C O U R S E M A T E R I A L

1.1 Sums and Products

Definitions: \[ \sum_{i=1}^{N} X_i = X_1 + X_2 + \ldots + X_N \]
\[ \prod_{i=1}^{N} X_i = X_1 \cdot X_2 \cdot \ldots \cdot X_N \]
\[ \sum_{i=1}^{N} \sum_{j=1}^{N} Z_{ij} = \sum_{i=1}^{N} (Z_{i1} + Z_{i2} + \ldots + Z_{iN}) \]
\[ = (Z_{11} + Z_{12} + \ldots + Z_{1N}) + \ldots + (Z_{N1} + Z_{N2} + \ldots + Z_{NN}) \]

Rule 1: \[ \sum_{i=1}^{N} k = N \cdot k \]
Rule 2: \[ \sum_{i=1}^{N} k \cdot X_i = k \sum_{i=1}^{N} X_i \]
Rule 3: \[ \sum_{i=1}^{N} (X_i + Y_i) = \sum_{i=1}^{N} X_i + \sum_{i=1}^{N} Y_i \]
Rule 4: \[ \sum_{i=1}^{N} (X_i - \bar{X}) = 0 \text{ mit } \bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \]
Rule 5: \[ \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y}) = \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X}) \cdot Y_i \]
\[ = \frac{1}{N} \sum_{i=1}^{N} X_i \cdot (Y_i - \bar{Y}) = \frac{1}{N} \sum_{i=1}^{N} X_i \cdot Y_i - \bar{X} \cdot \bar{Y} \]
Rule 6: \[ \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})^2 = \frac{1}{N} \sum_{i=1}^{N} X_i^2 - \bar{X}^2 \]
Rule 7: \[ \prod_{i=1}^{N} (X_i \cdot Y_i) = (\prod_{i=1}^{N} X_i) \cdot (\prod_{i=1}^{N} Y_i) \]
Rule 8: \[ \prod_{i=1}^{N} X_i^k = (\prod_{i=1}^{N} X_i)^k \]
Rule 9: \[ \sum_{i=1}^{N} \sum_{j=1}^{N} X_i \cdot Y_j = \sum_{j=1}^{N} \sum_{i=1}^{N} X_i \cdot Y_j \]
Rule 10: \[ \sum_{i=1}^{N} \sum_{j=1}^{N} X_i \cdot Y_j = (\sum_{i=1}^{N} X_i) \cdot (\sum_{j=1}^{N} Y_j) \]
Rule 11: \[ \sum_{i=1}^{N} \sum_{j=1}^{N} (Z_{ij} + A_{ij}) = \sum_{i=1}^{N} \sum_{j=1}^{N} Z_{ij} + \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} \]
1.2 Statistics and Probability Theory

Reference: WO Appendix B-C, Greene App. B-D

Random Variable (RV) $x$ taking values $x_i$

Probability distribution: $f(x_i) = \text{Prob}(x = x_i)$ for discrete RV

i) $0 \leq \text{Prob}(x = x_i) \leq 1$

ii) $\sum_{x_i} f(x_i) = 1$

Continuous RV: Density $f(x_i) \geq 0$

i) $\text{Prob}(a \leq x \leq b) = \int_a^b f(t)dt$

ii) $\int_{-\infty}^{\infty} f(t)dt = 1$

Cumulative distribution function CDF

$\text{Prob}(x \leq x_i) = F(x_i) = \begin{cases} 
\sum_{t \leq x_i} f(t) & : \text{discrete} \\
\int_{-\infty}^{x_i} f(t)dt & : \text{continuous}
\end{cases}$

For continuous case: $f(x_i) = \frac{dF(x_i)}{dx_i}$

Expected value (Mean):

$\mu \equiv E_x = \begin{cases} 
\sum_{x_i} x_i f(x_i) & : \text{discrete} \\
\int_{-\infty}^{\infty} t f(t)dt & : \text{continuous}
\end{cases}$

Variance:

$\sigma^2 \equiv \text{Var}(x) = E[(x - \mu)^2]$

$\sigma^2 = \begin{cases} 
\sum_{x_i} (x_i - \mu)^2 f(x_i) & : \text{discrete} \\
\int_{-\infty}^{\infty} (t - \mu)^2 f(t)dt & : \text{continuous}
\end{cases}$
Standard deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\text{Var}(x)}$$

Chebychev’s Inequality

$$\text{Prob}(|x - \mu| \geq \delta) \leq \frac{\sigma^2}{\delta^2}$$

$$\mathbb{E} g(x) = \left\{ \begin{array}{ll}
\sum x_i g(x_i) f(x_i) & : \text{discrete} \\
\int_{-\infty}^{\infty} g(t) f(t) dt & : \text{continuous}
\end{array} \right.$$  

In general: $$\mathbb{E} g(x) \neq g(\mathbb{E}(x))$$

Jensen’s inequality:

$$\mathbb{E} g(x) \leq g(\mathbb{E}(x)) \text{ for } g''(x) < 0$$  \text{concave}  

$$\mathbb{E} g(x) \geq g(\mathbb{E}(x)) \text{ for } g''(x) > 0$$  \text{convex}  

E.g. $$\mathbb{E} \log(x) \leq \log(\mathbb{E}(x))$$

Normal distribution

$$x \sim \mathcal{N}(\mu, \sigma^2) \text{ with density } f(x_i) = \frac{1}{\sqrt{2\pi \sigma}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

$$\mathbb{E} x = \mu \text{ and } \text{Var}(x) = \sigma^2$$
Standard Normal $z \sim N(0,1)$

Define density: $\phi(z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}}$

$F(z_i) = \Phi(z_i) = \int_{-\infty}^{z_i} \phi(t)dt = \int_{-\infty}^{z_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

$F_x(x_i) = \text{Prob}(x \leq x_i) = \text{Prob}\left(\frac{x - \mu}{\sigma} \leq \frac{x_i - \mu}{\sigma}\right)$

$= \text{Prob}\left(z \leq \frac{x_i - \mu}{\sigma}\right) = \Phi\left(\frac{x_i - \mu}{\sigma}\right)$

Skewness: $S \equiv E[(x-\mu)^3] = 0$ for normal distribution

Kurtosis: $E[(x-\mu)^4] = 3\sigma^4$ for normal distribution

Excess Kurtosis (relative to normal):

$$\frac{E[(x-\mu)^4]}{\sigma^4} - 3 = 0$$ for normal distribution

Chi-squared– ($\chi^2$), t– and F–distribution:

$\chi^2$–distribution: $z_1, \ldots, z_n$ independent $N(0,1)$

$$y = \sum_{j=1}^{n} z_j^2 \sim \chi^2_n$$–distributed with $n$ degrees of freedom
F- Distribution:

- \( y_1 \sim \chi^2_{n_1} \), \( y_2 \sim \chi^2_{n_2} \)
- \( y_1 \) and \( y_2 \) independent

\[
F(n_1, n_2) = \frac{y_1/n_1}{y_2/n_2} \sim F\text{-distributed with } n_1 \text{ degrees of freedom in numerator and } n_2 \text{ degrees of freedom in denominator}
\]

Stylized shape of probability density function of \( \chi^2_n \) or \( F(n_1, n_2) \)

t-distribution:

\[
t = \frac{z}{\sqrt{\frac{y}{n}}} \sim t_n \text{ distributed (t-distribution with n degrees of freedom)}
\]

\( z \sim N(0, 1) \), \( y \sim \chi^2_n \), and \( y, z \) independent
\( t_n \sim f_n(z_i) \rightarrow \phi(z_i) \text{ for } n \rightarrow \infty \)

**Note:** \( t^2 \sim F(1, n) \)

**Joint distribution:** \( x, y \) RV

\[
\text{Prob}(a \leq x \leq b, c \leq y \leq d) = \begin{cases} 
\sum_{a \leq x_i \leq b} \sum_{c \leq y_j \leq d} f(x_i, y_j) & : \text{discrete} \\
\int_a^b \int_c^d f(t, s) \, ds \, dt & : \text{continuous}
\end{cases}
\]

Probability density function: \( f(t, s) \geq 0 \)

\[
\sum_{x_i} \sum_{y_j} f(x_i, y_j) = 1 \quad \text{discrete}
\]

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t, s) \, ds \, dt = 1 \quad \text{continuous}
\]

Distribution function:

\[
F(x_i, y_j) = \text{Prob}(x \leq x_i, y \leq y_j) = \begin{cases} 
\sum_{x \leq x_i} \sum_{y \leq y_j} f(x_i, y_i) & : \text{discrete} \\
\int_{-\infty}^{x_i} \int_{-\infty}^{y_j} f(t, s) \, ds \, dt & : \text{continuous}
\end{cases}
\]
Expected value of function of \((x, y)\):

\[
E \ g(x, y) = \begin{cases} 
\sum \sum g(x_i, y_j)f(x_i, y_j) & : \text{discrete} \\
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(t, s)f(t, s) \, ds \, dt & : \text{continuous}
\end{cases}
\]

Covariance between \(x\) and \(y\):

\[
\sigma_{xy} \equiv Cov(x, y) = E[(x - Ex)(y - Ey)] = E xy - (Ex)(Ey)
\]

\(x, y\) independent:

\[
f(x_i, y_i) = f(x_i)f(y_i) \quad \Rightarrow \quad Cov(x, y) = 0
\]

Correlation:

\[
r_{xy} = \frac{Cov(x, y)}{\sqrt{Var(x) \cdot Var(y)}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}
\]

Rules:

\(a, b, c, d = \text{constants}\)

\[
E(ax + by + c) = a \ Ex + b \ Ey + c
\]

\[
Var(ax + by + c) = a^2 \ Var(x) + b^2 \ Var(y) + 2ab \ Cov(x, y)
\]

\[
Cov(ax + by, cx + dy) = ac \ Var(x) + bd \ Var(y) + (ad + bc) \ Cov(x, y)
\]

Conditional distribution:

\[
f(y = y_j|x = x_i) \equiv f(y_j|x_i) = \frac{f(x_i, y_j)}{f(x_i)}
\]

Conditional expectation:

\[
E(y|x = x_i) = \int_{-\infty}^{\infty} sf(y = s|x_i) \, ds
\]
Conditional variance:

\[ \text{Var}(y|x = x_i) = E[(y - E(y|x = x_i))^2|x = x_i] = \int_{-\infty}^{\infty} (s - E(y|x = x_i))^2 f(s|x_i) ds \]

The Role of Conditional Expectations in Econometrics

• \( y \) explained/dependent/response variable

• \( x = (x_1, \ldots, x_k) \) explanatory / independent variables, regressors, control variables, covariates

Structural conditional expectation (CE): \( E(y|w, c) \)

Based on random sample of \((y, w, c)\) we can estimate the effect of \( w \) on \( y \) holding \( c \) constant.

Complications arise when there is no random sample of \((y, w, c)\)

→ measurement error

→ simultaneous determination of \( y, w, c \)

→ some variables we would like to control for (elements of \( c \)) cannot be observed

\( \Rightarrow \) CE of interest involves data for which the econometrician cannot collect data or requires an experiment that cannot be carried out.

Identification assumptions:

→ Can recover structural CE of interest
Definition: $y$ (random variable) explained variable

$x \equiv (x_1, x_2, ..., x_k)$ \hspace{1em} (1 \times k)\text{-vector of explanatory variables}

$E(|y|) < \infty$

then function $\mu : \mathbb{R}^k \rightarrow \mathbb{R}$

$(CE) \quad E(y|x_1, x_2, ..., x_k) = \mu(x_1, x_2, ..., x_k) \text{ or } E(y|x) = \mu(x)$

Distinguish

$E(y|x)$ \hspace{3em} \text{random variable because } x \text{ is a random variable}

from

$E(y|x = x_0)$ \hspace{3em} \text{conditional expectation when } x \text{ takes specific value } x_0$

$\rightarrow$ Distinction most of the time not important

$\rightarrow$ Use $E(y|x)$ as short hand notation

Parametric model for $E(y|x)$ where $\mu(x)$ depends on a finite set of unknown parameters

Examples:

(i) $E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

(ii) $E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + \beta_4 x_1 x_2$

(iii) $E(y|x_1, x_2) = \exp[\beta_0 + \beta_1 \log(x_1) + \beta_2 x_2]$ \hspace{0.5em} with \hspace{0.5em} $y \geq 0, \hspace{0.5em} x_1 > 0$

(i) is linear in parameters and explanatory variables

(ii) is linear in parameters and nonlinear in explanatory variables

(iii) is nonlinear in both
Partial Effect:

- Continuous $x_i$, and differentiable $\mu$
  \[
  \Delta E(y|x) = \frac{\partial \mu}{\partial x_j} \Delta x_j \quad \text{holding } x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k \text{ fixed}
  \]
  \[
  \equiv \text{ ceteris paribus effect for properly specified population model}
  \]

- Discrete $x_j : x_{j,0} \rightarrow x_{j,1}$
  \[
  \Delta E(y|x) = E(y|x_1, \ldots, x_{j-1}, x_{j,1}, x_{j+1}, \ldots, x_k) - E(y|x_1, \ldots, x_{j-1}, x_{j,0}, x_{j+1}, \ldots, x_k)
  \]

Examples:

ad i) \[
\frac{\partial E(y|x)}{\partial x_1} = \beta_1 = \text{ constant}
\]

ad ii) \[
\frac{\partial E(y|x)}{\partial x_1} = \beta_1 + \beta_4 x_2 \quad \text{i.e. partial effect of } x_1 \text{ varies with } x_2
\]

ad iii) \[
\frac{\partial E(y|x)}{\partial x_1} = \exp[\beta_0 + \beta_1 \log(x_1) + \beta_2 x_2] \frac{\beta_1}{x_1} \rightarrow \text{ highly nonlinear}
\]

(Partial) Elasticity (only continuous case)

\[
\frac{\partial E(y|x)}{\partial x_j} \cdot \frac{x_j}{E(y|x)} = \frac{\partial \log[E(y|x)]}{\partial \log[x_j]}
\]

(Partial) Semielasticity:

\[
\frac{\partial E(y|x)}{\partial x_j} \cdot \frac{1}{E(y|x)} = \frac{\partial \log[E(y|x)]}{\partial x_j}
\]
Error form of models of conditional expectations

We can always write

\[ y = E(y|x) + u \quad \text{where} \quad u = y - E(y|x) \]

and it follows by definition

\[ E(u|x) = 0 \]

Implications:

1. \( E(u) = 0 \)

2. \( u \) is uncorrelated with any function of \( x_1, \ldots, x_k \)

Implication 1. follows from the law of iterated expectations

\[ \text{LIE : } E(y|x) = E[E(y|w)|x] \quad \text{if} \quad x = f(w) \]

i.e. \{Information set incorporated in \( x \} \subseteq \{Information set incorporated in \( w \}

i) \( E(y|x) = E[E(y|w)|x] \)

\[ \rightarrow \text{integrating out } w \text{ wrt } x: \int y f(y|x)dy = \int [\int y f(y|w, x)dy] f(w|x)dw \]

ii) \( E(y|x) = E[E(y|x)|w] \)

Knowing \( w \) implies knowing \( x \)

\[ \rightarrow \text{Routinely used in the course} \]

'The smaller information set always dominates'
Most important special case: \( w = (x, z) \)

\[
\frac{E(y|x)}{\mu_1(x)} = \frac{E[E(y|x, z)|x]}{\mu_2(x, z)}
\]

\[
\hat{\mu}_1 \text{ observed} = E[\mu_2(x, z)|x]
\]

Identification problem: Can we link the estimable \( \mu_1(x) \) to the structural \( \mu_2(x, z) \) which is the causal relationship of interest?

Therefore

\[ E(u) = E_x[E(u|x)] = E_x0 = 0 \]

and

\[ E(u|f(x)) = E[E(u|x)|f(x)] = E[0|f(x)] = 0 \]

which gives implication 2

Example:

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \]

with

\[ E(u|x_1, x_2) = 0 \]

implies: \( E(u) = 0, Cov(x_1, u) = 0, Cov(x_2, u) = 0 \) and \( u \) is also uncorrelated with \( x_1^2, x_2^2, x_1x_2, \exp(x_1) \) etc.

i.e. the functional form of \( E(y|x) \) is properly specified.

We have \( \beta_2 = \frac{\partial E(y|x_1, x_2)}{\partial x_2} \) because \( E(u|x_1, x_2) = 0 \), i.e. \( u \) is uncorrelated with any function of \( x_2 \). Thus \( \beta_2 \) describes the mean impact of \( x_2 \) on \( y \).

\[ E(u|x_1, x_2) = 0 \] sometimes called mean independence

We have:

Independence \( \Rightarrow \) Mean Independence \( \Rightarrow \) Uncorrelatedness

\( \Leftrightarrow \) \( \Leftrightarrow \)
Different nested sets of conditioning variables

\[ \begin{array}{cc}
\hat{x}, \hat{z} & \text{versus} & \hat{x} \\
\text{more information} & \text{less information}
\end{array} \]

\[ \begin{align*}
\mu_1(x) & = E(y|x) \\
\mu_2(x, z) & = E(y|x, z)
\end{align*} \]

By LIE, we have ('integrating \( z \) out')

\[ \begin{align*}
\mu_1(x) = E(y|x) & = E[E(y|x, z)|x] \\
& = E[\mu_2(x, z)|x]
\end{align*} \]

\[ \rightarrow \] allows to study effects of omitted regressors/unobserved components \( z \) on the relationship between \( y \) and \( x \).

Example: Wage Equation

\[ E(wage|educ, exper) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 educ \cdot exper \]

\[ = E(wage|educ, exper, exper^2, educ \cdot exper) \]

by LIE, i.e. it is redundant to condition on \( exper^2 \) and \( educ \cdot exper \).
Conditional Variance

The conditional variance of \( y \) given \( x \) is defined as

\[
\text{Var}(y|x) \equiv \sigma^2(x) \equiv E[(y - E(y|x))^2|x]
\]

\[
= E(y^2|x) - [E(y|x)]^2
\]

Note: \( \sigma^2(x) \) is a random variable when \( x \) is viewed as a random vector.

Properties:

\[
\text{Var}(a(x)y + b(x)|x) = [a(x)]^2\text{Var}(y|x)
\]

Decomposition of variance (corresponds to LIE)

\[
\text{Var}(y) = E[\text{Var}(y|x)] + \text{Var}(E(y|x))
\]

where \( \mu(x) = E(y|x) \).

Extension (further conditioning variable \( z \))

\[
\text{Var}(y|x) = E[\text{Var}(y|x, z)|x] + \text{Var}[E(y|x, z)|x]
\]

Consequently:

\[
E[\text{Var}(y|x)] \geq E[\text{Var}(y|x, z)]
\]

→ further conditioning variables \( z \) reduce the average conditional variances.
Probability Limit and Consistency of an Estimator

Motivation: For many econometric problems, the analytical properties of the estimator can only be determined asymptotically.

Definition 1: The probability limit $\theta$ of a sequence of random variables $\hat{\theta}_N$ results as the limit for $N$ going to infinity such that the probability that the absolute difference between $\hat{\theta}_N$ and $\theta$ is less than some small positive $\varepsilon$ goes to one. Mathematically this is expressed by

$$
\lim_{N \to \infty} P\{|\hat{\theta}_N - \theta| < \varepsilon\} = 1 \quad \text{for every} \quad \varepsilon > 0
$$

and abbreviated by $plim_{N \to \infty} \hat{\theta}_N = \theta$ (or $\hat{\theta}_N \xrightarrow{P} \theta$).

Definition 2: An estimator $\hat{\theta}_N$ for the true parameter value $\theta$ is (weakly) consistent, if

$$
plim_{N \to \infty} \hat{\theta}_N = \theta.
$$

Remarks: 1. The sample mean $\bar{Y}_N$ of a sequence of random variables $Y_i$ with expected value $E(Y_i) = \mu_Y$ is under very general conditions a consistent estimator of $\mu_Y$, d.h. $plim \bar{Y}_N = \mu_Y$.

2. For two sequences of random variables $\hat{\theta}_{1,N}$ and $\hat{\theta}_{2,N}$ it follows:

$$
plim (\hat{\theta}_{1,N} + \hat{\theta}_{2,N}) = plim \hat{\theta}_{1,N} + plim \hat{\theta}_{2,N},
$$

$$
plim (\hat{\theta}_{1,N} \cdot \hat{\theta}_{2,N}) = plim \hat{\theta}_{1,N} \cdot plim \hat{\theta}_{2,N},
$$

$$
plim \left( \frac{\hat{\theta}_{1,N}}{\hat{\theta}_{2,N}} \right) = \frac{plim \hat{\theta}_{1,N}}{plim \hat{\theta}_{2,N}}
$$

Slutzky’s Theorem: $plim g (\hat{\theta}_N) = g \left( plim \hat{\theta}_N \right)$ at continuity points of $g(.)$
Convergence and Asymptotic Orders of Magnitude

Motivation: For many semiparametric problems it is important to determine the speed of convergence, i.e. the asymptotic order of magnitude.

Definition 1 (Fixed Sequences): The sequence \( \{X_N\} \) of real numbers is said to be at most of order \( N^k \) and is denoted by

\[
X_N = O(N^k) \quad \text{if} \quad \lim_{N \to \infty} \frac{X_N}{N^k} = c
\]

for some constant \( c \).

Definition 2 (Fixed Sequences): The sequence \( \{X_N\} \) of real numbers is said to be of smaller order than \( N^k \) and is denoted by

\[
X_N = o(N^k) \quad \text{if} \quad \lim_{N \to \infty} \frac{X_N}{N^k} = 0
\]

Definition 3 (Stochastic Sequences): The sequence of random variables \( \{X_N\} \) is said to be at most of order \( N^k \) and is denoted by

\[
X_N = O_p(N^k)
\]

if for every \( \varepsilon > 0 \) there exist numbers \( C \) and \( \tilde{N} \) such that

\[
P \left\{ \left| \frac{X_N}{N^k} \right| > C \right\} < \varepsilon \quad \text{for all} \quad N > \tilde{N}.
\]

Definition 4 (Stochastic Sequences): The sequence of random variables \( \{X_N\} \) is said to be of smaller order than \( N^k \) and is denoted by

\[
X_N = o_p(N^k) \quad \text{if} \quad \plim_{N \to \infty} \frac{X_N}{N^k} = 0
\]

16
Chebychev’s Law of Large Numbers: Let the random variables \( \{X_i\} \) be uncorrelated with \( EX_i = \mu_i \) and \( Var(X_i) = \sigma_i^2 < \infty \) in a sample of size \( N \) (\( i = 1, \ldots, N \)). Then

\[
\bar{X}_N - \bar{\mu}_N \xrightarrow{P} 0
\]

if \( \bar{\sigma}^2 \to 0 \), as \( N \) goes to infinity where \( \bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i \) denotes the sample mean, \( \bar{\mu}_N = \frac{1}{N} \sum_{i=1}^{N} \mu_i \) and \( \bar{\sigma}^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sigma_i^2 = \frac{1}{N} \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \right) \).

Alternative Representation:
Under the above assumptions it follows that \( (\bar{X}_N - \bar{\mu}_N) = o_p(1) \)

Special Case: If \( \mu_i = \mu \) then \( \text{plim} \bar{X}_N = \mu \).

Lindberg–Levy’s Central Limit Theorem: Let \( \{X_i\} \) be a sequence of i.i.d. random variables such that \( EX_i = \mu \) and \( Var(X_i) = \sigma^2 < \infty \) in a sample of size \( N \) (\( i = 1, \ldots, N \)). Then

\[
\sqrt{N} \left( \frac{\bar{X}_N - \mu}{\sigma} \right) \xrightarrow{d} \mathcal{N}(0, 1) \quad (\text{i.e. } \bar{X}_N \text{ is } \sqrt{N} - \text{consistent}).
\]

Implication:
Under the above assumptions it follows that \( (\bar{X}_N - \mu) = O_p(N^{-1/2}) \).

Liapounov’s Central Limit Theorem: Let \( \{X_{N,i}\} \) be a sequence of independently distributed random variables with \( EX_{N,i} = \mu_{N,i} \) and \( Var(X_{N,i}) = \sigma_{N,i}^2 < \infty \) in a sample of size \( N \) (\( i = 1, \ldots, N \)). Let

\[
E|X_{N,i}|^{2+\delta} < \infty \text{ for some } \delta > 0. \text{ If } \lim_{N \to \infty} \sum_{i=1}^{N} \frac{E|X_{N,i}-\mu_{N,i}|^{2+\delta}}{\tilde{\sigma}_{N,i}^{2+\delta}} = 0, \text{ then } \frac{\sum_{i=1}^{N}(X_{N,i}-\mu_{N,i})}{\tilde{\sigma}_N} \xrightarrow{d} \mathcal{N}(0, 1) \text{ for } \tilde{\sigma}_N^2 = \sum_{i=1}^{N} \sigma_{N,i}^2.
\]

Implication:
Under the above assumptions it follows that \( \frac{\sum_{i=1}^{N}(X_{N,i}-\mu_{N,i})}{\tilde{\sigma}_N} = O_p(1) \).
1.3 Matrix Algebra

Reference: WO Appendix D

Matrix: Rectangular array of numbers

\[ A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{pmatrix} \quad n \times k \text{ matrix} \]

Transpose:

\[ A' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{nk} \end{pmatrix} \quad k \times n \text{ matrix} \]

\[(A + B)' = A' + B'\]

Inner Product: \( a' = (a_1, \ldots, a_n) \) and \( b' = (b_1, \ldots, b_n) \)

\[ a'b = a_1b_1 + \ldots + a_nb_n = b'a \]

Matrix Multiplication

\[ C_{n \times m} = A_{n \times k} \cdot B_{k \times m} \quad \Rightarrow \quad c_{ik} = a_{i \cdot} \cdot b_{\cdot k} \]

ith row of \( A \) \quad kth column of \( B \)

Identity matrix for \( n \in \mathbb{N} \):

\[ I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \quad I_nA = A \]

\[(AB)C = A(BC)\]

\[ A(B + C) = AB + AC \]

\[(AB)' = B'A'\]
Example: $n$ data points for $1 \times k$ vector $x_i = (x_{1i}, \ldots, x_{ki})$ (WO convention)

$$X = \begin{pmatrix}
  x_{11} & \cdots & x_{k1} \\
  \vdots & \ddots & \vdots \\
  x_{1n} & \cdots & x_{kn}
\end{pmatrix} \quad \text{n rows} \triangleq \text{observations}$$

Matrix product

$$X'X = \begin{pmatrix}
  x_{11} & \cdots & x_{1n} \\
  \vdots & \ddots & \vdots \\
  x_{k1} & \cdots & x_{kn}
\end{pmatrix} \cdot \begin{pmatrix}
  x_{11} & \cdots & x_{k1} \\
  \vdots & \ddots & \vdots \\
  x_{1n} & \cdots & x_{kn}
\end{pmatrix} = \begin{pmatrix}
  \sum_{i=1}^n x_{1i}^2 & \cdots & \sum_{i=1}^n x_{1i}x_{ki} \\
  \vdots & \ddots & \vdots \\
  \sum_{i=1}^n x_{ki}x_{1i} & \cdots & \sum_{i=1}^n x_{ki}^2
\end{pmatrix}$$

$$= \sum_{i=1}^n \begin{pmatrix}
  x_{1i} \\
  \vdots \\
  x_{ki}
\end{pmatrix} (x_{1i}, \ldots, x_{ki}) = \sum_{i=1}^n x_i'x_i \quad \leftarrow \text{summation notation}$$

Let $j_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ be a $n \times 1$ vector of ones, then $j_nj_n' = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

$$\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix},$$

and $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ $n \times 1$ vector, then

$$\frac{1}{n}j_nj_n'x = \frac{1}{n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \sum x_i \\ \vdots \\ \sum x_i \end{pmatrix} = \begin{pmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix} = j_n\bar{x}$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ sample average.
Deviations from sample average

\[ x - j_n \bar{x} = \begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} = x - \frac{1}{n} j_n j_n' x = \begin{pmatrix} I_n \\ \frac{-1}{n} j_n j_n' \end{pmatrix} x = M^0 x \]

where \( M^0 = I - \frac{1}{n} j_n j_n' \) is the matrix generating deviations from the mean (example of a projection matrix)

with

\[ M^0 j_n = \left( I_n - \frac{1}{n} j_n j_n' \right) j_n = j_n - \frac{1}{n} j_n j_n' j_n = j_n - j_n = 0 \]

since \( \frac{1}{n} j_n j_n' j_n = \frac{1}{n} n = 1 \).

\( M^0 \) is an example of a so called idempotent matrix, i.e. a square matrix \( M \) with \( M^2 = MM = M \).

When \( M \) is symmetric, it follows that \( M'M = M \)

Verify:

\[ M^0 M^0 = \left( I - \frac{1}{n} j_n j_n' \right) \left( I - \frac{1}{n} j_n j_n' \right) = I - \frac{1}{n} j_n j_n' - \frac{1}{n} j_n j_n' + \frac{1}{n^2} j_n j_n' j_n j_n' j_n = I - \frac{1}{n} j_n j_n' = M^0 \]

Sum of squared deviations:

\[ \sum_{i=1}^{n} (x_i - \bar{x})^2 = (M^0 x)'(M^0 x) = x'M^0'M^0 x = x'M^0 x = \sum_{i=1}^{n} x_i (x_i - \bar{x}) \]
Product of deviations of $x_i$ and $y_i$;

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = (M^0x)'(M^0y) = x'M^0'M^0y$$

$$= x'M^0y$$

$$= \sum x_i(y_i - \bar{y})$$

$$= \sum (x_i - \bar{x})y_i$$

Empirical Variance-Covariance-Matrix of $x, y$

$$\text{Cov}[(x, y)] = \begin{pmatrix}
\frac{1}{n} \sum (x_i - \bar{x})^2 & \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}) \\
\frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}) & \frac{1}{n} \sum (y_i - \bar{y})^2
\end{pmatrix}$$

$$= \frac{1}{n} \begin{pmatrix}
x'M^0x & x'M^0y \\
y'M^0x & y'M^0y
\end{pmatrix}$$

$$= \frac{1}{n} \begin{pmatrix}
x'M^0 \\
y'M^0
\end{pmatrix} (M^0x \quad M^0y)$$

$$= \frac{1}{n} \begin{pmatrix}
x' \\
y'
\end{pmatrix} M^0 (x \quad y)$$
Rank of a matrix $A$

= maximum number of linearly independent columns
= dimension of vector space spanned by column vectors
= maximum number of linearly independent rows
= dimension of vector space spanned by row vectors

A: $n \times k$ matrix $\rightarrow \rank(A) \leq \min(n, k)$

Properties:

i) $\rank(AB) \leq \min(\rank(A), \rank(B))$

ii) $\rank(A) = \rank(A'A) = \rank(AA')$

- Square $k \times k$ matrix $A$ has full rank if $\rank(A) = k$.
- $n \times k$ matrix $A$ with $n \geq k$ has full column rank if $\rank(A) = k$.
- $n \times k$ matrix $A$ with $n \leq k$ has full row rank if $\rank(A) = n$.

Inverse of a square matrix:

Let $A$ be a $k \times k$ matrix

Inverse $A^{-1}$ defined by $AA^{-1} = I$ or equivalently $A^{-1}A = I$

$A^{-1}$ exists, i.e. $A$ is invertible (or nonsingular) $\iff A$ has full rank.
Example: Diagonal matrix

\[
A := \begin{pmatrix}
 a_1 & 0 & \cdots & 0 \\
 0 & a_2 & \cdots & \vdots \\
 \vdots & \ddots & \ddots & 0 \\
 0 & \cdots & 0 & a_k
\end{pmatrix} = \text{diag}(a_1, \ldots, a_k)
\]

\[
\Rightarrow A^{-1} = \begin{pmatrix}
 \frac{1}{a_1} & 0 & \cdots & 0 \\
 0 & \frac{1}{a_2} & \cdots & \vdots \\
 \vdots & \ddots & \ddots & 0 \\
 0 & \cdots & 0 & \frac{1}{a_k}
\end{pmatrix}
\]

Inverse $A^{-1}$ exists if all $a_j \neq 0$ for $j = 1, \ldots, k$.

Properties:

i) $(A^{-1})^{-1} = A$

ii) $(A^{-1})' = (A')^{-1}$

iii) If $A$ is symmetric, then $A^{-1}$ is symmetric

iv) $(AB)^{-1} = B^{-1}A^{-1}$

v) $A = \begin{pmatrix}
 A_{11} & 0 \\
 0 & A_{22}
\end{pmatrix} \iff A^{-1} = \begin{pmatrix}
 A_{11}^{-1} & 0 \\
 0 & A_{22}^{-1}
\end{pmatrix}$ block diagonal

vi) Nonsingular matrix $B \rightarrow \text{rank}(AB) = \text{rank}(A)$
Eigenvalues (Characteristic Roots) and Eigenvectors: 

Eigenvalues $\lambda$ (scalars) and nonzero eigenvectors $c$ are the solution of $Ac = \lambda c$ for square $k \times k$ matrix $A$.

$$Ac = \lambda c \iff (A - \lambda I_n)c = 0$$

We are looking for the nontrivial solutions $c \neq 0$ which can be found by solving the characteristic equation involving the determinant

$$\det(A - \lambda I_n) = |A - \lambda I_n| = 0$$

for $\lambda$ and then finding some $c \neq 0$ for which $Ac = \lambda c$ (note $c$ is not unique!)

Properties:

i) $A$ has full rank ($A^{-1}$ exists) is equivalent to all eigenvalues are nonzero ($\lambda \neq 0$)

ii) If $A^{-1}$ exists, then its eigenvalues are the inverses of the eigenvalues of $A$

iii) Diagonal matrix

$$A = \begin{pmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & a_k \end{pmatrix}$$

Eigenvalues $\lambda_1 = a_1, \ldots, \lambda_k = a_k$

$$\begin{pmatrix} 1 \\ 0 \\ \ddots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \ddots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \ddots \\ 1 \end{pmatrix}$$

iv) $\det(A) = |A| = \prod_{j=1}^{k} \lambda_j$
Definitions:

- $A$ is called positive definite, if all eigenvalues are strictly positive ($\lambda_j > 0$)
- $A$ is called positive semidefinite, if all eigenvalues are nonnegative ($\lambda_j \geq 0$)
- $A$ is called negative definite, if all eigenvalues are strictly negative ($\lambda_j < 0$)
- $A$ is called negative semidefinite, if all eigenvalues are nonpositive ($\lambda_j \leq 0$)

Quadratic Form: $x'Ax$

- $A$ positive definite $\iff x'Ax > 0$ for all $x \neq 0$
- $A$ positive semidefinite $\iff x'Ax \geq 0$ for all $x \neq 0$
- $A$ negative definite $\iff x'Ax < 0$ for all $x \neq 0$
- $A$ negative semidefinite $\iff x'Ax \leq 0$ for all $x \neq 0$
Example:
x, y random variables with variance-covariance matrix
\[ V = \begin{pmatrix} \text{Var}(x) & \text{Cov}(x, y) \\ \text{Cov}(x, y) & \text{Var}(y) \end{pmatrix} \]

- \( V \) is always positive semidefinite.
- If \( x \) and \( y \) are not perfectly correlated, then \( V \) is positive definite.
- If \( x, y \) are jointly normally distributed \( \begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N} \left[ \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, V \right] \)
  then quadratic form
  \[ \begin{pmatrix} x - \mu_x, y - \mu_y \end{pmatrix} V^{-1} \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix} \sim \chi_2 \]
  if \( V \) has full rank.
- \( V^{-1} \): multivariate standardization.
- Since \( V \) is positive definite also \( V^{-1} \) is positive definite. Therefore
  \[ \begin{pmatrix} x & y \end{pmatrix} V^{-1} \begin{pmatrix} x \\ y \end{pmatrix} > 0 \text{ unless } \begin{pmatrix} x \\ y \end{pmatrix} = 0. \]

Trace of a matrix:

Square \( k \times k \) matrix \( A \)

\[ \text{tr}(A) = \sum_{j=1}^{k} a_{jj} \quad \text{sum of diagonal elements} \]

Properties:

i) \( \text{tr}(cA) = c \cdot \text{tr}(A) \) for scalar \( c \)

ii) \( \text{tr}(A') = \text{tr}(A) \)
iii) $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$

iv) $\text{tr}(AB) = \text{tr}(BA)$

v) $\text{tr}(A) = \sum_{j=1}^{k} \lambda_j$ trace of matrix equals the sum of its eigenvalues

**Kronecker Product:**

For $n \times k$ matrix $A$, $l \times m$ matrix $B$

$$A \otimes B = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nk} \end{bmatrix} \otimes B$$

$$= \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1k}B \\ a_{21}B & a_{22}B & \cdots & a_{2k}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nk}B \end{bmatrix}$$

$$\text{nl} \times \text{km} \text{ matrix}$$

Properties:

i) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

ii) $(A \otimes B)' = A' \otimes B'$

iii) $\text{tr}(A \otimes B) = \text{tr}(A) \cdot \text{tr}(B)$

iv) $(A \otimes B)(C \otimes D) = AC \otimes BD$ if $AC$, $BD$ is possible
Calculus and Matrix Algebra:

First and second order Taylor series approximation

- $y$ scalar
- $x = (x_1, \ldots, x_n)'$ $n \times 1$ vector
- $y = f(x)$ twice differentiable

Gradient:

\[
\nabla_x y := \frac{\partial y}{\partial x} = \frac{\partial f(x)}{\partial x} = \left( \begin{array}{c} \frac{\partial y}{\partial x_1} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{array} \right) = \left( \begin{array}{c} f_1 \\ \vdots \\ f_n \end{array} \right) \quad \text{column vector as convention}
\]

Hessian:

\[
H = \frac{\partial^2 y}{\partial x \partial x'} = \begin{bmatrix}
\frac{\partial^2 y}{\partial x_1^2} & \frac{\partial^2 y}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 y}{\partial x_1 \partial x_n} \\
\frac{\partial^2 y}{\partial x_2 \partial x_1} & \frac{\partial^2 y}{\partial x_2^2} & \cdots & \frac{\partial^2 y}{\partial x_2 \partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 y}{\partial x_n \partial x_1} & \frac{\partial^2 y}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 y}{\partial x_n^2}
\end{bmatrix} = [f_{ij}]
\]

First order Taylor series approximation in $x = (x_10, \ldots, x_n0)$

\[
y = f(x) \approx f(x_0) + \sum_{i=1}^{n} f_i(x_0)(x_i - x_{i0}) = f(x_0) + \left( \frac{\partial y}{\partial x} \bigg|_{x_0} \right)'(x - x_0)
\]
Second order approximation

\[ y = f(x) \approx f(x_0) + \sum_{i=1}^{n} f_i(x_0)(x_i - x_{i0}) + \]

\[ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij}(x_0) \cdot (x_i - x_{i0}) \cdot (x_j - x_{j0}) \]

\[ = f(x_0) + \left( \left. \frac{\partial y}{\partial x} \right|_{x_0} \right)' (x - x_0) + \frac{1}{2} (x - x_0)' H(x_0) (x - x_0) \]

inner product

quadratic form

Differentiation of inner products and quadratic forms:

i) \[ y = a'x = \sum_{i=1}^{n} a_i x_i = x'a \]

\[ \frac{\partial y}{\partial x} = \frac{\partial a'x}{\partial x} = \begin{pmatrix} a_1' \\ \vdots \\ a_n' \end{pmatrix} = a \]

ii) \[ z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} = Ax = \begin{pmatrix} \sum_{i=1}^{k} a_{1i}x_{1i} \\ \vdots \\ \sum_{i=1}^{k} a_{ni}x_{ni} \end{pmatrix} \]

A \( n \times k \) matrix, \( x \) \( k \times 1 \) vector, \( z \) \( n \times 1 \) vector

\[ \frac{\partial z}{\partial x} = \left( \frac{\partial z_1}{\partial x}, \ldots, \frac{\partial z_n}{\partial x} \right) = A' \quad \text{columnwise gradients of } z_1, \ldots, z_n \]

iii) \[ y = x'Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{ij} \quad \text{quadratic form} \]

a) \[ \frac{\partial y}{\partial x} = (A + A')x \]

If \( A \) is symmetric (\( A = A' \)), then \( \frac{\partial y}{\partial x} = 2Ax \)

b) \[ \frac{\partial y}{\partial A} = xx' = \begin{pmatrix} x_1^2 & \cdots & x_1x_n \\ \vdots & \ddots & \vdots \\ x_1x_n & \cdots & x_n^2 \end{pmatrix} \]

outer product, \( n \times n \)

matrix
Expected values and variances:

Let

- $a$ be a $k \times 1$ vector of constants
- $A$ a $n \times k$ matrix of constants, and
- $x$ a $k \times 1$ vector of random variables

then

$$E a'x = a'(E x) = \sum_{i=1}^{k} a_i E x_i$$

$$E Ax = A(E x) = \begin{bmatrix} \sum_{i=1}^{k} a_{1i} E x_i \\ \cdots \\ \sum_{i=1}^{k} a_{1i} E x_i \end{bmatrix}$$

$$Var(a'x) = a' Var(x)a = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij} Cov(x_i, x_j) \quad \leftarrow \text{quadratic form}$$

$$Var(Ax) = A Var(x) A'$$