CLASS NOTES UNIVERSITY OF CYPRUS:

SALIENT ASPECTS OF PROBABILITY THEORY PERTINENT TO CLASSICAL ECONOMETRICS

P.J. Dhrymes, April 8, 2005.[1]

The purpose of these notes is to set forth, in a convenient fashion, the essential results from probability theory necessary to understand classical econometrics. All references are to: *Topics in Advanced Econometrics: Probability Foundations, New York: Springer-Verlag, 1989* by P. J. Dhrymes.

Miscellaneous

Discussion of random variables (r.v.) takes place in a probability space $(\Omega, \mathcal{A}, \mathcal{P})$, where Ω is the **sample space**, $\mathcal{A}$ is the σ**-algebra** and $\mathcal{P}$ is the **probability measure**.

If X is an a.c. finite random variable,[2] i.e. if $P(A) = 0$ for $A = \{\omega : X(\omega) = \infty\}$, then

$$\frac{X}{b_n} \stackrel{\text{a.c}}{\rightarrow} 0,$$

where b_n is a sequence such that $\lim_{n\to\infty} |b_n| = \infty$.

Let $\{X_n : n \geq 1\}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and suppose that

$$S_n = \sum_{i=1}^{n} X_i, \quad \frac{S_n}{n} \stackrel{\text{p or a.c.}}{\rightarrow} 0,$$

then

$$\frac{X_n}{n} \stackrel{\text{p or a.c.}}{\rightarrow} 0.$$

[2] The notation a.c. means **almost certainly**; an alternative notation is a.s. which means **almost surely**.

See pp. 148-150.

Any random variable, as above, induces a σ-algebra, say, $\sigma(X)$, which is the smallest σ-algebra for studying X and is defined by

$$\sigma(X) = \{A : A = X^{-1}(B), B \ \in \mathcal{B}(R)\},$$

where R is the real line and $\mathcal{B}(R)$ is the Borel σ-algebra. pp.

The connection between random variables as used here and random variables as usually taught in elementary statistics courses is the following: for each random variable, X, defined on the probability space (Ω, $\mathcal{A}$, $\mathcal{P}$), consider its range, R, and the σ-algebra associated with it, the Borel σ-algebra, $\mathcal{B}(R)$. We then take the identity transformation (so that X is identified solely by the values it assumes)

$$X : R \rightarrow R$$

and we deal exclusively with the values assumed by the random variable; we assign probabilities to such sets (of values assumed by the random variable) by the rule: if $B \in \mathcal{B}(R)$, assign a probability to the event: X assumes a value in B, by

$$P(B) = \mathcal{P}(A), \quad A = X^{-1}(B).$$

It may be shown that P is a proper probability measure $P : \mathcal{B}(R) \rightarrow [0, 1]$, and that $(R, \mathcal{B}(R), P)$ is the probability space **induced** by X. In this context what distinguishes one variable from another is the probability measure they induce, P. Hence, the usual statement let X be a random variable with distribution, G. The connection between P and G is given by $G(x) = P(B)$ in the case of a set $B = (-\infty, x]$. pp.

Limits of sets

Let $\{A_n : n \geq 1\}$ be a sequence of sets defined on some σ-algebra $\mathcal{A}$. Define

$$B_n = \bigcup_{i=n}^{\infty} A_i, \quad C_n = \bigcap_{i=n}^{\infty} A_i, \quad B^* = \lim_{n\rightarrow\infty} B_n, \quad C_* = \lim_{n\rightarrow\infty} C_n,$$

and note that

$$B_n \supseteq A_n, \quad C_n \subseteq A_n,$$

and both are monotonic obeying $B_n \supseteq B_{n+1}$ and $C_n \subseteq C_{n+1}$, respectively. This means that the B-sequence is **non-increasing** while the C-sequence is **non-decreasing**. This justifies the definition of B^*, C^* as above since

$$B^* = \bigcap_{n=1}^{\infty} B_n, \; C^* = \bigcup_{n=1}^{\infty} C_n.$$

The set B^* is said to be the **limit superior**, and the set C_* is said to be the **limit inferior** of the sequence, and one writes

$$B^* = A^* = \limsup_{n\to\infty} A_n, \quad C_* = A_* = \liminf_{n\to\infty} A_n.$$

Evidently, $A^* \supseteq A_*$. **If** $A^* = A_*$, the common value of the limit inferior and limit superior is said to be the limit of the sequence and is denoted by $A = \lim_{n\to\infty} A_n$.

pp. 8- 10.

Note the notational equivalence:

$$\overline{A^*} = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \bar{A}_k = \bar{A}_*,$$

and conversely with the limit inferior, i.e. the complement of the limit inferior of the sequence $\{A_n : n \geq 1\}$ is the limit superior of the sequence $\{\bar{A}_n : n \geq 1\}$.

Remark 1: The intuitive meaning of the limit superior, A^*, is that it contains elements that belong to members of the sequence above *infinitely often*, a fact that is denoted by (the notation)

$$A^* = \{\omega : \omega \in A_n, \; i.o.\}$$

Modes of Convergence

Let $\{X_n : n \geq 0\}$ be a sequence of random variables defined on the probability space (Ω, $\mathcal{A}$, $\mathcal{P}$), we want to define statements like: The sequence converges to X_0. To this end define, for arbitrary integer r, the sets

$$A_{n,r} = \{\omega : |X_n(\omega) - X_0(\omega)| \geq \frac{1}{r}\}.$$

Convergence with Probability one, or almost sure (a.s), or almost certain (a.c.) convergence means the following: if, for given r, we denote by A_r^* the limit superior of the sequence, convergence with probability one means,

$$\mathcal{P}(A^*) = \lim_{r\to\infty} \mathcal{P}(A_r^*) = 0, \quad \text{because} \quad A^* = \bigcup_{r=1}^{\infty} A_r^*, \quad \text{and} \quad A_r^* \subseteq A_{r+1}^*.$$

For convenience in use we shall frequently denote this fact by

$$X_n \stackrel{\text{a.c. (or a.s.)}}{\rightarrow} X_0,$$

which will mean that for arbitrary r the probability attached to the limit superior A_r^* is arbitrarily close to zero. (pp.134-136)

Convergence in Probability

Convergence in probability, denoted by,

$$X_n \stackrel{\text{P}}{\rightarrow} X_0$$

simply means that, for any r there exists a number $q(r)$ such that

$$\lim_{n\to\infty} \mathcal{P}(A_{nr}) \leq q(r)$$

and $q(r)$ can be made arbitrarily close to zero by making the proper choice of r. Evidently, convergence a.c. implies convergence in probability. (pp. 134-136).

Convergence in Mean of order p, or $\mathbf{L}^p$

Let $\{X_n : n \geq 0\}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and suppose

$$E|X_n|^p < \infty, \quad n \geq 0, \quad p > 0.$$

The sequence converges X_0, **in mean of order p**, denoted by

$$X_n \stackrel{\text{L}^\text{p}}{\rightarrow} X_0$$

if and only if

$$\lim_{n\to\infty} E|X_n - X_0|^p = 0.$$

see pp. 156ff.

Convergence in Distribution

Let $\{X_n : n \geq 1\}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with respective cdf F_n; it converges **in distribution** to a random variable X_0 if and only if

$$F_n \xrightarrow{\text{c}} F,$$

where c indicates **complete convergence** as follows. Let $G_n,\ n \geq 1$ be a sequence of non-decreasing functions; it is said to **converge weakly** to a non-decreasing function G, denoted by $G_n \xrightarrow{\text{w}} G$ if and only if

$$\lim_{n\to\infty} G_n(x) = G(x), \qquad x \in C(G),$$

where $C(G)$ is the **set of continuity points** of G. If, in addition

$$\lim_{n\to\infty} G_n(\pm\infty) = G(\pm\infty) = \lim_{x\to\pm\infty} G(x),$$

the sequence is said to **converge completely**, a fact that is denoted by $G_n \xrightarrow{\text{c}} G$.

Since the cdfs (cumulative distribution functions) are non-decreasing functions, the explanation is complete, and convergence in distribution is denoted by

$$X_n \xrightarrow{\text{d}} X_0.$$

Remark 2. Note that, essentially, convergence in distribution **does not involve convergence to a random variable**, even though common usage of the term indicates that this is so. In fact convergence in distribution (or convergence in Law as in a more antique usage) indicates that the cdfs of the sequence converge completely and, thus, to a distribution function (cdf). Since to each cdf there corresponds (non-uniquely) a random variable, (cf: let X be a sequence of i.i.d. random variables) what we mean by the terminology is that the sequence converges **in an**

equivalence class of random variables having the distribution F. pp. 152-153.

Relationship among Modes of Convergence

Let $\{X_n : n \geq 0\}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$, then

(1).$X_n \stackrel{\text{a.c.}}{\to} X_0$ **implies** $X_n \stackrel{\text{P}}{\to} X_0$,

(2).$X_n \stackrel{\text{P}}{\to} X_0$ **implies** $X_n \stackrel{\text{d}}{\to} X_0$

(3).$X_n \stackrel{\text{L}^{\text{p}}}{\to} X_0$ **implies** $X_n \stackrel{\text{P}}{\to} X_0$

(4).$X_n \stackrel{\text{P}}{\to} X_0$ **does not imply** $X_n \stackrel{\text{L}^{\text{p}}}{\to} X_0$

(5).$X_n \stackrel{\text{P}}{\to} X_0$ **does not imply** $X_n \stackrel{\text{a.c.}}{\to} X_0$

(6).$X_n \stackrel{\text{L}^{\text{p}}}{\to} X_0$ **does not imply** $X_n \stackrel{\text{a.c.}}{\to} X_0$

(7).$X_n \stackrel{\text{d}}{\to} X_0$ **does not imply** $X_n \stackrel{P}{\to} X_0$.

Remark 3. Even though by (6) convergence in distribution does not imply convergence in probability for reasons given in Remark 2, there is one instance where this is so. It occurs when convergence in distribution is to a constant (degenerate). In **this case it does imply convergence in probability to the constant.** see pp. 165 and 262.

Applications of Modes of Convergence

Let $\{X_n : n \geq 0\}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and suppose

$$X_n \stackrel{\text{P}}{\to} X_0.$$

If $\{Y_n : n \geq 1\}$ is another sequence such that it converges in probability to Y_0 and, **if**, Y_0, X_0 are **equivalent** in the sense that they differ **only on a set of** $\mathcal{P}$-measure zero, then

$$|X_n - Y_n| \stackrel{\text{P}}{\to} 0.$$

see pp. 151-152.

Let $\{X_n, Y_n : n \geq 0\}$ be sequences of random variables defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and suppose

$$Y_n \xrightarrow{\text{d}} Y_0, \quad |X_n - Y_n| \xrightarrow{\text{P}} 0.$$

Then

$$X_n \xrightarrow{\text{d}} Y_0.$$

See pp. 161-162.

Let $\{X_n : n \geq 0\}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$, and let

$$\phi : R \rightarrow R$$

be a $\mathcal{B}(R)-$ **measurable function** whose discontinuities are contained in a set D of $\mathcal{P}$-measure zero. Then, provided $\phi(X_0)$ is well defined,

i. $X_n \xrightarrow{\text{P}} X_0$ implies $\phi(X_n) \xrightarrow{\text{P}} \phi(X_0)$;

ii. $X_n \xrightarrow{\text{a.c.}} X_0$ implies $\phi(X_n) \xrightarrow{\text{a.c.}} \phi(X_0)$;

iii. $X_n \xrightarrow{\text{d}} X_0$ implies $\phi(X_n) \xrightarrow{\text{d}} \phi(X_0)$.

See pp. 144-145, 147-148, 242-243.

A more useful form of item iii is as follows: For the sequence above, define the transformation

$$Y_n = \phi(X_n).$$

We show explicitly the form of the converged distribution of the Y-sequence. Let P_n and P_n^* be, respectively, the distributions induced by the elements of the X- and Y-sequences, respectively, and C an arbitrary set in $\mathcal{B}(R)$. If we can determine the value of $P_n^*(C)$, for arbitrary $C \in \mathcal{B}(R)$, we will have defined the distribution function of Y_n. But $Y_n \in C$ if and only if $X_n \in B = \phi^{-1}(C)$; now $X_n \in B$ with probability $P_n(B)$. Thus,

$$P_n^*(C) = P_n(B) = P_n[\phi^{-1}(C)] = P_n \circ \phi^{-1}(C),$$

where $P_n \circ \phi^{-1}$ is the **composition of** P_n and ϕ^{-1}. Since C is an arbitrary set in $\mathcal{B}(R)$, it follows

$$P_n^* = P_n \circ \phi^{-1},$$

and consequently

$$P_n \xrightarrow{\mathrm{w}} P_0, \quad \textbf{implies} \quad P_n^* \xrightarrow{\mathrm{w}} P_0 \circ \phi^{-1},$$

where the notation $\xrightarrow{\mathrm{w}}$ denotes weak convergence of measures which is roughly comparable, in this context, to complete convergence of cumulative distribution functions explained above. Note, in addition, that

$$P_0 \circ \phi^{-1} \quad \textbf{is the distribution of} \quad Y_0 = \phi(X_0).$$

An interesting by-product of this is the following very useful result: Let A_n, a_n be a suitable random matrix and vector, respectively, **converging in probability to** A, a, and let ξ_n be a sequence of random vectors converging in distribution to ξ_0, then

$$\zeta_n = A_n \xi_n + a_n \xrightarrow{\mathrm{d}} A\xi_0 + a.$$

See pp. 242-244

Laws of Large Numbers (LLN) and Central Limit Theorems (CLT)

Let $\{X_n : n \geq 1\}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and define

$$S_n = \sum_{i=1}^{n} X_i, \qquad Q_n = \frac{S_n - a_n}{b_n},$$

where $b_n > 0$, for all n, and the sequence tends to $+\infty$ and it, as well as a_n, are taken to be non-random. Both LLN and CLT describe properties of the sequence Q_n as $n \to \infty$.

Weak Law of Large Numbers (WLLN): If $Q_n \xrightarrow{\mathrm{P}} 0$ we say that the sequence obeys the WLLN.

Strong Law of Large Numbers (SLLN): If $Q_n \overset{\text{a.c.}}{\to} 0$, we say that the sequence obeys the SLLN.

Central Limit Theorem (CLT): if $Q_n \overset{\text{d}}{\to} \xi$ and ξ is a member of a well defined equivalence class of random variables, we say that the sequence obeys a CLT.

Criteria for applicability

Let $\{X_n : n \geq 1\}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and suppose the elements of the sequence are independent, identically distributed (i.i.d.) with mean μ (actually we need $E|X_1| < \infty$, but existence of variance is not required). Taking

$$a_n = n\mu, \qquad b_n = n,$$

we have that

$$Q_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) \overset{\text{a.c.}}{\to} 0,$$

and hence that it converges to zero in probability as well, so that it obeys both the WLLN and the SLLN.

See pp. 188-190.

For independent, not identically distributed random variables with finite variance we have Kolmogorov′s criterion for a SLLN. Taking

$$Q_n = \frac{1}{b_n}\sum_{i=1}^{n}(X_i - E(X_n))$$

we have that $Q_n \overset{\text{a.c.}}{\to} 0$, **provided**

$$\sum_{n=1}^{\infty}\left(\frac{\text{Var}(X_n)}{b_n^2}\right) < \infty.$$

Thus, for $b_n = n$ a sufficient condition for

$$Q_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_n)) \overset{\text{a.c.}}{\to} 0$$

is that

$$\sum_{n=1}^{\infty}\left(\frac{\text{Var}(X_n)}{n^2}\right) < \infty.$$

Evidently this is satisfied when $\mathrm{Var}(X_n)$ is (uniformly) bounded, or more generally when $\mathrm{Var}(X_n) \sim cn^{\alpha}$, for $\alpha \in [0,1)$. See pp. 186-188.

If the sequence above is merely a sequence of **uncorrelated** random variables with variance $\mathrm{Var}(X_n) \leq cn^{\alpha}$, and $\alpha \in [0, \frac{1}{2})$, then

$$Q_n = \frac{1}{n}\left(\sum_{i=1}^{n}[X_i - E(X_i)]\right) \stackrel{\text{a.c.}}{\to} 0$$

See pp. 191-193

Central Limit Theorems: Let $\{X_n : n \geq 1\}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and suppose they are independent identically distributed with mean μ and variance σ^2, then

$$Q_n = \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{n}} \stackrel{\text{d}}{\to} X \sim N(0, \sigma^2).$$

See p. 264

For the sequence above suppose that it is only one of **independent, not identically distributed** random variables, with mean μ_n and variance σ_{nn}, we have the following: put

$$z_n = \frac{S_n}{\sigma_n}, \quad S_n = \sum_{i=1}^{n}(X_i - \mu_i), \quad \sigma_n^2 = \sum_{i=1}^{n}\sigma_{ii}, \quad X_{in} = \frac{X_i - \mu_i}{\sigma_n},$$

and note that

$$z_n = \sum_{i=1}^{n} X_{in}, \quad \mathrm{Var}(X_{in}) = \frac{\sigma_{ii}}{\sigma_n^2} = \sigma_{in}^2.$$

If we denote the cdf of the X_i by F_i and the cdf of X_{in} by F_{in} we note that the mean of the last distribution is zero and its variance σ_{in}^2 which converges to zero with (the sample size) n. Moreover, define the **Lindeberg condition** by

$$\lim_{n\to\infty} W_n = 0, \quad W_n = \sum_{i=1}^{n}\int_{|\xi|>\frac{1}{r}} \xi^2 dF_{in}(\xi),$$

for arbitrary integer r. The Lindeberg central limit theorem (CLT) then states

Theorem (Lindeberg). Under the conditions above

$$z_n \xrightarrow{\text{d}} \zeta \sim N(0,1).$$

To see, roughly speaking, what the Lindeberg condition means, revert to the original variables and, for simplicity, assume the sequence has **zero means**. In this case the Lindeberg condition becomes

$$W_n = \frac{1}{\sigma_n^2}\sum_{i=1}^{n}\int_{|\xi|>\frac{\sigma_n}{r}} \xi^2 dF_i(\xi),$$

which says that as the sample increases the sum of the tails of the variance integral become negligible in comparison to the total variance.

Another way of displaying the Lindeberg CLT and the Lindeberg condition is to redefine X_{in} from

$$X_{in} = \frac{X_i - \mu_i}{\sigma_n} \quad \text{to} \quad X_{in} = \frac{X_i - \mu_i}{\sqrt{n}},$$

and put $z_n^* = \frac{S_n}{\sqrt{n}}$. This means that z_n^* is given by

$$z_n* = \frac{S_n}{\sigma_n}\frac{\sigma_n}{\sqrt{n}} = z_n\frac{\sigma_n}{\sqrt{n}},$$

so that

$$z_n^* \xrightarrow{\text{d}} \zeta\sigma \sim N(0,\sigma^2), \quad \sigma^2 = \lim_{n\to\infty}\frac{\sigma_n^2}{n},$$

provided the latter exists, i.e. $\sigma^2 < \infty$. See pp. 271-274.

Martingale Difference (MD) Central Limit Theorem: First, we define what martingales and martingale differences are. Let $\{X_n : n \geq 1\}$ be a sequence of random variables defined on the probability space $(\Omega,\ \mathcal{A},\ \mathcal{P})$ and consider the sequence of (sub) σ-algebras, $\mathcal{A}_{n-1} \subseteq \mathcal{A}_n, \quad n \in N$, where N is a subset of the integers in $(-\infty,\infty)$. The sequence $\{\mathcal{A}_n : n \in N\}$ is said to be a **stochastic basis** or a **filtration**. If, in the sequence above, X_n is $\mathcal{A}_n$-measurable, the sequence of pairs $\{(X_n, \mathcal{A}_n) : n \in N\}$ is said to be a **stochastic sequence**. If, in addition, $E|X_n| < \infty$, the sequence is said to be

i. a **martingale** if $E(X_{n+1}|\mathcal{A}_n) = X_n$;

ii. a **sub-martingale** if $E(X_{n+1}|\mathcal{A}_n) \geq X_n$;

iii. a **super-martingale** if $E(X_{n+1}|\mathcal{A}_n) \le X_n$;

iv. a **martingale difference** if $E(X_{n+1}|\mathcal{A}_n) = 0$.

In the above, the notation $|\mathcal{A}_n$ means conditioning on the (sub) σ-algebra in question, or the random variables generating them.

Theorem (MD CLT). In the notation of the previous theorem, let $\{X_{in} : i \le n\}$ be defined as

$$X_{in} = \frac{X_i - EX_i}{\sqrt{n}},$$

and suppose that, for each n, $\{(X_{in}, \mathcal{A}_{in}) : i \le n\}$ is a martingale difference sequence satisfying the Lindeberg condition, i.e.

$$\operatorname*{plim}_{n\to\,\infty} W_n = 0, \quad W_n = \sum_{i=1}^{n} E[X_{in}^2 I(|X_{in}| \ge \frac{1}{r})|\mathcal{A}_{i-1,n}],$$

where $I(|X_{in}| \ge \frac{1}{r})$ equals one, if $|X_{in}| \ge \frac{1}{r}$, and zero otherwise. The following statements are true

i. If $\sum_{i=1}^{n} E(X_{in}^2|\mathcal{A}_{i-1,n}) \xrightarrow{\text{P}} \sigma^2$, then $z_n = \sum_{i=1}^{n} X_{in}^2 \xrightarrow{\text{d}} \zeta \sim N(0, \sigma^2)$;

ii. If $\sum_{i=1}^{n} X_{in} \xrightarrow{\text{P}} \sigma^2$, then $z_n \xrightarrow{\text{d}} \zeta \sim N(0, \sigma^2)$.

See pp. 323-337.

Example. Consider, in the usual context, the model

$$y_t = \lambda y_{t-1} + u_t, \quad |\lambda| < 1, \quad \text{where } u_t,\ t = 1, 2, 3, \ldots, T, \tag{1}$$

is i.i.d., defined on the probability space (Ω , $\mathcal{A}$, $\mathcal{P}$) . The OLS estimator of λ is given by

$$\hat{\lambda} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} = \lambda + \frac{\sum_{t=2}^{T} u_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}.$$

Put

$$\hat{a}_T = \frac{\sum_{t=2}^{T} y_{t-1}^2}{T}, \quad \text{and note that} \quad \hat{a}_T \xrightarrow{\text{P}} \frac{\sigma^2}{(1-\lambda^2)}.$$

Therefore we can write

$$\sqrt{T}(\hat{\lambda} - \lambda) \sim \frac{1-\lambda^2}{\sigma^2} \sum_{t=2}^{T} X_{tT}, \qquad X_{tT} = \frac{y_{t-1} u_t}{\sqrt{T}}.$$

Now define the (sub) σ-algebras $\mathcal{A}_{tT} = \sigma(X_{sT} : s \leq t) \subseteq \mathcal{A}_{t+1,T}$, and note that X_{tT} is $\mathcal{A}_{tT}$-measurable. Moreover,

$$E(X_{tT}|\mathcal{A}_{t-1,T}) = \frac{1}{\sqrt{T}} y_{t-1} E u_t = 0, \quad E X_{tT}^2 | \mathcal{A}_{t-1,T} = \frac{1}{T} y_{t-1}^2 E u_t^2 = \frac{1}{T} y_{t-1}^2 \sigma^2,$$

so that

$$\sum_{t=2}^{T} E X_{tT}^2 | \mathcal{A}_{t-1,T} = \frac{1}{T} \sum_{t=2}^{T} y_{t-1}^2 \sigma^2 \xrightarrow{\mathrm{P}} \frac{\sigma^4}{1-\lambda^2}.$$

Consequently

$$\sqrt{T}(\hat{\lambda} - \lambda) \xrightarrow{\mathrm{d}} \zeta \sim N(0, 1-\lambda^2).$$

The reader may have noticed that we have given CLTs only for scalar random variables and may have wondered why we did not provide such results for sequences of random **vectors**. In fact, such results may be subsumed under the discussion of scalar CLTs as the following will make clear. First we note that the **characteristic function** (with parameter s) of a normal variable Y with mean μ and variance σ^2 is given by

$$\phi(s) = E e^{isY} = e^{is\mu - \frac{1}{2}s^2\sigma^2};$$

if Y is a **normal** random vector with mean vector μ and covariance matrix $\Sigma > 0$, its characteristic function, with parameter t, is given by

$$\phi(t) = E e^{it'Y} = e^{it'\mu - \frac{1}{2}t'\Sigma t}.$$

We now have

Theorem. Let X be a random vector, i.e.

$X : \Omega \to R^m, \ EX = \mu, \ \mathrm{Cov}(X) = \Sigma > 0$.

If for **arbitrary** conformable vector λ

$$\lambda' X \sim N(\lambda'\mu, \ \lambda'\Sigma\lambda),$$

then

$$X \sim N(\mu, \Sigma).$$

Proof: For $y(\lambda) = \lambda' X$, with arbitrary conformable λ, $y(\lambda)$ is, by the premise of the theorem, normal with mean $\lambda'\mu$ and variance $\lambda'\Sigma\lambda$. Its characteristic function, with parameter s, is therefore given by

$$\phi(s) = e^{is\lambda'\mu - \frac{1}{2}s^2\lambda'\Sigma\lambda}.$$

But taking $t = s\lambda$, we find

$$\phi(s) = \phi(t) = Ee^{it'X} = e^{it'\mu - \frac{1}{2}t'\Sigma t},$$

which is recognized as the characteristic function (with parameter t) of a multivariate normal with mean vector μ and covariance