Review of Statistical Concepts

Hüseyin Taştan[1]

[1]Department of Economics
Ph.D. Program - Advanced Econometrics
Yildiz Technical University

September 24, 2013

Contents

- Linear Transformations of Random Variables
- Linear Transformations of Normally Distributed Random Vectors
- Elements of Asymptotic Analysis (convergence in probability, convergence in distribution, consistency, asymptotic distribution, asymptotic efficiency)
- Stochastic Order Notation

Linear Transformations

- Let $\boldsymbol{x}$ be an $n \times 1$ vector of random variables with

$$\mathsf{E}(\boldsymbol{x}) = \mathsf{E}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \mathsf{E}(x_1) \\ \mathsf{E}(x_2) \\ \vdots \\ \mathsf{E}(x_n) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \equiv \boldsymbol{\mu}$$

$$\begin{aligned} \mathsf{Var}(\boldsymbol{x}) &= \mathsf{E}[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top] \\ &= \boldsymbol{\Sigma} \quad (n \times n, \text{square, symmetric, positive definite}) \\ &= \mathsf{E}(\boldsymbol{x}\boldsymbol{x}^\top) - \boldsymbol{\mu}\boldsymbol{\mu}^\top \end{aligned}$$

Linear Transformations

Elements of covariance matrix

$$\boldsymbol{\Sigma} =$$

$$\begin{bmatrix} \mathsf{E}[(x_1-\mu_1)^2] & \mathsf{E}[(x_1-\mu_1)(x_2-\mu_2)] & \dots & \mathsf{E}[(x_1-\mu_1)(x_n-\mu_n)] \\ \mathsf{E}[(x_2-\mu_2)(x_1-\mu_1)] & \mathsf{E}[(x_2-\mu_2)^2] & \dots & \mathsf{E}[(x_2-\mu_2)(x_n-\mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{E}[(x_n-\mu_n)(x_1-\mu_1)] & \mathsf{E}[(x_n-\mu_n)(x_2-\mu_2)] & \dots & \mathsf{E}[(x_n-\mu_n)^2] \end{bmatrix}$$

$$= \begin{bmatrix} \mathsf{Var}(x_1) & \mathsf{Cov}(x_1,x_2) & \dots & \mathsf{Cov}(x_1,x_n) \\ \mathsf{Cov}(x_2,x_1) & \mathsf{Var}(x_2) & \dots & \mathsf{Cov}(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{Cov}(x_n,x_1) & \mathsf{Cov}(x_n,x_2) & \dots & \mathsf{Var}(x_n) \end{bmatrix}$$

Always square, symmetric, positive (semi) definite, nonsingular.

5

Linear Transformations

- Also let $\boldsymbol{a}$ be an n-vector of constants.
- Define the linear transformation $y = \boldsymbol{a}^\top \boldsymbol{x}$. Then,

$$\mathsf{E}(y) = \mathsf{E}(\boldsymbol{a}^\top \boldsymbol{x}) = \boldsymbol{a}^\top \mathsf{E}(\boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{\mu}$$

and

$$\begin{aligned}
\mathsf{Var}(y) &= \mathsf{Var}(\boldsymbol{a}^\top \boldsymbol{x}) = \mathsf{E}\left[\left(\boldsymbol{a}^\top \boldsymbol{x} - \mathsf{E}(\boldsymbol{a}^\top \boldsymbol{x})\right)\left(\boldsymbol{a}^\top \boldsymbol{x} - \mathsf{E}(\boldsymbol{a}^\top \boldsymbol{x})\right)^\top\right] \\
&= \mathsf{E}\left[\boldsymbol{a}^\top (\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{a}\right] \\
&= \boldsymbol{a}^\top \mathsf{E}\left[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top\right] \boldsymbol{a} \\
&= \boldsymbol{a}^\top \boldsymbol{\Sigma} \boldsymbol{a}
\end{aligned}$$

6

Linear Transformations

- Let $\boldsymbol{A}$ be an $n \times n$ matrix of constants.
- Also define the linear transformation $\boldsymbol{A}\boldsymbol{x}$. Then,

$$\mathsf{E}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{\mu}$$

and

$$\begin{aligned}
\mathsf{Var}(\boldsymbol{A}\boldsymbol{x}) &= \mathsf{E}\left[(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}\boldsymbol{\mu})(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}\boldsymbol{\mu})^\top\right] \\
&= \mathsf{E}\left[\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu}))^\top\right] \\
&= \mathsf{E}\left[\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{A}^\top\right] \\
&= \boldsymbol{A}\mathsf{E}\left[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top\right] \boldsymbol{A}^\top \\
&= \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top
\end{aligned}$$

7

Generating Normal Variates in MATLAB

- Generating $m \times 1$ vector of iid standard normal random variables

```
randn(m,1)
```

- Generating $m \times n$ matrix of iid standard normal random variables, note that covariance matrix is identity

```
randn(m,n)
```

- Generating $n \times 1$ vector from $X \sim N(\mu, \sigma^2)$

```
X = mu + sigma * randn(n,1)
```

8

Multivariate Normal Distribution

Let $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]^\top$ be a vector of random variables from a multivariate normal distribution with mean

$$E(\boldsymbol{X}) = E\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \boldsymbol{\mu}$$

and covariance matrix

$$\mathsf{var}(\boldsymbol{X}) = E\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^\top\right] = \boldsymbol{\Sigma}$$

Multivariate Normal Distribution

$\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ pdf:

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \left(\frac{1}{2\pi}\right)^{n/2} |\det \boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

Covariance matrix $\boldsymbol{\Sigma}$ is always positive definite, thus using Cholesky factorization:

$$\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$$

where $\boldsymbol{L}$ is a lower-triangular matrix. Using the transformation

$$\boldsymbol{Z} = L^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \Rightarrow \boldsymbol{X} = \boldsymbol{L}\boldsymbol{Z} + \boldsymbol{\mu}$$

will have a multivariate standard normal distribution:

$$f(\boldsymbol{z}) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

Multivariate Normal Distribution

$$\boldsymbol{X} = \boldsymbol{L}\boldsymbol{Z} + \boldsymbol{\mu}$$

For example let $\boldsymbol{X} = [X_1, X_2]^\top$ and

$$\boldsymbol{\mu} = \begin{bmatrix} 5 \\ 10 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 2 & -1.5 \\ -1.5 & 4 \end{bmatrix}$$

```
>> mu=[5;10];
>> sigma=[2 -1.5;-1.5 4];
>> L = chol(sigma)

L =

    1.4142   -1.0607
         0    1.6956

>> Z=randn(10000,2);
>> X=Z*L+repmat(mu',10000,1);  % generate 10000 observations
```


Multivariate Normal Distribution

```
>> mean(X)

ans =

    4.9879   10.0071

>> cov(X)

ans =

    1.9797   -1.4896
   -1.4896    3.9479

>> corr(X)

ans =

    1.0000   -0.5328
   -0.5328    1.0000

>> plotmatrix(X)
```


Multivariate Normal Distribution

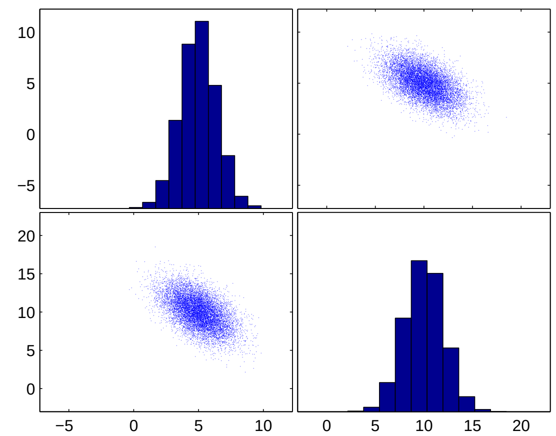

Figure : Two simulated normal variables

Multivariate Normal Distribution

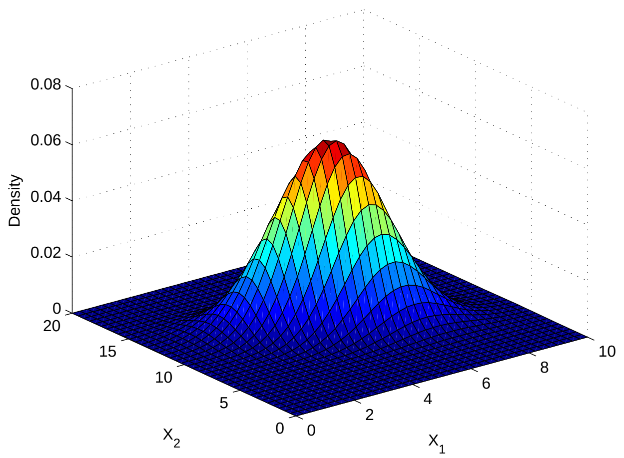

Figure : Bivariate Normal Probability Density Function

Lognormal Distribution

If $X \sim$ lognormal(μ, σ^2) then

$$\log(X) \sim \mathsf{N}(\mu, \sigma^2)$$

Let Z be a standard normal random variable. Then

$$X = e^{\mu + \sigma Z}$$

follows a lognormal distribution with mean μ and σ^2. Mean and variance are

$$\mathsf{E}(X) = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

$$\mathsf{Var}(X) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)$$

In MATLAB

```
>> X=exp(mu+sigma*Z);
```

Standard Distributions in MATLAB

Distribution	pdf	cdf	Random number generator
Standard Normal	`normpdf`	`normcdf`	`randn`
Standard Uniform	`unifpdf`	`unifcdf`	`rand`
Uniform(a, b)	`unifpdf`	`unifcdf`	`unifrnd`
Student's t	`tpdf`	`tcdf`	`trnd`
F	`fpdf`	`fcdf`	`frnd`
Chi-square	`chi2pdf`	`chi2cdf`	`chi2rnd`
Exponential	`exppdf`	`expcdf`	`exprnd`
Beta	`betapdf`	`betacdf`	`betarnd`
Gamma	`gampdf`	`gamcdf`	`gamrnd`
Lognormal	`lognpdf`	`logncdf`	`lognrnd`
Discrete Uniform	`unidpdf`	`unidcdf`	`unidrnd`
Binomial	`binopdf`	`binocdf`	`binornd`
Poisson	`poisspdf`	`poisscdf`	`poissrnd`

There are several other distributions that MATLAB can sample from. See the manual for a complete list.

Linear Transformations of Normal Vectors

Let $n \times 1$ vector of random variables be distributed normally: $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the linear combination

$$\boldsymbol{Ax} + \boldsymbol{b} \sim N\left(\boldsymbol{A\mu} + \boldsymbol{b},\ \boldsymbol{A}\Sigma\boldsymbol{A}^\top\right)$$

where $\boldsymbol{A}$ is full rank matrix of known constants.

Quadratic Forms

Theorem (Distribution of an Idempotent Quadratic Form in a Standard Normal Vector)

If $\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{I})$ and $\boldsymbol{A}$ is idempotent then

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \sim \chi^2_\nu, \quad \nu = \mathit{rank}(\boldsymbol{A})$$

Note that for an idempotent matrix rank($\boldsymbol{A}$) =trace($\boldsymbol{A}$).

For example

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \boldsymbol{x}^\top \boldsymbol{M}_0 \boldsymbol{x} \sim \chi^2_{n-1}$$

where

$$\boldsymbol{M}_0 = \boldsymbol{I} - \frac{1}{n}\boldsymbol{\imath}\boldsymbol{\imath}^\top$$

and $\boldsymbol{\imath}$ is n-vector of ones and $\bar{x}$ is the sample mean. Notice that diagonal elements of $\boldsymbol{M}_0$ are all $1 - 1/n$, thus $tr(\boldsymbol{M}_0) = n - 1$ which is also the rank.

Quadratic Forms

Previous theorem may be generalized.
If $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then

$$\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}) \sim N(\mathbf{0}, \boldsymbol{I})$$

In the univariate case this transformation is just familiar standardization. The quadratic form

$$(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \sim \chi^2_n$$

Any positive definite matrix can be decomposed into two matrices such that

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}$$

(e.g. Cholesky factorization). Inverting we have

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2}$$

Elements of Asymptotic Analysis

Definition (Limit of a Sequence)

The sequence $\{a_n\}$ has a limit a denoted

$$a_n \to a, \quad n \to \infty$$

if for all $\varepsilon > 0$ there is some $n_\varepsilon < \infty$ such that for all $n \geq n_\varepsilon$

$$|a_n - a| \leq \varepsilon$$

where $\varepsilon > 0$ can be chosen arbitrarily small.

The definition above is valid for deterministic sequences.

Elements of Asymptotic Analysis

For sequences of random variables we use the concept of convergence in probability.

Definition (Convergence in Probability)

A sequence of random variables x_n converges in probability to c if for any $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(|x_n - c| > \varepsilon\right) = 0$$

or

$$\lim_{n\to\infty} \mathbb{P}\left(|x_n - c| < \varepsilon\right) = 1$$

If this condition is satisfied then

$$\operatorname{plim}(x_n) = c, \quad \text{or,} \quad x_n \xrightarrow{p} c$$

For example a sequence $\{x_n\}$ converges to 0 if

$$\lim_{n\to\infty} \mathbb{P}\left(|x_n| < \varepsilon\right) = 1$$

Elements of Asymptotic Analysis

Definition (Convergence in Quadratic Mean (Mean Square))

A sequence of random variables x_n with

$$\mathsf{E}(x_n) = \mu_n, \quad \mathsf{Var}(x_n) = \sigma_n^2$$

converges in mean square if

$$\text{plim}(\mu_n) = c$$

and

$$\text{plim}(\sigma_n^2) = 0$$

Convergence in mean square implies convergence in probability, but reverse is not true.

Elements of Asymptotic Analysis

Example

Let $\bar{x}_n$ be a sequence of sample means based on identically and independently distributed (iid) samples from a population with mean μ and variance σ^2. Then it can be shown that

$$\mathsf{E}(\bar{x}_n) = \mu, \quad \mathsf{Var}(\bar{x}_n) = \frac{\sigma^2}{n}$$

Obviously since

$$\text{plim}(\mathsf{Var}(\bar{x}_n)) = 0$$

$\bar{x}_n$ converges in mean square to μ, i.e.,

$$\text{plim}(\bar{x}_n) = \mu$$

Elements of Asymptotic Analysis

Theorem (Markov's Inequality)

Let w be a nonnegative random variable and let $\varepsilon > 0$ be a positive constant. Then

$$\mathbb{P}(w \geq \varepsilon) \leq \frac{E(w)}{\varepsilon}$$

Proof: see Greene p.1068

Theorem (Chebyshev's Inequality)

Let u be a random variable and also let $\varepsilon > 0$ and c be constants. Then

$$\mathbb{P}(|u - c| \geq \varepsilon) \leq \frac{E[(u-c)^2]}{\varepsilon^2}$$

If $c = E(u)$ then

$$\mathbb{P}(|u - E(u)| \geq \varepsilon) \leq \frac{Var(u)}{\varepsilon^2}$$

Proof: see Greene p.1068

Elements of Asymptotic Analysis

Theorem (Weak Law of Large Numbers (WLLN))

If y_i, $i = 1, 2, \ldots, n$, are iid and $E(|y|) < \infty$ then as $n \to \infty$

$$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} \xrightarrow{p} E(y)$$

Elements of Asymptotic Analysis

Definition (Consistency)

An estimator $\hat{\theta}_n$ of the parameter θ is consistent if

$$\hat{\theta}_n \xrightarrow{p} \theta$$

as $n \to \infty$. If this condition is satisfied we write

$$\text{plim}(\hat{\theta}_n) = \theta.$$

In most cases convergence in-mean-square is sufficient to show consistency.

Elements of Asymptotic Analysis

Theorem (Slutsky's Theorem)

Let $g(\hat{\theta}_n)$ be a continuous function that does not depend on n directly. Then

$$\text{plim}\left(g(\hat{\theta}_n)\right) = g\left(\text{plim}(\hat{\theta}_n)\right).$$

Rules of operations with plim can easily be deduced from the Slutsky's Theorem.

Elements of Asymptotic Analysis

Definition (Convergence in Distribution)

Let x_n be a sequence of random variables with cumulative density function (cdf) $F_n(x)$. If

$$\lim_{n\to\infty} |F_n(x) - F(x)| = 0$$

at all continuity points then x_n converges in distribution to x:

$$x_n \xrightarrow{d} x, \quad \Leftrightarrow \quad F_n(x_n) \to F(x)$$

As an example consider the limiting distribution of t_{n-1}.

Elements of Asymptotic Analysis

Theorem (Central Limit Theorem (Lindberg-Levy))

Let $\{x_1, x_2, \ldots, x_n\}$ be a random sample from a population with mean $\mu < \infty$ and variance $0 < \sigma^2 < \infty$. Then, regardless of the distribution of the population

$$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

or equivalently

$$\frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$

Elements of Asymptotic Analysis

Definition (Asymptotic Distribution)

Let $\hat{\boldsymbol{\theta}}_n$ be an estimator of $\boldsymbol{\theta}$. Also, suppose that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V})$$

Then, the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$ is given by

$$\hat{\boldsymbol{\theta}}_n \overset{a}{\sim} N\left(\boldsymbol{\theta}, \frac{1}{n}\boldsymbol{V}\right)$$

where

$$\mathsf{Avar}(\hat{\boldsymbol{\theta}}_n) = \frac{1}{n}\boldsymbol{V}$$

is the asymptotic variance-covariance matrix.

Asymptotic Distribution

Example

Consider the distribution of sample mean of a random sample from a population with mean μ and variance σ^2. By the CLT we know that

$$\frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$

which implies that

$$\sqrt{n}(\bar{x}_n - \mu) \overset{a}{\sim} N(0, \sigma^2)$$

$$(\bar{x}_n - \mu) \overset{a}{\sim} N(0, \sigma^2/n)$$

$$\bar{x}_n \overset{a}{\sim} N(\mu, \sigma^2/n)$$

Elements of Asymptotic Analysis

Definition (Asymptotic Efficiency)

Let $\tilde{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\theta}}_n$ be two consistent estimators of the parameter vector $\boldsymbol{\theta}$ with

$$\hat{\boldsymbol{\theta}}_n \overset{a}{\sim} N\left(\boldsymbol{\theta}, \frac{1}{n}\boldsymbol{V}\right)$$

$$\tilde{\boldsymbol{\theta}}_n \overset{a}{\sim} N\left(\boldsymbol{\theta}, \frac{1}{n}\boldsymbol{W}\right)$$

If the difference between covariance matrices

$$\boldsymbol{W} - \boldsymbol{V}$$

is some positive definite matrix, then $\hat{\boldsymbol{\theta}}_n$ is asymptotically more efficient than $\tilde{\boldsymbol{\theta}}_n$.

Asymptotic Efficiency

Example

Consider the sample mean and the sample median (m_n) as estimators for the population mean. We know both are consistent with asymptotic distributions:

$$\bar{x}_n \overset{a}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$m_n \overset{a}{\sim} N\left(\mu, \frac{\pi}{2}\frac{\sigma^2}{n}\right)$$

Obviously $\bar{x}_n$ is asymptotically more efficient than m_n since $\bar{x}_n$ has a smaller variance.

Stochastic Order Notation

- How to represent convergence to zero or stochastic boundedness of random vectors and variables?
- Two commonly used notation is $o(\cdot)$ (small oh) and $O(\cdot)$ (big oh).
- Let x_n and a_n, $n = 1, 2, \ldots$, be two non-random sequences. The notation

$$x_n = o(1) \ \text{ implies } x_n \to 0$$

as $n \to \infty$.
- Similarly

$$x_n = o(a_n) \ \text{ implies } a_n^{-1} x_n \to 0$$

as $n \to \infty$.

Stochastic Order Notation

- For example, let

$$h_n = n^{1.4}, \quad n = 1, 2, \ldots$$

- Obviously

$$h_n = o(n^{1.5})$$

- Because

$$n^{-1.5} h_n = n^{-0.1} \to 0.$$

Stochastic Order Notation

- Big Oh notation, on the other hand, implies that the sequence is bounded.
- The notation

$$x_n = O(1)$$

implies that there exists $M < \infty$ such that $|x_n| \leq M$ for all n (x_n is bounded uniformly in n).
- The notation

$$x_n = O(a_n)$$

is equivalent to

$$a_n^{-1} x_n = O(1).$$

Stochastic Order Notation

- For example, the sequence

$$h_n = n^2 + 2n + 1, \quad n = 1, 2, \ldots$$

is

$$h_n = O(n^2)$$

- Because

$$n^{-2} h_n = 1 + \frac{2}{n} + \frac{1}{n^2} \to 1.$$

- But

$$h_n = o(n^{2+\delta}), \quad \text{for any } \ \delta > 0$$

Stochastic Order Notation

- We can define similar concepts for random variables and vectors.
- Let z_n and a_n be sequences of random variables.
- The notation, small-oh-p-one

$$z_n = o_p(1)$$

means that

$$z_n \xrightarrow{p} 0 \text{ as } n \to \infty$$

- If

$$a_n^{-1} z_n = o_p(1)$$

- then we write

$$z_n = o_p(a_n).$$

- For example, of $\hat{\theta}_n$ is a consistent estimator of θ then

$$\hat{\theta}_n = \theta + o_p(1).$$

Stochastic Order Notation

- The notation

$$z_n = O_p(1)$$

means that z_n is bounded in probability.
- Similarly

$$z_n = O_p(a_n)$$

implies

$$a_n^{-1} z_n = O_p(1).$$

- $O_p(1)$ is weaker than $o_p(1)$ in the sense that $z_n = o_p(1)$ implies $z_n = O_p(1)$ but not the reverse.
- If $z_n = O_p(a_n)$ then $z_n = o_p(b_n)$ for any b_n such that $a_n/b_n \to 0$.

Stochastic Order Notation

- If a random vector converges in distribution $z_n \xrightarrow{d} z$, e.g. $z \overset{a}{\sim} N(\mathbf{0}, \boldsymbol{V})$, then $z_n = O_p(1)$.
- For example

$$\bar{x}_n = \mu + O_p(n^{-1/2}).$$

- Rules for manipulating $o_p(1)$ and $O_p(1)$ sequences:

$$\begin{aligned}
o_p(1) + o_p(1) &= o_p(1) \\
o_p(1) + O_p(1) &= O_p(1) \\
O_p(1) + O_p(1) &= O_p(1) \\
o_p(1) o_p(1) &= o_p(1) \\
o_p(1) O_p(1) &= o_p(1) \\
O_p(1) O_p(1) &= O_p(1)
\end{aligned}$$