Lecture 5: Sampling and Large-Sample Distribution Theory

PhD Mathematics II: Probability

Juan Llavador Peralt

IIES

November 12, 2024

Outline

Sample

- Let $X_1, X_2, \ldots, X_n$ be a collection of **independent** and **identically** distributed RVs
- $X_1 \sim f_X(x|\theta)$, where θ is a parameter vector (in the Gaussian case, $\theta = (\mu, \sigma^2)$)
- Then $X_1, X_2, \ldots, X_n$ is a **random sample** of size n from the distribution $f_X(x|\theta)$
- A **realization** of the sample is $x_1, x_2, \ldots, x_n$
- The **joint density** of the sample is:

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n|\theta) = \prod_{i=1}^{n} f_{X_1}(x_i|\theta)$$

Statistics

- Let $X_1, X_2, \ldots, X_n$ be **random sample**
- A **statistic** is a function of the sample:

$$T = g(X_1, X_2, \ldots, X_n)$$

- It does not depend on the parameter θ
- Examples:
  - Sample mean: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
  - Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
  - Minimum: $X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}$
  - Maximum: $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$
- Question: Is $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])^2$ a statistic?

The Sample Mean

- The sample mean is a statistic: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and hence a RV
- What is the **expectation** of the sample mean?

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\cdot n\mathbb{E}[X_1] = \mathbb{E}[X_1]$$

- What is the **variance** of the sample mean?

$$\mathsf{Var}[\bar{X}] = \frac{1}{n^2}\sum_{i=1}^{n}\mathsf{Var}(X_i) = \frac{1}{n}\mathsf{Var}(X_1)$$

The Sample Variance

- The sample variance is a statistic: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ and hence a RV
- What is the **expectation** of the sample variance?
- First, note that we can rewrite S^2 as:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

- Note that $\mathbb{E}[X_i^2] = \mathsf{Var}(X_i) + \mathbb{E}[X_i]^2$
- And $\mathbb{E}[\bar{X}^2] = \mathsf{Var}(\bar{X}) + \mathbb{E}[\bar{X}]^2 = \mathsf{Var}(X_1)/n + \mathbb{E}[X_1]^2$
- So we have:

$$\mathbb{E}[S^2] = \frac{1}{n-1}\left(n(\mathsf{Var}(X_1) + \mathbb{E}[X_1]^2) - n(\mathsf{Var}(X_1)/n + \mathbb{E}[X_1]^2)\right) = \mathsf{Var}(X_1)$$

Markov's Inequality

- Let X be a non-negative RV with finite expectation $\mathbb{E}[X] < \infty$
- Then, for any $a > 0$, we have:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

- Proof (for X continuous):

$$\begin{aligned}
\mathbb{P}(X \geq a) &= \int_a^\infty f_X(x) d(x) \\
&\leq \int_a^\infty \frac{x}{a} f_X(x) d(x) \\
&= \frac{1}{a} \int_a^\infty x f_X(x) d(x) \\
&\leq \frac{1}{a} \mathbb{E}[X]
\end{aligned}$$

Chebychev's Inequality

- Let X be a RV with finite mean μ and variance σ^2
- Then, for any $k > 0$, we have:

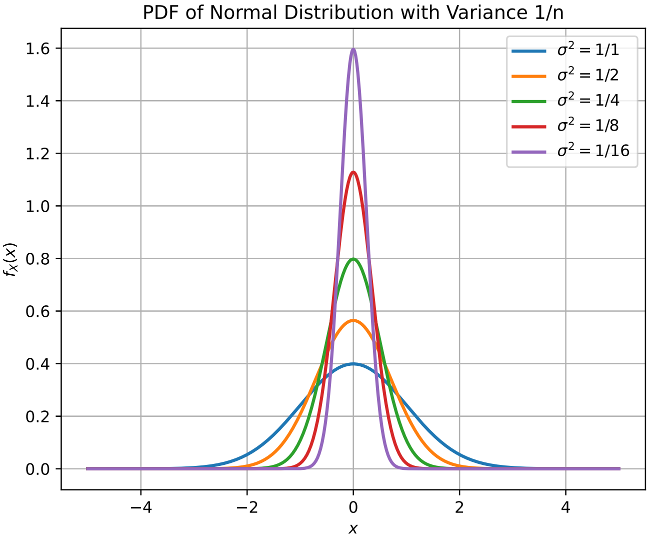
$$\mathbb{P}(|X - \mu| \geq \sigma k) \leq \frac{1}{k^2}$$

- Proof:

$$\begin{aligned}
\mathbb{P}(|X - \mu| \geq \sigma k) &= \mathbb{P}((X - \mu)^2 \geq \sigma^2 k^2) \\
&\leq \frac{\mathbb{E}[(X - \mu)^2]}{\sigma^2 k^2} \\
&= \frac{1}{k^2}
\end{aligned}$$

Convergence for Random Variables

Consider a sequence of $\mathcal{N}(0, 1/n)$



Convergence in Probability

- Let $X_1, X_2, \ldots$ be a sequence of RVs and X be another RV
- We say that X_n **converges in probability** to X if:

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0, \quad \forall \epsilon > 0$$

- We write: $X_n \xrightarrow{p} X$ or plim$X_n = X$
- Note that it is equivalent to:

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| < \epsilon) = 1, \quad \forall \epsilon > 0$$

- Often $X = \mu \in \mathbb{R}$ is a constant
- Used to establish the **consistency** of an estimator (more of this in Econometrics I)

Weak Law of Large Numbers

- Let $X_1, X_2, \ldots$ be a sequence of i.i.d. RVs with $\mathbb{E}[X_1] = \mu$ and Var$[X_1] = \sigma^2 < \infty$
- Let $S_n/n = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Then, $S_n/n \xrightarrow{p} \mu$
- Proof:

$$\begin{aligned}
\mathbb{P}(|S_n/n - \mu| \geq \epsilon) &= \mathbb{P}\left(|S_n - n\mu| \geq \epsilon \frac{\sigma}{\sqrt{n}} \frac{\sqrt{n}}{\sigma}\right) \\
&\leq \frac{\sigma^2}{n\epsilon^2} \to 0, \quad \text{as } n \to \infty
\end{aligned}$$

Convergence Almost Surely

- Let $X_1, X_2, \ldots$ be a sequence of RVs and X be another RV
- We say that X_n **converges almost surely** to X if:

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$$

- We write:

$$X_n \xrightarrow{a.s.} X$$

- Resembles pointwise convergence of functions
  - In analysis, $f_n(x) \to f(x)$ pointwise if $\lim_{n\to\infty} f_n(x) = f(x)$ for all x
  - Here, $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for all ω except for those with measure zero

Strong Law of Large Numbers

- Let $X_1, X_2, \dots$ be a sequence of i.i.d. RVs with $\mathbb{E}[X_1] = \mu < \infty$
- Let $S_n = \sum_{i=1}^n X_i$
- Then,
$$S_n/n \xrightarrow{a.s.} \mu$$
- That is
$$\mathbb{P}(\{\omega : \lim_{n\to\infty} S_n(\omega)/n = \mu\}) = 1$$
- Intutively, SLLN suggest $\bar{X}_n, \bar{X}_{n+1}, \dots$ will be simulatenously close to μ
- WLLN suggests that each $\bar{X}_n$ will be close to μ

Convergence in Mean-Square

- Let $X_1, X_2, \ldots$ be a sequence of RVs and X be another RV
- We say that X_n **converges in mean-square** to X if:

$$\lim_{n \to \infty} \mathbb{E}[(X_n - X)^2] = 0$$

- We write:

$$X_n \xrightarrow{m.s.} X$$

Convergence in Distribution

- Let $X_1, X_2, \ldots$ be a sequence of RVs and X be another RV
- We say that X_n **converges in distribution** to X if:

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x), \quad \text{for all } x \text{ where } F_X(x) \text{ is continuous}$$

- We write:

$$X_n \xrightarrow{d} X$$

- Used to establish the **asymptotic normality** of an estimator (more of this in Econometrics I)

Convergence of Moment Generating Functions

- Let $X_1, X_2, \ldots$ be a sequence of RVs and X be another RV
- Let $M_{X_n}(t)$ and $M_X(t)$ be the MGFs of X_n and X, respectively
- We say that X_n **converges in distribution** to X if:

$$\lim_{n\to\infty} M_{X_n}(t) = M_X(t), \quad \text{for all } t \text{ in a neighborhood of zero}$$

- Sometimes called "Levy's Continuity Theorem"
- This is the case in which the MGF exists, but this extends to characteristic functions too

Central Limit Theorem (Lindeberg-Levy)

- Let $X_1, X_2, \ldots$ be a sequence of i.i.d. RVs with $\mathbb{E}[X_1] = \mu$ and $\text{Var}[X_1] = \sigma^2 < \infty$
- Let $S_n = \sum_{i=1}^{n} X_i$
- Then, the distribution of the standardized sum converges to the standard normal distribution:
$$\sqrt{n}\frac{S_n/n - \mu}{\sigma} \xrightarrow{d} Z \sim \mathcal{N}(0, 1)$$
- Proof: Use the characteristic function

Proof when the MGF exists 1/2

- Consider the MGF of the standard normal: $M_Z(t) = \exp\left(\frac{t^2}{2}\right)$
- We will show that $M_{\sqrt{n}\frac{S_n/n-\mu}{\sigma}}(t) \to M_Z(t)$
- Define $Y_i = \frac{X_i-\mu}{\sigma}$, then $\sqrt{n}\frac{S_n/n-\mu}{\sigma} = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i$
- Since X_i are i.i.d., Y_i are i.i.d. with $\mathbb{E}[Y_i] = 0$ and Var$[Y_i] = 1$
- The MGF of $\frac{S_n-n\mu}{\sigma\sqrt{n}}$ is:

$$M_{\frac{S_n-n\mu}{\sigma\sqrt{n}}}(t) = \left[M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

Proof when the MGF exists 2/2

- Taylor expand the MGF of Y_1 around 0:

$$M_{Y_1}\left(\frac{t}{\sqrt{n}}\right) = \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{t}{\sqrt{n}}\right)^k \mathbb{E}[Y_1^k]$$

- Note that $\mathbb{E}[Y_1] = 0$ and $\mathbb{E}[Y_1^2] = 1$
  Then

$$M_{Y_i}(t) = 1 + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + o\left[\left(\frac{t}{\sqrt{n}}\right)^2\right]$$

- Where $o(t^2/n)$ means that $\lim_{n\to\infty} \frac{o(t^2/n)}{(t^2/n)} = 0$

- Then,

$$M_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) = \left[1 + \frac{1}{n}\left(\frac{1}{2}t^2 + no\left[\left(\frac{t}{\sqrt{n}}\right)^2\right]\right)\right]^n \to \exp\left(\frac{t^2}{2}\right), \quad \text{as } n \to \infty$$

Aside: Multivariate Gaussian Distribution

- Let $X_1, X_2, \dots, X_n$ be random variables and $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ a random vector
- $\mathbb{E}[\mathbf{X}] = \mu$ is an $n \times 1$ vector and Var$[\mathbf{X}] = \Sigma$ is an $n \times n$ matrix
- We say that $\mathbf{X}$ follows a **multivariate Gaussian distribution** $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$ if:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)$$

- We say that $X_1, X_2, \dots, X_n$ are **jointly Gaussian**

Aside: Joint Normality and Independence

- Normally and independent distributed RVs are jointly Gaussian
- The converse is not true: jointly Gaussian does not imply independence
- In general, uncorrelated RVs are not independent
- However, if $X_1, X_2, \ldots, X_n$ are **jointly Gaussian** and uncorrelated, then they are independent
- This does not hold if the RVs are not **jointly** Gaussian

Multivariate CLT

- Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be a sequence of k-dimensional i.i.d. RVs
- With $\mathbb{E}[\mathbf{X}_1] = \mu$ and Var$[\mathbf{X}_1] = \Sigma$ is symmetric and positive definite
- Define the random vector $\mathbf{Z}_i = \Sigma^{-1/2}(\mathbf{X}_i - \mu)$
- Then:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{Z}_i \xrightarrow{d} \mathcal{N}(0, I_k)$$

On the assumptions of the CLT

- We have assumed that X_i are i.i.d. with finite mean and variance
- **Lyapunov's and Lindeberg-Feller CLT** relax the identically distributed assumption
- Can also be extended to some cases with dependent RVs
- There is also a generalized CLT for the case with infinite variance (but it does not converge to a normal distribution)

Relationship Between Modes of Convergence

- Almost sure convergence implies convergence in probability:

$$X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{p} X$$

- Convergence in mean-square implies convergence in probability:

$$X_n \xrightarrow{m.s.} X \Rightarrow X_n \xrightarrow{p} X$$

- Convergence in probability implies convergence in distribution:

$$X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$$

- When X is a constant, convergence in distribution implies convergence in probability:

$$X_n \xrightarrow{d} \mu \Rightarrow X_n \xrightarrow{p} \mu$$

Convergence in Mean-Square implies Convergence in Probability

- Let $X_n \xrightarrow{m.s.} X$
- Then using Markov inequality:

$$\mathbb{P}(|X_n - X| \geq \epsilon) \leq \frac{\mathbb{E}[(X_n - X)^2]}{\epsilon^2} \to 0, \quad \text{as } n \to \infty$$

Main Convergence Theorem

- Suppose that X_n converges to X in a certain mode
- Question: What can we say about $g(X_n)$?
- Suppose $X_1, X_2, \ldots$ is a sequence of random vectors $\mathbb{R}^k$ and $g : \mathbb{R}^k \to \mathbb{R}^s$ is a **continuous** function
- Then:

$$X_n \xrightarrow{a.s.} X \Rightarrow g(X_n) \xrightarrow{a.s.} g(X)$$
$$X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$$
$$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$$

Slutsky-Cramer's Convergence Theorem

- For $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$, we have:

$$
\begin{gathered}
X_n + Y_n \xrightarrow{d} X + c \\
X_n Y_n \xrightarrow{d} Xc \\
X_n / Y_n \xrightarrow{d} X/c, \quad \text{if } c \neq 0
\end{gathered}
$$

- Cramer-Wold Theorem: If $X_n \xrightarrow{d} X$, then for any $a \in \mathbb{R}^k$, $a' X_n \xrightarrow{d} a' X$

Algebra of Probability Limits

- For a continuous function $g(x) : \mathbb{R}^k \to \mathbb{R}^s$ and suppose that $\text{plim}X_n$ exists:

$$\text{plim}g(X_n) = g(\text{plim}X_n)$$

- Some direct implications:
  - If $\text{plim}X_n = \mu$ and $\text{plim}Y_n = \nu$, then:

$$\text{plim}(X_n + Y_n) = \mu + \nu$$
$$\text{plim}(X_n Y_n) = \mu\nu$$
$$\text{plim}(X_n / Y_n) = \mu/\nu,$$

  - If W_n is an invertible matrix and $\text{plim}W_n = \Omega$, then:

$$\text{plim}(W_n^{-1}) = \Omega^{-1}, \quad \text{if } \Omega \text{ is invertible}$$

  - If X_n, Y_n are random matrices and $\text{plim}X_n = A$ and $\text{plim}Y_n = B$, then:

$$\text{plim}(X_n Y_n) = AB$$

- If the plim is a constant, then g only needs to be continuous at that point

Delta Method

- Let X_n be a sequence of RVs with

$$X_n \xrightarrow{p} X$$
$$\sqrt{n}(X_n - X) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

- Let $g : \mathbb{R} \to \mathbb{R}$ be a function such that $g'(X) \neq 0$

- Then:

$$\sqrt{n}(g(X_n) - g(X)) \xrightarrow{d} \mathcal{N}(0, \sigma^2[g'(X)]^2)$$

- To see where this comes from, consider the Taylor expansion of $g(X_n)$:

$$g(X_n) = g(X) + g'(X)(X_n - X) + o_p(|X_n - X|)$$

Delta Method in a Multivariate Case

- Let $X_n \in \mathbb{R}^k$ be a sequence of RVs with

$$X_n \xrightarrow{d} X$$
$$\sqrt{n}(X_n - X) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

- Let $g : \mathbb{R}^k \to \mathbb{R}$ be a function such that the Jacobian matrix $Dg(X)$ is non-singular
- Then:

$$\sqrt{n}(g(X_n) - g(X)) \xrightarrow{d} \mathcal{N}(0, Dg(X)\Sigma Dg(X)^T)$$

- Where $Dg(X)$ is the Jacobian matrix of g evaluated at X

Example: Ratio of two means

- Suppose we have two random samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$
- Assume $\begin{pmatrix} \bar{X} \\ \bar{Y} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$, with $\mu_Y \neq 0$
- And $\sqrt{n} \begin{pmatrix} \bar{X} & - & \mu_X \\ \bar{Y} & & \mu_Y \end{pmatrix} \xrightarrow{d} \mathcal{N} \left(\begin{matrix} 0 \\ 0 \end{matrix}, \Sigma \right)$
- Then, by the Delta Method:

$$\sqrt{n} \left(\frac{\bar{X}}{\bar{Y}} - \frac{\mu_X}{\mu_Y} \right) \xrightarrow{d} \mathcal{N} \left(0, \begin{pmatrix} 1/\mu_Y & -\mu_X/\mu_Y^2 \end{pmatrix} \Sigma \begin{pmatrix} 1/\mu_Y \\ -\mu_X/\mu_Y^2 \end{pmatrix} \right)$$

Important Distributions for Inference

Chi-Squared Distribution

- Let $Z_1, Z_2, \ldots, Z_n \sim \mathcal{N}(0,1)$ be i.i.d. RVs
- Then, $X = \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$ (chi-squared with n degrees of freedom)

Student's t-Distribution

- Let $Z \sim \mathcal{N}(0,1)$ and $X \sim \chi^2(n)$ be independent RVs
- Then, $X = \frac{Z}{\sqrt{X/n}} \sim t(n)$ (Student's t with n degrees of freedom)
- As $n \to \infty$, $t(n) \xrightarrow{d} \mathcal{N}(0,1)$

F-Distribution

- Let $X \sim \chi^2(n)$ and $Y \sim \chi^2(m)$ be independent RVs
- Then, $X = \frac{X/n}{Y/m} \sim F(n,m)$ (F-distribution with n, m degrees of freedom)

Example: Chi-Squared Distribution

- $X_n \xrightarrow{d} X \sim \mathcal{N}(0,1) \implies X_n^2 \xrightarrow{d} X^2 \sim \chi^2(1)$
- $\mathbf{X}_n \xrightarrow{d} \mathcal{N}(0,\Sigma) \implies \mathbf{X}_n^T\mathbf{X}_n \xrightarrow{d} \chi^2(k)$ where k is the number of elements in $\mathbf{X}_n$
- $\sqrt{n}\left(\bar{X}-\mu\right) \xrightarrow{d} \mathcal{N}(0,\sigma^2) \implies \left(\frac{\sqrt{n}\left(\bar{X}-\mu\right)}{\sigma}\right)^2 \xrightarrow{d} \chi^2(1)$
- $\sqrt{n}\left(\bar{\mathbf{X}}-\mu\right) \xrightarrow{d} \mathcal{N}(0,\Sigma) \implies \left(\sqrt{n}\left(\bar{\mathbf{X}}-\mu\right)\right)^T \Sigma^{-1} \left(\sqrt{n}\left(\bar{\mathbf{X}}-\mu\right)\right) \xrightarrow{d} \chi^2(k)$