Elements of Probability Theory

CHUNG-MING KUAN

Department of Finance

National Taiwan University

December 5, 2009

Lecture Outline

Lecture Outline (cont'd)

5. Law of Large Numbers

6. Central Limit Theorem

7. Stochastic Processes
   - Brownian motion
   - Weak Convergence

8. Functional Central Limit Theorem

Probability Space and σ-Algebra

- A probability space is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$, where
  1. Ω is the outcome space, whose elements ω are outcomes of the random experiment,
  2. $\mathcal{F}$ is a a σ-algebra, a collection of subsets of Ω,
  3. $\mathbb{P}$ is a a probability measure assigned to the elements in $\mathcal{F}$.
- $\mathcal{F}$ is a σ-algebra if
  1. $\Omega \in \mathcal{F}$,
  2. if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$,
  3. if $A_1, A_2, \cdots \in \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.
- By (2), $\Omega^c = \emptyset \in \mathcal{F}$. From de Mongan's law,

$$\left(\bigcup_{n=1}^{\infty} A_n \right)^c = \bigcap_{n=1}^{\infty} A_n^c \in \mathcal{F}.$$

Probability Measure

- $\mathbb{P}: \mathcal{F} \mapsto [0,1]$ is a real-valued set function such that
  1. $\mathbb{P}(\Omega) = 1$,
  2. $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$.
  3. if $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint, then $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.
- $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, $\mathbb{P}(A) \leq \mathbb{P}(B)$ if $A \subseteq B$, and

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

- If $\{A_n\}$ is an increasing (decreasing) sequence in $\mathcal{F}$ with the limiting set A, then $\lim_{n \to \infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Borel Field

- Let $\mathcal{C}$ be a collection of subsets of Ω. The σ-algebra generated by $\mathcal{C}$, $\sigma(\mathcal{C})$, is the intersection of all σ-algebras that contain $\mathcal{C}$ and hence the smallest σ-algebra containing $\mathcal{C}$.
- When $\Omega = \mathbb{R}$, the Borel field, $\mathcal{B}$, is the σ-algebra generated by all open intervals (a, b) in $\mathbb{R}$.
- Note that (a, b), $[a, b]$, $(a, b]$, and $(-\infty, b]$ can be obtained from each other by taking complement, union and/or intersection. For example,

$$(a, b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right), \qquad (a, b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right].$$

  Thus, the collection all open intervals (closed intervals, half-open intervals or half lines) generates the same Borel field.

- The Borel field on $\mathbb{R}^d$, $\mathcal{B}^d$, is generated by all open hypercubes:

$$(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d).$$

- $\mathcal{B}^d$ can be generated by all closed hypercubes:

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d],$$

  or by

$$(-\infty, b_1] \times (-\infty, b_2] \times \cdots \times (-\infty, b_d].$$

- The sets that generate the Borel field $\mathcal{B}^d$ are all Borel sets.

Random Variable

- A random variable z defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $z\colon \Omega \mapsto \mathbb{R}$ such that for every B in the Borel field $\mathcal{B}$, its inverse image is in $\mathcal{F}$:

$$z^{-1}(B) = \{\omega\colon z(\omega) \in B\} \in \mathcal{F}.$$

  That is, z is a $\mathcal{F}/\mathcal{B}$-measurable (or simply $\mathcal{F}$-measurable) function.

- Given ω, the resulting value $z(\omega)$ is known as a realization of z.
- A $\mathbb{R}^d$ valued random variable (random vector) $\mathbf{z}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is: $\mathbf{z}\colon \Omega \mapsto \mathbb{R}^d$ such that for every $B \in \mathcal{B}^d$,

$$\mathbf{z}^{-1}(B) = \{\omega\colon \mathbf{z}(\omega) \in B\} \in \mathcal{F};$$

  i.e., $\mathbf{z}$ is a $\mathcal{F}/\mathcal{B}^d$-measurable function.

Borel Measurable

- All the inverse images of random vector $\mathbf{z}$, $\mathbf{z}^{-1}(B)$, form a σ-algebra, denoted as $\sigma(\mathbf{z})$.
  - It is known as the σ-algebra generated by $\mathbf{z}$, or the information set associated with $\mathbf{z}$.
  - It is the smallest σ-algebra in $\mathcal{F}$ such that $\mathbf{z}$ is measurable.
- A function $g\colon \mathbb{R} \mapsto \mathbb{R}$ is $\mathcal{B}$-measurable or Borel measurable if

$$\{\zeta \in \mathbb{R}\colon g(\zeta) \leq b\} \in \mathcal{B}.$$

- For random variable z defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and Borel measurable function $g(\cdot)$, $g(z)$ is a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$. The same conclusion holds for d-dimensional random vector $\mathbf{z}$ and $\mathcal{B}^d$-measurable function $g(\cdot)$.

Distribution Function

- The joint distribution function of $\mathbf{z}$ is a non-decreasing, right-continuous function $F_{\mathbf{z}}$ such that for $\zeta = (\zeta_1, \ldots, \zeta_d)' \in \mathbb{R}^d$,

$$F_{\mathbf{z}}(\zeta) = \mathbb{P}\{\omega \in \Omega\colon z_1(\omega) \leq \zeta_1, \ldots, z_d(\omega) \leq \zeta_d\},$$

  with

$$\lim_{\zeta_1 \to -\infty, \ldots, \zeta_d \to -\infty} F_{\mathbf{z}}(\zeta) = 0, \qquad \lim_{\zeta_1 \to \infty, \ldots, \zeta_d \to \infty} F_{\mathbf{z}}(\zeta) = 1.$$

- The marginal distribution function of the i^{th} component of $\mathbf{z}$ is

$$F_{z_i}(\zeta_i) = \mathbb{P}\{\omega \in \Omega\colon z_i(\omega) \leq \zeta_i\} = F_{\mathbf{z}}(\infty, \ldots, \infty, \zeta_i, \infty, \ldots, \infty).$$

Independence

- y and z are (pairwise) independent iff for any Borel sets B_1 and B_2,

$$\mathbb{P}(y \in B_1 \text{ and } z \in B_2) = \mathbb{P}(y \in B_1)\, \mathbb{P}(z \in B_2).$$

- A sequence of random variables $\{z_i\}$ is totally independent if

$$\mathbb{P}\left(\bigcap_{\text{all } i} \{z_i \in B_i\}\right) = \prod_{\text{all } i} \mathbb{P}(z_i \in B_i).$$

Lemma 5.1

Let $\{z_i\}$ be a sequence of independent random variables and h_i, $i = 1, 2, \ldots$ be Borel-measurable functions. Then $\{h_i(z_i)\}$ is also a sequence of independent random variables.

Expectation

- The expectation of Z_i is the Lebesgue integral of z_i wrt to $\mathbb{P}$:

$$\mathbb{E}(z_i) = \int_\Omega z_i(\omega) \,\mathrm{d}\,\mathbb{P}(\omega).$$

In terms of its distribution function,

$$\mathbb{E}(z_i) = \int_{\mathbb{R}^d} \zeta_i \,\mathrm{d}\,F_{\mathbf{z}}(\zeta) = \int_{\mathbb{R}} \zeta_i \,\mathrm{d}\,F_{z_i}(\zeta_i).$$

- For Borel measurable function $g(\cdot)$ of $\mathbf{z}$,

$$\mathbb{E}[g(\mathbf{z})] = \int_\Omega g(\mathbf{z}(\omega)) \,\mathrm{d}\,\mathbb{P}(\omega) = \int_{\mathbb{R}^d} g(\zeta) \,\mathrm{d}\,F_{\mathbf{z}}(\zeta).$$

For example, the covariance matrix of $\mathbf{z}$ $\mathbb{E}(\mathbf{z}\mathbf{z}')$.

A function g is convex on a set S if for any $a \in [0,1]$ and any x, y in S,

$$g\left(ax + (1-a)y\right) \leq ag(x) + (1-a)g(y);$$

g is concave on S if the inequality above is reversed.

Lemma 5.2 (Jensen)

Let g be a convex function on the support of z. For an integrable random variable z such that $g(z)$ is integrable, $g(\mathbb{E}(z)) \leq \mathbb{E}[g(z)]$; the inequality reverses if g is concave.

L_p-Norm

- For random variable z with finite p th moment, its L_p-norm is: $\|z\|_p = [\mathbb{E}(z^p)]^{1/p}$.
- The inner product of square integrable random variables z_i and z_j is:

$$\langle z_i, z_j \rangle = \mathbb{E}(z_i z_j).$$

  The L_2-norm of z_i can be obtained as $\|z_i\|_2 = \langle z_i, z_i \rangle^{1/2}$.
- For any $c > 0$ and $p > 0$, note that

$$c^p \, \mathbb{P}(|z| \geq c) = c^p \int \mathbf{1}_{\{\zeta : |\zeta| \geq c\}} dF_z(\zeta) \leq \int_{\{\zeta : |\zeta| \geq c\}} |\zeta|^p dF_z(\zeta) \leq \mathbb{E}\,|z|^p,$$

  where $\mathbf{1}_A$ is the indicator function of the event A.

Lemma 5.3 (Markov)

Let z be a random variable with finite p th moment. Then,

$$\mathbb{P}(|z| \geq c) \leq \frac{\mathbb{E}\,|z|^p}{c^p},$$

where c is a positive real number.

- For $p = 2$, Markov's inequality is also known as Chebyshev's inequality.
- Markov's inequality is trivial if c is small such that $\mathbb{E}\,|z|^p / c^p > 1$. When c becomes large, the probability that z assumes very extreme values will be vanishing at the rate c^{-p}.

Lemma 5.4 (Hölder)

Let y be a random variable with finite p th moment ($p > 1$) and z a random variable with finite q^{th} moment ($q = p/(p-1)$). Then,

$$\mathbb{E}\,|yz| \leq \|y\|_p\,\|z\|_q.$$

Since $|\,\mathbb{E}(yz)| \leq \mathbb{E}\,|yz|$, we also have:

Lemma 5.5 (Cauchy-Schwatz)

Let y and z be two square integrable random variables. Then,

$$|\,\mathbb{E}(yz)| \leq \|y\|_2\,\|z\|_2.$$

Let $y = 1$ and $x = z^p$. For $q > p$ and $r = q/p$, by Hölder's inequality,

$$\mathbb{E}\,|z^p| \le \|x\|_r\,\|y\|_{r/(r-1)} = [\mathbb{E}(z^{pr})]^{1/r} = [\mathbb{E}(z^q)]^{p/q}.$$

Lemma 5.6 (Liapunov)

Let z be a random variable with finite q th moment. Then for $p < q$, $\|z\|_p \le \|z\|_q$.

Lemma 5.7 (Minkowski)

Let z_i, $i = 1, \ldots, n$, be random variables with finite p th moment ($p \ge 1$). Then, $\|\sum_{i=1}^n z_i\|_p \le \sum_{i=1}^n \|z_i\|_p$.

When $n = 2$, this is just the triangle inequality for L_p-norms.

Conditional Distributions

- Given $A, B \in \mathcal{F}$, suppose we know B has occurred. Given the outcome space is restricted to B, the likelihood of A is characterized by the conditional probability: $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\,\mathbb{P}(B)$.
- The conditional density function of $\mathbf{z}$ given $\mathbf{y} = \eta$ is

$$f_{\mathbf{z}|\mathbf{y}}(\zeta \mid \mathbf{y} = \eta) = \frac{f_{\mathbf{z},\mathbf{y}}(\zeta, \eta)}{f_{\mathbf{y}}(\eta)}.$$

- $f_{\mathbf{z}|\mathbf{y}}(\zeta \mid \mathbf{y} = \eta)$ is clearly non-negative. Also

$$\int_{\mathbb{R}^d} f_{\mathbf{z}|\mathbf{y}}(\zeta \mid \mathbf{y} = \eta) d\zeta = \frac{1}{f_{\mathbf{y}}(\eta)} \int_{\mathbb{R}^d} f_{\mathbf{z},\mathbf{y}}(\zeta, \eta) d\zeta = \frac{1}{f_{\mathbf{y}}(\eta)} f_{\mathbf{y}}(\eta) = 1.$$

That is, $f_{\mathbf{z}|\mathbf{y}}(\zeta \mid \mathbf{y} = \eta)$ is a legitimate density function.

- Given the conditional density function $f_{\mathbf{z}|\mathbf{y}}$, for $A \in \mathcal{B}^d$,

$$\mathbb{P}(\mathbf{z} \in A \mid \mathbf{y} = \eta) = \int_A f_{\mathbf{z}|\mathbf{y}}(\zeta \mid \mathbf{y} = \eta) d\zeta.$$

  This probability is defined even when $\mathbb{P}(\mathbf{y} = \eta)$ is zero.

- When $A = (-\infty, \zeta_1] \times \cdots \times (-\infty, \zeta_d]$, the conditional distribution function is

$$F_{\mathbf{z}|\mathbf{y}}(\zeta \mid \mathbf{y} = \eta) = \mathbb{P}(z_1 \le \zeta_1, \ldots, z_d \le \zeta_d \mid \mathbf{y} = \eta).$$

- When $\mathbf{z}$ and $\mathbf{y}$ are independent, the conditional density (distribution) reduces to the unconditional density (distribution).

- Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, the conditional expectation $\mathbb{E}(\mathbf{z} \mid \mathcal{G})$ is the integrable and $\mathcal{G}$-measurable random variable satisfying

$$\int_G \mathbb{E}(\mathbf{z} \mid \mathcal{G}) \, \mathrm{d}\,\mathbb{P} = \int_G \mathbf{z} \, \mathrm{d}\,\mathbb{P}, \quad \forall G \in \mathcal{G}.$$

- Suppose that $\mathcal{G}$ is the trivial σ-algebra $\{\Omega, \emptyset\}$, then $\mathbb{E}(\mathbf{z} \mid \mathcal{G})$ must be a constant c, so that

$$\mathbb{E}(\mathbf{z}) = \int_\Omega \mathbf{z} \, \mathrm{d}\,\mathbb{P} = \int_\Omega \mathbf{c} \, \mathrm{d}\,\mathbb{P} = \mathbf{c}.$$

- Consider $\mathcal{G} = \sigma(\mathbf{y})$, the σ-algebra generated by $\mathbf{y}$.

$$\mathbb{E}(\mathbf{z} \mid \mathbf{y}) = \mathbb{E}[\mathbf{z} \mid \sigma(\mathbf{y})],$$

  which is interpreted as the prediction of $\mathbf{z}$ given all the information associated with $\mathbf{y}$.

By definition,

$$\mathbb{E}[\mathbb{E}(\mathbf{z} \mid \mathcal{G})] = \int_\Omega \mathbb{E}(\mathbf{z} \mid \mathcal{G}) \, \mathrm{d} \, \mathbb{P} = \int_\Omega \mathbf{z} \, \mathrm{d} \, \mathbb{P} = \mathbb{E}(\mathbf{z});$$

That is, only a smaller σ-algebra matters in conditional expectation.

Lemma 5.9 (Law of Iterated Expectations)

Let $\mathcal{G}$ and $\mathcal{H}$ be two sub-σ-algebras of $\mathcal{F}$ such that $\mathcal{G} \subseteq \mathcal{H}$. Then for the integrable random vector $\mathbf{z}$,

$$\mathbb{E}[\mathbb{E}(\mathbf{z} \mid \mathcal{H}) \mid \mathcal{G}] = \mathbb{E}[\mathbb{E}(\mathbf{z} \mid \mathcal{G}) \mid \mathcal{H}] = \mathbb{E}(\mathbf{z} \mid \mathcal{G}).$$

- If $\mathbf{z}$ is $\mathcal{G}$-measurable, then $\mathbb{E}[g(\mathbf{z})\mathbf{x} \mid \mathcal{G}] = g(\mathbf{z}) \, \mathbb{E}(\mathbf{x} \mid \mathcal{G})$ with prob. 1.

Lemma 5.11

Let z be a square integrable random variable. Then

$$\mathbb{E}[z - \mathbb{E}(z \mid \mathcal{G})]^2 \leq \mathbb{E}(z - \tilde{z})^2,$$

for any $\mathcal{G}$-measurable random variable $\tilde{z}$.

Proof: For any square integrable, $\mathcal{G}$-measurable random variable $\tilde{z}$,

$$\mathbb{E}\big([z - \mathbb{E}(z \mid \mathcal{G})]\tilde{z}\big) = \mathbb{E}\big([\mathbb{E}(z \mid \mathcal{G}) - \mathbb{E}(z \mid \mathcal{G})]\tilde{z}\big) = 0.$$

It follows that

$$\begin{aligned} \mathbb{E}(z - \tilde{z})^2 &= \mathbb{E}[z - \mathbb{E}(z \mid \mathcal{G}) + \mathbb{E}(z \mid \mathcal{G}) - \tilde{z}]^2 \\ &= \mathbb{E}[z - \mathbb{E}(z \mid \mathcal{G})]^2 + \mathbb{E}[\mathbb{E}(z \mid \mathcal{G}) - \tilde{z}]^2 \\ &\geq \mathbb{E}[z - \mathbb{E}(z \mid \mathcal{G})]^2. \quad \square \end{aligned}$$

- The conditional variance-covariance matrix of $\mathbf{z}$ given $\mathbf{y}$ is

$$\begin{aligned}\operatorname{var}(\mathbf{z} \mid \mathbf{y}) &= \mathbb{E}\big([\mathbf{z} - \mathbb{E}(\mathbf{z} \mid \mathbf{y})][\mathbf{z} - \mathbb{E}(\mathbf{z} \mid \mathbf{y})]' \mid \mathbf{y}\big) \\ &= \mathbb{E}(\mathbf{z}\mathbf{z}' \mid \mathbf{y}) - \mathbb{E}(\mathbf{z} \mid \mathbf{y})\, \mathbb{E}(\mathbf{z} \mid \mathbf{y})',\end{aligned}$$

  which leads to decomposition of *analysis of variance*:

$$\operatorname{var}(\mathbf{z}) = \mathbb{E}[\operatorname{var}(\mathbf{z} \mid \mathbf{y})] + \operatorname{var}\big(\mathbb{E}(\mathbf{z} \mid \mathbf{y})\big).$$

- **Example 5.12**: Suppose that

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\mathbf{y}} \\ \boldsymbol{\mu}_{\mathbf{x}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{yy}} & \boldsymbol{\Sigma}_{\mathbf{xy}}' \\ \boldsymbol{\Sigma}_{\mathbf{xy}} & \boldsymbol{\Sigma}_{\mathbf{xx}} \end{bmatrix} \right).$$

  Then,

$$\begin{aligned}\mathbb{E}(\mathbf{y} \mid \mathbf{x}) &= \boldsymbol{\mu}_{\mathbf{y}} - \boldsymbol{\Sigma}_{\mathbf{xy}}'\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}), \\ \operatorname{var}(\mathbf{y} \mid \mathbf{x}) &= \operatorname{var}(\mathbf{y}) - \operatorname{var}\big(\mathbb{E}(\mathbf{y} \mid \mathbf{x})\big) = \boldsymbol{\Sigma}_{\mathbf{yy}} - \boldsymbol{\Sigma}_{\mathbf{xy}}'\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\Sigma}_{\mathbf{xy}}.\end{aligned}$$

Almost Sure Convergence

A sequence of random variables, $\{z_n(\cdot)\}_{n=1,2,\ldots}$, is such that for a given ω, $z_n(\omega)$ is a realization of the random element ω with index n, and that for a given n, $z_n(\cdot)$ is a random variable.

Almost Sure Convergence

Suppose $\{z_n\}$ and z are all defined on $(\Omega, \mathcal{F}, \mathbb{P})$. $\{z_n\}$ is said to converge to z almost surely if, and only if,

$$\mathbb{P}(\omega\colon z_n(\omega) \to z(\omega) \text{ as } n \to \infty) = 1,$$

denoted as $z_n \xrightarrow{\text{a.s.}} z$ or $z_n \to z$ a.s. (with prob. 1).

Lemma 5.13

Let $g\colon \mathbb{R} \mapsto \mathbb{R}$ be a function continuous on $S_g \subseteq \mathbb{R}$. If $z_n \xrightarrow{\text{a.s.}} z$, where z is a random variable such that $\mathbb{P}(z \in S_g) = 1$, then $g(z_n) \xrightarrow{\text{a.s.}} g(z)$.

Proof: Let $\Omega_0 = \{\omega\colon z_n(\omega) \to z(\omega)\}$ and $\Omega_1 = \{\omega\colon z(\omega) \in S_g\}$. Thus, for $\omega \in (\Omega_0 \cap \Omega_1)$, continuity of g ensures that $g(z_n(\omega)) \to g(z(\omega))$. Note that

$$(\Omega_0 \cap \Omega_1)^c = \Omega_0^c \cup \Omega_1^c,$$

which has probability zero because $\mathbb{P}(\Omega_0^c) = \mathbb{P}(\Omega_1^c) = 0$. (Why?) It follows that $\Omega_0 \cap \Omega_1$ has probability one, showing that $g(z_n) \to g(z)$ with probability one. □

Convergence in Probability

Convergence in Probability

$\{z_n\}$ is said to converge to z in probability if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\omega\colon |z_n(\omega) - z(\omega)| > \epsilon) = 0,$$

or equivalently, $\lim_{n\to\infty} \mathbb{P}(\omega\colon |z_n(\omega) - z(\omega)| \le \epsilon) = 1$. This is denoted as $z_n \xrightarrow{\mathbb{P}} z$ or $z_n \to z$ in probability.

Note: In this definition, the events $\Omega_n(\epsilon) = \{\omega\colon |z_n(\omega) - z(\omega)| \le \epsilon\}$ may vary with n, and convergence is referred to the probability of such events: $p_n = \mathbb{P}(\Omega_n(\epsilon))$, rather than the random variables z_n.

Almost sure convergence implies convergence in probability.

To see this, let Ω_0 denote the set of ω such that $z_n(\omega) \to z(\omega)$. For $\omega \in \Omega_0$, there is some m such that ω is in $\Omega_n(\epsilon)$ for all $n > m$. That is,

$$\Omega_0 \subseteq \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \Omega_n(\epsilon) \in \mathcal{F}.$$

As $\cap_{n=m}^{\infty} \Omega_n(\epsilon)$ is non-decreasing in m, it follows that

$$\begin{aligned} \mathbb{P}(\Omega_0) &\leq \mathbb{P}\left(\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \Omega_n(\epsilon) \right) \\ &= \lim_{m \to \infty} \mathbb{P}\left(\bigcap_{n=m}^{\infty} \Omega_n(\epsilon) \right) \leq \lim_{m \to \infty} \mathbb{P}\big(\Omega_m(\epsilon)\big). \end{aligned}$$

- **Example 5.15**

  Let $\Omega = [0, 1]$ and $\mathbb{P}$ be the Lebesgue measure. Consider the sequence of intervals $\{I_n\}$ in $[0, 1]$: $[0, 1/2)$, $[1/2, 1]$, $[0, 1/3)$, $[1/3, 2/3)$, $[2/3, 1]$, $\ldots$, and let $z_n = \mathbf{1}_{I_n}$. When n tends to infinity, I_n shrinks toward a singleton. For $0 < \epsilon < 1$, we have

  $$\mathbb{P}(|z_n| > \epsilon) = \mathbb{P}(I_n) \to 0,$$

  which shows $z_n \xrightarrow{\mathbb{P}} 0$. On the other hand, each $\omega \in [0, 1]$ must be covered by infinitely many intervals, so that $z_n(\omega) = 1$ for infinitely many n. This shows that $z_n(\omega)$ does not converge to zero. □

  Note: Convergence in probability permits z_n to deviate from the probability limit infinitely often, but almost sure convergence does not, except for those ω in the set of probability zero.

Lemma 5.16

Let $\{z_n\}$ be a sequence of square integrable random variables. If $\mathbb{E}(z_n) \to c$ and $\mathrm{var}(z_n) \to 0$, then $z_n \xrightarrow{\mathbb{P}} c$.

Lemma 5.17

Let $g\colon \mathbb{R} \mapsto \mathbb{R}$ be a function continuous on $S_g \subseteq \mathbb{R}$. If $z_n \xrightarrow{\mathbb{P}} z$, where z is a random variable such that $\mathbb{P}(z \in S_g) = 1$, then $g(z_n) \xrightarrow{\mathbb{P}} g(z)$.

Proof: By the continuity of g, for each $\epsilon > 0$, we can find a $\delta > 0$ s.t.

$$\{\omega : |z_n(\omega) - z(\omega)| \leq \delta\} \cap \{\omega : z(\omega) \in S_g\} \subseteq \{\omega : |g(z_n(\omega)) - g(z(\omega))| \leq \epsilon\}.$$

Taking complementation of both sides, we have

$$\mathbb{P}(|g(z_n) - g(z)| > \epsilon) \leq \mathbb{P}(|z_n - z| > \delta) \to 0.$$

Lemma 5.13 and Lemma 5.17 are readily generalized to $\mathbb{R}^d$-valued random variables. For instance, $\mathbf{z}_n \xrightarrow{\text{a.s.}} \mathbf{z}$ ($\mathbf{z}_n \xrightarrow{\mathbb{P}} \mathbf{z}$) implies

$$z_{1,n} + z_{2,n} \xrightarrow{\text{a.s.}} (\xrightarrow{\mathbb{P}})\, z_1 + z_2,$$

$$z_{1,n} z_{2,n} \xrightarrow{\text{a.s.}} (\xrightarrow{\mathbb{P}})\, z_1 z_2,$$

$$z_{1,n}^2 + z_{2,n}^2 \xrightarrow{\text{a.s.}} (\xrightarrow{\mathbb{P}})\, z_1^2 + z_2^2,$$

where $z_{1,n}, z_{2,n}$ are two elements of $\mathbf{z}_n$ and z_1, z_2 are the corresponding elements of $\mathbf{z}$. Also, provided that $z_2 \neq 0$ with probability one,

$$z_{1,n}/z_{2,n} \xrightarrow{\text{a.s.}} (\xrightarrow{\mathbb{P}})\, z_1/z_2.$$

Convergence in Distribution

$\{z_n\}$ is said to converge to z in distribution, denoted as $z_n \xrightarrow{D} z$, if

$$\lim_{n\to\infty} F_{z_n}(\zeta) = F_z(\zeta),$$

for every continuity point ζ of F_z.

- We also say that z_n is asymptotically distributed as F_z, denoted as $z_n \overset{A}{\sim} F_z$; F_z is thus known as the limiting distribution of z_n.
- Cramér-Wold Device. Let $\{\mathbf{z}_n\}$ be a sequence of random vectors in $\mathbb{R}^d$. Then $\mathbf{z}_n \xrightarrow{D} \mathbf{z}$ if and only if $\boldsymbol{\alpha}'\mathbf{z}_n \xrightarrow{D} \boldsymbol{\alpha}'\mathbf{z}$ for every $\boldsymbol{\alpha} \in \mathbb{R}^d$ such that $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$.

Lemma 5.19

If $z_n \xrightarrow{\mathbb{P}} z$, then $z_n \xrightarrow{D} z$. For a constant c, $z_n \xrightarrow{\mathbb{P}} c$ iff $z_n \xrightarrow{D} c$.

Proof: For some arbitrary $\epsilon > 0$ and a continuity point ζ of F_z, we have

$$\begin{aligned}
&\mathbb{P}(z_n \leq \zeta) = \\
&\quad \mathbb{P}(\{z_n \leq \zeta\} \cap \{|z_n - z| \leq \epsilon\}) + \mathbb{P}(\{z_n \leq \zeta\} \cap \{|z_n - z| > \epsilon\}) \\
&\qquad \leq \mathbb{P}(z \leq \zeta + \epsilon) + \mathbb{P}(|z_n - z| > \epsilon).
\end{aligned}$$

Similarly, $\mathbb{P}(z \leq \zeta - \epsilon) \leq \mathbb{P}(z_n \leq \zeta) + \mathbb{P}(|z_n - z| > \epsilon)$. If $z_n \xrightarrow{\mathbb{P}} z$, then by passing to the limit and noting that ϵ is arbitrary,

$$\lim_{n \to \infty} \mathbb{P}(z_n \leq \zeta) = \mathbb{P}(z \leq \zeta).$$

That is, $F_{z_n}(\zeta) \to F_z(\zeta)$. The converse is not true in general, however.

Theorem 5.20 (Continuous Mapping Theorem)

Let $g\colon \mathbb{R} \mapsto \mathbb{R}$ be a function continuous almost everywhere on $\mathbb{R}$, except for at most countably many points. If $z_n \xrightarrow{D} z$, then $g(z_n) \xrightarrow{D} g(z)$.

For example, $z_n \xrightarrow{D} \mathcal{N}(0,1)$ implies $z_n^2 \xrightarrow{D} \chi^2(1)$.

Theorem 5.21

Let $\{y_n\}$ and $\{z_n\}$ be two sequences of random vectors such that $y_n - z_n \xrightarrow{\mathbb{P}} 0$. If $z_n \xrightarrow{D} z$, then $y_n \xrightarrow{D} z$.

Theorem 5.22

If y_n converges in probability to a constant c and z_n converges in distribution to z, then $y_n + z_n \xrightarrow{D} c + z$, $y_n z_n \xrightarrow{D} cz$, and $z_n/y_n \xrightarrow{D} z/c$ if $c \neq 0$.

Non-Stochastic Order Notations

Order notations are used to describe the behavior of real sequences.

- b_n is (at most) of order c_n, denoted as $b_n = O(c_n)$, if there exists a $\Delta < \infty$ such that $|b_n|/c_n \leq \Delta$ for all sufficiently large n.
- b_n is of smaller order than c_n, denoted as $b_n = o(c_n)$, if $b_n/c_n \to 0$.
- An $O(1)$ sequence in bounded; an $o(1)$ sequence converges to zero. The product of $O(1)$ and $o(1)$ sequences is $o(1)$.

Theorem 5.23

(a) If $a_n = O(n^r)$ and $b_n = O(n^s)$, then $a_n b_n = O(n^{r+s})$, $a_n + b_n = O(n^{\max(r,s)})$.

(b) If $a_n = o(n^r)$ and $b_n = o(n^s)$, then $a_n b_n = o(n^{r+s})$, $a_n + b_n = o(n^{\max(r,s)})$.

(c) If $a_n = O(n^r)$ and $b_n = o(n^s)$, then $a_n b_n = o(n^{r+s})$, $a_n + b_n = O(n^{\max(r,s)})$.

Stochastic Order Notations

The order notations defined earlier easily extend to describe the behavior of sequences of random variables.

- $\{z_n\}$ is $O_{\text{a.s.}}(c_n)$ (or $O(c_n)$ almost surely) if z_n/c_n is $O(1)$ a.s.
- $\{z_n\}$ is $O_{\mathbb{P}}(c_n)$ (or $O(c_n)$ in probability) if for every $\epsilon > 0$, there is some Δ such that $\mathbb{P}(|z_n|/c_n \geq \Delta) \leq \epsilon$, for all n sufficiently large.
- Lemma 5.23 holds for stochastic order notations. For example, $y_n = O_{\mathbb{P}}(1)$ and $z_n = o_{\mathbb{P}}(1)$, then $y_n z_n$ is $o_{\mathbb{P}}(1)$.
- It is very restrictive to require a random variable being bounded almost surely, but a well defined random variable is typically bounded in probability, i.e., $O_{\mathbb{P}}(1)$.

Let $\{z_n\}$ be a sequence of random variables such that $z_n \xrightarrow{D} z$ and ζ be a continuity point of F_z. Then for any $\epsilon > 0$, we can choose a sufficiently large ζ such that $\mathbb{P}(|z| > \zeta) < \epsilon/2$. As $z_n \xrightarrow{D} z$, we can also choose n large enough such that

$$\mathbb{P}(|z_n| > \zeta) - \mathbb{P}(|z| > \zeta) < \epsilon/2,$$

which implies $\mathbb{P}(|z_n| > \zeta) < \epsilon$. We have proved:

Lemma 5.24

Let $\{z_n\}$ be a sequence of random variables such that $z_n \xrightarrow{D} z$. Then $z_n = O_{\mathbb{P}}(1)$.

Law of Large Numbers

- When a law of large numbers holds almost surely, it is a strong law of large numbers (SLLN); when a law of large numbers holds in probability, it is a weak law of large numbers (WLLN).
- A sequence of random variables obeys a LLN when its sample average essentially follows its mean behavior; random irregularities (deviations from the mean) are "wiped out" in the limit by averaging.
- Kolmogorov's SLLN : Let $\{z_t\}$ be a sequence of *i.i.d.* random variables with mean μ_o. Then, $T^{-1} \sum_{t=1}^{T} z_t \xrightarrow{\text{a.s.}} \mu_o$.
- Note that i.i.d. random variables need not obey Kolmogorov's SLLN if they do not have a finite mean, e.g., i.i.d. Cauchy random variables.

Theorem 5.26 (Markov's SLLN)

Let $\{z_t\}$ be a sequence of independent random variables such that for some $\delta > 0$, $\mathbb{E}\,|z_t|^{1+\delta}$ is bounded for all t. Then,

$$\frac{1}{T}\sum_{t=1}^{T}[z_t - \mathbb{E}(z_t)] \xrightarrow{\text{a.s.}} 0.$$

- Note that here z_t need not have a common mean, and the average of their means need not converge.
- Compared with Kolmogorov's SLLN, Markov's SLLN requires a stronger moment condition but not identical distribution.
- A LLN usually obtains by regulating the moments of and dependence across random variables.

Example 5.27 Suppose that $y_t = \alpha_o y_{t-1} + u_t$ with $|\alpha_o| < 1$. Then, $\operatorname{var}(y_t) = \sigma_u^2/(1-\alpha_o^2)$, and $\operatorname{cov}(y_t, y_{t-j}) = \alpha_o^j \frac{\sigma_u^2}{1-\alpha_o^2}$. Thus,

$$\begin{aligned}\operatorname{var}\left(\sum_{t=1}^{T} y_t\right) &= \sum_{t=1}^{T} \operatorname{var}(y_t) + 2\sum_{\tau=1}^{T-1} (T-\tau)\operatorname{cov}(y_t, y_{t-\tau}) \\ &\leq \sum_{t=1}^{T} \operatorname{var}(y_t) + 2T\sum_{\tau=1}^{T-1} |\operatorname{cov}(y_t, y_{t-\tau})| = O(T),\end{aligned}$$

so that $\operatorname{var}\left(T^{-1}\sum_{t=1}^{T} y_t\right) = O(T^{-1})$. As $\mathbb{E}(T^{-1}\sum_{t=1}^{T} y_t) = 0$,

$$\frac{1}{T}\sum_{t=1}^{T} y_t \xrightarrow{\mathbb{P}} 0.$$

by Lemma 5.16. That is, $\{y_t\}$ obeys a WLLN.

Lemma 5.28

Let $y_t = \sum_{i=0}^{\infty} \pi_i u_{t-i}$, where u_t are i.i.d. random variables with mean zero and variance σ_u^2. If $\sum_{i=-\infty}^{\infty} |\pi_i| < \infty$, then $T^{-1} \sum_{t=1}^{T} y_t \xrightarrow{\text{a.s.}} 0$.

- In Example 5.27, $y_t = \sum_{i=0}^{\infty} \alpha_o^i u_{t-i}$ with $|\alpha_o| < 1$, so that $\sum_{i=0}^{\infty} |\alpha_o^i| < \infty$
- Lemma 5.28 is quite general and applicable to processes that can be expressed as an MA process with absolutely summable weights, e.g., weakly stationary AR(p) processes.
- For random variables with strong correlations over time, the variation of their partial sums may grow too rapidly and cannot be eliminated by simple averaging.

Example 5.29: For the sequences $\{t\}$ and $\{t^2\}$,

$$\sum_{t=1}^{T} t = T(T+1)/2, \quad \sum_{t=1}^{T} t^2 = T(T+1)(2T+1)/6.$$

Hence, $T^{-1}\sum_{t=1}^{T} t$ and $T^{-1}\sum_{t=1}^{T} t^2$ both diverge.

Example 5.30: u_t are i.i.d. with mean zero and variance σ_u^2. Consider now $\{tu_t\}$, which does not have bounded $(1+\delta)$ th moment and does not obey Markov's SLLN. Moreover,

$$\operatorname{var}\left(\sum_{t=1}^{T} tu_t\right) = \sum_{t=1}^{T} t^2 \operatorname{var}(u_t) = \sigma_u^2 \frac{T(T+1)(2T+1)}{6},$$

so that $\sum_{t=1}^{T} tu_t = O_{\mathbb{P}}(T^{3/2})$. It follows that $T^{-1}\sum_{t=1}^{T} tu_t = O_{\mathbb{P}}(T^{1/2})$. That is, $\{tu_t\}$ does not obey a WLLN.

Example 5.31: y_t is a random walk: $y_t = y_{t-1} + u_t$. For $s < t$,

$$y_t = y_s + \sum_{i=s+1}^{t} u_i = y_s + v_{t-s},$$

where v_{t-s} is independent of y_s and $\mathrm{cov}(y_t, y_s) = \mathbb{E}(y_s^2) = s\sigma_u^2$. Thus,

$$\mathrm{var}\left(\sum_{t=1}^{T} y_t\right) = \sum_{t=1}^{T} \mathrm{var}(y_t) + 2\sum_{\tau=1}^{T-1} \sum_{t=\tau+1}^{T} \mathrm{cov}(y_t, y_{t-\tau}) = O(T^3),$$

for $\sum_{t=1}^{T} \mathrm{var}(y_t) = \sum_{t=1}^{T} t\sigma_u^2 = O(T^2)$ and

$$2\sum_{\tau=1}^{T-1} \sum_{t=\tau+1}^{T} \mathrm{cov}(y_t, y_{t-\tau}) = 2\sum_{\tau=1}^{T-1} \sum_{t=\tau+1}^{T} (t-\tau)\sigma_u^2 = O(T^3).$$

Then, $\sum_{t=1}^{T} y_t = O_{\mathbb{P}}(T^{3/2})$ and $T^{-1}\sum_{t=1}^{T} y_t$ diverges in probability.

Example 5.32: y_t is the random walk in Example 5.31. Then, $\mathbb{E}(y_{t-1}u_t) = 0$, $\operatorname{var}(y_{t-1}u_t) = \mathbb{E}(y_{t-1}^2)\,\mathbb{E}(u_t^2) = (t-1)\sigma_u^4$, and for $s < t$,

$$\operatorname{cov}(y_{t-1}u_t, y_{s-1}u_s) = \mathbb{E}(y_{t-1}y_{s-1}u_s)\,\mathbb{E}(u_t) = 0.$$

This yields

$$\operatorname{var}\left(\sum_{t=1}^{T} y_{t-1}u_t\right) = \sum_{t=1}^{T} \operatorname{var}(y_{t-1}u_t) = \sum_{t=1}^{T}(t-1)\sigma_u^4 = O(T^2),$$

and $\sum_{t=1}^{T} y_{t-1}u_t = O_{\mathbb{P}}(T)$. As $\operatorname{var}(T^{-1}\sum_{t=1}^{T} y_{t-1}u_t)$ converges to $\sigma_u^4/2$, rather than 0, $\{y_{t-1}u_t\}$ does not obey a WLLN, even though its partial sums are $O_{\mathbb{P}}(T)$.

Central Limit Theorem (CLT)

Lemma 5.35 (Lindeberg-Lévy's CLT)

Let $\{z_t\}$ be a sequence of i.i.d. random variables with mean μ_o and variance $\sigma_o^2 > 0$. Then, $\sqrt{T}(\bar{z}_T - \mu_o)/\sigma_o \xrightarrow{D} \mathcal{N}(0, 1)$.

- i.i.d. random variables need not obey this CLT if they do not have a finite variance, e.g., $t(2)$ r.v.
- Note that $\bar{z}_T$ converges to μ_o in probability, and its variance σ_o^2/T vanishes when T tends to infinity. A normalizing factor $T^{1/2}$ suffices to prevent a degenerate distribution in the limit.
- When $\{z_t\}$ obeys a CLT, $\bar{z}_T$ is said to converge to μ_o at the rate $T^{-1/2}$, and $\bar{z}_T$ is understood as a root-T consistent estimator.

Lemma 5.36 (Liapunov's CLT)

Let $\{z_{Tt}\}$ be a triangular array of independent random variables with mean μ_{Tt} and variance $\sigma^2_{Tt} > 0$ such that $\bar{\sigma}^2_T = \frac{1}{T}\sum_{t=1}^{T} \sigma^2_{Tt} \to \sigma^2_o > 0$. If for some $\delta > 0$, $\mathbb{E}\,|z_{Tt}|^{2+\delta}$ are bounded, then $\sqrt{T}(\bar{z}_T - \bar{\mu}_T)/\sigma_o \xrightarrow{D} \mathcal{N}(0,1)$.

- A CLT usually requires stronger conditions on the moment of and dependence across random variables than those needed to ensure a LLN.
- Moreover, every random variable must also be asymptotically negligible, in the sense that no random variable is influential in affecting the partial sums.

Example 5.37: $\{u_t\}$ is a sequence of independent random variables with mean zero, variance σ_u^2, and bounded $(2+\delta)$ th moment. we know $\mathrm{var}(\sum_{t=1}^T tu_t)$ is $O(T^3)$, which implies that variance of $T^{-1/2}\sum_{t=1}^T tu_t$ is diverging at the rate $O(T^2)$. On the other hand, observe that

$$\mathrm{var}\left(\frac{1}{T^{1/2}}\sum_{t=1}^T \frac{t}{T}u_t\right) = \frac{T(T+1)(2T+1)}{6T^3}\sigma_u^2 \to \frac{\sigma_u^2}{3}.$$

It follows that

$$\frac{\sqrt{3}}{T^{1/2}\sigma_u}\sum_{t=1}^T \frac{t}{T}u_t \xrightarrow{D} \mathcal{N}(0,1).$$

These results show that $\{(t/T)u_t\}$ obeys a CLT, whereas $\{tu_t\}$ does not.

Example 5.38: y_t is a random walk: $y_t = y_{t-1} + u_t$, where u_t are i.i.d. with mean zero and variance σ_u^2. We know y_t do not obey a LLN and hence do not obey a CLT.

CLT for Triangular Array

$\{z_{Tt}\}$ is a triangular array of random variables and obeys a CLT if

$$\frac{1}{\sigma_o \sqrt{T}} \sum_{t=1}^{T} [z_{Tt} - \mathbb{E}(z_{Tt})] = \frac{\sqrt{T}(\bar{z}_T - \bar{\mu}_T)}{\sigma_o} \xrightarrow{D} \mathcal{N}(0,1),$$

where $\bar{z}_T = T^{-1} \sum_{t=1}^{T} z_{Tt}$, $\bar{\mu}_T = \mathbb{E}(\bar{z}_T)$, and

$$\sigma_T^2 = \text{var}\left(T^{-1/2} \sum_{t=1}^{T} z_{Tt} \right) \to \sigma_o^2 > 0.$$

- Consider an array of square integrable random vectors $\mathbf{z}_{Tt}$ in $\mathbb{R}^d$. Let $\bar{\mathbf{z}}_T$ denote the average of $\mathbf{z}_{Tt}$, $\bar{\boldsymbol{\mu}}_T = \mathbb{E}(\bar{\mathbf{z}}_T)$, and

$$\boldsymbol{\Sigma}_T = \operatorname{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{z}_{Tt}\right) \to \boldsymbol{\Sigma}_o,$$

a positive definite matrix. Using the Cramér-Wold device, $\{\mathbf{z}_{Tt}\}$ is said to obey a multivariate CLT, in the sense that

$$\boldsymbol{\Sigma}_o^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\mathbf{z}_{Tt} - \mathbb{E}(\mathbf{z}_{Tt})] = \boldsymbol{\Sigma}_o^{-1/2}\sqrt{T}(\bar{\mathbf{z}}_T - \bar{\boldsymbol{\mu}}_T) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{I}_d),$$

if $\{\boldsymbol{\alpha}'\mathbf{z}_{Tt}\}$ obeys a CLT, for any $\boldsymbol{\alpha} \in \mathbb{R}^d$ such that $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$.

Stochastic Processes

- A d-dimensional stochastic process with the index set $\mathcal{T}$ is a measurable mapping $\mathbf{z}\colon \Omega \mapsto (\mathbb{R}^d)^{\mathcal{T}}$ such that

$$\mathbf{z}(\omega) = \{\mathbf{z}_t(\omega),\ t \in \mathcal{T}\}.$$

  For each $t \in \mathcal{T}$, $\mathbf{z}_t(\cdot)$ is a $\mathbb{R}^d$-valued r.v.; for each ω, $\mathbf{z}(\omega)$ is a sample path (realization) of $\mathbf{z}$, a $\mathbb{R}^d$-valued function on $\mathcal{T}$.

- The finite-dimensional distributions of $\{\mathbf{z}(t,\cdot),\ t \in \mathcal{T}\}$ is

$$\mathbb{P}(\mathbf{z}_{t_1} \leq \mathbf{a}_1, \ldots, \mathbf{z}_{t_n} \leq \mathbf{a}_n) = F_{t_1,\ldots,t_n}(\mathbf{a}_1, \ldots, \mathbf{a}_n).$$

  - $\mathbf{z}$ is stationary if $F_{t_1,\ldots,t_n}$ are invariant under index displacement.
  - $\mathbf{z}$ is Gaussian if $F_{t_1,\ldots,t_n}$ are all (multivariate) normal.

Brownian motion

The process $\{w(t),\ t \in [0, \infty)\}$ is the standard Wiener process (standard Brownian motion) if it has continuous sample paths almost surely and satisfies:

1. $\mathbb{P}\big(w(0) = 0\big) = 1$.
2. For $0 \le t_0 \le t_1 \le \cdots \le t_k$,

   $$\mathbb{P}\big(w(t_i) - w(t_{i-1}) \in B_i,\ i \le k\big) = \prod_{i \le k} \mathbb{P}\big(w(t_i) - w(t_{i-1}) \in B_i\big),$$

   where B_i are Borel sets.
3. For $0 \le s < t$, $w(t) - w(s) \sim \mathcal{N}(0, t - s)$.

Note: w here has independent and Gaussian increments.

- $w(t) \sim \mathcal{N}(0, t)$ such that for $r \leq t$,

$$\operatorname{cov}\big(w(r),\, w(t)\big) = \mathbb{E}\big[w(r)\big(w(t) - w(r)\big)\big] + \mathbb{E}\big[w(r)^2\big] = r.$$

- The sample paths of w are a.s. continuous but highly irregular (nowhere differentiable).

  To see this, note $w_c(t) = w(c^2 t)/c$ for $c > 0$ is also a standard Wiener process. (Why?) Then, $w_c(1/c) = w(c)/c$. For a large c such that $w(c)/c > 1$, $\frac{w_c(1/c)}{1/c} = w(c) > c$. That is, the sample path of w_c has a slope larger than c on a very small interval $(0, 1/c)$.

- The difference quotient:

$$[w(t+h) - w(t)]/h \sim \mathcal{N}(0,\ 1/|h|)$$

  can not converge to a finite limit (as $h \to 0$) with a positive prob.

The d-dimensional, standard Wiener process $\mathbf{w}$ consists of d mutually independent, standard Wiener processes, so that for $s < t$, $\mathbf{w}(t) - \mathbf{w}(s) \sim \mathcal{N}(\mathbf{0},\ (t-s)\,\mathbf{I}_d)$.

Lemma 5.39

Let $\mathbf{w}$ be the d-dimensional, standard Wiener process.

1. $\mathbf{w}(t) \sim \mathcal{N}(\mathbf{0},\ t\,\mathbf{I}_d)$.
2. $\text{cov}(\mathbf{w}(r), \mathbf{w}(t)) = \min(r, t)\,\mathbf{I}_d$.

The Brownian bridge $\mathbf{w}^0$ on $[0, 1]$ is $\mathbf{w}^0(t) = \mathbf{w}(t) - t\mathbf{w}(1)$. Clearly, $\mathbb{E}[\mathbf{w}^0(t)] = \mathbf{0}$, and for $r < t$,

$$\text{cov}\big(\mathbf{w}^0(r), \mathbf{w}^0(t)\big) = \text{cov}\big(\mathbf{w}(r) - r\mathbf{w}(1), \mathbf{w}(t) - t\mathbf{w}(1)\big) = r(1-t)\,\mathbf{I}_d.$$

Weak Convergence

$\mathbb{P}_n$ converges weakly to $\mathbb{P}$, denoted as $\mathbb{P}_n \Rightarrow \mathbb{P}$, if for every bounded, continuous real function f on S,

$$\int f(s) \, \mathrm{d}\mathbb{P}_n(s) \to \int f(s) \, \mathrm{d}\,\mathbb{P}(s),$$

where $\{\mathbb{P}_n\}$ and $\mathbb{P}$ are probability measures on $(S, \mathcal{S})$.

- When $\mathbf{z}_n$ and $\mathbf{z}$ are all $\mathbb{R}^d$-valued random variables, $\mathbb{P}_n \Rightarrow \mathbb{P}$ reduces to the usual notion of convergence in distribution: $\mathbf{z}_n \xrightarrow{D} \mathbf{z}$.
- When $\mathbf{z}_n$ and $\mathbf{z}$ are d-dimensional stochastic processes with the distributions induced by $\mathbb{P}_n$ and $\mathbb{P}$, $\mathbf{z}_n \xrightarrow{D} \mathbf{z}$, also denoted as $\mathbf{z}_n \Rightarrow \mathbf{z}$, implies that all the finite-dimensional distributions of $\mathbf{z}_n$ converge to the corresponding distributions of $\mathbf{z}$.

Continuous Mapping Theorem

Lemma 5.40 (Continuous Mapping Theorem)

Let $g\colon \mathbb{R}^d \mapsto \mathbb{R}$ be a function continuous almost everywhere on $\mathbb{R}^d$, except for at most countably many points. If $\mathbf{z}_n \Rightarrow \mathbf{z}$, then $g(\mathbf{z}_n) \Rightarrow g(\mathbf{z})$.

Proof: Let S and S' be two metric spaces with Borel σ-algebras $\mathcal{S}$ and $\mathcal{S}'$ and $g\colon S \mapsto S'$ be a measurable mapping. For $\mathbb{P}$ on $(S, \mathcal{S})$, define $\mathbb{P}^*$ on $(S', \mathcal{S}')$ as

$$\mathbb{P}^*(A') = \mathbb{P}(g^{-1}(A')), \qquad A' \in \mathcal{S}'.$$

For every bounded, continuous f on S', $f \circ g$ is also bounded and continuous on S. $\mathbb{P}_n \Rightarrow \mathbb{P}$ now implies that

$$\int f \circ g(s) \,\mathrm{d}\,\mathbb{P}_n(s) \to \int f \circ g(s) \,\mathrm{d}\,\mathbb{P}(s),$$

which is equivalent to $\int f(a) \,\mathrm{d}\,\mathbb{P}_n^*(a) \to \int f(a) \,\mathrm{d}\,\mathbb{P}^*(a)$, proving $\mathbb{P}_n^* \Rightarrow \mathbb{P}^*$.

Functional Central Limit Theorem (FCLT)

- ζ_i are i.i.d. with mean zero and variance σ^2. Let $s_n = \zeta_1 + \cdots + \zeta_n$ and $z_n(i/n) = (\sigma\sqrt{n})^{-1} s_i$.
- For $t \in [(i-1)/n, i/n)$, the constant interpolations of $z_n(i/n)$ is

$$z_n(t) = z_n((i-1)/n) = \frac{1}{\sigma\sqrt{n}} s_{[nt]},$$

where $[nt]$ is the the largest integer less than or equal to nt.

- From Lindeberg-Lévy's CLT,

$$\frac{1}{\sigma\sqrt{n}} s_{[nt]} = \left(\frac{[nt]}{n}\right)^{1/2} \frac{1}{\sigma\sqrt{[nt]}} s_{[nt]} \xrightarrow{D} \sqrt{t}\,\mathcal{N}(0,\ 1),$$

which is just $\mathcal{N}(0,\ t)$, the distribution of $w(t)$.

- For $r < t$, we have

$$(z_n(r), z_n(t) - z_n(r)) \xrightarrow{D} \big(w(r), w(t) - w(r)\big),$$

and hence $(z_n(r), z_n(t)) \xrightarrow{D} (w(r), w(t))$. This is easily extended to establish convergence of any finite-dimensional distributions and leads to the functional central limit theorem.

Lemma 5.41 (Donsker)

Let ζ_t be i.i.d. with mean μ_o and variance $\sigma_o^2 > 0$ and

$$z_T(r) = \frac{1}{\sigma_o\sqrt{T}} \sum_{t=1}^{[Tr]} (\zeta_t - \mu_o), \quad r \in [0, 1].$$

Then, $z_T \Rightarrow w$ as $T \to \infty$.

- Let ζ_t be r.v.s with mean μ_t and variance $\sigma_t^2 > 0$. Define long-run variance of ζ_t as

$$\sigma_*^2 = \lim_{T\to\infty} \operatorname{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\zeta_t\right),$$

$\{\zeta_t\}$ is said to obey an FCLT if $z_T \Rightarrow w$ as $T \to \infty$, where

$$z_T(r) = \frac{1}{\sigma_*\sqrt{T}}\sum_{t=1}^{[Tr]}(\zeta_t - \mu_t), \quad r \in [0,1].$$

- In the multivariate context, FCLT is $\mathbf{z}_T \Rightarrow \mathbf{w}$ as $T \to \infty$, where

$$\mathbf{z}_T(r) = \frac{1}{\sqrt{T}}\boldsymbol{\Sigma}_*^{-1/2}\sum_{t=1}^{[Tr]}(\boldsymbol{\zeta}_t - \boldsymbol{\mu}_t), \quad r \in [0,1],$$

$\mathbf{w}$ is the d-dimensional, standard Wiener process, and

$$\boldsymbol{\Sigma}_* = \lim_{T\to\infty}\frac{1}{T}\,\mathbb{E}\left[\left(\sum_{t=1}^{T}(\boldsymbol{\zeta}_t - \boldsymbol{\mu}_t)\right)\left(\sum_{t=1}^{T}(\boldsymbol{\zeta}_t - \boldsymbol{\mu}_t)\right)'\right],$$

Example 5.43

- $y_t = y_{t-1} + u_t$, $t = 1, 2, \ldots$, with $y_0 = 0$, where u_t are i.i.d. with mean zero and variance σ_u^2.
- By Donsker's FCLT, the partial sum $y_{[Tr]} = \sum_{t=1}^{[Tr]} u_t$ is such that

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} y_t = \sigma_u \sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} \frac{1}{\sqrt{T}\sigma_u} y_{[Tr]} \, \mathrm{d}r \Rightarrow \sigma_u \int_0^1 w(r) \, \mathrm{d}r,$$

- This result also verifies that $\sum_{t=1}^{T} y_t$ is $O_{\mathbb{P}}(T^{3/2})$. Similarly,

$$\frac{1}{T^2} \sum_{t=1}^{T} y_t^2 = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{y_t}{\sqrt{T}} \right)^2 \Rightarrow \sigma_u^2 \int_0^1 w(r)^2 \, \mathrm{d}r,$$

so that $\sum_{t=1}^{T} y_t^2$ is $O_{\mathbb{P}}(T^2)$.