\frac{n-2}{n-4}$$

• When $n \to +\infty$, we have $\mathcal{T}(n) \to \mathcal{N}(0,1)$. In practice, $n = 30$.

III. Basics of Estimation

Main goal: Often, we want to characterize the relationship between two variables (e.g. the market's and Vodafone's returns) or to understand the behavior of a random variable (e.g. the return of an asset) in order to forecast it, etc.... For this purpose, we need to know some parameters of interest. For instance, the conditional distribution of Vodafone's return given the market's return will depend on some parameters θ that are **unknown**. Therefore, we have to use the **data** to extract the information about θ. More precisely, we will propose a function of the data, called an **estimator** of θ and denoted $\hat{\theta}$, that **approximates** θ. We call this method **estimation**. We will study the properties of $\hat{\theta}$.

Definition: Random Sample. A sample of n observations on one or more variables, denoted $x_1, x_2, ..., x_n$ is a **random sample** if the n observations are drawn independently from the same population, or probability distribution, $F(x;\theta)$. The random sample is said to be **independent, identically distributed** and denoted **i.i.d.**

When one studies time series, one considers **consecutive** observation of the same variable (like daily returns of an stock). In that case, the sample is not i.i.d.

Definition: Statistic. Any function of the data $x_1, x_2, ..., x_n$ is called a **statistic**.

Definition: Empirical Distribution. Let $(x_1, ..., x_n)$ be a random sample. The empirical distribution is the distribution that assigns probability $1/n$ to each x_i, i.e.

$$P_n[X = x_i] = \frac{1}{n}.$$

Definition: Sample Moments. The moments of the empirical distributions are called the sample moments: for any function $h(X)$, its sample moments is

$$E_n[h(X)] \equiv \frac{1}{n}\sum_{i=1}^{n} h(x_i).$$

Definition: Sample Frequency. A sample frequency is the observed frequency of an event $X \in A$.

Note that $P(X \in A) = E[\mathbf{1}_{X\in A}]$ where $\mathbf{1}_B$ is the indicator function, i.e. $\mathbf{1}_B = 1$ when the event B holds and $\mathbf{1}_B = 0$ otherwise. Hence,

$$P_n[X \in A] = E_n[\mathbf{1}_{X\in A}] = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_i\in A}.$$

Application: Histogram, i.e., the empirical density function.

Properties of estimators.

• An estimator is a random variable given that it is a function of the data which are random. Therefore, we will try to characterize its density function, expectation, variance, etc...

• In general, it is very hard to characterize the distribution of an estimator for a given sample size n. We will need the use of the asymptotic theory (or large sample theory).

• Consequently, we will focus for the moment on the expectation and the variance of an estimator.

• A desirable property of any estimator is to be unbiased, i.e., if you repeat the experiment, on average, you will obtain the unknown parameter θ.

Definition: Unbiased Estimator. $\hat{\theta}$ is unbiased when

$$E[\hat{\theta}] = \theta.$$

Definition: Bias of an Estimator. The bias of an estimator $\hat{\theta}$ is defined as

$$\text{Biais}[\hat{\theta}] = E[\hat{\theta}] - \theta.$$

Examples: Assume that $E[X] = \mu$ and $Var[X] = \sigma^2$.

- The sample average is an unbiased estimator of $\mu = E[X]$.

$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n}\sum_{i=1}^{n} E[X] = E[X] = \mu.$$

- The sample variance is biased. One can show that

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2\right] = \frac{n-1}{n}\sigma^2.$$

This leads to the introduction of a new unbiased estimator of σ^2 (used in practice and in softwares) given by

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2.$$

• Another desirable property of an unbiased estimator is relate to its variance. We would prefer a small variance.

Example: variance of the sample mean.

$$Var[\bar{X}] = \frac{\sigma^2}{n}$$

Definition: Efficiency. Consider two unbiased estimator $\hat{\theta}_1$ and $\hat{\theta}_2$ of θ. $\hat{\theta}_1$ will be called more efficient $\hat{\theta}_2$ when

$$Var[\hat{\theta}_1] < Var[\hat{\theta}_2].$$

When θ is vector, the previous definition means that the matric $Var[\hat{\theta}_2] - Var[\hat{\theta}_1]$ is a positive definite matrix.

• It is also of interest to compare estimators that are biased. In this case, we need a criterion that combines the bias and the variance of the estimator.

Definition: Mean-squared error of an estimator. It is defined as

$$\begin{aligned} MSE[\hat{\theta}] &\equiv E[(\hat{\theta}-\theta)^2] = Var[\hat{\theta}] + (Bias[\hat{\theta}])^2 \text{ when } \theta \text{ is a scalar} \\ &\equiv E[(\hat{\theta}-\theta)(\hat{\theta}-\theta)'] = Var[\hat{\theta}] + (Bias[\hat{\theta}])(Bias[\hat{\theta}])' \text{ when } \theta \text{ is a vector.} \end{aligned}$$

We always prefer the estimator that has the smallest MSE (when it is possible to have one).

Example: When $X \sim \mathcal{N}(\mu, \sigma^2)$, one can show that

$$Var[\hat{V}ar[X]] = \frac{2(n-1)\sigma^4}{n^2} \;<\; Var[s^2] = \frac{2\sigma^4}{n-1}$$

while

$$MSE[\hat{V}ar[X]] - MSE[s^2] = \sigma^4 \left[\frac{2n-1}{n^2} - \frac{2}{n-1}\right] > 0.$$

In other words, we prefer the estimator s^2.

• There are other properties of estimators that are important, in particular their distributions in finite sample or asymptotically (when the sample size $n \to +\infty$).

IV. Basics on Finite Sample Distribution of estimators.

• Ideally, we would like to know the finite sample distribution (i.e. the exact distribution) of the estimator $\hat{\theta}$ to fully know the properties of the estimator. Unfortunately, there are few examples where one can obtain the exact distribution. Instead of the exact distribution, one will use the asymptotic theory to approximate the distribution of the estimator.

• An important example where one can derive the exact distribution of some estimators is the normal case.

The normal case: Assume that x_1, x_2, ..., x_n, are i.i.d. and follow a $\mathcal{N}(\mu, \sigma^2)$. Then one can show:

1. $$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$$

2. $$\frac{s^2}{\sigma^2} \sim \frac{\chi^2(n-1)}{n-1}$$

3. The random variables $\bar{X}$ and s^2 are independent

4. $$T = \sqrt{n}\,\frac{(\bar{X}-\mu)}{s} \sim \mathcal{T}(n-1).$$

Point Estimation and Interval Estimation.

• So far, we were interested in estimating θ, i.e., we focused on **point estimation**.

• A different approach is the **interval estimation**: We want to characterize an interval (or a set) such that the probability that the unknown parameter θ is in the interval equals given number, say 95%, and denoted $1-\alpha$.

• For this purpose, we need a **pivotal statistic**, i.e., a statistic that its distribution is known and does not depend on any parameter.

Example:

1. $\sqrt{n}(\bar{X}-\mu)$ is not a pivotal statistic because its distribution is $\mathcal{N}(0,\sigma^2)$ where σ^2 is unknown.
2. The "T" statistic $\sqrt{n}\,\dfrac{(\bar{X}-\mu)}{s}$ is pivotal because it distribution is $\mathcal{T}(n-1)$.
3. $\dfrac{s^2}{\sigma^2}$ is a pivotal statistic because it distribution $\dfrac{\chi^2(n-1)}{n-1}$ is known.

- We have

$$P\left[\left|\sqrt{n}\ \frac{(\bar{X}-\mu)}{s}\right| \leq \mathcal{T}(n-1)_{\alpha/2}\right] = 1-\alpha$$

where $\mathcal{T}(n-1)_\alpha$ is the $1-\alpha$ quantile of the $\mathcal{T}(n-1)$ distribution $(0<\alpha<1)$.

- This probability statement is equivalent to

$$P\left[\mu \in \left[\bar{X}-\mathcal{T}(n-1)_{\alpha/2}\frac{s}{\sqrt{n}}\ ,\ \bar{X}+\mathcal{T}(n-1)_{\alpha/2}\frac{s}{\sqrt{n}}\right]\right] = 1-\alpha.$$

In other words, the interval $\left[\bar{X}-\mathcal{T}(n-1)_{\alpha/2}\frac{s}{\sqrt{n}}\ ,\ \bar{X}+\mathcal{T}(n-1)_{\alpha/2}\frac{s}{\sqrt{n}}\right]$ contains μ with probability $1-\alpha$. It is called a **confidence interval**.

- One can do the same analysis for σ^2 by using the pivotal statistic $\dfrac{s^2}{\sigma^2}$.

- **Example**: $n=25$, $\bar{X}=1.63$ and $s=0.51$. One has

$$P(-2.064 \leq \frac{5(\bar{X}-\mu)}{s} \leq 2.064) = 0.95.$$

Hence, the 95% confidence interval of μ is $[1.4195, 1.8405]$.

V. Basics on Testing Hypotheses

Main goal: Often, financial theory and economic theory imply restrictions that the data should follow. Likewise, when one wants to forecast a variable y, one wants to know whether it is useful to include another variable x. In both case, one has to test the corresponding hypothesis.

- The classical testing procedures are based on constructing a statistic from a random sample that will enable us to decide with reasonable confidence whether or not the data in the sample would have been generated by a hypothesized population. This procedure involves a statement of hypothesis, usually in terms of **null** or maintained hypothesis and **an alternative**, conventionally denoted by H_0 and H_1, respectively.

- Classical (or Neyman-Person) methodology involves splitting sample space into two regions: if the test statistic falls in **rejection region** (also known as critical region), then null hypothesis H_0 is rejected; if the test statistic falls in **acceptance region** , then H_0 is not rejected.

- In order to do a tests, one has to find a pivotal statistic.

• Since the sample is random, the test statistic is also random. The same test procedure can lead to different conclusions in different samples. There are two ways such a procedure can be error:

1. Type I error. The procedure may lead to rejection of the null hypothesis when it is true.
2. Type II error. The procedure may fail to reject the null hypothesis when it is false.

Definition 1 *The probability of type I error is the* **size of the test**. *This is conventionally denoted by* α *and is also called significance level. In other words,*

$$\alpha = P[\text{ Reject the null} \mid \text{the null is true}].$$

The type I error could be eliminated by making the rejection region very small. By eliminating the probability of type I error, that is making it unlikely that the hypothesis is rejected, we must increase the probability of a type II error.

Definition 2 *The* **power of a test** *is the probability that it will correctly lead to rejection of a false null hypothesis:*

$$\text{power} = 1 - \beta = 1 - \mathrm{P}(\text{type II error})$$

with

$$\beta = \mathrm{P}(\text{type II error}) = \mathrm{P}[\text{ Does not Reject the null} \mid \text{the null does not hold}].$$

Definition 3 *A test is* **most powerful** *if it has greater power than any other test of the same size.*

Definition 4 **A test is consistent** *if its power goes to one as the sample size grows to infinity.*

Rejection region and acceptance region.

• These regions are characterized by using a pivotal statistic.

• One should characterize the rejection region, which depends on the alternative. We will use an example to show how one should proceed.

The normal case: Assume that x_1, x_2, ..., x_n, are i.i.d. and follow a $\mathcal{N}(\mu, \sigma^2)$. Then one has the pivotal statistic

$$T = \sqrt{n}\,\frac{(\bar{X} - \mu)}{s} \sim \mathcal{T}(n-1).$$

One should rewrite the pivotal statistic **under the null**. In our example, the null is $H_0 : \mu = \mu^0$, so the pivotal statistic becomes

$$T = \sqrt{n}\,\frac{(\bar{X} - \mu^0)}{s} \sim \mathcal{T}(n-1).$$

First case: A two sided test. $H_0 : \mu = \mu^0,\ H_a : \mu \neq \mu^0$. One has to consider the pivotal statistic written **under the null** and find out the area where the statistic is unlikely to follow the desirable distribution. Here, if the data are generated under the alternative, the absolute value of T should be large, meaning that $\bar{X}$ is far from μ^0. In other words, the rejection region is

$$\left| T = \sqrt{n}\ \frac{(\bar{X} - \mu^0)}{s} \right| > \mathcal{T}(n-1)_{\alpha/2}.$$

Second case: A one sided test. $H_0 : \mu = \mu^0,\ H_a : \mu > \mu^0$. One has to consider the pivotal statistic written **under the null** and find out the area where the statistic is unlikely to follow the desirable distribution. Here, if the data are generated under the alternative, the statistic T should be **positive and large**. Hence, the rejection region is

$$T = \sqrt{n}\ \frac{(\bar{X} - \mu^0)}{s} > \mathcal{T}(n-1)_{\alpha}.$$

• **Example**: $n = 25,\ \bar{X} = 1.63$ and $s = 0.51$.

- $H_0 : \mu = 1.5,\ H_a : \mu \neq 1.5$. One gets $T = 1.27$.

- $H_0 : \mu = 1.5,\ H_a : \mu > 1.5$. One gets $T = 1.27$.

- $H_0 : \mu = 1.3,\ H_a : \mu < 1.3$. One gets $T = 3.23$

• **P-value**. Softwares provide the p-value, which corresponds to the probability of the rejection region when the T statistic is its boundary. One should reject the null when the p-value is smaller than α.

Connection between interval estimation and hypothesis testing.

There is a link between interval estimation and the hypothesis test. The confidence interval give a range of plausible values for the parameter. Therefore, it stands to reason that if a hypothesized value of the parameter does not fall in this range of plausible values, the data are not consistent with the hypothesis and it should be rejected. Consider testing

$$H_0 : \theta = \theta_0 \;\text{ against }\; H_1 : \theta \neq \theta_0.$$

We know from the previous lecture how to build a confidence interval of θ. If θ^0 does not belong to the confidence set, then we reject the null. Otherwise, we do not reject it.

- **Example**: $n = 25$, $\bar{X} = 1.63$ and $s = 0.51$. The 95% confidence interval of μ is $[1.4195, 1.8405]$.

a) $H_0 : \mu = 1.5$, $H_a : \mu \neq 1.5$. At the 5 percent significance test, we do not reject the null given that 1.5 is in the confidence interval $[1.4195, 1.8405]$.

b) $H_0 : \mu = 2$, $H_a : \mu \neq 2$. At the 5 percent significance test, we reject the null given that 2 is not in the confidence interval $[1.4195, 1.8405]$.

VI. Asymptotic Theory

Main Goal: We saw in the previous lectures that in most examples that do not involve normal distributions, one can not derive the **finite sample distribution** or **exact distribution** of an estimator. Consequently, one can not for instance build confidence intervals. In this lecture, we will try to approximate the finite sample distribution of an estimator by allowing the sample size to go to infinity. We will derive the **asymptotic distribution** of an estimator. One can view this approach as a Taylor expansion type method for statistics.

We will start the lecture by studying the desirable properties of an estimator when $n \to +\infty$.

Here, we have a sequence of estimator $\hat{\theta}_n$. We will now use the index n given that the sample size varies. As in real analysis, we want to have $\hat{\theta}_n$ **converges** to θ when $n \to +\infty$. We need however to define the term **converges**, or more precisely, to define in which sense $\hat{\theta}_n$ **converges** to θ when $n \to +\infty$.

Concept of Convergence.

Convergence in probability . The sequence of random variables Z_n converges in probability to a random variable (or a constant) Z if, for any $\epsilon > 0$,

$$Prob(\mid Z_n - Z \mid > \epsilon) \to 0 \text{ as } n \to \infty.$$

We also say

$$Z_n \to_p x \text{ as } n \to \infty \text{ or } \text{Plim}_{n\to\infty} Z_n = Z.$$

In many cases of interest, Z will be a non-stochastic constant.

The concept easily extends to vectors and matrices. For instance, for random matrices Z_n, Z,

$$Z_n \to_p Z \text{ as } n \to \infty,$$

mean that convergence occurs element by element. One can also consider the norm of $Z_n - Z$ and has the following definition: $\text{Plim}_{n\to+\infty} Z_n = Z$ when for any $\varepsilon > 0$ one has

$$Prob(||Z_n - Z|| > \epsilon) \to 0 \text{ as } n \to \infty.$$

Consistent estimator: An estimator θ_n of θ is called consistent if Plim $\theta_n = \theta$.

Convergence in Mean-Square or L^2. The sequence of random variables Z_n converges in mean-square to a random variable (or a constant) Z if

$$E[(Z_n - Z)^2] \to 0 \text{ as } n \to \infty \text{ (scalar case)}$$

$$E[||Z_n - Z||^2] \to 0 \text{ as } n \to \infty \text{ (multivariate and matricx cases)}$$

Connection between convergence in probability and convergence in mean-square:

$$\textbf{Convergence in mean} - \textbf{square} \Longrightarrow \textbf{Convergence in probability}.$$

Application: sample mean. $x_1, x_2, ..., x_n$ i.i.d. from a distribution with mean μ and variance σ^2. The sample mean is given by

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

We know that $\quad E[\bar{X}_n] = \mu \;$ and $\; Var[\bar{X}_n] = \dfrac{\sigma^2}{n}.$

Hence,

$$E[(\bar{X}_n - \mu)^2] = \frac{\sigma^2}{n} \to 0 \text{ as } n \to +\infty,$$

i.e., $\bar{X}_n$ converges in mean-square to μ. Consequently, $\bar{X}_n$ converges probability to μ, i.e., $\bar{X}_n$ is a consistent estimator of μ.

There are other types of convergence, although are less used in practice.

Almost sure convergence. The sequence of random variables Z_n converges almost surely to a random variable (or a constant) Z if, for any $\epsilon > 0$,

$$Prob(\sup_{i \geq n} ||Z_i - Z|| > \epsilon) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

We write $Z_n \rightarrow_{as} Z$.

The almost sure convergence is a very strong concept and it implies all the others concept of convergence.

Convergence in p-th mean or L^p. The sequence of random variables Z_n converges in L^p to a random variable (or a constant) Z if

$$E[|Z_n - Z|^p] \rightarrow 0 \text{ as } n \rightarrow \infty \text{ (scalar case)}$$
$$E[||Z_n - Z||^p] \rightarrow 0 \text{ as } n \rightarrow \infty \text{ (multivariate and matricx cases)}$$

Assume that $p > q$. Then,

$$\textbf{Convergence in } L^p \Longrightarrow \textbf{Convergence in } L^q.$$

We will now study the asymptotic distribution of the estimators. We need to introduce the concept of of **convergence in distribution**.

Convergence in distribution. Let Z_n be a sequence of random variables, each with a cdf $F_n(z)$. Then Z_n converges in distribution to the random variable Z with cdf $F(z)$ if for any z

$$F_n(z) \to F(z) \text{ as } n \to \infty.$$

We shall indicate this as

$$Z_n \to_d Z \text{ as } n \to \infty.$$

We shall call $F(z)$ the **asymptotic distribution** of Z_n.

Remark: The convergence in probability implies the convergence in distribution, but not the opposite. The opposite holds when Z is non-stochastic.

Two very important results are the following.

Slutzky Theorem. If $Z_n \to_p Z$ then $g(Z_n) \to_p g(Z)$ for any **continuous** function $g(\cdot)$.

Continuous mapping Theorem. If $Z_n \to_d Z$ then $g(Z_n) \to_d g(Z)$ for any **continuous** function $g(\cdot)$.

These extends to vectors and matrices, implying the following.

For convergence in probability. If Z_n and Y_n are r.v. with $\text{plim} Z_n = Z$, $\text{plim} Y_n = Y$, then

$$\begin{aligned}
&\text{plim}(Z_n + Y_n) = Z + Y, \\
&\text{plim} Z_n Y_n = ZY, \\
&\text{plim} Z_n / Y_n = Z/Y \ \ for \ \ Y \neq 0.
\end{aligned}$$

If Z_n is random matrix and plim $Z_n = Z$ with constant and invertible Z, then

$$\text{plim } Z_n^{-1} = Z^{-1}.$$

For convergence in distribution. If $Z_n \rightarrow_d Z$ and $Y_n \rightarrow_p Y$, then

$$Z_n Y_n \rightarrow_d ZY,$$

$$Z_n / Y_n \rightarrow_d Z/Y \quad for \quad Y \neq 0.$$

If Z_n is random vector and $Z_n \rightarrow_d X$ and Y_n is random matrix with $Y_n \rightarrow_p Y$ then

$$Y_n Z_n \rightarrow_d YZ.$$

Cramer-Wold device. If $Z_n \rightarrow_d Z$ then the scalar $Y_n \equiv c' Z_n \rightarrow_d c' Z \equiv y$ for any constant vector c.

Law of Large Number

By Law of Large Number (LLN) we mean a set of results which states the conditions under which the simple average of the observations of a sample $(x_1, ..., x_n)$ goes, either in probability (weak LLN) or almost surely (strong LLN), to the unconditional mean.

For example Khinchine's LLN says that

$$\bar{X}_n \rightarrow_p \mu$$

when x_i are $i.i.d.$ with finite mean $\mu = Ex_i$.

The various LLNs differ in terms of the different strength of the underlying assumptions. Another one is *Chebychev* LLN: if x_i are such that $\mu_i = Ex_i < \infty$, $var(x_i) = \sigma_i^2 < \infty$ and

$$\frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 \rightarrow 0$$

then

$$\bar{x} - \mu_n \rightarrow_p 0$$

where $\mu_n = 1/n \sum_{i=1}^{n} \mu_i$.

Central Limit Theorem

By Central Limit Theorem (CLT) we mean the set of results which state the conditions under which a sequence of rv, suitably normalized, converges to a standard normal. A simple example is the **Lindberg-Levy** CLT: if x_i are $i.i.d.$ with finite mean μ and finite variance σ^2 then

$$\sqrt{n}(\bar{X}_n - \mu) \rightarrow_d N(0, \sigma^2).$$

Note that the marginal distribution of each x_i can be anything you like, as long as the first two moments are constant and finite.

Application: One can show that

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X}_n)^2 \rightarrow_p \sigma^2.$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X}_n)^2 \rightarrow_p \sigma^2.$$

Hence,

$$\sqrt{n}\frac{(\bar{X}_n - \mu)}{\hat{\sigma}} \rightarrow_d N(0, 1).$$

and

$$\sqrt{n}\frac{(\bar{X}_n - \mu)}{s} \rightarrow_d N(0, 1).$$

Another CLT very much used is the multivariate **Lindberg-Feller** CLT: if X_i is a sequence of random variables such that

1) all mixed third moments of the sequence are finite, that is $E \mid X_{ai}X_{bj}X_{ck} \mid < \infty$,

2)

$$\frac{1}{n}\sum_{i=1}^{n} Q_i \to Q > 0 \text{ (i.e., definite – positive)}$$

where $Q_i = Var(X_i)$.

3)

$$\left(\sum_{i=1}^{n} Q_i\right)^{-1} Q_j \to 0 \text{ for every } j.$$

Then, by defining μ_n as

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i,$$

one gets

$$\sqrt{n}(\bar{X}_n - \mu_n) \to_d N(0, Q).$$

Delta method.

Assume that $\theta \in \mathbb{R}^k$ is a vector and that

$$\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, Q).$$

Consider a function $g(\cdot) : \mathbb{R}^k \to \mathbb{R}^q$ which is C^1, i.e., it is differentiable and its derivative is a continuous function. Then one has

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \to_d N\left(0, \frac{\partial g}{\partial \theta'}(\theta)\ Q\ \frac{\partial g'}{\partial \theta}(\theta)\right).$$

Useful result when one is interested in non-linear transforms of parameters.

Unbiased estimator, Confidence set, testing, etc...

All the analysis that we done for **n fixed** can be extended to the case $n \to +\infty$. One should add the term "asymptotic".

Asymptotically Unbiased estimator:

$$E[\hat{\theta}_n] - \theta \to 0 \text{ when } n \to +\infty.$$

Asymptotically Efficient Estimator: $\hat{\theta}_{1,n}$ is asymptotically more efficient that $\hat{\theta}_{2,n}$ when

$$n(Var[\hat{\theta}_{2,n}] - Var[\hat{\theta}_{1,n}]) \to Q \text{ when } n \to +\infty$$

where Q is definite positive. Here we add n in front of $(Var[\hat{\theta}_{2,n}] - Var[\hat{\theta}_{1,n}])$ because the two variances converge to 0.

Asymptotic pivotal statistic.

Asymptotic level of a test.

Asymptotic power of a test.

Testing hypothesis by using the asymptotic distribution.

Application: Sample mean of $x_1, ...x_n$, i.i.d. from a $\mathcal{N}(\mu, \sigma^2)$. n=25. One approximates a $\mathcal{T}(24)$ by a $\mathcal{N}(0,1)$: "T=1.96" instead of "T=2.064".