Advanced Econometrics

Prof. Bernd Fitzenberger, Ph.D.

Humboldt-University Berlin – Summer Semester 2017

Course Outline

0. Introductory Material

1. Review Linear Regression Model for Cross-Sectional Data

2. System Estimation, Linear Panel Data Models

3. Nonlinear Least Squares and Maximum Likelihood

4. Binary Response Models and Limited Dependent Variables

5. Linear Quantile Regression

0. Introductory Material

Section Outline

0.1. Matrix Algebra

Reference: Greene (2008) App. A

Matrix: Rectangular array of numbers

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

Transpose:

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

$$(A+B)' = A' + B'$$

Inner Product:

for $a' = (a_1, \ldots, a_n)$ and $b' = (b_1, \ldots, b_n)$

$$a'b = a_1 b_1 + \ldots + a_n b_n = b'a$$

Matrix Multiplication:

$$\underbrace{C}_{n\times m} = \underbrace{A}_{n\times k}\underbrace{B}_{k\times m} \qquad \Rightarrow \qquad c_{ik} = \underbrace{a_{i\cdot}}_{\nearrow}{}' \underbrace{b_{\cdot k}}_{\nwarrow}$$

ith row of A kth column of B

Identity matrix for $n \in \mathbb{N}$:

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \qquad I_n A = A$$

Rules for Matrix Multiplication:

$$
\begin{aligned}
(AB)C &= A(BC) \\
A(B+C) &= AB + AC \\
(AB)' &= B'A'
\end{aligned}
$$

Example: n data points for $1 \times k$ vector $x_i = (x_{1i}, \dots, x_{ki})$ (WO convention)

$$
X = \begin{pmatrix} x_{11} & \cdots & x_{k1} \\ \dots & & \dots \\ x_{1n} & \cdots & x_{kn} \end{pmatrix} \qquad \text{n rows} \hat{=} \text{observations}
$$

Matrix product:

$$
\begin{aligned}
X'X &= \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \cdots & & \cdots \\ x_{k1} & \cdots & x_{kn} \end{pmatrix} \cdot \begin{pmatrix} x_{11} & \cdots & x_{k1} \\ \cdots & & \cdots \\ x_{1n} & \cdots & x_{kn} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{n} {x_{1i}}^2 & \cdots & \sum_{i=1}^{n} x_{1i}x_{ki} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ki}x_{1i} & \cdots & \sum_{i=1}^{n} {x_{ki}}^2 \end{pmatrix} \\
&= \sum_{i=1}^{n} \begin{pmatrix} x_{1i} \\ \vdots \\ x_{ki} \end{pmatrix} (x_{1i}, \ldots, x_{ki}) = \sum_{i=1}^{n} x_i' x_i \qquad \leftarrow \text{summation notation}
\end{aligned}
$$

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

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

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

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

Deviations from sample average

$$x - j_n\bar{x} = \begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} = x - \frac{1}{n} j_n j_n' x = \Bigg(\underbrace{I_n}_{\text{identity matrix}} - \frac{1}{n} j_n j_n' \Bigg) x = M^0 x$$

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 $\frac{1}{n} j_n' j_n = \frac{1}{n} n = 1$.

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

When M is symmetric, it follows that $M'M = M$.

<u>Verify</u>:

$$\begin{aligned} M^0 M^0 &= \left(I - \frac{1}{n} j_n j_n'\right)\left(I - \frac{1}{n} j_n j_n'\right) \\ &= I - \frac{1}{n} j_n j_n' - \frac{1}{n} j_n j_n' + \frac{1}{n^2} j_n \underbrace{j_n' j_n}_{n} j_n' \\ &= I - \frac{1}{n} j_n j_n' = M^0 \end{aligned}$$

Sum of squared deviations:

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = (M^0 x)'(M^0 x) = x' M^{0\prime} M^0 x = x' M^0 x = \sum_{i=1}^{n} x_i (x_i - \bar{x})$$

Product of deviations of x_i and y_i:

$$\begin{aligned}
\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) &= (M^0 x)'(M^0 y) = x' M^{0\prime} M^0 y \\
&= x' M^0 y \\
&= \sum x_i (y_i - \bar{y}) \\
&= \sum (x_i - \bar{x}) y_i
\end{aligned}$$

Empirical Variance-Covariance-Matrix of x, y

$$
\begin{aligned}
\mathrm{Cov}\left[(x,y)\right] &= \begin{pmatrix} \frac{1}{n}\sum(x_i-\bar{x})^2 & \frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y}) \\ \frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y}) & \frac{1}{n}\sum(y_i-\bar{y})^2 \end{pmatrix} \\
&= \frac{1}{n}\begin{pmatrix} x'M^0x & x'M^0y \\ y'M^0x & y'M^0y \end{pmatrix} \\
&= \frac{1}{n}\begin{pmatrix} x'M^0 \\ y'M^0 \end{pmatrix}\begin{pmatrix} M^0x & M^0y \end{pmatrix} \\
&= \frac{1}{n}\begin{pmatrix} x' \\ y' \end{pmatrix} M^0 \begin{pmatrix} x & y \end{pmatrix}
\end{aligned}
$$

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 $\rightarrow$ $\text{rank}(A) \leq \min(n, k)$

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) $\Leftrightarrow$ A has full rank.

Example: Diagonal matrix

$$A := \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_k \end{pmatrix} = \mathrm{diag}(a_1, \ldots, a_k) \Rightarrow \quad A^{-1} = \begin{pmatrix} \frac{1}{a_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{a_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{a_k} \end{pmatrix}$$

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

vi) Nonsingular matrix $B \rightarrow rank(AB) = rank(A)$

Eigenvalues (Characteristic Roots) and Eigenvectors:

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

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

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 λ 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

Diagonal matrix

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

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

Eigenvectors $$\begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ \cdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 1 \end{pmatrix}$$

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

<u>Definition</u>:

- 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 ($\lambda_j \leq 0$)

Spectral decomposition of a symmetric matrix:

A $k \times k$ symmetric matrix A has k distinct orthogonal eigenvectors $c_1, c_2, \ldots, c_k$ and k not necessarily distinct, real eigenvalues $\lambda_1, \ldots, \lambda_k$.

We have $Ac_j = \lambda_j c_j$ which is summarized in $AC = C\Lambda$ where $C = [c_1 \cdots c_k]$ eigenvectors as columns

and $\Lambda = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_k \end{pmatrix}$ diagonal matrix with eigenvalues.

Orthogonality of eigenvectors: $c_i' c_j = 0$ for $i \neq j$ and normalization $c_i' c_i = 1$

$$CC' = C'C = I_n \quad \text{and} \quad C' = C^{-1}$$

This implies:

Diagonalization $\qquad C'AC = C'C\Lambda = \Lambda$

Spectral Decomposition $\qquad A = CC'ACC' = C\Lambda C' = \sum_{j=1}^{k} \lambda_j \, c_j c_j'$

The Generalized Inverse of a Matrix

- Case when A is not invertible because A is not a square matrix or A is not singular!

Definition: A generalized inverse of A is another matrix A^+ that satisfies

1. $AA^+A = A$
2. $A^+AA^+ = A^+$
3. A^+A is symmetric
4. AA^+ is symmetric

Note:

- A unique matrix that satisfies 1.–4. is called the Moore-Penrose inverse
- If A^{-1} exists, then $A^+ = A^{-1}$

Two cases: **Case A** (no square matrix $k < n$) and **Case B** (symmetric square matrix)

Case A: Let A be an $n \times k$ matrix with $k < n$ and $rank(A) = r \leq k$

1.) $r = k \Leftrightarrow A$ does have full column rank $\Leftrightarrow (A'A)^{-1}$ exists
Moore-Penrose inverse is

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

Verify 1.–4.:

1. $AA^{+}A = A(A'A)^{-1}A'A = A$
2. $A^{+}AA^{+} = (A'A)^{-1}A'AA^{+} = A^{+}$
3. $A^{+}A = (A'A)^{-1}A'A = I$ symmetric
4. $(A(A'A)^{-1}A')' = A''(A'A)^{-1}A' = A(A'A)^{-1}A'$ symmetric

2.) $rank(A) = r < k$
Use r nonzero characteristic roots of $A'A$ and associated eigenvectors in matrix C_1, then

$$A'A = C_1\Lambda_1^{-1}C_1' \quad \text{spectral decompose}$$

The Moore-Penrose inverse is

$$A^{+} = C_1\Lambda_1^{-1}C_1'A'$$

where $r \times r$ diagonal matrix $\Lambda_1 = diag(\lambda_1, \ldots, \lambda)$ of nonzero eigenvalues.

Case B: If A is symmetric $(n = k)$, then

$$A^+ = C_1 \Lambda_1^{-1} C_1'$$

where Λ_1 is a diagonal matrix containing the nonzero eigenvalues of A and C_1 the associated orthonormalized eigenvectors.

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 $\begin{pmatrix} x \\ y \end{pmatrix} \sim \text{N}\left[\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, V \right]$ then quadratic form $\begin{pmatrix} x & y \end{pmatrix} V^{-1} \begin{pmatrix} x \\ y \end{pmatrix} \sim \chi^2_2$-distributed, if V has full rank.
- V^{-1}: multivariate standardization.
- Since V is positive definite also V^{-1} is positive definite and therefore $\begin{pmatrix} x & y \end{pmatrix} V^{-1} \begin{pmatrix} x \\ y \end{pmatrix} > 0$ unless $\begin{pmatrix} x \\ y \end{pmatrix} = 0$.

Trace of a matrix:

Square $k \times k$ matrix A

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

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

$$\underbrace{A \otimes B}_{(nl)\times(km)} \text{ matrix} = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nk} \end{bmatrix} \otimes B$$

$$= \underbrace{\left.\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1k}B \\ a_{21}B & a_{22}B & \cdots & a_{2k}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nk}B \end{bmatrix}\right\} n \cdot l \quad \text{rows}}_{k \cdot m \quad \text{columns}}$$

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

Gradient:

$$\nabla_x y := \underbrace{\frac{\partial y}{\partial x}}_{n\times 1 \text{ vector}} = \frac{\partial f(x)}{\partial x} = \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{pmatrix} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$$

column vector as convention

Hessian:

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

First order Taylor series approximation in $x = (x_{10}, \ldots, x_{n0})$

$$y = f(x) \approx f(x_0) + \sum_{i=1}^{n} f_i(x_0)(x_i - x_{i0}) = f(x_0) + \left(\left. \frac{\partial y}{\partial x} \right|_{x_0} \right)' (x - x_0)$$

Second order approximation

$$\begin{aligned} y = f(x) &\approx f(x_0) + \sum_{i=1}^{n} f_i(x_0)(x_i - x_{i0}) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij}(x_0) \cdot (x_i - x_{i0}) \cdot (x_j - x_{j0}) \\ &= f(x_0) + \underbrace{\left(\left. \frac{\partial y}{\partial x} \right|_{x_0} \right)' (x - x_0)}_{\text{inner product}} + \frac{1}{2} \underbrace{(x - x_0)' H(x_0)(x - x_0)}_{\text{quadratic form}} \end{aligned}$$

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

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

iii) $y = x'Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{ij}$ 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) $\frac{\partial y}{\partial A} = xx' = \begin{pmatrix} x_1^2 & \cdots & x_1 x_n \\ \vdots & \ddots & \vdots \\ x_1 x_n & \cdots & x_n^2 \end{pmatrix}$ 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

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

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

$$Var(a'x) = a' Var(x) a = \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j Cov(x_i, x_j) \geq 0 \quad \leftarrow \text{ quadratic form}$$

$Var(x)$ must be positive semidefinite

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

0.2. Statistics and Probability Theory

Reference: WO 2+3, Greene App. B-D

Random Variable (RV) x taking values x_i

Probability distribution: $f(x_i) = Prob(x = x_i)$ for discrete RV

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

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

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

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

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

Cumulative Distribution Function CDF

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

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

Expected value (Mean):

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

Variance:

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

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

Standard deviation:

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

Chebychev‘s Inequality:

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

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

In general: $Eg(x) \neq g(E(x))$

Jensen‘s inequality:

$$Eg(x) \leq g(E(x)) \quad \text{for} \quad \underbrace{g''(x) < 0}_{concave}$$

$$Eg(x) \geq g(E(x)) \quad \text{for} \quad \underbrace{g''(x) > 0}_{convex}$$

E.g. $\quad E\log(x) \leq \log(E(x))$

Normal distribution

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

$$Ex = \mu \quad \text{and} \quad Var(x) = \sigma^2$$

Standard Normal $z \sim N(0,1)$

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

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

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

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

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

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

Excess Kurtosis (relative to normal):

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

Chi-squared– (χ^2) , t– and F–distributions

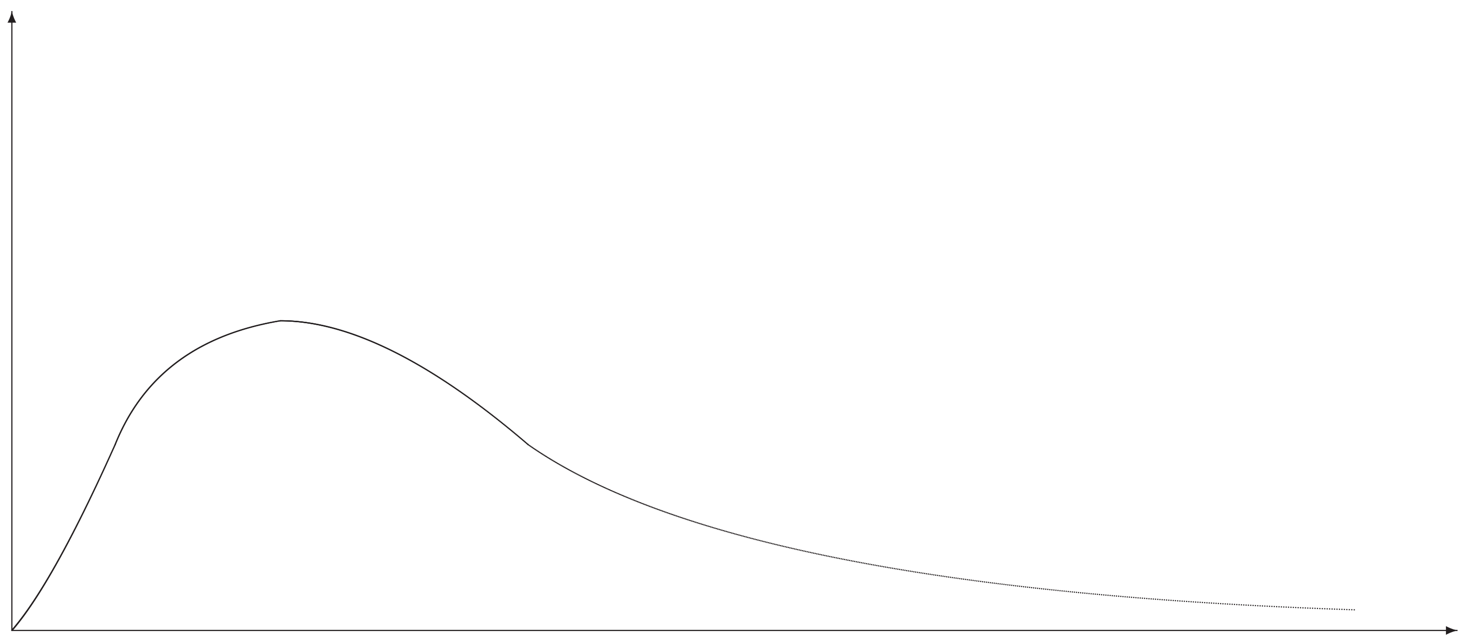
χ^2–distribution: $z_1, \ldots\ldots, z_n$ independent $N(0,1)$

$$y = \sum_{j=1}^{n} z_j^2 \quad \sim \quad \chi_n^2\text{–distributed with } n \text{ degrees of freedom}$$

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–distributed with n_1 degrees of freedom in numerator and n_2 degrees of freedom in denominator

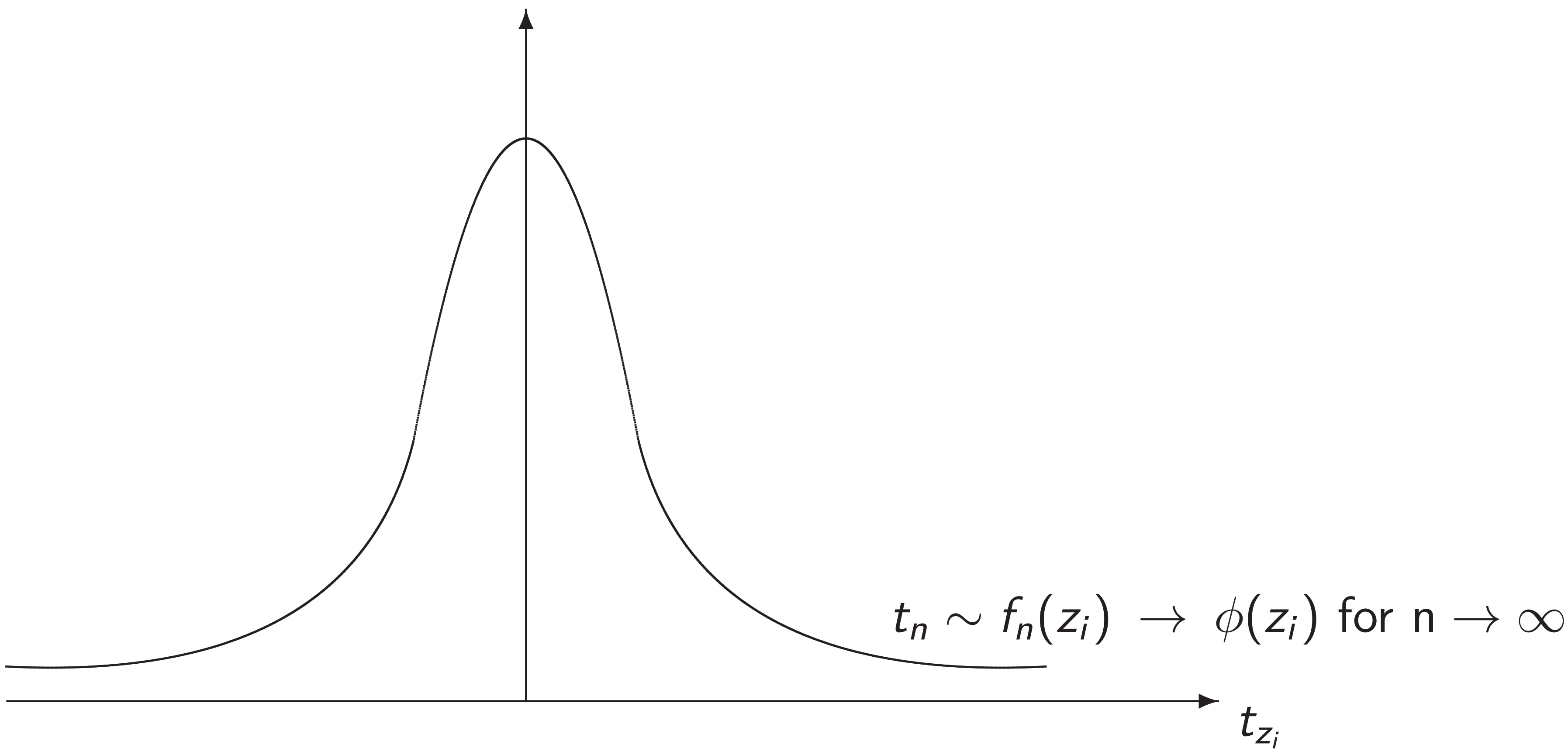


stylized shape of probability density function of χ^2_n or $F(n_1, n_2)$

t–distribution:

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

$z \sim N(0,1)$, $y \sim \chi^2_n$, and y, z independent



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

Joint distribution: x, y RV

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

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

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

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t,s) \, ds \, dt = 1 \qquad \text{continuous}$$

Distribution function:

$$F(x_i, y_j) = Prob(x \le x_i, y \le y_j) = \begin{cases} \sum_{x \le x_i} \sum_{y \le y_j} f(x_i, y_i) & : \text{discrete} \\ \int_{-\infty}^{x_i} \int_{-\infty}^{y_j} f(t,s) ds \, dt & : \text{continuous} \end{cases}$$

Expected value of function of (x, y):

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

Covariance between x and y:

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

x, y independent :

$$f(x_i, y_i) \quad = \quad f(x_i) f(y_i) \quad \overset{\Rightarrow}{\not\Leftarrow} \quad Cov(x,y) = 0$$

Correlation:

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

Rules:

$a, b, c, d =$ constants

$$\begin{aligned} E(ax + by + c) &= a\,Ex + b\,Ey + c \\ Var(ax + by + c) &= a^2 Var(x) + b^2 Var(y) + 2abCov(x, y) \\ Cov(ax + by, cx + dy) &= acVar(x) + bdVar(y) + (ad + bc)Cov(x, y) \end{aligned}$$

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:

$$E(y|x = x_i) = \int_{-\infty}^{\infty} s \underbrace{f(y = s|x_i)}_{\equiv f(s|x_i)} ds$$

Conditional variance:

$$\begin{aligned} Var(y|x = x_i) &= E[(y - E(y|x = x_i))^2 | x = x_i] \\ &= \int_{-\infty}^{\infty} (s - E(y|x = x_i))^2 \, f(s|x_i) ds \end{aligned}$$

0.3. Asymptotics

Motivation:

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

Probability Limit and Consistency of an Estimator

Definition 1:

The **probability limit** θ 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 θ is less than some small positive ε goes to one. Mathematically this is expressed by

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

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

Definition 2:

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

$$\operatorname*{plim}_{N\to\infty} \hat{\theta}_N = \theta \,.$$

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 μ_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:

$plim\,g\left(\hat{\theta}_N\right) \;=\; g\left(plim\,\hat{\theta}_N\right)$ 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

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

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

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

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

$$X_N = o_p(N^k) \quad \text{if} \quad \underset{N\to\infty}{plim} \frac{X_N}{N^k} = 0 \quad .$$

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

$$\bar{X}_N - \bar{\mu}_N \quad \stackrel{P}{\to} \quad 0$$

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 $\left(\bar{X}_N - \bar{\mu}_N\right) = o_p(1)$

Special Case: If $\quad \mu_i = \mu \quad$ then $\quad plim \bar{X}_N = \mu \quad$.

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

$$\sqrt{N}\,\frac{(\bar{X}_N - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1) \qquad \text{(i.e. } \bar{X}_N \text{ is } \sqrt{N} - \text{consistent).}$$

Implication:

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

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 $Var(X_{N,i}) = \sigma^2_{N,i} < \infty$ in a sample of size N ($i = 1, \ldots, N$).

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

Implication:

Under the above assumptions it follows that $\frac{\sum_{i=1}^{N}(X_{N,i}-\mu_{N,i})}{\tilde{\sigma}_N} = O_p(1)$