Technical Report RM 703, Department of Statistics and Probability
Michigan State University

Weak Convergence of Stochastic Processes

V. Mandrekar

Abstract

The purpose of this course was to present results on weak convergence and invariance principle with statistical applications. As various techniques used to obtain different statistical applications, I have made an effort to introduce students to embedding technique of Skorokhod in chapter 1 and 5. Most of the material is from the book of Durrett [3]. In chapter 2, we relate this convergence to weak convergence on C [0,1] following the book of Billingsley [1]. In addition, we present the work in [1] on weak convergence on $D[0,1]$ and $D[0,\infty)$ originated in the work of Skorokhod. In particular, we present the interesting theorem of Aldous for determining compactness in $D[0,\infty)$ as given in [1]. This is then exploited in chapter 4 to obtain central limit theorems for continuous semi-martingale due to Lipster and Shiryayev using ideas from the book of Jacod and Shiryayev [5]. As an application of this work we present the work of R. Gill [4], Kaplan-Meier estimate of life distributin with censored data using techniques in [2]. Finally in the last chapter we present the work on empirical processes using recent book of Van der Vaart and Wellner [6].

I thank Mr. J. Kim for taking careful notes and typing them.

Finally, I dedicated these notes to the memory of A. V. Skorokhod from whom I learned a lot.

2 Weak Convergence In Metric Spaces

2.1 Cylindrical Measures

Let $\{X_t, t \in T\}$ be family of random variable on a probability space $(\Omega, \mathcal{F}, P)$. Assume that X_t takes values in $(\mathcal{X}_t, \mathcal{A}_t)$.

For any finite set $S \subset T$,

$$\mathcal{X}_S = \prod_{t \in S} \mathcal{X}_t, \mathcal{A}_S = \bigotimes_{t \in S} \mathcal{A}_t, Q_S = P \circ (X_t, t \in S)^{-1}$$

where Q_S is the induced measure. Check if $\Pi_{S'S} : \mathcal{X}_{S'} \to \mathcal{X}_S$ for $S \subset S'$, then

$$Q_S = Q_{S'} \circ \Pi_{S'S}^{-1} \tag{1}$$

Suppose we are given a family $\{Q_S, S \subset T \text{ finite dimensional}\}$ where Q_S on $(\mathcal{X}_S, \mathcal{A}_S)$. Assume they satisfy (1). Then, there exists probability measure on $(\mathcal{X}_T, \mathcal{A}_T)$ such that

$$Q \circ \Pi_S^{-1} = Q_S$$

where $\mathcal{X}_T = \prod_{t \in T} \mathcal{X}_t, \mathcal{A}_T = \sigma\Big(\bigcup_{S \subset T} \mathcal{C}_S\Big), \mathcal{C}_S = \Pi_S^{-1}(\mathcal{A}_S)$.

Remark. For $S \subset T$ and $C \in \mathcal{C}_S$, define

$$\begin{aligned} Q_0(C) &= Q_S(A) \\ C &= \Pi_S^{-1}(A) \\ \mathcal{C}_S &= \Pi_S^{-1}(\mathcal{A}_S) \end{aligned}$$

We can define Q_0 on $\bigcup_{S \subset T} \mathcal{C}_S$. Then, for $C \in \mathcal{C}_S$ and $\mathcal{C}_{S'}$,

$$Q_0(C) = Q_S(A) = Q_{S'}(A),$$

and hence Q_0 is well-defined.

Note that

$$\mathcal{C}_{S_1} \cup \mathcal{C}_{S_2} \cup \cdots \cup \mathcal{C}_{S_k} \subset \mathcal{C}_{S_1 \cup \cdots \cup S_k}$$

Q_0 is finitely additive on $\bigcup_{S \subset T} \mathcal{C}_S$. We have to show the countable additivity.

Definition 2.1 *A collection of subsets $\mathcal{K} \subset \mathcal{X}$ is called a compact class if for every sequence $\{C_k\} \subset \mathcal{K}$, for all $n < \infty$,*

$$\bigcap_{k=1}^{n} C_k \neq \emptyset \Longrightarrow \bigcap_{k=1}^{\infty} C_k \neq \emptyset$$

Exercise 1. Every subcollection of compact class is compact.

Exercise 2. If $\mathcal{X}$ and $\mathcal{Y}$ are two spaces and $T : \mathcal{X} \to \mathcal{Y}$, and $\mathcal{K}$ is compact class in $\mathcal{Y}$, then $T^{-1}(\mathcal{K})$ is compact class in $\mathcal{X}$.

Definition 2.2 *A finitely additive measure μ on $(\mathcal{X}, \mathcal{A}_0)$ is called compact if there exists a compact class $\mathcal{K}$ such that for every $A \in \mathcal{A}_0$ and $\epsilon > 0$, there exists $C_\epsilon \in \mathcal{K}$, and $A_\epsilon \in \mathcal{A}_0$ such that*

$$A_\epsilon \subset C_\epsilon \subset A \quad \textit{and} \quad \mu(A - A_\epsilon) < \epsilon$$

We call $\mathcal{K}$ is $\mu-$approximable $\mathcal{A}_0$.

Lemma 2.1 *Every compact finitely additive measure is countably additive.*

Proof. Suppose $(\mathcal{X}, \mathcal{A}_0, \mu)$ is given. There exists a compact class $\mathcal{K}$ which is $\mu-$approximable $\mathcal{A}_0$. Let $\{A_n\} \subset \mathcal{A}_0$ such that $A_n \searrow \emptyset$. We need to show that $\mu(A_n) \searrow 0$. For given $\epsilon > 0$, let $B_n \in \mathcal{A}_0$ and $C_n \in \mathcal{K}$ such that

$$B_n \subset C_n \subset A_n, \text{ and } \mu(A_n - B_n) < \frac{\epsilon}{2^n}$$

Suppose $\mu(A_n)$ does not go to 0, i.e., for all n, $\mu(A_n) > \epsilon$. Since we know that

$$\mu\Big(A_n - \bigcap_{k=1}^{n} B_k\Big) = \mu\Big(\bigcap_{k=1}^{n} A_k\Big) - \mu\Big(\bigcap_{k=1}^{n} B_k\Big) < \frac{\epsilon}{2},$$

we conclude that for all n,

$$\mu\Big(\bigcap_{k=1}^{n} B_k\Big) > \frac{\epsilon}{2}$$

Next for all n

$$\bigcap_{k=1}^{n} B_k \neq \emptyset,$$

and hence we have for all n

$$\bigcap_{k=1}^{n} C_k \neq \emptyset,$$

which implies

$$\bigcap_{k=1}^{\infty} C_k \neq \emptyset$$

since $C_k \in \mathcal{K}$, and $\mathcal{K}$ is compact. Therefore, it follows that

$$\bigcap_{k=1}^{\infty} A_k \supset \bigcap_{k=1}^{\infty} C_k \neq \emptyset$$

implies

$$\lim_{n \to \infty} A_n \neq \emptyset,$$

which is a contradiction.

2.2 Kolmogorov Consistency Theorem

In this section we show that, given finite dimensional family satisfying (1.1) for random variables we can construct all of them on the same probability space. Suppose $S \subset T$ is finite subset and Q_S is measure on $(\mathcal{X}_S, \mathcal{A}_S)$ satisfying consistency condition (1-1). Let $(\mathcal{X}_{\{t\}}, \mathcal{A}_{\{t\}}, Q_{\{t\}})$ be compact probability measure space. For each $t \in T$ with compact class $\mathcal{K}_t \subset \mathcal{A}_t$, $Q_{\{t\}}$ approximates $\mathcal{A}_{\{t\}}$. Then, there exists a unique probability measure Q_0 on $(\mathcal{X}_T, \mathcal{A}_T)$ such that

$$\Pi_S : \mathcal{X}_T \to \mathcal{X}_S, \text{ and } Q_0 \circ \Pi_S^{-1} = Q_S$$

Remark. For $(\underbrace{\prod_{t \in T} \mathcal{X}_t}_{\Omega}, \underbrace{\bigotimes_{t \in T} \mathcal{A}_t}_{\mathcal{F}}, Q_0)$ as probability space

$$X_t(\omega) = \omega(t), \quad \bigcup_{S \subset T} \mathcal{C}_S = \mathcal{C}_0$$

Proof)
Define

$$\mathcal{D} = \{C : C = \Pi_t^{-1}(K), K \in \mathcal{K}_t, t \in T\}$$

Let

$$\{\Pi_{t_i}^{-1}(C_{t_i}), i = 1, 2...\}$$

be a countable families of sets and

$$B_t = \bigcup_{t_i = t} \Pi_{t_i}^{-1}(C_{t_i})$$

If the countable intersection is empty, then B_{t_0} is empty for some t_0. Since $\mathcal{K}_{t_0}$ is a compact class and all $C_{t_i} \in \mathcal{K}_{t_0}$, we get a finite set of t_i's ($t_i = t_0$). Let's call it J for which

$$\bigcup_{t_i \in J} C_{t_i} = \emptyset \Longrightarrow \bigcup_{t_i \in J} \Pi_{t_i}^{-1}(C_{t_i}) = \emptyset$$

Since $\mathcal{D}$ is a compact class, $\mathcal{K}$ as a countable intersections of sets in $\mathcal{D}$ os a compact class. We shall show that Q_0 is a compact measure, i.e., $\mathcal{K}$ Q_0-approximates $\mathcal{C}_0$. Take $C \in \mathcal{C}_0$ and $\epsilon > 0$. For some $S \subset T$,

$$C = \Pi_S^{-1}(B)$$

Choose a rectangle

$$\prod_{t \in S} (A_t) \subset B$$

so that for $A_t \in \mathcal{A}_t$

$$\begin{aligned} Q_S(B - \prod_{t \in S} A_t) &< \frac{\epsilon}{2} \\ Q_0(\Pi_S^{-1}(B) - \Pi_S^{-1}(\prod_{t \in S} A_t)) &< \frac{\epsilon}{2} \end{aligned}$$

For each t, choose $K_t \in \mathcal{K}_t$ such that $K_t \subset A_t$ and

$$Q_t(A_t) < Q_t(K_t) + \frac{\epsilon}{\text{cardinality}(S)}$$

Let

$$K = \bigcup_{t \in S} \Pi_t^{-1}(K_t) \text{ for } K_t \in \mathcal{K}_t$$

Then, $K \subset C$ and

$$\begin{aligned} Q_0\Big(\Pi_S^{-1}(B) - \Pi_S^{-1}(\prod_{t \in S} K_t)\Big) &= Q_0\Big(\Pi_S^{-1}(B - \prod_{t \in S} A_t)\Big) \\ &\quad + Q_0\Big(\Pi_S^{-1}(\prod_{t \in S} A_t) - \Pi_S^{-1}(\prod_{t \in S} K_t)\Big) \\ &< \epsilon \end{aligned}$$

Q_0 extends to a countable additive measure on $\sigma(\mathcal{C})$. Call it Q. Consider now $\Omega = \mathcal{X}_t$, $\sigma(\mathcal{C})$, and Q and define $X_t(\omega) = \omega(t)$.

($\mathcal{X}_t = R$, R^d, or complete separate metric space.)

Example 1. $T = N, \mathcal{X}_t = R$ with $Q_{\{t\}}$ probability measure on $\mathcal{B}(\mathcal{R})$. Suppose

$$Q_n = \bigotimes_{t \in \{1,2,\ldots,n\}} Q_{\{t\}}$$

Then, there exists $\{X_n, n \in N\}$ of random variables defined on R^∞.

Example 2. $T = [0, 1]$.

Let $\{C(t,s), t, s \in T\}$ be a set of real valued function with $C(t,s) = C(s,t)$ and

$$\sum_{t,s \in S} a_t a_s C(t,s) \geq 0$$

for S finite.($\{a_t, t \in S\} \subset R$) Let Q_S be a probability measure with characteristic function for $\mathbf{t} \in R^d$

$$\phi_{Q_S}(\mathbf{t}) = \exp\left(i\mathbf{a}'\mathbf{t} - \frac{1}{2}\mathbf{t}' \sum_S \mathbf{t}\right) \tag{2}$$

where $\sum_S$ is the covariance matrix given by

$$\sum_S = \Big(C(u,v)\Big)_{u,v\in S}$$

Q_S satisfies the condition of Kolmogorov Consistency Theorem. Therefore, there exists $\{X_t, t \in T\}$ a family of random variables such that joint distribution is Q_S.

Example. Take $t, s \in [0,1]$ and $C(t,s) = \min(t,s)$.

$$\begin{aligned} C(t,s) &= \int_0^1 1_{[0,t]}(u) \cdot 1_{[0,s]}(u) du \\ &= \min(t,s) \end{aligned}$$

Let $S = \{t_1, ..., t_n\}, \{a_1, ..., a_n\} \subset R$. Then,

$$\begin{aligned} \sum_{i,j} a_i a_j C(t_i, t_j) &= \sum_{i,j} a_i a_j \int_0^1 1_{[0,t_i]}(u) \cdot 1_{[0,t_j]}(u) du \\ &= \int_0^1 \big(\sum_{i=1}^n a_i 1_{[0,t_i]}(u)\big)^2 du \\ &\geq 0 \end{aligned}$$

since $\sum\sum a_i a_j C(t_i, t_j)$ is non-negative definite. Therefore, there exists a Gaussian Process with covariance $C(t,s) = \min(t,s)$.

2.3 The finite-dimensional family for Brownian Motion

Given Covariance function $C(t,s) = C(s,t)$ with $s, t \in T$, and for all $\{a_1, ..., a_k\}$ and $\{t_1, ..., t_k\} \subset T$, such that

$$\sum_{k,j} a_k a_j C(t_k, t_j) \geq 0$$

then there exists a Gaussian Process $\{X_t, t \in T\}$ with $EX_t = 0$ for all t and $C(t,s) = E(X_t X_s)$.

Example.
$C(t,s) = \min(t,s)$, and $T = [0,1]$.

$$\begin{aligned} \min(t,s) &= \int_0^1 1_{[o,t]}(u) \cdot 1_{[o,s]}(u) du \\ &= EX_t X_s \end{aligned}$$

There exists $\{X_t, t \in T\}$ such that $C(t,s) = E(X_t X_s) = \min(t,s)$.

Since X_t is Gaussian, we know that

$$X_t \in L^2(\Omega, \mathcal{F}, P)$$

Let

$$M(X) = \overline{\mathrm{SP}}^{L^2}(X_t, t \in T)$$

Consider the map

$$I(X_t) = 1_{[0,t]}(u) \in L_2([0,1]), \text{ (Lebesue measure)}$$

I is an isometry. Therefore, for $(t_1, ..., t_n)$ with $t_1 \leq ... \leq t_n$

$$\begin{aligned} I(X_{t_k} - X_{t_{k-1}}) &= I(X_{t_k}) - I(X_{t_{k-1}}) \text{ (because } I \text{ is a linear map)} \\ &= 1_{[0,t_k]}(u) - 1_{[0,t_{k-1}]}(u) \\ &= 1_{(t_{k-1},t_k]}(u) \end{aligned}$$

For $k \neq j$

$$\begin{aligned} E(X_{t_k} - X_{t_{k-1}})(X_{t_j} - X_{t_{j-1}}) &= \int_0^1 1_{(t_{k-1},t_k]}(u) \cdot 1_{(t_{j-1},t_j]}(u) du \\ &= 0 \end{aligned}$$

$X_{t_k} - X_{t_{k-1}}$ is independent of $X_{t_j} - X_{t_{j-1}}$ if $(t_{k-1}, t_k] \cap (t_{j-1}, t_j] = \emptyset$ because

$$(t_{k-1}, t_k] \cap (t_{j-1}, t_j] = \emptyset \Longrightarrow E(X_{t_k} - X_{t_{k-1}})(X_{t_j} - X_{t_{j-1}}) = 0$$

$\{X_t, t \in T\}$ is an independent increment process.

Given $t_0 = 0, X_0 = 0,\ t_0 \le t_1 \le ... \le t_n$, we have

$$P(X_{t_k} - X_{t_{k-1}} \le x_k, k = 1, 2..., n) = \prod_{k=1}^{n} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(t_k - t_{k-1})}} e^{-\frac{y_k^2}{2(t_k - t_{k-1})}} dy_k$$

By using transformation

$$\begin{aligned} Y_{t_1} &= X_{t_1} \\ Y_{t_2} &= X_{t_2} - X_{t_1} \\ \vdots &= \vdots \\ Y_{t_n} &= X_{t_n} - X_{t_{n-1}} \end{aligned}$$

we can compute joint density of $(X_{t_1}, ..., X_{t_n})$.

$$f_{(X_{t_1},...,X_{t_n})}(x_1, ..., x_n) = \frac{1}{\prod_{k=1}^{n} \sqrt{2\pi(t_k - t_{k-1})}} \exp[-\frac{1}{2} \sum_{k=1}^{n} \frac{(x_k - x_{k-1})^2}{(t_k - t_{k-1})}]$$

Define

$$p(t, x, B) = \frac{1}{\sqrt{2\pi t}} \int_B e^{-\frac{(y-x)^2}{t}} dy, \quad t \ge 0$$

and

$$\widetilde{p}(t, s, x, B) = p(t - s, x, B)$$

Exercise 1. Prove for $0 = t_0 \le t_1 \le ... \le t_n$

$$Q_{X_{t_1},...,X_{t_n}}(B_1 \times \cdots \times B_n) = \int_{B_n} \cdots \int_{B_1} p(t_1 - t_0, dy_1) p(t_2 - t_1, y_1, dy_2) \cdots p(p(t_n - t_{n-1}, y_{n-1}, dy_n))$$

Suppose we are given transition function $p(t, s, x, B)$ with $x \le t$. Assume that $s \le t \le u$

$$p(u, s, x, B) = \int p(t, s, x, dy) p(s, t, y, B) \text{ (C-Kolmogorov Condition)}$$

Then, for $0 = t_0 \le t_1 \le ... \le t_n$

$$Q^x_{t_1,...,t_n}(B_1 \times \cdots \times B_n) = \int_{B_n} \cdots \int_{B_1} p(t_1, t_0, x, dy_1) p(t_2, t_1, y_1, dy_2) \cdots p(p(t_n, t_{n-1}, y_{n-1}, dy_n))$$

(Use Fubini's Theorem.) For this consistent family, there exists a stochastic process with Q as finite dimensional distributions.

Exercise 2. Check that $\tilde{p}(t, s, x, B)$ satisfies the above condition.

$$\begin{aligned} Q^x(X_s \in B_1, X_t \in B_2) &= \int_{B_1} p(s, 0, x, dy) p(t, s, y, B_2) \\ &= \int_{B_1} Q^x(X_t \in B_2 | X_s = y) Q^x \circ X_s^{-1}(dy) \end{aligned}$$

where

$$Q^x(X_{t_n} \in B_n | X_{t_1}, ..., X_{t_n}) = p(t_n, t_{n-1}, X_{t_{n-1}}, B_n) \text{ (Markov)}$$

The Gaussian Process with covariance

$$\min(t, s), \quad t, s, \in [0, 1]$$

has independent increments and is Markov.

Remarks. Consider

- $X(\omega) \in R^{[0,1]}$
- $C[0, 1] \notin \sigma(C(R^{[0,1]}))$

$C[0, 1]$ is not measurable. But $C[0, 1]$ has Q_0^*(C[0,1]) = 1 (outer measure).

2.4 Properties of Brownian Motion

Definition 2.3 *A Gaussian process is a stochastic process $\{X_t, t \in T\}$, for which any finite linear combination of samples has a joint Gaussian distribution. More accurately, any linear functional applied to the sample function X_t will give a normally distributed result. Notation-wise, one can write $X \sim GP(m, K)$, meaning the random function X is distributed as a GP with mean function m and covariance function K.*

Remark.

$$\begin{aligned} X &\sim N(\mu_X, \sigma_X^2) \\ Y &\sim N(\mu_Y, \sigma_Y^2) \\ Z &= X + Y \end{aligned}$$

Then, $Z \sim N(\mu_Z, \sigma_Z^2)$ where

$$\mu_Z = \mu_X + \mu_Y \text{ and } \sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho_{X,Y}\sigma_X\sigma_Y$$

Proposition 2.1 *Given covariance $C(t, s) = C(s, t)$ with $s, t \in T$ and for all $\{a_1, ..., a_k\} \subset R$ and $\{t_1, ..., t_k\} \subset T$*

$$\sum_{k,j} a_k a_j C(t_k, t_j) \geq 0$$

then there exists a Gaussian Process such that

$$\{X_t, t \in T\} \textit{ with for all } t,\ EX_t = 0,\ C(t, s) = E(X_t, X_s)$$

Example: $C(t, s) = \min(t, s)$ and $T = [0, 1]$.
In this example,

$$\begin{aligned} \min(t, s) &= \int_0^1 1_{[0,t]}(u) 1_{[0,s]}(u) du \\ &= EX_t X_s \end{aligned}$$

So, there exists $\{X_t, t \in T\}$, which is Gaussian Process, such that for all t

$$EX_t = 0, \text{ and } C(t, s) = E(X_t, X_s) = \min(t, s)$$

Since X_t is Gaussian, we know that

$$X_t \in L^2(\Omega, \mathcal{F}, P)$$

Let

$$M(X) = \overline{SP}^{L^2}(X_t, t \in T) \text{ (SP means "Span")}$$

Consider the map $I : M(X) \to \overline{SP}\{1_{[0,t](u)}, t \in [0,1]\}$ such that

$$I(X_t) = 1_{[0,t]}(u)$$

and

$$I\Big(\sum a_k X_{t_k}\Big) = \sum a_k I(X_{t_k})$$

Proposition 2.2 *I is a linear map.*

Proof. Easy.

Proposition 2.3 *I is an isometry*

Proof. Since $\{X_t, t \in T\}$ is Gaussian,

$$\begin{aligned} Var(X_{t_k} - X_{t_{k-1}}) &= Var(X_{t_k}) + Var(X_{t_k}) - 2Cov(X_{t_k}, X_{t_{k-1}}) \\ &= C(t_k, t_k) + C(t_{k-1}, t_{k-1}) - 2C(t_{k-1}, t_k) \\ &= t_k + t_{k-1} - 2t_{k-1} \\ &= t_k - t_{k-1} \end{aligned}$$

Therefore,

$$\begin{aligned} ||X_{t_k} - X_{t_{k-1}}||^2_{L_2} &= \int_{[0,1]} (X_{t_k} - X_{t_{k-1}})^2 dP \\ &= E(X_{t_k} - X_{t_{k-1}})^2 \\ &= Var(X_{t_k} - X_{t_{k-1}}) \\ &= t_k - t_{k-1} \end{aligned}$$

Also,

$$\begin{aligned} ||I(X_{t_k}) - I(X_{t_{k-1}})||^2_{L_2} &= ||1_{[0,t_k]}(u) - 1_{[0,t_{k-1}]}(u)||^2_{L_2} \\ &= ||1_{(t_{k-1},t_k]}(u)||^2_{L_2} \\ &= t_k - t_{k-1} \\ &= ||X_{t_k} - X_{t_{k-1}}||^2_{L_2} \end{aligned}$$

This completes the proof.

Suppose that $t_2 \leq ... \leq t_k$. Then,

$$\begin{aligned} I(X_{t_k} - X_{t_{k-1}}) &= I(X_{t_k}) - I(X_{t_{k-1}}) \text{ (since } I \text{ is an linear map)} \\ &= 1_{[0,t_k]}(u) - 1_{[0,t_{k-1}]}(u) \\ &= 1_{(t_{k-1},t_k]}(u) \end{aligned}$$

$X_{t_k} - X_{t_{k-1}}$ is independent of $X_{t_j} - X_{t_{j-1}}$ if

$$(t_{k-1}, t_k] \cap (t_{j-1}, t_j] = \emptyset$$

Proposition 2.4 *If $X_{t_k} - X_{t_{k-1}}$ is independent of $X_{t_j} - X_{t_{j-1}}$, then*

$$E(X_{t_k} - X_{t_{k-1}})(X_{t_j} - X_{t_{j-1}}) = 0$$

Proof. For $k \neq j$

$$\begin{aligned}
& E(X_{t_k} - X_{t_{k-1}})(X_{t_j} - X_{t_{j-1}}) \\
= & \int_0^1 1_{(t_{k-1}, t_k]}(u) 1_{(t_{j-1}, t_j]}(u) du \\
= & \ 0
\end{aligned}$$

Suppose $\{X_t, t \in T\}$ is an independent increment process such that $t_0 = 0$, $X_0 = 0$, and $t_0 \leq t_1 \leq ... \leq t_n$. Then, $X_{t_k} - X_{t_{k-1}}$ is Gaussian with mean 0 and variance $t_k - t_{k-1}$.

$$\begin{aligned}
P(X_{t_k} - X_{t_{k-1}} \leq x_k, k = 1, 2..., n) &= \prod_{k=1}^{n} P(X_{t_k} - X_{t_{k-1}} \leq x_k) \\
&= \prod_{k=1}^{n} \int_{-\infty}^{x_k} \frac{1}{\sqrt{2\pi(t_k - t_{k-1})}} e^{-\frac{y_k^2}{2(t_k - t_{k-1})}} dy_k
\end{aligned}$$

Let

$$\begin{aligned}
Y_{t_1} &= X_{t_1} \\
Y_{t_2} &= X_{t_2} - X_{t_1} \\
&\vdots \\
Y_{t_n} &= X_{t_n} - X_{t_{n-1}}
\end{aligned}$$

Then,

$$f_{X_{t_1}, ..., X_{t_n}}(x_1, ..., x_n) = \frac{1}{\prod_{k=1}^{n}} \sqrt{2\pi(t_k - t_{k-1})} \exp\left[-\frac{1}{2} \sum_{k=1}^{n} \frac{(x_k - x_{k-1})^2}{(t_k - t_{k-1})} \right]$$

2.5 Kolmogorov Continuity Theorem.

For each t, if $P(\tilde{X}_t = X_t) = 1$, then we say the finite dimensional distributions of $\tilde{X}_t$ and X_t are the same and $\{\tilde{X}_t\}$ is a version of $\{X_t\}$.

Proposition 2.5 *Let $\{X_t, t \in [0,1]\}$ be a stochastic process with*

$$E|X_t - X_s|^\beta \leq C|t-s|^{1+\alpha} \text{ with } C, \alpha, \beta > 0$$

Then, there exists a version of $\{X_t, t \in [0,1]\}$ which has a continuous sample paths.

Corollary 2.1 *The Gaussian Process with covariance function*

$$C(t,s) = \min(t,s), \;\; t, s \in [0,1]$$

(has independent increment and is Markovian) has a version which is continuous.

$$\begin{aligned}
\underbrace{E(X_t - X_s)^2}_{=|t-s|} &= EX_t^2 - 2EX_tX_s + EX_s^2 \\
&= t - 2s + s \\
&= |t-s| \\
E(X_t - X_s)^4 &= 3[E(X_t - X_s)^2]^2 \\
&= 3|t-s|^2
\end{aligned}$$

We shall denote the continuous version by W_t, called Wiener process BM.

Proof. Take $0 < \gamma < \frac{\alpha}{\beta}$ and $\delta > 0$ such that

$$(1-\delta)(1+\alpha-\beta\gamma) > 1+\delta$$

For $0 \leq i \leq j \leq 2^n$ and $|j-i| \leq 2^{n\delta}$,

$$\begin{aligned}
\sum_{i,j} P\Big(|X_{j2^{-n}} - X_{i2^{-n}}| > [(j-i)2^{-n}]^\gamma\Big) &\leq C\sum_{i,j}[(j-i)2^{-n}]^{-\beta\gamma+(1+\delta)} \text{ (by Chevyshev)} \\
&= C_2 2^{n[(1+\delta)-(1-\delta)(1+\alpha-\beta\gamma)]} \\
&< \infty
\end{aligned}$$

where $(1+\delta) - (1-\delta)(1+\alpha-\beta\gamma) = -\mu$.

Then, by Borell-Cantelli Lemma,

$$P\Big(|X_{j2^{-n}} - X_{i2^{-n}}| > [(j-i)2^{-n}]^\gamma\Big) = 0$$

i.e., there exists $n_0(\omega)$ such that for all $n \geq n_0(\omega)$

$$|X_{j2^{-n}} - X_{i2^{-n}}| \leq [(j-i)2^{-n}]^\gamma$$

Let $t_1 < t_2$ be rational numbers in $[0,1]$ such that

$$\begin{aligned} t_2 - t_1 &\leq 2^{-n_0(1-\delta)} \\ t_1 &= i2^{-n} - 2^{-p_1} - ... - 2^{-p_k} \; (n < p_1 < ... < p_k) \\ t_2 &= j2^{-n} - 2^{-q_1} - ... - 2^{-q_k} \; (n < q_1 < ... < q_k) \\ t_1 &\leq i2^{-n} \leq j2^{-n} \leq t_2 \end{aligned}$$

Let

$$h(t) = t^\gamma \text{ for } 2^{-(n+1)(1-\delta)} \leq t \leq 2^{-n(1-\delta)}$$

Then,

$$\begin{aligned} \left|X_{t_1} - X_{i2^{-n}}\right| &\leq C_1 h(2^{-n}) \\ \left|X_{t_2} - X_{j2^{-n}}\right| &\leq C_2 h(2^{-n}) \\ \left|X_{t_2} - X_{t_1}\right| &\leq C_3 h(t_2 - t_1) \end{aligned}$$

and

$$\left|X_{i2^{-n}-2^{-p_1}-...-2^{-p_k}} - X_{i2^{-n}-2^{-p_1}-...-2^{-p_{k-1}}}\right| \leq 2^{-p_k\gamma}$$

Under this condition, process is uniformly continuous on rational numbers in $[0,1]$.

Let

$$\psi : (\Omega_Q, \mathcal{C}_{\Omega_Q}) \to (C[0,1], \sigma(C[0,1]))$$

Then, ψ extends uniformly continuous functions on rational to continuous function on $[0,1]$.

Let P is a measure generated by the same finite dimensional on rationals. Then

$$\tilde{P} = P \circ \psi^{-1}$$

is the measure of $\tilde{X}_t$, version of X_t. In Gaussian, there exists a version of continuous sample path. In case of Brownian motion, there exists a version. We call it $\{W_t\}$.

$\{W_{t+t_0} - W_{t_0}, t \in [0, \infty]\}$ is a Weiner Process.

2.6 Exit time for Brownian Motion and Skorokhod Theorem

Let

$$\mathcal{F}_t^W = \sigma(W_s, s \le t)$$

τ is called "stopping time" if

$$\{\tau \le t\} \in \mathcal{F}_t^W$$

Define

$$\mathcal{F}_\tau = \{A : A \cap \{\tau \le t\} \in \mathcal{F}_t^W\}$$

Then $\mathcal{F}_\tau$ is a $\sigma-$field.

$\frac{W_t}{\sqrt{t}}$ is a standard normal variable.

Define

$$T_a = \inf\{t : W_t = a\}$$

Then, $T_a < \infty$ a.e. and is a stopping time w.r.t $\{\mathcal{F}_t^W\}$.

Theorem 2.1 *Let $a < x < b$. Then*

$$P_x(T_a < T_b) = \frac{b - x}{b - a}$$

Remark. W_t is Gaussian and has independent increment. Also, for $s \le t$

$$E(W_t - W_s | \mathcal{F}_s) = 0$$

and hence $\{W_t\}$ is Martingale.

Proof. Let $T = T_a \wedge T_b$. We know $T_a, T_b < \infty$ a.e., and

$$W_{T_a} = a, \text{ and } W_{T_b} = b$$

Since $\{W_t\}$ is MG,

$$\begin{aligned} E_x W_T &= E_x W_0 \\ &= x \\ &= aP(T_a < T_b) + b(1 - P(T_a < T_b)) \end{aligned}$$

Therefore,

$$P_x(T_a < T_b) = \frac{b - x}{b - a}$$

$\{W_t, t \in [0, \infty]\}$ is a Weiner Process, starting from x with $a < x < b$. Recall

$$\begin{aligned}
T_a &= \inf\{t : W_t = a\} \\
T_b &= \inf\{t : W_t = b\} \\
T &= T_a \wedge T_b \\
E_x W_T &= E_x W_0 \\
P_x(T_a < T_b) &= \frac{b - x}{b - a}
\end{aligned}$$

We know that

$$\begin{aligned}
E((W_t - W_s)^2 | \mathcal{F}_s^W) &= E(W_t - W_s)^2 \\
&= (t - s)
\end{aligned}$$

Also,

$$\begin{aligned}
E((W_t - W_s)^2 | \mathcal{F}_s^W) &= E(W_t^2 | \mathcal{F}_s^W) - 2E(W_t W_s | \mathcal{F}_s^W) + E(W_s^2 | \mathcal{F}_s^W) \\
&= E(W_t^2 | \mathcal{F}_s^W) - W_s^2 \\
&= E(W_t^2 - W_s^2 | \mathcal{F}_s^W) \\
&= (t - s)
\end{aligned}$$

Therefore,

$$E(W_t^2 - t | \mathcal{F}_s^W) = W_s^2 - s$$

and hence $\{(W_t^2 - t), t \in [0, \infty]\}$ is a martingale.

Suppose that $x = 0$ and $a < 0 < b$. Then $T = T_a \wedge T_b$ is a finite stopping time. Therefore,

$$T \wedge t$$

is also stopping time.

$$E(W_{T\wedge t}^2 - T \wedge t) = 0$$

$$E_0(W_T^2) = E_0 T$$

$$\begin{aligned}
EW_T^2 &= ET \\
&= a^2 P(T_a < T_b) + b^2(1 - P(T_a < T_b)) \\
&= -ab
\end{aligned}$$

Suppose X has two values a, b with $a < 0 < b$ and

$$\begin{aligned}
EX &= aP(X = a) + bP(X = b) \\
&= 0
\end{aligned}$$

Remark.
$$P(X=a)=\frac{b}{b-a} \text{ and } P(X=b)=-\frac{a}{b-a}$$

Let $T=T_a\wedge T_b$. Then, W_T has the same distribution as X. We denote

$$\mathcal{L}(W_T)=\mathcal{L}(X)$$

or

$$W_T=_{\mathcal{D}} X$$

2.6.1 Skorokhod Theorem.

Let X be random variable with $EX=0$ and $EX^2<\infty$. Then, there exists a $\mathcal{F}_t^W-$stopping time T such that

$$\mathcal{L}(W_T)=\mathcal{L}(X) \text{ and } ET=EX^2$$

Proof. Let $F(x)=P(X\leq x)$.

$$\begin{aligned} EX=0 \quad &\Rightarrow \quad \int_{-\infty}^{0} udF(u)+\int_{0}^{\infty} vdF(v)=0 \\ &\Rightarrow \quad -\int_{-\infty}^{0} udF(u)=\int_{0}^{\infty} vdF(v)=C \end{aligned}$$

Let ψ be a bounded function with $\psi(0)=0$. Then,

$$\begin{aligned} & C\int_{R}\psi(x)dF(x) \\ = \quad & C\Big(\int_{0}^{\infty}\psi(v)dF(v)+\int_{-\infty}^{0}\psi(u)dF(u)\Big) \\ = \quad & \int_{0}^{\infty}\psi(v)dF(v)\int_{-\infty}^{0}-udF(u)+\int_{-\infty}^{0}\psi(u)dF(u)\int_{0}^{\infty}vdF(v) \\ = \quad & \int_{0}^{\infty}dF(v)\int_{-\infty}^{0}dF(u)(v\psi(u)-u\psi(v)) \end{aligned}$$

Therefore,

$$\begin{aligned} \int_{R}\psi(x)dF(x) \quad &= \quad C^{-1}\int_{0}^{\infty}dF(v)\int_{-\infty}^{0}dF(u)(v\psi(u)-u\psi(v)) \\ &= \quad C^{-1}\int_{0}^{\infty}dF(v)\int_{-\infty}^{0}dF(u)(v-u)\Big[\frac{v}{v-u}\psi(u)-\frac{u}{v-u}\psi(v)\Big] \end{aligned}$$

Consider (U,V) be a random vector in R^2 such that

$$P\Big[(U,V)=(0,0)\Big]=F(\{0\})$$

and for $A \subset (-\infty, 0) \times (0, \infty)$

$$P((U,V) \in A) = C^{-1} \int\int_A dF(u)dF(v)(v-u)$$

If $\psi = 1$,

$$\begin{aligned}
P((U,V) \in (-\infty,0) \times (0,\infty)) &= C^{-1} \int_0^\infty dF(v) \int_{-\infty}^0 dF(u)(v-u) \\
&= C^{-1} \int_0^\infty dF(v) \int_{-\infty}^0 dF(u)(v-u) \Big[\frac{v}{v-u}\psi(u) - \frac{u}{v-u}\psi(v) \Big] \\
&= \int_R \psi(x) dF(x) \\
&= \int_R dF(x) \\
&= 1
\end{aligned}$$

and hence, P is a probability measure.

Let $u < 0 < v$ such that

$$\mu_{U,V}(\{u\}) = \frac{v}{v-u} \text{ and } \mu_{U,V}(\{v\}) = -\frac{u}{v-u}$$

Then, by Fubini,

$$\int \psi(x) dF(x) =$$

On product space $\Omega \times \Omega'$, let

$$\begin{aligned}
W_t(\omega, \omega') &= W_t(\omega) \\
(U,V)(\omega,\omega') &= (U,V)(\omega')
\end{aligned}$$

$T_{U,V}$ is not a stopping time on $\mathcal{F}_t^W$.

We know that if $U = u$ and $V = v$

$$\mathcal{L}(T_{U,V}) = \mathcal{L}(X)$$

Then,

$$\begin{aligned}
ET_{U,V} &= E_{U,V} E(T_{U,V} | U,V) \\
&= -EUV \\
&= C^{-1} \int_{-\infty}^0 dF(u)(-u) \int_0^\infty dF(v) v(v-u) \\
&= -\int_0^\infty dF(v)(-u) \Big[C^{-1} \int_0^\infty v dF(v) - u \Big] \\
&= EX^2
\end{aligned}$$

Next time, we will show that

$$W(t+\tau) - W(\tau)$$

is again Brownian motion.

2.7 Embedding of sums of i.i.d. random variable in Brownian Motion

Let $t_0 \in [0, \infty)$. Then,

$$\{W(t+t_0) - W(t_0), t \geq 0\}$$

is a Brownian motion and independent of $\mathcal{F}_{t_0}$.

Let τ be a stopping time w.r.t $\mathcal{F}_t^W$. Then,

$$W_t^*(\omega) = W_{\tau(\omega)+t}(\omega) - W_{\tau(\omega)}(\omega)$$

is a Brownian Motion w.r.t $\mathcal{F}_\tau^W$ where

$$\mathcal{F}_\tau^W = \{B \in \mathcal{F} : B \cup \{\tau \leq t\} \in \mathcal{F}_t^W\}$$

Let V_0 be countable. Then,

$$\{\omega : W_t^*(\omega) \in B\} = \bigcup_{t_0 \in V_0} \{\omega : W(t+t_0) - W(t_0) \in B, \tau = t_0\}$$

For $A \in \mathcal{F}_\tau^W$,

$$\begin{aligned} P\Big[\{(W_{t_1}^*, ..., W_{t_k}^*) \in B_k\} \cap A\Big] &= \sum_{t_0 \in V_0} P\Big[\{(W_{t_1}, ..., W_{t_k}) \in B_k\} \cap A \cap \{\tau = t_0\}\Big] \\ &= \sum_{t_0 \in V_0} P\Big((W_{t_1}^*, ..., W_{t_k}^*) \in B_k\Big) \cdot P(A \cap \{\tau = t_0\}) \\ & \text{(because of independence)} \\ &= P\Big((W_{t_1}^*, ..., W_{t_k}^*) \in B_k\Big) \sum_{t_0 \in V_0} P(A \cap \{\tau = t_0\}) \\ &= P\Big((W_{t_1}^*, ..., W_{t_k}^*) \in B_k\Big) P(A) \end{aligned}$$

Let τ be any stopping time such that

$$\tau_n = \begin{cases} 0, & \text{if } \tau\text{=}0; \\ \frac{k}{2^n}, & \text{if } \frac{k-1}{2^n} < \tau \leq \frac{k}{2^n}. \end{cases}$$

If $\frac{k}{2^n} \leq t < \frac{k+1}{2^n}$,

$$\{\tau_n \leq t\} = \Big\{\tau \leq \frac{k}{2^n}\Big\} \in \mathcal{F}_{\frac{k}{2^n}} \subset \mathcal{F}_t$$

Claim. $\mathcal{F}_\tau \subset \mathcal{F}_{\tau_n}$.

Proof. Suppose $C \in \mathcal{F}_\tau = \{B : B \cap \{\tau \le t\} \in \mathcal{F}_t\}$. Then, since $\frac{k}{2^n} \le t$,

$$C \cap \{\tau \le t\} = C \cap \left\{\tau \le \frac{k}{2^n}\right\} \in \mathcal{F}_{\frac{k}{2^n}}$$

This completes the proof by continuity.

$W_t^n = W_{\tau_n + t} - W_{\tau_n}$ is a Brownian Motion for each n, independent of $\mathcal{F}_\tau$.

Theorem 2.2 *Let $X_1, ..., X_n$ be i.i.d. with $EX_i = 0$, $EX_i^2 < \infty$ for all i. Then there exists a sequence of stopping time $T_0 = 0, T_1, ..., T_n$ such that*

$$\mathcal{L}(S_n) = \mathcal{L}(W_{T_n})$$

where $(T_i - T_{i-1})$ are i.i.d.

Proof. $(U_1, V_1), ..., (U_n, V_n)$ i.i.d. as (U, V) and independent of W_t.

$$T_0 = 0, T_k = \inf\{t \ge T_{k-1}, W_{T_k} - W_{T_{k-1}} \in (U_k, V_k)\}$$

$T_k - T_{k-1}$ are i.i.d. and

$$\begin{array}{rcl} \mathcal{L}(X_1) & = & \mathcal{L}(W_{T_1}) \\ \mathcal{L}(X_2) & = & \mathcal{L}(W_{T_2} - W_{T_1}) \\ \vdots & \vdots & \vdots \\ \mathcal{L}(X_n) & = & \mathcal{L}(W_{T_n} - W_{T_{n-1}}) \end{array}$$

Hence $\mathcal{L}(S_n) = \mathcal{L}(W_{T_n})$. Now

$$\begin{aligned} \mathcal{L}\Big(\frac{S_n}{\sqrt{n}}\Big) &= \mathcal{L}\Big(\frac{W(T_n)}{\sqrt{n}}\Big) \\ &= \mathcal{L}\Big(\frac{W(T_n/n \cdot n)}{\sqrt{n}}\Big) \\ &= \mathcal{L}\left(W\Big(\frac{T_n}{n}\Big)\right) \quad \Big(\text{since } \frac{W(nt)}{\sqrt{n}} \approx W(t)\Big) \end{aligned}$$

Assume $EX_1^2 = 1$. Then

$$\frac{T_n}{n} \to_{a.s.} E(T_1) = EX_1^2 = 1,$$

and hence

$$\frac{S_n}{\sqrt{n}} \to W(1). \quad \text{(CLT)}$$

2.8 Donsker's Theorem

$X_{n,1}, ..., X_{n,n}$ are i.i.d. for each n with $EX_{n,m} = 0$, $EX_{n,m}^2 < \infty$, and $S_{n,m} = X_{n,1} + \cdots + X_{n,m} = W_{\tau_m^n}$ where τ_m^n is stopping time and W is Brownian motion. Define

$$S_{n,u} = \begin{cases} S_{n,m}, & \text{if } u = m \in \{0, 1, 2, ..., n\}; \\ \text{linear}, & \text{if } u \in [m-1, m]. \end{cases}$$

Lemma 2.2 *If $\tau_{[ns]}^n \to s$ for $s \in [0, 1]$, then with $||||$ as supnorm*

$$||S_{n,[n\cdot]} - W(\cdot)|| \to 0 \textit{ in probability.}$$

Proof. For given $\epsilon > 0$, there exists $\delta > 0(1/\delta$ is an integer) such that

$$P\Big(|W_t - W_s| < \epsilon, \text{ for all } t, s \in [0, 1], |t - s| < 2\delta\Big) > 1 - \epsilon \tag{3}$$

τ_m^n is increasing in m. For $n \geq N\delta$,

$$P\Big(|\tau_{nk\delta}^n - k\delta| < \delta, k = 1, 2, ..., \frac{1}{\delta}\Big) \geq 1 - \epsilon$$

since $\tau_{[ns]}^n \to s$. For $s \in ((k-1)\delta, k\delta)$, we have

$$\begin{array}{rcl} \tau_{[ns]}^n - s & \geq & \tau_{[n(k-1)\delta]}^n - k\delta \\ \tau_{[ns]}^n - s & \leq & \tau_{[nk\delta]}^n - (k-1)\delta \end{array}$$

Combining these, we have for $n \geq N\delta$

$$P\Big(\sup_{0 \leq s \leq 1} |\tau_{[ns]}^n - s| < 2\delta\Big) > 1 - \epsilon. \tag{4}$$

For ω in event in (3) and (4), we get for $m \leq n$

$$\Big|\underbrace{W_{\tau_m^n}}_{=S_{n,m}} - W_{\frac{m}{n}}\Big| < \epsilon$$

For $t = \frac{m+\theta}{n}$ with $0 < \theta < 1$,

$$\begin{array}{rcl} \Big|S_{n,[nt]} - W_t\Big| & \leq & (1-\theta)\Big|S_{n,m} - W_{\frac{m}{n}}\Big| + \theta\Big|S_{n,m+1} - \dfrac{W_{m+1}}{n}\Big| \\ & & +(1-\theta)\Big|W_{\frac{m}{n}} - W_t\Big| + \theta\Big|\dfrac{W_{n+1}}{n} - W_t\Big| \end{array}$$

For $n \geq N_\delta$ with $\frac{1}{n} < 2\delta$,

$$P\Big(||S_{n,ns} - W_s||_\infty \geq 2\epsilon\Big) < 2\epsilon.$$

We now derive some consequences of Donsker theorem.

Theorem 2.3 *Let f be bounded and continuous function on* $[0,1]$. *Then*

$$Ef\Big(S_{n,[n\cdot]}\Big) \to Ef(W(\cdot))$$

Proof. For fixed $\epsilon > 0$, define

$$G_\delta = \{W, W' \in C[0,1] : ||W - W'||_\infty < \delta \text{ implies } |f(W) - f(W')| < \epsilon\}$$

Observe that $G_\delta \uparrow C[0,1]$ as $\delta \downarrow 0$. Then,

$$\left| Ef\Big(S_{n,[n\cdot]}\Big) - Ef(W(\cdot)) \right| \le \epsilon + 2M\Big(P(G_\delta^c) + P\Big(||S_{n,[n\cdot]} - W(\cdot)|| > \delta\Big)\Big)$$

Since $P(G_\delta^c) \to 0$ and $P\Big(||S_{n,[n\cdot]} - W(\cdot)|| > \delta\Big) \to 0$ by (2.2). In particular

$$\max_t \left| \frac{S_{[nt]}}{\sqrt{n}} \right| \to \max_t |W(t)| \text{ in distribution}$$

and

$$\max_{1\le m\le n} \left| \frac{S_m}{\sqrt{n}} \right| \to \max_t |W(t)| \text{ in distribution}$$

Let

$$R_n = 1 + \max_{1\le m\le n} S_m - \min_{1\le m\le n} S_m$$

Then

$$\frac{R_n}{\sqrt{n}} \Rightarrow_{weakly} \max_{0\le t\le 1} W(t) - \min_{0\le t\le 1} W(t)$$

We now derive from Donsker theorem invariance principle for U-statistics.

$$\prod_{i=1}^{[nt]} \left(1 + \frac{\theta X_i}{\sqrt{n}}\right) = \sum_{k=1}^{[nt]} n^{-\frac{k}{2}} \sum_{1\le i_1\le\cdots\le i_k\le n} X_{i_1}\cdots X_{i_k}$$

where X_i are i.i.d. and $EX_i^2 < \infty$. Next,

$$\begin{aligned} \log\left[\prod_{i=1}^{[nt]} \left(1 + \frac{\theta X_i}{\sqrt{n}}\right)\right] &= \sum_{i=1}^{[nt]} \log\left(1 + \frac{\theta X_i}{\sqrt{n}}\right) \\ &= \theta \sum_{i=1}^{[nt]} \frac{X_i}{\sqrt{n}} - \frac{\theta^2}{2} \underbrace{\sum_{i=1}^{[nt]} \frac{X_i^2}{n}}_{\to t \text{ by SNNL}} + \frac{\theta^3}{3} \sum_{i=1}^{[nt]} \underbrace{\frac{X_i^3}{n\sqrt{n}}}_{\to 0} - \underbrace{\cdots}_{\to 0} \\ &\Rightarrow \theta W(t) - \frac{\theta^2}{2} t, \end{aligned}$$

and hence

$$\prod_{i=1}^{[nt]} \left(1 + \frac{\theta X_i}{\sqrt{n}}\right) \Rightarrow e^{\theta W(t) - \frac{\theta^2}{2} t}$$

2.9 Empirical Distribution Function

Let us define empirical distribution function

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le x), \quad x \in R$$

Then Glivenko-Cantelli lemma says

$$\sup_x |\hat{F}_n(x) - F(x)| \to_{a.s.} 0$$

Assume F is continuous. Let $U_i = F(X_i)$. For $y \in [0,1]$, define

$$\hat{G}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(F(X_i) \le y).$$

Then by 1-1 transformation,

$$\sqrt{n}\sup_x |\hat{F}_n(x) - F(x)| = \sqrt{n}\sup_{y\in[0,1]} |\hat{G}_n(y) - y|$$

Let $U_1, U_2, ..., U_n$ be uniform distribution and let $U_{(i)}$ be order statistic such that

$$U_{(1)}(\omega) \le \cdots \le U_{(n)}(\omega).$$

Next,

$$f\Big(U_{(1)}, ..., U_{(n)}\Big) = f\Big(U_{\pi(1)}, ..., U_{\pi(n)}\Big)$$

$$\begin{aligned} f_{U_{\pi(1)},...,U_{\pi(n)}}(u_1, ..., u_n) &= f_{U_1,...,U_n}(u_1, ..., u_n) \\ &= \begin{cases} 1, & \text{if } \mathbf{u} \in [0,1]^n; \\ 0, & \text{if } \mathbf{u} \notin [0,1]^n. \end{cases} \end{aligned}$$

For bounded g,

$$Eg\Big(U_1, ..., U_n\Big) = \sum_{\pi\in\Pi} \int_{u_1<u_2<\cdots<u_n} g_{U_{\pi(1)},...,U_{\pi(n)}}(u_1, ..., u_n) f\Big(u_{\pi(1)}, ..., u_{\pi(n)}\Big) du_1 \cdots du_n$$

So we get

$$f_{U_1,...,U_n}(u_1, ..., u_n) = \begin{cases} n!, & \text{if } u_1 < u_2 < \cdots < u_n; \\ 0, & \text{otherwise.} \end{cases}$$

Theorem 2.4 *Let e_j be i.i.d. exponential distribution with failure rate λ. Then*

$$\mathcal{L}\Big(U_1, ..., U_n\Big) = \mathcal{L}\left(\frac{Z_1}{Z_{n+1}}, ..., \frac{Z_n}{Z_{n+1}}\right)$$

where

$$Z_i = \sum_{j=1}^{i} e_j.$$

Proof. First we have

$$f_{e_1,\ldots,e_{n+1}}(u_1,...,u_{n+1}) = \begin{cases} \lambda^{n+1}e^{-\sum_{i=1}^{n+1}u_i}, & \text{if } u_i \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

Let $s_i - s_{i-1} = u_i$ for $i = 1, 2, ..., n+1$. Then

$$f_{Z_1,\ldots,Z_{n+1}}(s_1,...,s_{n+1}) = \prod_{i=1}^{n+1} \lambda e^{-(s_i - s_{i-1})}$$

Use transformation. Let

$$\begin{aligned} v_i &= \frac{s_i}{s_{i+1}} \text{ for } i \leq n \\ v_{n+1} &= s_{n+1} \end{aligned}$$

Then

$$\begin{aligned} f_{V_1,\ldots,V_{n+1}}(v_1,...,v_{n+1}) &= \prod_{i=1}^{n+1} \Big(\lambda e^{-\lambda v_{n+1}(v_i - v_{i-1})}\Big)\lambda e^{-\lambda v_{n+1}(1-v_n)} \\ &= \lambda^{n+1}e^{-\lambda v_n v_{n+1} - \lambda v_n + \lambda v_n v_{n+1}} v_{n+1}^n \end{aligned}$$

$$\begin{aligned} D_n &= \sqrt{n} \max_{1\leq m\leq n} \left| \frac{Z_m}{Z_{n+1}} - \frac{m}{n} \right| \\ &= \frac{n}{Z_{n+1}} \max_{1\leq m\leq n} \left| \frac{Z_m}{\sqrt{n}} - \frac{m}{n}\frac{Z_{n+1}}{\sqrt{n}} \right| \\ &= \frac{n}{Z_{n+1}} \max_{1\leq m\leq n} \left| \frac{Z_m - m}{\sqrt{n}} - \frac{m}{n}\frac{Z_{n+1} - n}{\sqrt{n}} \right| \\ &= \frac{n}{Z_{n+1}} \max_{1\leq m\leq n} \left| W_n(t) - t\Big(W_n(1) + \frac{Z_{n+1} - Z_n}{\sqrt{n}}\Big) \right| \end{aligned}$$

where

$$W_n(t) = \begin{cases} \frac{Z_m - m}{\sqrt{n}}, & \text{if } t = \frac{m}{n}; \\ \text{linear} , & \text{between.} \end{cases}$$

We know that $n/Z_{n+1} \to_{a.s.} 1$ and

$$E\Big(\frac{Z_{n+1} - Z_n}{\sqrt{n}}\Big)^2 = \frac{1}{n}Ee_n^2 \to 0,$$

and hence by Chebyshev's inequality,

$$\frac{Z_{n+1} - Z_n}{\sqrt{n}} \to_p 0.$$

Since $\max(\cdot)$ is a continuous function and

$$\Big(W_n(\cdot) - \cdot W_n(1)\Big) \Rightarrow_D \Big(W(\cdot) - \cdot W(1)\Big),$$

we have

$$\begin{aligned} D_n &= \underbrace{\frac{n}{Z_{n+1}}}_{\to_{a.s}1} \max_{1\le m\le n} \left| W_n(t) - t\Big(W_n(1) + \underbrace{\frac{Z_{n+1}-Z_n}{\sqrt{n}}}_{\to_p 0}\Big)\right| \\ &\Rightarrow_D \max_{1\le m\le n} \Big| \underbrace{W(t) - tW(1)}_{\text{Brownian Bridge}} \Big| \end{aligned}$$

$$P\Big(W_{t_1} \le x_1, W_{t_2} \le x_2, ..., W_{t_k} \le x_k, W(1) = 0\Big) = P\Big(W_{t_1} \le x_1, W_{t_2} \le x_2, ..., W_{t_k} \le x_k\Big) \cdot P\Big(W(1) = 0\Big)$$

$$\begin{aligned} P\Big(W_{t_1} \le x_1, W_{t_2} \le x_2, ..., W_{t_k} \le x_k \Big| W(1) = 0\Big) &= \frac{P\Big(W_{t_1} \le x_1, W_{t_2} \le x_2, ..., W_{t_k} \le x_k, W(1) = 0\Big)}{P(W(1) = 0)} \\ &= P\Big(W_{t_1} \le x_1, W_{t_2} \le x_2, ..., W_{t_k} \le x_k\Big) \end{aligned}$$

$\{W_t^0\}$ is called Brownian Bridge if

$$\begin{aligned} EW_t^0 W_s^0 &= E(W_t - tW(1))(W_s - sW(1)) \\ &= \min(t,s) - st - ts + ts \\ &= s(1-t) \end{aligned}$$

for $s \le t$.

2.10 Weak Convergence of Probability Measures on Polish Space

Let $(\mathcal{X}, \rho)$ be a complete separable metric space. $\{P_n\}$, a sequence of probability measure on $\mathcal{B}(\mathcal{X})$ converges weakly to P if for all bounded continuous function on $\mathcal{X}$

$$\int f dP_n \longrightarrow \int f dP$$

and we write $P_n \Rightarrow P$.

Theorem 2.5 *Every probability measure P on $(S, \mathcal{S})$ is regular; that is, for every $\mathcal{S}-$set A and every ϵ there exist a closed set F and an open set G such that $F \subset A \subset G$ and $P(G - F) < \epsilon$.*

Proof. Denote the metric on S by $\rho(x, y)$ and the distance from x to A by $\rho(x, A) = \inf\{\rho(x, y) : y \in A\}$. If A is closed, we can take $F = A$ and $G = A^\delta = \{x : \rho(x, A) < \delta\}$ for some δ, since the latter sets decrease to A as $\delta \downarrow 0$. Hence we need only show that the class $\mathcal{G}$ of $\mathcal{S}-$sets with the asserted property is a $\sigma-$field. Given sets A_n in $\mathcal{G}$, choose closed sets F_n and open sets G_n such that $F_n \subset A_n \subset G_n$ and $P(G_n - F_n) < \epsilon/2^{n+1}$. If $G = \bigcup_n G_n$, and if $F = \bigcup_{n \leq n_n} F_n$, with n_0 so chosen that

$$P\left(\bigcup_n F_n - F\right) < \frac{\epsilon}{2},$$

then $F \subset \bigcup_n A_n \subset G$ and $P(G - F) < \epsilon$. Thus $\mathcal{G}$ is closed under the formation of countable unions; since it is obviously closed under complementation, $\mathcal{G}$ is a $\sigma-$field.

(2.5) implies that P is completely determined by the values of $P(F)$ for closed sets F. The next theorem shows that P is also determined by the values of $\int f dP$ for bounded, continuous f. The proof depends on approximating the indicator I_F by such an f, and the function $f(x) = (1 - \rho(x, F)/\epsilon)^+$ works. It is bounded, and it is continuous, even uniformly continuous, because $|f(x) - f(y)| \leq \rho(x, y)/\epsilon$. And $x \in F$ implies $f(x) = 1$, while $x \notin F^\epsilon$ implies $\rho(x, F) \geq \epsilon$ and hence $f(x) = 0$. Therefore,

$$I_F(x) \leq f(x) = (1 - \rho(x, F)/\epsilon)^+ \leq I_{F^\epsilon}(x). \tag{5}$$

Theorem 2.6 *Probability measures P and Q on $\mathcal{S}$ coincide if and only if $\int f dP = \int f dQ$ for all bounded, uniformly continuous real functions f.*

Proof. ($\Rightarrow$) Trivial.
($\Leftarrow$) For the bounded, uniformly continuous f of (5), $P(F) \leq \int f dP = \int f dQ \leq Q(F^\epsilon)$. Letting $\epsilon \downarrow 0$ gives $P(F) = Q(F)$, provided F is closed. By symmetry and (2.5), $P = Q$.

The following notion of tightness plays a fundamental role both in the theory of weak convergence and in its applications. A probability measure P on $(S, \mathcal{S})$ is *tight* if fir each ϵ there exists a compact set K such that $P(K) \geq 1-\epsilon$. By (2.5), P is tight if and only if for each $A \in \mathcal{S}$

$$P(A) = \sup\{P(K) : K \subset A, \quad K \text{ is compact.}\}$$

Theorem 2.7 *If S is separable and complete, then each probability measure on $(S, \mathcal{S})$ is tight.*

Proof. Since S is separable, there is, for each k, a sequence $A_{k1}, A_{k2}, \ldots$ of open $1/k-$balls covering S. Choose n_k large enough that

$$P\Big(\bigcup_{i \leq n_k} A_{ki}\Big) > 1 - \frac{\epsilon}{2^k}.$$

By the completeness hypothesis, the totally bounded set

$$\bigcap_{k \geq 1} \bigcup_{i \leq n_k} A_{ki}$$

has compact closure K. But clearly $P(K) > 1-\epsilon$. This completes the proof.

The following theorem provides useful conditions equivalent to weak convergence; any of them could serve as the definition. A set A in $\mathcal{S}$ whose boundary ∂A satisfies $P(\partial A) = 0$ is called $P-$*continuity set*. Let P_n, P be probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

Theorem 2.8 ***(The Portmanteau Theorem)*** *The followings are equivalent.*

1. *for bounded and continuous f*

$$\lim_{n\to\infty} \int f dP_n = \int f dP$$

2. *for closed set F*

$$\limsup_{n\to\infty} P_n(F) \leq P(F)$$

3. *For open set G*

$$\liminf_{n\to\infty} P_n(G) \geq P(G)$$

4. *for all set A with $P(\partial A) = 0$*

$$\lim_{n\to\infty} P_n(A) = P(A)$$

Proof. (1)$\to$ (2)
Let $f_k(x) = \psi_k(\rho(x, F))$. First of all, we know that

$$f_k(x) \searrow 1_F(x)$$

Then, for any $\delta > 0$, there exists K such that for all $k \geq K$

$$\begin{aligned}
\limsup_{n\to\infty} P_n(F) &= \limsup_{n\to\infty} \int 1_F dP_n \\
&\leq \limsup_{n\to\infty} \int f_k dP_n \\
&= \lim_{n\to\infty} \int f_k dP_n \\
&= \int f_k dP \\
&\leq P(F) + \delta
\end{aligned}$$

The last inequality follows from the fact that

$$\int_F f_n dP \searrow P(F)$$

As a result, for all $\delta > 0$, we have

$$\limsup_{n\to\infty} P_n(F) \leq P(F) + \delta$$

(2) $\to$ (3)

Let $G = F^c$. Then, it follows directly.

(4) $\to$ (1)
Let f be approximation of f_n which satisfy (4). Then it follows.

Theorem 2.9 *A necessary and sufficient condition for $P_n \Rightarrow P$ is that each subsequence $\{P_{n_i}\}$ contain a further subsequence $\{P_{n_{i(m)}}\}$ converging weakly to P.*

Proof. The necessary is easy. As for sufficiency, if P_n doesn't converge weakly to P, then there exists some bounded and continuous f such that $\int f dP_n$ doesn't converge to $\int f dP$. But then, for some positive ϵ and some subsequence P_{n_i},

$$\left| \int f dP_{n_i} - \int f dP \right| > \epsilon$$

for all i, and no further subsequence can converge weakly to P.

Suppose that h maps $\mathcal{X}$ into another metric space $\mathcal{X}'$, with metric ρ' and Borel $\sigma-$field $\mathcal{B}(\mathcal{X})$. If h is measurable $\mathcal{X}/\mathcal{X}'$, then each probability P on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ induces on $(\mathcal{X}', \mathcal{B}(\mathcal{X}'))$ a probability $P \circ h^{-1}$ defined as usual by $P \circ h^{-1}(A) = P(h^{-1}(A))$. We need conditions under which $P_n \Rightarrow P$ implies $P_n \circ h^{-1} \Rightarrow P \circ h^{-1}$. One such condition is that h is continuous: If f is bounded and continuous on $\mathcal{X}'$, then fh is bounded and continuous on $\mathcal{X}$, and by change of variable, $P_n \Rightarrow P$ implies

$$\int_{\mathcal{X}'} f(y) P_n \circ h^{-1}(dy) = \int_{\mathcal{X}} f(h(x)) P_n(dx) \to \int_{\mathcal{X}} f(h(x)) P(dx) = \int_{\mathcal{X}'} f(y) P \circ h^{-1}(dy) \tag{6}$$

Theorem 2.10 *Let $(\mathcal{X}, \rho)$ and $(\mathcal{X}', \rho')$ be two polish space and*

$$h : \mathcal{X} \to \mathcal{X}'$$

with $P(D_h) = 0$. Then, $P_n \Rightarrow P$ implies

$$P_n \circ h^{-1} \Rightarrow P \circ h^{-1}.$$

Proof. Since

$$h^{-1}(F) \subset \overline{h^{-1}(F)} \subset D_h \cup h^{-1}(F),$$

$$\begin{aligned}
\limsup_{n\to\infty} P_n(h^{-1}(F)) &\leq \limsup_{n\to\infty} P_n(\overline{h^{-1}(F)}) \\
&\leq P\Big(D_h \cup h^{-1}(F)\Big) \\
&\leq P(h^{-1}(F)) \text{ (since } D_h \text{ is set of zero)}
\end{aligned}$$

Therefore, for all closed set F,

$$\limsup_{n\to\infty} P_n \circ h^{-1}(F) \leq P \circ h^{-1}(F)$$

and hence, by (2.8), the proof is completed.

Let X_n and X are random variables($\mathcal{X}-$ valued). Then, we say $X_n \to_{\mathcal{D}} X$ if $P \circ X_n^{-1} \Rightarrow P \circ X^{-1}$.

Observation. If $X_n \to_{\mathcal{D}} X$ and $\rho(X_n, Y_n) \to_p 0$, then

$$Y_n \to_{\mathcal{D}} X$$

Remark. We use the following property of limsup and liminf.

$$\limsup_{n\to\infty}(a_n + b_n) \le \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n$$

and

$$\liminf_{n\to\infty}(a_n + b_n) \ge \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n$$

Proof. Consider closed set F. Let $F^\epsilon = \{x : \rho(x, F) \le \epsilon\}$. Then, $F^\epsilon \searrow F$ as $\epsilon \to 0$ and

$$\begin{aligned} \{X_n \notin F^\epsilon\} &= \{\omega : X_n(\omega) \notin F^\epsilon\} \\ &= \{\omega : \rho(X_n(\omega), F) > \epsilon\} \end{aligned}$$

Therefore,

$$\begin{aligned} \omega \in \{X_n \notin F^\epsilon\} \cap \{\rho(X_n, Y_n) < \epsilon\} &\Rightarrow \rho(X_n(\omega), F) > \epsilon \text{ and } \rho(X_n(\omega), Y_n(\omega)) < \epsilon \\ &\Rightarrow \\ &\Rightarrow \rho(Y_n(\omega), F) > 0 \text{ (draw graph.)} \\ &\Rightarrow Y_n(\omega) \notin F \\ &\Rightarrow \omega \in \{Y_n \notin F\} \end{aligned}$$

Thus,

$$\{X_n \notin F^\epsilon\} \cap \{\rho(X_n, Y_n) < \epsilon\} \subset \{Y_n \notin F\}$$

Therefore,

$$P(Y_n \in F) \le P(\rho(X_n, Y_n) > \epsilon) + P(X_n \in F^\epsilon)$$

Let $P_n^Y = P \circ Y_n^{-1}$ and $P^X = P \circ X^{-1}$. Then, for all $\epsilon > 0$

$$\begin{aligned} \limsup_{n\to\infty} P_n(F) &= \limsup_{n\to\infty} P(Y_n \in F) \\ &\le \limsup_{n\to\infty} P(\underbrace{\rho(X_n, Y_n)}_{\to_p 0} > \epsilon) + \limsup_{n\to\infty} P(X_n \in F^\epsilon) \\ &= \limsup_{n\to\infty} P(X_n \in F^\epsilon) \\ &= P(X \in F^\epsilon) \text{ (since } X_n \Rightarrow_{\mathcal{D}} X) \end{aligned}$$

Therefore, for all closed set F, we have

$$\limsup_{n\to\infty} P_n^Y(F) \le P^X(F)$$

and hence, by (2.5),

$$P_n^Y \Longrightarrow P^X,$$

which implies $Y_n \Rightarrow_{\mathcal{D}} X$.

We say that a family of probability measure $\Pi \subset \mathcal{P}(\mathcal{X})$ is tight if given $\epsilon > 0$, there exists compact K_ϵ such that

$$P(K_\epsilon) > 1 - \epsilon \text{ for all } P \in \Pi$$

2.10.1 Prokhorov Theorem

Definition 2.4 Π *is relatively compact if for* $\{P_n\} \subset \Pi$*, there exists a subsequence* $\{P_{n_i}\} \subset \Pi$ *and probability measure* P*(not necessarily an element of* Π*) such that*

$$P_{n_i} \Rightarrow P$$

Even though $P_{n_i} \Rightarrow P$ makes no sense if $P(\mathcal{X}) < 1$, it is to be emphasized that we do require $P(\mathcal{X}) = 1$-we disallow any escape of mass, as discussed below. For the most part we are concerned with the relative compactness of sequences $\{P_n\}$; this means that every subsequence $\{P_{n_i}\}$ contains a further subsequence $\{P_{n_{i(m)}}\}$ such that $P_{n_{i(m)}} \Rightarrow P$ for some probability measure P.

Example. Suppose we know of probability measures P_n and P on $(C, \mathcal{C})$ that the finite-dimensional distributions of P_n converges weakly to those of P: $P_n\pi^{-1}_{t_1,\ldots,t_k} \Rightarrow P\pi^{-1}_{t_1,\ldots,t_k}$ for all k and all $t_1, ..., t_k$. Notice that P_n need not converge weakly to P. Suppose, however, that we also know that $\{P_n\}$ is relatively compact . Then each $\{P_n\}$ contains some $\{P_{n_{i(m)}}\}$ converging weakly to some Q. Since the mapping theorem then gives $P_{n_{i(m)}}\pi^{-1}_{t_1,\ldots,t_k} \Rightarrow Q\pi^{-1}_{t_1,\ldots,t_k}$ and since $P_n\pi^{-1}_{t_1,\ldots,t_k} \Rightarrow P\pi^{-1}_{t_1,\ldots,t_k}$ by assumption, we have $P\pi^{-1}_{t_1,\ldots,t_k} = Q\pi^{-1}_{t_1,\ldots,t_k}$ for all $t_1, ..., t_k$. Thus the finite-dimensional distributions of P and Q are identical, and since the class $\mathcal{C}_f$ of finite-dimensional sets is a separating class, $P = Q$. Therefore, each subsequence contains a further subsequence converging weakly to P-not to some fortuitous limit, but specifically to P. It follows by (2.9) that the entire sequence $\{P_n\}$ converges weakly to P. Therefore: If $\{P_n\}$ is relatively compact and the finite-dimensional distributions of P_n converge weakly to those of P, then $P_n \Rightarrow P$. This idea provides a powerful method for proving weak convergence in C and other function spaces. Not that, if $\{P_n\}$ does converge weakly to P, then it is relatively compact, so that this is not too strong a condition.

Theorem 2.11 *Suppose* $(\mathcal{X}, \rho)$ *is a Polish space and* $\Pi \subset \mathcal{P}(\mathcal{X})$ *is relatively compact, then it is tight.*

This is the converse half of Prohorov's theorem. It contains (2.7), since a Π consisting of a single measure is obviously relatively compact. Although this converse puts things in perspective, the direct half is what is essential to the applications.

Proof. Consider open sets, $G_n \nearrow \mathcal{X}$. For each $\epsilon > 0$ there exists n, such that for all $P \in \Pi$

$$P(G_n) > 1 - \epsilon$$

Otherwise, for each n we can find P_n such that $P_n(G_n) < 1 - \epsilon$. Then by by relative compactness, there exists $\{P_{n_i}\} \subset \Pi$ and probability measure $Q \in \Pi$

such that $P_{n_i} \Rightarrow Q$. Thus,

$$\begin{aligned}
Q(G_n) &\leq \liminf_{i\to\infty} P_{n_i}(G_n) \\
&\leq \liminf_{i\to\infty} P_{n_i}(G_{n_i}) \text{ (since } n_i \geq n \text{ and hence } G_n \subset G_{n_i}) \\
&< 1-\epsilon
\end{aligned}$$

Since $G_n \nearrow \mathcal{X}$,

$$\begin{aligned}
1 &= Q(\mathcal{X}) \\
&= \lim_{n\to\infty} Q(G_n) \\
&< 1-\epsilon
\end{aligned}$$

which is contradiction. Let A_{k_m}, $m = 1, 2, ...$ be open ball with radius $\frac{1}{k_m}$, covering $\mathcal{X}$(separability). Then, there exists n_k such that for all $P \in \Pi$

$$P\Big(\bigcup_{i\leq n_k} A_{k_i}\Big) > 1 - \frac{\epsilon}{2^k}$$

Then, let

$$K_\epsilon = \overline{\bigcap_{k\geq 1}\bigcup_{i\leq n_k} A_{k_i}}$$

where $\bigcap_{k\geq 1}\bigcup_{i\leq n_k} A_{k_i}$ is totally bounded set. Then, K_ϵ is compact(completeness), and $P(K_\epsilon) > 1-\epsilon$.

Remark. The last inequality is from the following. Let B_i be such that $P(B_i) > 1 - \frac{\epsilon}{2^i}$. Then,

$$\begin{aligned}
P(B_i) > 1 - \frac{\epsilon}{2^i} &\Rightarrow P(B_i^c) \leq \frac{\epsilon}{2^i} \\
&\Rightarrow P(\cup_{i=1}^\infty B_i^c) \leq \epsilon \\
&\Rightarrow P(\cap_{i=1}^\infty B_i) > 1-\epsilon
\end{aligned}$$

2.11 Tightness and compactness in weak convergence

Theorem 2.12 *If Π is tight, then for $\{P_n\} \subset \Pi$, there exists a subsequence $\{P_{n_i}\} \subset \{P_n\}$ and probability measure P such that*

$$P_{n_i} \Rightarrow P$$

Proof. Choose compact $K_1 \subset K_2 \subset ...$ such that for all n

$$P_n(K_u) > 1 - \frac{1}{u}$$

from tightness condition. Look at $\bigcup_u K_u$. We know that there exists a countable family of open sets, $\mathcal{A}$, such that if $x \in \cup_u K_u$ and G is open, then

$$x \in A \subset \bar{A} \subset G$$

for some $A \in \mathcal{A}$. Let

$$\mathcal{H} = \{\emptyset\} \cup \{ \text{ finite union of sets of the form } \bar{A} \cap K_u \, u \geq 1, A \in \mathcal{A}\}$$

Then, $\mathcal{H}$ is a countable family. Using Cantor Diagonalization method, there exists $\{n_i\}$ such that for all $H \in \mathcal{H}$,

$$\alpha(H) = \lim_{i \to \infty} P_{n_i}(H)$$

Our aim is to construct a probability measure P such that for all open set G

$$P(G) = \sup_{H \subset G} \alpha(H) \tag{7}$$

Suppose we showed (7) above. Consider an open set G. Then, for $\epsilon > 0$, there exists $H_\epsilon \subset G$ such that

$$\begin{aligned}
P(G) &= \sup_{H \subset G} \alpha(H) \\
&< \alpha(H_\epsilon) + \epsilon \\
&= \lim_i P_{n_i}(H_\epsilon) + \epsilon \\
&= \liminf_i P_{n_i}(H_\epsilon) + \epsilon \\
&\leq \liminf_i P_{n_i}(G) + \epsilon
\end{aligned}$$

and hence, for all open set G,

$$P(G) \leq \liminf_i P_{n_i}(G),$$

which is equivalent to $P_{n_i} \Rightarrow P$.

Observe $\mathcal{H}$ is closed under finite union and

1. $\alpha(H_1) \le \alpha(H_2)$ if $H_1 \subset H_2$
2. $\alpha(H_1 \cup H_2) = \alpha(H_1) + \alpha(H_2)$ if $H_1 \cap H_2 = \emptyset$
3. $\alpha(H_1 \cup H_2) \le \alpha(H_1) + \alpha(H_2)$

Define for open set G

$$\beta(G) = \sup_{H \subset G} \alpha(H) \tag{8}$$

Then, $\alpha(\emptyset) = \beta(\emptyset) = 0$ and β is monotone.

Define for $M \subset \mathcal{X}$

$$\gamma(M) = \inf_{M \subset G} \beta(G)$$

Then,

$$\begin{aligned} \gamma(M) &= \inf_{M \subset G} \beta(G) \\ &= \inf_{M \subset G} \big(\sup_{H \subset G} \alpha(H) \big) \\ \gamma(G) &= \inf_{G \subset G'} \beta(G') \\ &= \beta(G) \end{aligned}$$

M is $\gamma-$measurable if for all $L \subset \mathcal{X}$

$$\gamma(L) \ge \gamma(M \cap L) + \gamma(M \cap L^2)$$

We shall prove that γ is outer measure, and hence open and closed sets are $\beta-$measurable.

$\gamma-$measurable sets M form a $\sigma-$field, $\mathcal{M}$ and

$$\gamma\Big|_{\mathcal{M}}$$

is a measure.

Claim. Each closed set is in $\mathcal{M}$ and

$$P = \gamma\Big|_{\mathcal{B}(\mathcal{X})}$$

so that for open set G

$$P(G) = \gamma(G) = \beta(G)$$

Note that P is a probability measure. K_u has finite covering of sets in $\mathcal{A}$ when $K_u \in \mathcal{H}$.

$$1 \ \ge \ P(\mathcal{X})$$

$$
\begin{aligned}
&= \beta(\mathcal{X}) \\
&= \sup_u \alpha(K_u) \\
&= \sup_u \left(1 - \frac{1}{u}\right) \\
&= 1
\end{aligned}
$$

Step 1. If $F \subset G$ (F is closed and G is open), and if $F \subset H$ for some $H \in \mathcal{H}$ then there exists some $H_0 \in \mathcal{H}$ such that

$$H \subset H_0 \subset G$$

Proof. Consider $x \in F$ and $A_x \in \mathcal{A}$ such that

$$x \in A_x \subset \bar{A}_x \subset G$$

Since F is closed subset of compact, F is compact. Since A_x covers F, there exists finite subcovers $A_{x_1}, A_{x_2}, ..., A_{x_k}$. Take

$$H_0 = \bigcup_{i=1}^{k} \left(\bar{A_{x_i}} \cap K_u\right)$$

Step 2 β is finitely sub-additive on open set. Suppose $H \subset G_1 \cup G_2$, and $H \in \mathcal{H}$. Let

$$
\begin{aligned}
F_1 &= \{x \in H : \rho(x, G_1^c) \geq \rho(x, G_2^c)\} \\
F_2 &= \{x \in H : \rho(x, G_2^c) \geq \rho(x, G_1^c)\}
\end{aligned}
$$

If $x \in F_1$ but not in G_1, then $x \in G_2$, and hence $x \in H$. Suppose x is not in G_2. Then, $x \in G_2^c$ and hence, $\rho(x, G_2^c) > 0$. Therefore,

$$
\begin{aligned}
0 &= \rho(x, G_1^c) \text{ (since } x \in G_1^c) \\
&< \underbrace{\rho(x, G_2^c)}_{>0}
\end{aligned}
$$

which contradicts that $x \in F_1$, and hence contradicts $\rho(x, G_1^c) \geq \rho(x, G_2^c)$. Similarly, if $x \in F_2$ but not in G_2, then $x \in G_1$. Therefore, $F_1 \subset G_1$ and $F_2 \subset G_2$. Since F_i's are closed, by Step 1, there exist H_1 and H_2 such that

$$F_1 \subset H_1 \subset G_1$$

and

$$F_2 \subset H_2 \subset G_2$$

Therefore,

$$
\begin{aligned}
\alpha(H) &\leq \alpha(H_1) + \alpha(H_2) \\
\beta(G) &\leq \beta(G_1) + \beta(G_2)
\end{aligned}
$$

Step 3. β is countably sub-additive on open set $H \subset \bigcup_n G_n$ where G_n is open set.

Since H is compact(union of compacts), there exist a finite subcovers, i.e., there exists n_0 such that

$$H \subset \bigcup_{n \leq n_0} G_n$$

and

$$\begin{aligned} \alpha(H) &\leq \beta(H) \\ &\leq \beta\Big(\bigcup_{n \leq n_0} G_n\Big) \\ &= \sum_{n \leq n_0} \beta(G_n) \\ &= \sum_n \beta(G_n) \end{aligned}$$

Therefore,

$$\begin{aligned} \beta\Big(\bigcup_n G_n\Big) &= \sup_{H \subset \cup_n G_n} \alpha(H) \\ &\leq \sup_{H \subset \cup_n G_n} \sum_n \beta(G_n) \\ &= \sum_n \beta(G_n) \end{aligned}$$

Step 4. γ is an outer measure. We know γ is monotonic by definition and is countably sub-additive. Given $\epsilon > 0$ and subsets $\{M_n\} \subset \mathcal{X}$, choose open sets G_n, $M_n \subset G_n$ such that

$$\begin{aligned} \beta(G_n) &\leq \gamma(M_n) + \frac{\epsilon}{2^n} \\ \gamma\Big(\bigcup_n M_n\Big) &\leq \beta\Big(\bigcup_n G_n\Big) \\ &= \sum \beta(G_n) \\ &= \sum_n \gamma(M_n) + \epsilon \end{aligned}$$

Step 5. F is closed G is open.

$$\beta(G) \geq \gamma(F \cap G) + \gamma(F^c \cap G)$$

Choose $\epsilon > 0$ and $H_1 \in \mathcal{H}$, $H_1 \subset F^c \cap G$ such that

$$\alpha(H_1) > \beta(G \cap F^c) - \epsilon$$

Chose H_0 such that

$$\alpha(H_0) > \beta(H_1^c \cap G) - \epsilon$$

Then, $H_0, H_1 \subset G$, and $H_0 \cap H_1 = \emptyset$,

$$\begin{aligned}
\beta(G) &\geq \alpha(H_0 \cup H_1) \\
&= \alpha(H_0) + \alpha(H_1) \\
&> \beta(H_1^c \cap G) + \beta(F^c \cap G) - 2\epsilon \\
&\geq \gamma(F \cap G) + \gamma(F^c \cap G) - 2\epsilon
\end{aligned}$$

Step 6. If $F \in \mathcal{M}$ then F are all closed. If G is open and $L \subset G$, then,

$$\beta(G) \geq \gamma(F \cap L) + \gamma(F^c \cap L)$$

Then,

$$\begin{aligned}
& \inf \beta(G) \geq \inf \big(\gamma(F \cap L) + \gamma(F^c \cap L)\big) \\
\Longrightarrow \quad & \gamma(L) \geq \gamma(F \cap L) + \gamma(F^c \cap L)
\end{aligned}$$

3 Weak Convergence on $C[0,1]$ and $D[0,\infty)$

3.1 Structure of Compact sets in $C[0,1]$

Let $\mathcal{X}$ be complete separable metric space. We showed that Π is tight iff Π is relatively compact. Consider P_n is measure on $C[0,1]$ and let

$$\pi_{t_1,...,t_k}(x) = (x(t_1), ..., x(t_k))$$

and suppose that

$$P_n \circ \pi^{-1}_{t_1,...,t_k} \Longrightarrow P \circ \pi^{-1}_{t_1,...,t_k}$$

does not imply

$$P_n \Longrightarrow P$$

on $C[0,1]$. However, $P_n \circ \pi^{-1}_{t_1,...,t_k} \Rightarrow P \circ \pi^{-1}_{t_1,...,t_k}$ and $\{P_n\}$ is tight. Then, $P_n \Rightarrow P$.

Proof. Since tightness implies(as we proved), there exists a probability measure Q and subsequence $\{P_{n_i}\}$ such that

$$P_{n_i} \Longrightarrow Q$$

$$P \circ \pi^{-1}_{t_1,...,t_k} = Q \circ \pi^{-1}_{t_1,...,t_k}$$

giving $P = Q$. Hence all limit points of subsequences of P_n is P, i.e., $P_n \Rightarrow P$.

Arzela-Ascoli Theorem

Definition 3.1 *The uniform norm (or sup norm) assigns to real- or complex-valued bounded functions f defined on a set S the non-negative number*

$$||f||_\infty = ||f||_{\infty,S} = \sup\{|f(x)| : x \in S\}$$

This norm is also called the supremum norm, the Chebyshev norm, or the infinity norm. The name "uniform norm" derives from the fact that a sequence of functions $\{f_n\}$ converges to f under the metric derived from the uniform norm if and only if f_n converges to f uniformly.

Theorem 3.1 *The set $A \subset C[0,1]$ is relative compact in sup topology if and only if*

$$\text{(i)} \ \sup_{x \in A} |x(0)| < \infty$$

$$\text{(ii)} \ \lim_{\delta} \left(\sup_{x \in A} w_x(\delta) \right) = 0$$

Remark(Modulus of Continuity). Here

$$w_x(\delta) = \sup_{|s-t|\leq\delta} |x(t) - x(s)|$$

Proof.
Consider function

$$f : C[0,1] \to R$$

such that $f(x) = x(0)$.

Claim. f is continuous.

Proof of Claim. We want to show that for $\epsilon > 0$, there exists δ such that

$$||x-y||_\infty = \sup_{t\in A}\{|x(t) - y(t)|\} < \delta \longrightarrow |f(x) - f(y)| = |x(0) - y(0)| < \epsilon$$

Given $\epsilon > 0$, let $\delta = \epsilon$. Then, we are done.

Since A is compact, continuous mapping $x \mapsto x(0)$ is bounded. Therefore,

$$\sup_{x\in A} |x(0)| < \infty$$

$w_x\left(\frac{1}{n}\right)$ is continuous in x uniformly on A and hence

$$\lim_{n\to\infty} w_x\left(\frac{1}{n}\right) = 0$$

Suppose (i) and (ii) hold. Choose k large enough so that

$$\sup_{x\in A} w_x\left(\frac{1}{k}\right) = \sup_{x\in A}\Big(\sup_{|s-t|\leq\frac{1}{k}} |x(s) - x(t)|\Big)$$

is finite. Since

$$|x(t)| < |x(0)| + \sum_{i=1}^{k} \Big| x\big(\frac{it}{k}\big) - x\big(\frac{(i-1)t}{k}\big)\Big|$$

we have

$$\alpha = \sup_{0\leq t\leq 1}\Big(\sup_{x\in A} |x(t)|\Big) < \infty$$

Choose $\epsilon > 0$ and finite $\epsilon-$covering H of $[-\alpha, \alpha]$. Choose k large enough so that

$$w_x\left(\frac{1}{k}\right) < \epsilon$$

Take B to be finite set of polygonal functions on $C[0,1]$ that are linear on $\left[\frac{i-1}{k}, \frac{i}{k}\right]$ and takes the values in H at end points.

If $x \in A$ and $\left|x\left(\frac{1}{k}\right)\right| \leq \alpha$ so that there exists a point $y \in B$ such that

$$\left|x\left(\frac{i}{k}\right) - y\left(\frac{i}{k}\right)\right| < \epsilon, \quad i = 1, 2, ..., k$$

then

$$\left|y\left(\frac{i}{k}\right) - x(t)\right| < 2\epsilon \quad \text{for } t \in \left[\frac{i-1}{k}, \frac{i}{k}\right]$$

$y(t)$ is convex combination of $y\left(\frac{i}{k}\right), y\left(\frac{i-1}{k}\right)$, so it is within 2ϵ of $x(t)$, $\rho(x,y) < 2\epsilon$, B is finite, B is $2\epsilon-$covering of A.

Theorem 3.2 *$\{P_n\}$ is tight on $C[0,1]$ if and only if*

1. *For each $\eta > 0$, there exists a and n_0 such that for $n \geq n_0$*

$$P_n(\{x : |x(0)| > a\}) > \eta$$

2. *For each $\epsilon, \eta > 0$, there exists $0 < \delta < 1$ and n_0 such that for $n \geq n_0$*

$$P_n(\{x : w_x(\delta) \geq \epsilon\}) < \eta$$

Proof. Since $\{P_n\}$ is tight, given $\delta > 0$, choose K compact such that

$$P_n(K) > 1 - \eta$$

Note that by Arzela-Ascoli Theorem, for large a

$$K \subset \{x : |x(0)| \leq a\}$$

and for small δ

$$K \subset \{x : w_x(\delta) \leq \epsilon\}$$

Now, $C[0,1]$ is complete separable metric space. So each n, P_n is tight. Given $\eta > 0$, there exists a such that

$$P_n(\{x : |x(0)| > a\}) < \eta$$

and $\epsilon, \eta > 0$, there exists $\delta > 0$ such that

$$P_n(\{x : w_x(\delta) \geq \epsilon\}) < \eta$$

This happens for P_n where $n \leq n_0$ with n_0 is finite. Assume (i) and (ii) holds for all n. Given η, choose a so that

$$B = \{x : |x(0)| \leq a\},$$

a satisfies for all n

$$P_n(B) > 1 - \eta,$$

and choose δ_k such that

$$B_k = \left\{ x : w_x(\delta_k) \leq \frac{1}{k} \right\},$$

with

$$P_n(B_k) > 1 - \frac{\eta}{2^k}$$

Let $K = \overline{A}$ where

$$A = B \cap \left(\bigcap_k B_k \right)$$

K is compact, and by Arzelar-Ascoli theorem,

$$P_n(K) > 1 - 2\eta$$

3.2 Invariance principle for sums of i.i.d random variables

Let X_i's be i.i.d with $EX_i = 0$ and $EX_i^2 = \sigma^2$. Define

$$W_t^n(\omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) + (nt - [nt]) \frac{1}{\sigma^2\sqrt{n}} X_{[nt]+1}$$

Consider linear interpolation

$$W_{\frac{k}{n}} = \frac{S_k}{\sqrt{n}} \quad \text{where } W^n \in C[0,1] \text{ a.e. } P_n$$

Let

$$\psi_{nt} = (nt - [nt]) \cdot \frac{X_{[nt+1]}}{\sigma\sqrt{n}}$$

Claim For fixed t, by Chevyshev's inequality, as $n \to \infty$

$$\psi_{nt} \to 0$$

Proof of Claim.

$$\begin{aligned}
P(|\psi_{nt}|\epsilon) &= P\left(|X_{[nt+1]}| > \frac{\sigma\sqrt{n}\epsilon}{(nt - [nt])} \right) \\
&\leq \frac{E|X_{[nt+1]}|^2}{\frac{\sigma^2 n\epsilon^2}{(nt-[nt])^2}} \\
&= \frac{(nt - [nt])^2}{n\epsilon^2} \\
&\leq \frac{1}{n\epsilon^2} \to 0 \qquad (9)
\end{aligned}$$

By CLT,

$$\frac{S_{[nt]}}{\sigma\sqrt{[nt]}} \Rightarrow_{\mathcal{D}} N(0,1)$$

Since $\frac{[nt]}{n} \to t$, by CLT,

$$\frac{S_{[nt]}}{\sigma\sqrt{n}} = \frac{S_{[nt]}}{\sigma\sqrt{[nt]}} \times \frac{\sqrt{[nt]}}{\sqrt{n}} \Rightarrow_{\mathcal{D}} \sqrt{t}Z$$

by Slutsky's equation. Therefore,

$$W_t^n \Rightarrow_{\mathcal{D}} \sqrt{t}Z$$

Then,

$$\begin{aligned}
(W_s^n, W_t^n - W_s^n) &= \frac{1}{\sigma\sqrt{n}}\Big(S_{[ns]}, S_{[nt]} - S_{[ns]}\Big) + \Big(\psi_{ns}, \psi_{nt} - \psi_{ns}\Big) \\
&\Longrightarrow_{\mathcal{D}} (N_1, N_2)
\end{aligned}$$

Since $S_{[ns]}$ and $S_{[nt]} - S_{[ns]}$ are independent, N_1 and N_2 are independent normal with variance s and $t-s$. Thus,

$$\begin{aligned}
(W_s^n, W_t^n) &= (W_s^n, (W_t^n - W_s^n) + W_s^n) \\
&\Rightarrow_{\mathcal{D}} (N_1, N_1 + N_2)
\end{aligned}$$

We considered 2 dimensional. We can take $k-$dimensional. Similar argument shows that

$$(W_1^n, ..., W_k^n) \Rightarrow_{\mathcal{D}} \text{ finite dimensional distribution of Brownian Motion.}$$

Now, we have to show that P_n is tight. Recall Arzela-Ascoli Theorem.

Theorem 3.3 *$\{P_n\}$ is tight on $C[0,1]$ if and only if*

1. *For each $\eta > 0$, there exists a and n_0 such that for $n \geq n_0$*

$$P_n(\{x : |x(0)| > a\}) > \eta$$

2. *For each $\epsilon, \eta > 0$, there exists $0 < \delta < 1$ and n_0 such that for $n \geq n_0$*

$$P_n(\{x : w_x(\delta) \geq \epsilon\}) < \eta$$

Theorem 3.4 *Suppose* $0 = t_0 < t_1 < \cdots < t_\nu = 1$ *and*

$$\min_{1<i<\nu}(t_i - t_{i-1}) \geq \delta \tag{10}$$

Then for arbitrary x,

$$w_x(\delta) \leq 3 \max_{1 \leq i \leq \nu} \Big(\sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})| \Big) \tag{11}$$

and for any P on $C[0,1]$

$$P(x : w_x(\delta) \geq 3\epsilon) \leq \sum_{i=1}^{\nu} P(x : \sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})| \geq \epsilon) \tag{12}$$

Proof. Let m denote the maximum in (11), i.e.,

$$m = \max_{1 \leq i \leq \nu} \Big(\sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})| \Big)$$

If s, t lie in $I_i = [t_{i-1}, t_i]$. Then

$$\begin{aligned} |x(s) - x(t)| &\leq |x(s) - x(t_{i-1})| + |x(t) - x(t_{i-1})| \\ &\leq 2m \end{aligned}$$

Suppose s, t lie in adjoining intervals I_{i-1} and I_i. Then,

$$\begin{aligned} |x(s) - x(t)| &\leq |x(s) - x(t_{i-1})| + |x(t_i) - x(t_{i-1})| + |x(t) - x(t_i)| \\ &\leq 3m \end{aligned}$$

Since

$$\min_{1 < i < \nu} (t_i - t_{i-1}) \geq \delta \tag{13}$$

for s and t to be such that $|s - t| < \delta$, s and t should lie in the same interval or adjoining intervals. Therefore,

$$\begin{aligned} w_x(\delta) &= \sup_{|s-t| \leq \delta} |x(t) - x(s)| \\ &\leq \max \left\{ \sup_{s,t \in \text{same interval}} |x(t) - x(s)|, \sup_{s,t \in \text{adjoining interval}} |x(t) - x(s)| \right\} \\ &\leq 3m \end{aligned}$$

This proves (11). Note that if $X \geq Y$, then

$$P(X > a) \geq P(Y > a)$$

Therefore,

$$\begin{aligned} P(x : w_x(\delta) > 3\epsilon) &\leq P\Big(3 \max_{1 \leq i \leq \nu} \Big(\sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})| \Big) > 3\epsilon\Big) \\ &= P\left(x : \max_{1 \leq i \leq \nu} \Big(\sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})| \Big) > \epsilon\right) \\ &= \sum_{i=1}^{\nu} P\Big(x : \sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})| > \epsilon\Big) \end{aligned}$$

This proves the theorem.

Condition (ii) of Arzela-Ascoli theorem holds if for each ϵ, η, there exists $\delta \in (0,1)$ and n_0 such that for all $n \geq n_0$

$$\frac{1}{\delta} P_n \left(x : \sup_{t \leq s \leq t+\delta} \left| x(s) - x(t) \right| \geq \epsilon \right) > \eta$$

Now apply Theorem (3.4) with $t_i = i\delta$ for $i < \nu = [1/\delta]$. Then by (12) we get condition (ii) of Arzela-Ascoli theorem holds.

3.3 Invariance principle for sums of stationary sequences

Definition 3.2 *$\{X_n\}$ is stationary if for any m,*

$$(X_{i_i}, ..., X_{i_k}) =^{\mathcal{D}} (X_{i_i+m}, ..., X_{i_k+m})$$

Lemma 3.1 *Suppose $\{X_n\}$ is stationary and W^n is defined as above. If*

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \lambda^2 P\Big(\max_{k\le n} |S_k| > \lambda\sigma\sqrt{n}\Big) = 0$$

then, W^n is tight.

Proof. Since $W_0^n = 0$, the condition (i) of Arzela-Ascoli theorem is satisfied. Let P_n is induced measure of W^n, i.e.,

$$P\Big(w(W^n, \delta) \ge \epsilon\Big) = P(w_{W^n}(\delta) \ge \epsilon)$$

We shall show that for all $\epsilon > 0$

$$\lim_{\delta\to 0} \limsup_{n\to\infty} P\Big(w(W^n, \delta) \ge \epsilon\Big) = 0$$

If

$$\min(t_t - t_{i-1}) \ge \delta$$

then, by Theorem (3.4)

$$P\Big(w(W^n, \delta) \ge 3\epsilon\Big) \le \sum_{i=1}^{\nu} P\Big(\sup_{t_{i-1}\le s\le t_i} \big|W_s^n - W_t^n\big| \ge \epsilon\Big)$$

Take $t_i = \frac{m_i}{n}$, $0 = m_0 < m_1 < \cdots < m_\nu = n$. W_t^n is polygonal and hence,

$$\sup_{t_{i-1}\le s\le t_i} \big|W_s^n - W_t^n\big| = \max_{m_{i-1}\le k\le m_i} \frac{|S_k - S_{m_{i-1}}|}{\sigma\sqrt{n}}$$

Therefore,

$$\begin{aligned}
P\Big(w(W^n, \delta) \ge 3\epsilon\Big) &\le \sum_{i=1}^{\nu} P\Big(\max_{m_{i-1}\le k\le m_i} \frac{|S_k - S_{m_{i-1}}|}{\sigma\sqrt{n}} \ge \epsilon\Big) \\
&\le \sum_{i=1}^{\nu} P\Big(\max_{m_{i-1}\le k\le m_i} |S_k - S_{m_{i-1}}| \ge \sigma\sqrt{n}\epsilon\Big) \\
&= \sum_{i=1}^{\nu} P\Big(\max_{k\le m_i - m_{i-1}} |S_k| \ge \sigma\sqrt{n}\epsilon\Big) \text{ (by stationarity)}
\end{aligned}$$

This inequality holds if

$$\frac{m_i}{n} - \frac{m_{i-1}}{n} \ge \delta \text{ for } 1 < i < \nu$$

Take $m_i = im$ for $0 \le i < \nu$ and $m_\nu = n$. For $i < \nu$ choose δ such that

$$m_i - m_{i-1} = m \ge n\delta$$

Let $m = [n\delta]$, $\nu = \left[\frac{n}{m}\right]$. Then,

$$m_\nu - m_{\nu-1} \le m$$

and

$$\nu = \left[\frac{n}{m}\right] \longrightarrow \frac{1}{\delta} \quad \text{where } \frac{1}{2\delta} < \frac{1}{\delta} < \frac{2}{\delta}$$

Therefore, for large n

$$\begin{aligned} P\Big(w(W^n, \delta) \ge 3\epsilon\Big) &\le \sum_{i=1}^{\nu} P\Big(\max_{k \le m_i - m_{i-1}} |S_k| \ge \sigma\sqrt{n}\epsilon\Big) \\ &\le \nu P\Big(\max_{k \le m} |S_k| \ge \sigma\sqrt{n}\epsilon\Big) \\ &\le \frac{2}{\delta} P\Big(\max_{k \le m} |S_k| \ge \sigma\sqrt{n}\epsilon\Big) \end{aligned}$$

Take $\lambda = \frac{\epsilon}{\sqrt{2\delta}}$. Then,

$$P\Big(w(W^n, \delta) \ge 3\epsilon\Big) \le \frac{4\lambda^2}{\epsilon^2} P\Big(\max_{k \le m} |S_k| \ge \lambda\sigma\sqrt{n}\Big)$$

By the condition of the Lemma, given $\epsilon, \eta > 0$, there exists $\lambda > 0$ such that

$$\frac{4\lambda^2}{\epsilon^2} \limsup_{n \to \infty} P\Big(\max_{k \le n} |S_k| > \lambda\sigma\sqrt{n}\Big) < \eta$$

Now, for fixed λ, δ, let $m \to \infty$ with $n \to \infty$

Look at X_k i.i.d. Then,

$$\lim_{\lambda \to 0} \limsup_{n \to \infty} P\Big(\max_{k \le n} |S_k| > \lambda\sigma\sqrt{n}\Big) = 0$$

We know that

$$P\Big(\max_{u \le m} |S_u| > \alpha\Big) \le 3 \max_{u \le m} P\Big(|S_u| > \frac{\alpha}{3}\Big)$$

To show

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} \lambda^2 P\Big(\max_{k \le n} |S_k| > \lambda\sigma\sqrt{n}\Big) = 0$$

we assume that X_i i.i.d. normal, and hence, $S_k/\sqrt{k}$ is asymptotically normal, N. Since we know

$$P(|N| > \lambda) \le \frac{EN^4}{\lambda^4} = \frac{3\sigma^4}{\lambda^4},$$

we have for $k \leq n$ $\left(\frac{\sqrt{n}}{\sqrt{k}} > 1\right)$

$$\begin{aligned} P(|S_k| > \lambda\sigma\sqrt{n}) &= P(\sqrt{k}|N| > \lambda\sigma\sqrt{n}) \\ &\leq \frac{3}{\lambda^4\sigma^4} \end{aligned}$$

k_λ is large and $k_\lambda \leq k \leq n$. Then,

$$\begin{aligned} P(|S_k| > \lambda\sigma\sqrt{n}) &\leq P(|S_k| > \lambda\sigma\sqrt{k}) \\ &\leq \frac{3}{\lambda^4} \end{aligned}$$

Also,

$$\begin{aligned} P(|S_k| > \lambda\sigma\sqrt{n}) &\leq \frac{E|S_k|^2/\sigma^2}{\lambda^2 n} \\ &\leq \frac{k_\lambda}{\lambda^2 n} \end{aligned}$$

and hence,

$$\max_{k\leq n} P(|S_k| > \lambda\sigma\sqrt{n}) \leq \max\left\{\frac{3}{\lambda^2}, \frac{k_\lambda}{\lambda^2 n}\right\}$$

3.4 Weak Convergence on Skorokhod Space

3.4.1 The Space $D[0,1]$

Let

$$x : [0,1] \to R$$

be right-continuous with left limit exists such that

1. for $0 \leq t < 1$
$$\lim_{s\searrow t} x(s) = x(t+) = x(t)$$

2. for $0 < t \leq 1$
$$\lim_{s\nearrow t} x(s) = x(t-)$$

We say that $x(t)$ has discontinuity of the first kind at t if left and right limit exist.

For $x \in D$ and $T \subset [0,1]$

$$w_x(T) = w(x,T) = \sup_{s,t\in T} |x(s) - x(t)|$$

We define Modulus of continuity

$$w_x(\delta) = \sup_{0\leq t\leq 1-\delta} w_x([t, t+\delta))$$

Lemma 3.2 *(D1) For each $x \in D$, and $\epsilon > 0$ there exist points $0 = t_0 < t_1 < \cdots < t_\nu = 1$ and $w_x([t_{i-1}, t_i)) < \epsilon$.*

Proof. Call t_i's above δ-sparse. If $\min_i\{(t_i - t_{i-1})\} \geq \delta$, define for $0 < \delta < 1$

$$w'_x(\delta) = w'(x, \delta) = \inf_{\{t_i\}} \max_{1 \leq i \leq \nu} w_x([t_{i-1}, t_i))$$

If we prove the above Lemma, we get $x \in D$

$$\lim_{\delta \to 0} w'_x(\delta) = 0$$

If $\delta < \frac{1}{2}$, we can split $[0, 1)$ into subintervals $[t_{i-1}, t_i)$ such that

$$\delta < (t_i - t_{i-1}) \leq 2\delta$$

and hence,

$$w'_x(\delta) \leq w_x(2\delta)$$

Let define jump function

$$j(x) = \sup_{0 \leq t \leq 1} |x(t) - x(t-)|$$

We shall prove that

$$w_x(\delta) \leq 2w'_x(\delta) + j(x)$$

Choose $\delta-$sparse sequence $\{t_i\}$ such that

$$w_x([t_{i-1}, t_i)) < w'_x(\delta) + \epsilon$$

We can do this from the definition

$$w'_x(\delta) = w'(x, \delta) = \inf_{\{t_i\}} \max_{1 \leq i \leq \nu} w_x([t_{i-1}, t_i))$$

If $|s - t| < \delta$, then $s, t \in [t_{i-1}, t_i)$ or belongs to adjoining intervals. Then,

$$|x(s) - x(t)| \begin{cases} w'_x(\delta) + \epsilon, & \text{if } s, t \text{ belong to the same interval;} \\ 2w'_x(\delta) + \epsilon + j(x), & \text{if } s, t \text{ belong to adjoining intervals.} \end{cases}$$

If x is continuous, $j(x) = 0$ and hence,

$$w_x(\delta) \leq 2w'_x(\delta)$$

3.4.2 Skorokhod Topology

Let Λ be the class of strictly increasing function on $[0, 1]$ and $\lambda(0) = 0, \lambda(1) = 1$. Define

$$d(x, y) = \inf\{\epsilon : \exists \lambda \in \Lambda \text{ such that } \sup_t |\lambda(t) - t| < \epsilon \text{ and } \sup_t |x(\lambda(t)) - y(t)| < \epsilon\}$$

$d(x, y) = 0$ implies there exists $\lambda_n \in \Lambda$ such that $\lambda_n(t) \to t$ uniformly and $x(\lambda(t)) \to y(t)$ uniformly. Therefore,

$$\begin{aligned}
||\lambda - I|| &= \sup_{t\in[0,1]} |\lambda(t) - t| \\
||x - y \circ \lambda|| &= \sup_{t\in[0,1]} |x(t) - y(\lambda(t))| \\
d(x, y) &= \inf_{\lambda} \Big(||\lambda - I|| \vee ||x - y \circ \lambda|| \Big)
\end{aligned}$$

If $\lambda(t) = t$, then

1. $d(x, y) = \sup |x(t) - y(t)| < \infty$ since we showed $|x(s) - x(t)| \le w'_x(\delta) < \infty$.
2. $d(x, y) = d(y, x)$.
3. $d(x, y) = 0$ only if $x(t) = y(t)$ or $x(t) = y(t-)$.

If $\lambda_1, \lambda_2 \in \Lambda$ and $\lambda_1 \circ \lambda_2 \in \Lambda$

$$||\lambda_1 \circ \lambda_2 - I|| \le ||\lambda_1 - I|| + ||\lambda_2 - I||$$

If $\lambda_1, \lambda_2 \in \Lambda$, then the followings hold:

1. $\lambda_1 \circ \lambda_2 \in \Lambda$
2. $||\lambda_1 \circ \lambda_2 - I|| \le ||\lambda_1 - I|| + ||\lambda_2 - I||$
3. $||x - z \circ (\lambda_1 \circ \lambda_2)|| \le ||x - y \circ \lambda_2|| + ||y - z \circ \lambda_1||$
4. $d(x, z) \le d(x, y) + d(y, z)$

Therefore, Skorokhod Topology is given by d.

Remark. d_0 is equivalent to d, but (D, d_0) is complete.

Choose $\lambda \in \Lambda$ near identity. Then for t, s close, $\frac{\lambda(t)-\lambda(s)}{t-s}$ is close to 1. Therefore,

$$||\lambda||^0 = \sup_{s<t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right| \in (0, \infty)$$

3.5 Metric on $D[0, 1]$ to make it complete

Let $\lambda \in \Lambda$ (λ is non-decreasing, $\lambda(0) = 0$, and $\lambda(1) = 1$). Recall

$$||\lambda||^0 = \sup_{s<t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right| \in (0, \infty)$$

Consider d^0

$$\begin{aligned} d^0(x,y) &= \inf\{\epsilon > 0 : \exists \lambda \in \Lambda \text{ with } ||\lambda||^0 < \epsilon \text{ and } \sup_t |x(t) - y(\lambda(t))| < \epsilon\} \\ &= \inf_{\lambda \in \Lambda} \{||\lambda||^0 \vee ||x - y \circ \lambda||\} \end{aligned}$$

since, for $u > 0$,

$$|u - 1| \leq e^{|\log u|} - 1,$$

we have

$$\begin{aligned} \sup_{0 \leq t \leq 1} |\lambda(t) - t| &= \sup_{0 \leq t \leq 1} t\Big|\frac{\lambda(t) - \lambda(0)}{t - 0} - 1\Big| \\ &= e^{||\lambda||^0} - 1 \end{aligned}$$

For any ν, $\nu \leq e^\nu - 1$, and hence

$$d(x,y) \leq e^{d^0(x,y)} - 1.$$

Thus, $d^0(x_n, y) \to 0$ implies $d(x_n, y) \to 0$.

Lemma 3.3 *(**D2**) If $x, y \in D[0,1]$ and $d(x,y) < \delta^2$, then $d^0(x,y) \leq 4\delta + w'_x(\delta)$.*

Proof. Take $\epsilon < \delta$ and $\{t_i\}$ $\delta-$sparse with

$$w_x([t_{i-1}, t_i)) < w'_x(\delta) + \epsilon \quad \forall i$$

We can do this from definition of $w'_x(\delta)$. Choose $\mu \in \Lambda$ such that

$$\sup_t |x(t) - y(\mu(t))| = \sup_t |x(\mu^{-1}(t)) - y(t)| < \delta^2 \tag{14}$$

and

$$\sup_t |\mu(t) - t| < \delta^2 \tag{15}$$

This follows from $d(x,y) < \delta^2$. Take λ to agree with μ at points t_i and linear between.

$\mu^{-1} \circ \lambda$ fixes t_i and is increasing in t. Also, $(\mu^{-1} \circ \lambda)(t)$ lies in the same interval $[t_{i-1}, t_i)$. Thus, from (14),

$$\begin{aligned} |x(t) - y(\lambda(t))| &\leq \underbrace{|x(t) - x((\mu^{-1} \circ \lambda)(t))|}_{\leq w'_x(\delta) + \epsilon} + \underbrace{|x((\mu^{-1} \circ \lambda)(t)) - y(\lambda(t))|}_{< \delta^2} \quad \text{(by Triangle Inequality)} \\ &= w'_x(\delta) + \epsilon + \delta^2 \end{aligned}$$

$\delta < \frac{1}{2} < 4\delta + w'_x(\delta)$. λ agrees with μ at t_i's. Then by (15) and $(t_i - t_{i-1}) > \delta$ ($\delta-$sparse)

$$\begin{aligned} |(\lambda(t_i) - \lambda(t_{i-1})) - (t_i - t_{i-1})| &< 2\delta^2 \\ &< 2\delta(t_i - t_{i-1}) \end{aligned}$$

and

$$|(\lambda(t) - \lambda(s)) - (t - s)| \leq 2\delta|t - s|$$

for $t, s \in [t_{i-1}, t_i)$ by polygonal property. Now, we take a care of adjoining interval. For u_1, u_2, u_3

$$|(\lambda(u_3)-\lambda(u_1))-(u_3-u_1)| \leq |(\lambda(u_3)-\lambda(u_2))-(u_3-u_2)|+|(\lambda(u_2)-\lambda(u_1))-(u_2-u_1)|$$

If t and s are in adjoining intervals, we get the same bound. Since for $u < \frac{1}{2}$

$$|\log(1 \pm u)| \leq 2u,$$

we have

$$\log(1 - 2\delta) \leq \log \frac{\lambda(t) - \lambda(s)}{t - s} \leq \log(1 + 2\delta)$$

Therefore,

$$||\lambda||^0 = \sup_{s<t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right| < 4\delta$$

and hence, d^0 and d are equivalent. Now, we shall show that D^0 is separable and is complete.

Consider $\sigma = \{s_u\}$ with $0 = s_0 < \cdots < s_k = 1$ and define $A_\sigma : D \to D$ by

$$(A_\sigma x)(t) = x(s_{u-1})$$

for $t \in [s_{u-1}, s_u)$ with $1 \leq u \leq k$ with $(A_\sigma x)(s_k) = x(1)$.

Lemma 3.4 *(D3) If* $\max(s_u - s_{u-1}) \leq \delta$, *then*

$$d(A_\sigma x, x) \leq \delta \vee w'_x(\delta)$$

Proof. Let $A_\sigma x \equiv \hat{x}$. Let $\zeta(t) = s_{u-1}$ if $t \in [s_{u-1}, s_u)$ with $\zeta(1) = s_k = 1$. Then, $\hat{x}(t) = x(\zeta(t))$. Given $\epsilon > 0$, find $\delta-$sparse set $\{t_i\}$ such that

$$w_x([t_{i-1}, t_i)) < w'_x(\delta) + \epsilon$$

for all i. Let $\lambda(t_i)$ be defined by

1. $\lambda(t_0) = s_0$
2. $\lambda(t_i) = s_v$ if $t_i \in [s_{v-1}, s_v)$ where

$$t_i - t_{i-1} > \delta \geq s_v - s_{v-1}$$

Then, $\lambda(t_i)$ is increasing. Now, extend it to $\lambda \in \Lambda$ by linear interpolation.

Claim.

$$\begin{aligned} ||\hat{x}(t) - x(\lambda^{-1}(t)) &= |x(\zeta(t)) - x(\lambda^{-1}(t))| \\ &< w'_x(\delta) + \epsilon \end{aligned}$$

Clearly, if $t = 0$, or $t = 1$, it is true. Let us look at $0 < t < 1$. First we observe that $\zeta(t), \lambda^{-1}(t)$ lie in the same interval $[t_{i-1}, t_i)$.(We will prove it.) This follows if we shows

$$t_j \leq \zeta(t) \text{ iff } t_j \leq \lambda^{-1}(t)$$

or equivalently,

$$t_j > \zeta(t) \text{ iff } t_j > \lambda^{-1}(t)$$

This is true for $t_j = 0$. Suppose $t_j \in (s_{v-1}, s_v]$ and $\zeta(t) = s_i$ for some i. By the definition of $\zeta(t)$ $t \leq \zeta(t)$ is equivalent to $s_v \leq t$. Since $t_j \in (s_{v-1}, s_v]$, $\lambda(t_j) = s_v$. This completes the proof.

3.6 Separability of Skorokhod space

d^0- convergence is stronger than $d-$ convergence.

Theorem 3.5 *(**D1**) The space (D, d) is separable, and hence, so is (D, d^0).*

Proof. Let B_k be the set of functions taking constant rational value on $\left[\frac{u-1}{k}, \frac{u}{k}\right]$ and taking rational value at 1. Then, $B = \cup_k B_k$ is countable. Given $x \in D$, $\epsilon > 0$, choose k such that $\frac{1}{k} < \epsilon$ and $w_x\left(\frac{1}{k}\right)$. Apply Lemma D3 with $\sigma = \left\{\frac{u}{k}\right\}$. Note that $A_\sigma x$ has finite many values and

$$d(x, A_\sigma x) < \epsilon$$

Since $A_\sigma x$ has finitely many real values, we can find $y \in B$ such that given $d(x, y) < \epsilon$,

$$d(A_\sigma x, y) < \epsilon$$

Now, we shall prove the completeness.

Proof) We take d^0-Cauchy sequence. Then it contains a d^0-convergent subsequence. If $\{x_k\}$ is Cauchy, then there exists $\{y_n\} = \{x_{k_n}\}$ such that

$$d^0(y_n, y_{n+1}) < \frac{1}{2^n}$$

There exists $\mu_n \in \Lambda$ such that

1. $||\mu_n||^0 < \frac{1}{2^n}$

2.

$$\begin{aligned} \sup_t |y_n(t) - y_{n+1}(\mu_n(t))| &= \sup_t |y_n(\mu_n^{-1}(t)) - y_{n+1}(t)| \\ &< \frac{1}{2^n} \end{aligned}$$

We have to find $y \in D$ and $\lambda_n \in \Lambda$ such that

$$||\lambda_n||^0 \to 0$$

and

$$y_n(\lambda_n^{-1}(t)) \to y(t)$$

uniformly.

Heuristic.(not a proof) Suppose $y_n(\lambda_n^{-1}(t)) \to y(t)$. Then, by (2), $y_n(\mu_n^{-1}(\lambda_{n+1}^{-1}(t)))$ is within $\frac{1}{2^n}$ of $y_{n+1}(\lambda_{n+1}^{-1}(t))$. So, $y_n(\lambda_n^{-1}(t)) \to y(t)$ uniformly.

Find λ_n such that

$$y_n(\mu_n^{-1}(\lambda_{n+1}^{-1}(t))) = y_n(\lambda_n^{-1}(t))$$

i.e.,

$$\mu_n^{-1} \circ \lambda_{n+1}^{-1} = \lambda_n^{-1}$$

Thus,

$$\begin{aligned} \lambda_n &= \lambda_{n+1}\mu_n \\ &= \lambda_{n+2}\mu_{n+1}\mu_n \\ &\vdots \\ &= \cdots \mu_{n+2}\mu_{n+1}\mu_n \end{aligned}$$

Proof. Since

$$e^u - 1 \leq 2u$$

for $0 \leq u \leq \frac{1}{2}$, we have

$$\sup_t |\lambda(t) - t| \leq e^{||\lambda||^0} - 1$$

Therefore,

$$\begin{aligned} \sup_t |(\mu_{n+m+1}\mu_{n+m} \cdots \mu_n)(t) - (\mu_{n+m}\mu_{n+m-1} \cdots \mu_n)(t)| &\leq \sup_s |\mu_{n+m+1}(s) - s| \\ &\leq 2||\mu_{n+m+1}||^0 \\ &= \frac{1}{2^{n+m}} \end{aligned}$$

For fixed n

$$(\mu_{n+m}\mu_{n+m-1}\cdots\mu_n)(t)$$

converges uniformly in t as n goes to ∞. Let

$$\lambda_n(t) = \lim_{m\to\infty}(\mu_{n+m}\mu_{n+m-1}\cdots\mu_n)(t)$$

Then, λ_n is continuous and non-decreasing with $\lambda_n(0) = 0$ and $\lambda_n(1) = 1$. We have to prove $||\lambda_n||^0$ is finite. Then, λ_n is strictly increasing.

$$\begin{aligned}
\left|\log\frac{(\mu_{n+m}\mu_{n+m-1}\cdots\mu_n)(t)-(\mu_{n+m}\mu_{n+m-1}\cdots\mu_n)(s)}{t-s}\right| &\leq ||\mu_{n+m}\mu_{n+m-1}\cdots\mu_n||^0 \\
&\quad (\text{ since } \lambda_n\in\Lambda, ||\lambda_n||^0<\infty) \\
&\leq ||\mu_{n+m}||^0+\cdots+||\mu_n||^0 \\
&\quad (\text{ since } ||\lambda_1\lambda_2||^0\leq||\lambda_1||^0+||\lambda_2||^0) \\
&< \frac{1}{2^{n-1}}
\end{aligned}$$

Let $m\to\infty$. Then, $||\lambda_n||^0 < \frac{1}{2^{n-1}}$ is finite, and hence, λ_n is strictly increasing. Now, by (2),

$$\begin{aligned}
\sup_t|y_n(\lambda_n^{-1}(t))-y_n(\lambda_{n+1}^{-1}(t))| &\leq \sup_s|y_n(s)-y_{n+1}(\mu_n(s))| \\
&< \frac{1}{2^n}
\end{aligned}$$

Therefore, $\{y_n(\lambda_n^{-1}(t))\}$ is Cauchy under supnorm and

$$y_n(\lambda_n^{-1}(t))\to y(t)\in D$$

and hence converges in d^0.

3.7 Tightness in Skorokhod space

We turn now the problem of characterizing compact sets in D. We will prove an analogue of the Arzelà-Ascoli theorem.

Theorem 3.6 *A necessary and sufficient condition for a set A to be relatively compact in the Skorohod topology is that*

$$\sup_{x\in A}||x||<\infty \tag{16}$$

and

$$\lim_{\delta\to0} sup_{x\in A}w'_x(\delta)=0. \tag{17}$$

Proof of Sufficiency. Let

$$\alpha = \sup_{x \in A} ||x||.$$

Given $\epsilon > 0$, choose a finite $\epsilon-$net H in $[-\alpha, \alpha]$ and choose δ so that $\delta < \epsilon$ and $w'_x(\delta) < \epsilon$ for all x in A. Apply Lemma 3 for any $\sigma = \{s_u\}$ satisfying $\max(s_u - s_{u-1}) < \delta$: $x \in A$ implies $d(x, A_\sigma x)$. Take B to be the finite set of y that assume on each $[s_{u-1}, s_u)$ a constant value from H and satisfy $y(1) \in H$. Since B contains a y for which $d(x, A_\sigma x)$, it is a finite $2\epsilon-$net for A in the sense of d. Thus A is totally bounded in the sense of d. But we must show that A is totally bounded in the sense of d^0, since this is the metric under which D is complete. Given (a new) ϵ, choose a new δ so that $0 < \delta \leq 1/2$ and so that $4\delta + w'_x(\delta) < \epsilon$ holds for all x in A. We have already seen that A is $d-$totally bounded, and so there exists a finite set B' that is a δ^2-net for A in the sense of d. But then, by Lemma 2, B' is an $\epsilon-$net for A in the sense of d^0.

The proof of necessity requires a lemma and a definition.

Definition 3.3 *In any metric space, f is upper semi-continuous at x, if for all $\epsilon > 0$, there exists $\delta > 0$ such that*

$$\rho(x, y) < \delta \Rightarrow f(y) < f(x) + \epsilon$$

Lemma 3.5 *For fixed δ, $w'(x, \delta)$ is upper-semicontinuous in x.*

Proof. Let x, δ, and ϵ be given. Let $\{t_i\}$ be a $\delta-$spars set such that $w_x[t_{i-1}, t_i) < w'_x(\delta) + \epsilon$ for each i. Now choose η small enough that $\delta + 2\eta < \min(t_i - t_{i-1})$ and $\eta < \epsilon$. Suppose that $d(x, y) < \eta$. Then for some λ in Λ, we have

$$\sup_t |y(t) - x(\lambda t)| < \eta$$

$$\sup_t |\lambda^{-1} t - t| < \eta$$

Let $s_i = \lambda^{-1} t_i$. Then $s_i - s_{i-1} > t_i - t_{i-1} - 2\eta > \delta$. Moreover, if s and t both lies in $[s_{i-1}, s_i)$, then λs and λt both lie in $[t_{i-1}, t_i)$, and hence $|y(s) - y(t)| < |x(\lambda s) - x(\lambda t)| + 2\eta \leq w'_x(\delta) + \epsilon + 2\eta$. Thus $d(x, y) < \eta$ implies $w'_y(\delta) < w'_x(\delta) + 3\epsilon$.

Definition 3.4 *($d-$bounded) A is $d-$bounded if diameter is bounded, i.e.,*

$$\textit{diameter}(A) = \sup_{x,y \in A} d(x, y) < \infty$$

Proof of Necessity in Theorem (3.6). If A^- is compact, then it is $d-$bounded, and since $\sup_t |x(t)|$ is the $d-$distance from x to the 0-function, (16) follows. By Lemma 1, $w'(x, \delta)$ goes to 0 with δ for each x. But since $w'(\cdot, \delta)$ is upper-semiconinutous, the convergence is uniform on compact sets.

Theorem (3.6), which characterizes compactness in D, gives the following result. Let $\{P_n\}$ be a sequence of probability measure on $(D, \mathcal{D})$.

Theorem 3.7 *The sequence $\{P_n\}$ is tight if and only if these two conditions hold:*
We have

$$\lim_{a\to\infty} \limsup_n P_n\Big(\{x : ||x|| \geq a\}\Big) = 0 \tag{18}$$

(ii) for each ϵ,

$$\lim_{\delta\to\infty} \limsup_n P_n\Big(\{x : w'_x(\delta) \geq \epsilon\}\Big) = 0. \tag{19}$$

Proof. Conditions (i) and (ii) here are exactly conditions (i) and (ii) of Azela-Ascoli theorem with $||x||$ in place of $|x(0)|$ and w' in place of w. Since D is separable and complete, a single probability measure on D is tight, and so the previous proof goes through.

3.8 The Space $D[0,\infty)$

Here we extend the Skorohod theory to the space $D_\infty = D[0,\infty)$ of cadlag functions on $[0,\infty)$, a space more natural than $D = D[0,1]$ for certain problems.

In addition to D_∞, consider for each $t > 0$ the space $D_t = D[0,t]$ of cadlag functions on $[0,t]$. All the definitions for D_1 have obvious analogues for D_t : $\sup_{s\le t}|x(s)|$, Λ_t, $||\lambda||_t^0$, d_t^0, d_t. And all the theorems carry over from D_1 to D_t in an obvious way. If x is an element of D_∞, or if x is an element of D_u and $t < u$, then x can also be regarded as an element of D_t by restricting its domain of definition. This new cadlag function will be denoted by the same symbol; it will always be clear what domain is intended.

One might try to define Skorohod convergence $x_n \to x$ in D_∞ by requiring that $d_t^0(x_n, x) \to 0$ for each finite, positive t. But in a natural theory, $x_n = I_{[0,1-1/n]}$ will converge to $x = I_{[0,1]}$ in D_∞, while $d_1^0(x_n, x) = 1$. The problem here is that x is discontinuous at 1, and the definition must accommodate discontinuities.

Lemma 3.6 *Let x_n and x be elements of D_u. If $d_u^0(x_n, x) \to 0$ and $m < u$, and if x is continuous at m, then $d_m^0(x_n, x) \to 0$.*

Proof. We can work with the metrics d_u and d_m. By hypothesis, there are elements λ_n of Λ_u such that

$$||\lambda_n - I||_u \to 0$$

and

$$||x_n - x\lambda_n||_u \to 0.$$

Given ϵ, choose δ so that $|t-m| \le 2\delta$ implies $|x(t) - x(m)| < \epsilon/2$. Now choose n_0 so that, if $n \ge n_0$ and $t \le u$, then $|\lambda_n t - t| < \delta$ and $|x_n(t) - x(\lambda_n t)| < \epsilon/2$. Then, if $n \ge n_0$ and $|t-m| <\le \delta$, we have $|\lambda_n t - m| \le |\lambda_n t - t| + |t-m| < 2\delta$ and hence $|x_n(t) - x(m)| \le |x_n(t) - x(\lambda_n t)| + |x(\lambda_n t) - x(m)| < \epsilon$. Thus

$$\sup_{|t-m|\le\delta} |x(t) - x(m)| < \epsilon, \qquad \sup_{|t-m|\le\delta} |x_n(t) - x(m)| < \epsilon, \qquad \text{for } n \ge n_0. \tag{20}$$

If
(i) $\lambda_n m < m$, let $p_n = m - \frac{1}{n}$;
(ii) $\lambda_n m > m$, let $p_n = \lambda^{-1}\left(m - \frac{1}{n}\right)$;
(iii) $\lambda_n m = m$, let $p_n = m$.

Then,
(i) $|p_n - m| = \frac{1}{n}$;
(ii) $|p_n - m| =\le |\lambda_n^{-1}(m - n^{-1}) - (m - n^{-1})| + \frac{1}{n}$;

(iii) $|p_n - m| = m$.
Therefore, $p_n \to m$. Since

$$|\lambda_n p_n - m| \leq |\lambda_n p_n - p_n| + |p_n - m|,$$

we also have $\lambda_n p_n \to m$. Define $\mu_n \in \Lambda_n$ so that $\mu_n t = \lambda_n t$ on $[0, p_n]$ and $\mu_n m = m$; and interpolate linearly on $[p_n, m]$. Since $\mu_n m = m$ and μ_n is linear over $[p_n, m]$, we have $|\mu_n t - t| \leq |\lambda_n p_n - p_m|$ there, and therefore, $\mu_n t \to t$ uniformly on $[0, m]$. Increase the n_0 of (20) so that $p_n > m - \delta$ and $\lambda_n p_n > m - \delta$ for $n \geq n_0$. If $t \leq p_n$, then $|x_n(t) - x(\mu_n t)| = |x_n(t) - x(\lambda_n t)| \leq ||x_n - x\lambda_n||_u$. On the other hand, if $p_n \leq t \leq m$ and $n \geq n_0$, then $m \geq t \geq p_n > m - \delta$ and $m \geq \mu_n t \geq \mu_n p_n = \lambda_n p_n > m - \delta$, and therefore, by (20), $|x_n(t) - x(\mu_n t)| \leq |x_n(t) - x(m)| + |x(m) - x(\mu_n t)| < 2\epsilon$. Thus, $|x_n(t) - x(\mu_n t)| \to 0$ uniformly on $[0, m]$.

The metric on D_∞ will be defined in terms of the metrics $d_m^0(x, y)$for integral m, but before restricting x and y to $[0, m]$, we transform them in shcu a way that they are continuous at m. Define

$$g_n(t) = \begin{cases} 1, & \text{if } t \leq m - 1; \\ m - t, & \text{if } m - 1 \leq t \leq m; \\ 0, & t \geq m. \end{cases} \tag{21}$$

For $x \in D_\infty$, let x^m be the element of D_∞ defined by

$$x^m(t) = g_m(t)x(t), \quad t \geq 0 \tag{22}$$

And now take

$$d_\infty^0(x, y) = \sum_{m=1}^{\infty} 2^{-m}(1 \wedge d_m^0(x^m, y^m)). \tag{23}$$

If $d_\infty^0(x, y) = 0$, then $d_m^0(x, y) = 0$ and $x^m = y^m$ for all m, and this implies $x = y$. The other properties being easy to establish, d_∞^0 is a metric on D_∞; it defines the Skorohod topology there. If we replace d_m^0 by d_m in (23), we have a metric d_∞ equivalent to d_∞^0.

Let Λ_∞ be the set of continuous, increasing maps of $[0, \infty)$ onto itself.

Theorem 3.8 *There is convergence $d_\infty^0(x_n, x) \to 0$ in D_∞ if and only if there exist elements λ_n of Λ_∞ such that*

$$\sup_{t < \infty} |\lambda_n t - t| \to 0 \tag{24}$$

and for each m,

$$\sup_{t \leq m} |x_n(\lambda_n t) - x(t)| \to 0. \tag{25}$$

Proof. Suppose that $d^0_\infty(x_n, x)$ and $d_\infty(x_n, x)$ go to 0. Then there exist elements λ^m_n of Λ_m such that

$$\epsilon^m_n = ||I - \lambda^m_n||_m \vee ||x^m_n \lambda^m_n - x^m||_m \to 0$$

for each m. Choose l_m so that $n \geq l_m$ implies $\epsilon^m_n < 1/m$. Arrange that $l_m < l_{m+1}$, and for $l_m \leq n < l_{m+1}$, let $m_n = m$. Since $l_m \leq n < l_{m+1}$, we have $m_n \to n$ and $\epsilon^{m_n}_n < 1/m_n$. Define

$$\lambda_n t = \begin{cases} \lambda^{m_n}_n t, & \text{if } t \leq m_n; \\ t + \lambda^{m_n}_n(m_n) - m_n, & \text{if } t \geq m_n. \end{cases}$$

Then $|\lambda_n t - t| < 1/m_n$ for $t \geq m_n$ as well as for $t \leq m_n$, and therefore,

$$\sup_t |\lambda_n t - t| \leq \frac{1}{m_n} \to 0.$$

Hence, 24. Fix c. If n is large enough, then $c < m_n - 1$, and so

$$||x_n \lambda_n - x||_c = ||x^{m_n}_n \lambda^{m_n}_n - x^{m_n}||_c \leq \frac{1}{m_n} \to 0,$$

which is equivalent to (25).

Now suppose that (24) and (25) hold. Fix m. First,

$$x^m_n(\lambda_n t) = g_m(\lambda_n t) x_n(\lambda_n t) \to g_m(t) x(t) = x^m(t) \tag{26}$$

holds uniformly on $[0, m]$. Define p_n and μ_n as in the proof of Lemma 1. As before, $\mu_n t \to t$ uniformly on $[0, m]$. For $t \leq p_n$, $|x^m(t) - x^m_n(\mu_n t)| = |x^m(t) - x^m_n(\lambda_n t)|$, and this goes to 0 uniformly by 26. For the case $p_n \leq t \leq m$, first note that $|x^m(u)| \leq g_m(u)||x||_m$ for all $u \geq 0$ and hence,

$$|x^m(t) - x^m_n(\mu_n t)| \leq g_m(t)||x||_m + g_m(\mu_n t)||x_n||_m. \tag{27}$$

By (24), for large n we have $\lambda_n(2m) > m$ and hence $||x_n||_m \leq ||x_n \lambda_n||_{2m}$; and $||x_n \lambda_n||_{2m} \to ||x_n||_{2m}$ by (25). This means that $||x_n||_m$ is bounded(m is fixed). Given ϵ, choose n_0 so that $n \geq n_0$ implies that p_n and $\mu_n p_n$ both lies in $(m - \epsilon, m]$, an interval on which g_m is bounded by ϵ. If $n \geq n_0$ and $p_n \leq t \leq m$, then t and $\mu_n t$ both lie is $(m - \epsilon, m]$, and it follows by (27) that $|x^m(t) - x^m_n(\mu_n t)| \leq \epsilon\big(||x||_m + ||x_n||_m\big)$. Since $||x_n||_m$ is bounded, this implies that $|x^m(t) - x^m_n(\mu_n t)| \to 0$ holds uniformly on $[p_n, m]$ as well as on $[0, p_n]$. Therefore, $d^0_m(x^m_n, x^m) \to 0$ for each m and hence $d^0_\infty(x_n, x)$ and $d_\infty(x_n, x)$ go to 0. This completes the proof.

Theorem 3.9 *There is convergence $d^0_\infty(x_n, x) \to 0$ in D_∞ if and only if $d^0_t(x_n, x) \to 0$ for each continuity point t of x.*

Proof. If $d^0_\infty(x_n, x) \to 0$, then $d^0_\infty(x^m_n, x^m) \to 0$ for each m. Given a continuity point t of x, fix an integer m for which $t < m - 1$. By Lemma 1(with t and m in the roles of m and u) and the fact that y and y^m agree on $[0, t]$, $d^0_t(x_n, x) = d^0_t(x^m_n, x^m) \to 0$.

To prove the reverse implication, choose continuity points t_m of x in such a way that $t_m \uparrow \infty$. The argument now follows the first part of the proof of (3.8). Choose elements λ^m_n of Λ_{t_m} in such a way that

$$\epsilon^m_n = ||\lambda^m_n - I||_{t_m} \vee ||x_n\lambda^m_n - x||_{t_m} \to 0$$

for each m. As before, define integers m_n in such a way that $m_n \to \infty$ and $\epsilon^{m_n}_n < 1/m_n$, and this time define $\lambda_n \in \Lambda_\infty$ by

$$\lambda_n t = \begin{cases} \lambda^{m_n}_n t, & \text{if } t \leq t_{m_n}; \\ t, & \text{if } t \geq t_{m_n}. \end{cases}$$

The $|\lambda_n t - t| \leq 1/m_n$ for all t, and if $c < t_{m_n}$, then $||x_n\lambda_n - x||_c = ||x_n\lambda^{m_n}_n - x||_c \leq 1/m_n \to 0$. This implies that (24) and ((25)) hold, which in turn implies that $d^0_\infty(x_n, x) \to 0$. This completes the proof.

3.8.1 Separability and Completeness

For $x \in D_\infty$, define $\psi_m x$ as x^m restricted to $[0, m]$. Then, since $d^0_m(\psi_m x_n, \psi_m x) = d^0_m(x^m_n, x^m)$, ψ_m is a continuous map of D_∞ into D_m. In the product space $\Pi = D_1 \times D_2 \times \cdots$, the metric

$$\rho(\alpha, \beta) = \sum_{m=1}^{\infty} 2^{-m}\big(1 \wedge d^0_m(\alpha_m, \beta_m)\big)$$

defines the product topology, that of coordinatewise convergence. Now define $\psi : D_\infty \to \Pi$ by $\psi x = (\psi_1 x, \psi_2 x, ...)$:

$$\psi_m : D_\infty \to D_m, \quad \psi : D_\infty \to \Pi$$

Then $d^0_\infty(x, y) = \rho(\psi x, \psi y) : \psi$ is an isometry of D_∞ into Π.

Lemma 3.7 *The image ψD_∞ is closed in* Π.

Proof. Suppose that $x_n \in D_\infty$ and $\alpha \in \Pi$, and $\rho(\psi x_n, \alpha) \to 0$; then $d^0_m(x^m_n, \alpha_m) \to 0$ for each m. We must find an x in D_∞ such that $\alpha = \psi x$-that is , $\alpha_m = \psi_m x$ for each m. Let T be the dense set of t such that, for every $m \geq t$, α_m is continuous at t. Since $d^0_m(x^m_n, \alpha_m) \to 0$, $t \in T \cap [0, m]$ implies $x^m_n(t) = g_n(t)x_n(t) \to \alpha_m(t)$.

This means that, for every t in T, the limit $x(t) = \lim x_n(t)$ exists(consider an $m > t+1$, so that $g_n(t) = 1$). Now $g_m(t)x(t) = \alpha_m(t)$ on $T \cap [0, m]$. It follows that $x(t) = \alpha_m(t)$ on $T \cap [0, m-1]$, so that x can be extended to a cadlag function on each $[0, m-1]$ and then to a cadlag function on $[0, \infty]$. And now, by right continuity, $g_m(t)x(t) = \alpha_m(t)$ on $[0, m]$, or $\psi_m x = x^m = \alpha_m$. This completes the proof.

Theorem 3.10 *The space D_∞ is separable and complete.*

Proof. Since Π is separable and complete, so are the closed subspace ψD_∞ and its isometric copy D_∞. This completes the proof.

3.8.2 Compactness

Theorem 3.11 *A set A is relatively compact in D_∞ if and only if, for each m, $\psi_m A$ is relatively compact in D_m.*

Proof. If A is relatively compact, then $\bar{A}$ is compact and hence the continuous image $\psi_m \bar{A}$ is also compact. But then, $\psi_m A$, as a subset of $\psi_m \bar{A}$, is relatively compact.

Conversely, if each $\psi_m A$ is relatively compact, then each $\overline{\psi_m A}$ is compact, and therefore $B = \overline{\psi_1 A} \times \overline{\psi_2 A} \times \cdots$ and $E = \psi D_\infty \cap B$ are both compact in Π. But $x \in A$ implies $\psi x \in \overline{\psi_m A}$ for each m, so that $\psi x \in B$. Hence $\psi A \subset E$, which implies that ψA is totally bounded and so is its isometric image A. This completes the proof.

For an explicit analytical characterization of relative compactness, analogous to the Arzela-Ascoli theorem, we need to adapt the $w'(x, \delta)$ to D_∞. For an $x \in D_m$ define

$$w'_m(x, \delta) = \inf \max_{1 \le i \le v} w(x, [t_{i-1}, t_i)), \tag{28}$$

where the infimum extends over all decompositions $[t_{i-1}, t_i), 1 \le i \le v$, of $[0, m)$ such that $t_t - t_{i-1} > \delta$ for $1 \le i \le v$. Note that the definition does not require $t_v - t_{v-1} > \delta$; Although 1 plays a special role in the theory of D_1, the integers m should play no special role in the theory of D_∞.

The exact analogue of $w'(x, \delta)$ is (28), but with the infimum extending only over the decompositions satisfying $t_t - t_{i-1} > \delta$ for $i = v$ as well as for $i < v$. Call this $\bar{w}_m(x, \delta)$. By an obvious extension, a set B in D_m is relatively compact if and only if $\sup_x ||x||_m < \infty$ and $\lim_\delta \sup_x \bar{w}(x, \delta) = 0$. Suppose that $A \subset D_\infty$, and transform the two conditions by giving $\psi_m A$ the role of B. By (3.11), A is relatively compact if and only if , for every m

$$\sup_{x\in A} ||x^m||_m < \infty \tag{29}$$

and

$$\lim_{\delta\to 0} \sup_{x\in A} \bar{w}_m(x^m,\delta) = 0. \tag{30}$$

The next step is to show that (29) and (30) are together equivalent to the condition that, for every m,

$$\sup_{x\in A} ||x||_m < \infty \tag{31}$$

and

$$\lim_{\delta\to 0} \sup_{x\in A} w'_m(x,\delta) = 0. \tag{32}$$

The equivalence of (29) and (31) follows easily because $||x^m||_m \le ||x||_m \le ||x^{m+1}||_{m+1}$. Suppose (31) and (32) both hold, and let K_m be the supremum in (31). If $x \in A$ and $\delta < 1$, then we have $|x^m(t)| \le K_m\delta$ for $m-\delta \le t < m$. Given ϵ, choose δ so that $K_m\delta < \epsilon/4$ and the supremum in (32) is less than $\epsilon/2$. If $x \in A$ and $m-\delta$ lies in the interval $[t_{j-1}, t_j)$ of the corresponding partition, replace the intervals $[t_{i-1}, t_i)$ for $i \ge j$ by the single interval $[t_{j-1}, m)$. This new partition shows that $\bar{w}_m(x,\delta)$. Hence (30).

That (30) implies (32) is clear because $w'_m(x,\delta) \le \bar{w}_m(x,\delta)$: An infimum increases if its range is reduced. This gives us the following criterion.

Theorem 3.12 *A set $A \in D_\infty$ is relatively compact if and only if (31) and (32) hold for all m.*

3.8.3 Tightness

Theorem 3.13 *The sequence $\{P_n\}$ is tight if and only if there two conditions hold:*
(i) For each m

$$\lim_{a\to\infty} \limsup_n P_n\Big(\{x : ||x||_m \ge a\}\Big) = 0. \tag{33}$$

(ii) For each m and ϵ,

$$\lim_{\delta} \limsup_n P_n\Big(\{x : w'_m(x,\delta) \ge \epsilon\}\Big) = 0. \tag{34}$$

And there is the corresponding corollary. Let

$$j_m(x) = \sup_{t\le m} |x(t) - x(t-)|. \tag{35}$$

Corollary 3.1 *Either of the following two conditions can be substituted for (i) in (3.13):*
(i') For each t in a set T that is dense in $[0, \infty)$,

$$\lim_{a \to \infty} \limsup_n P_n\Big(\{x : |x(t)| \geq a\}\Big) = 0. \tag{36}$$

(ii') The relation (36) holds for $t = 0$, and for each m,

$$\lim_{a \to \infty} \limsup_n P_n\Big(\{x : j_m(x) \geq a\}\Big) = 0. \tag{37}$$

Proof. The proof is almost the same as that for the corollary to (3.7).

Assume (ii) and (i'). Choose points t_i such that $0 = t_0 < t_1 < \cdots < t_v = m$, $t_i - t_{i-1} > \delta$ for $1 \leq i \leq v - 1$, and $w_x[t_{i-1}, t_i) < w'_m(x, \delta) + 1$ for $1 \leq i \leq v$. Choose from T points s_j such that $0 = s_0 < s_1 < \cdots < s_k = m$ and $s_j - s_{j-1} < \delta$ for $1 \leq j \leq k$. Let $m(x) = \max_{0 \leq j \leq k} |x(s_j)|$. If $t_v - t_{v-1} > \delta$, then $||x||_m \leq m(x) + w'_m(x, \delta) + 1$, just as before. If $t_v - t_{v-1} \leq \delta$(and $\delta < 1$, so that $t_{v-1} > m - 1$), then $||x||_{m-1} \leq m(x) + w'_m(x, \delta) + 1$. The old argument now gives (33), but with $||x||_m$ replaced by $||x||_{m-1}$, which is just as good.

In the proof that (ii) and (i') imply (i), we have $(v - 1)\delta \leq m$ instead of $v\delta <!$. But then, $v \leq m\delta^{-1} + 1$, and the old argument goes through. This completes the proof.

3.8.4 Aldous's Tightness Criterion

Equation

$$\lim_{a\to\infty} \limsup_n P\Big(||X^n||_m\Big) = 0 \tag{38}$$

Consider two conditions.

Condition 1^0. For each ϵ, η, m, there exist a δ_0 and an n_0 such that, if $\delta \le \delta_0$ and $n \ge n_0$, and if τ is a discrete X^n-stopping time satisfying $\tau \le m$, then

$$P\Big(\big|X^n_{\tau+\delta} - X^n_\tau\big| \ge \epsilon\Big) \le \eta. \tag{39}$$

Condition 2^0. For each ϵ, η, m, there exist a δ and an n_0 such that, if $n \ge n_0$, and if τ_1 and τ_2 are a discrete X^n-stopping time satisfying $0 \le \tau_1 \le \tau_2 \le m$, then

$$P\Big(\big|X^n_{\tau_2} - X^n_{\tau_1}\big| \ge \epsilon, \tau_2 - \tau_1 \le \delta\Big) \le \eta. \tag{40}$$

Theorem 3.14 *Conditions* 1^0 *and* 2^0 *are equivalent.*

Proof. Note that $\tau + \delta$ is a stopping time since

$$\{\tau + \delta \le t\} = \{\tau \le t - \delta\} \in \mathcal{F}^{X_n}_t.$$

In condition 2, put $\tau_2 = \tau$, $\tau_1 = \tau$. Then it gives condition 1. For the converse, suppose that $\tau \le m$ and choose δ_0 so that $\delta \le 2\delta_0$ and $n \ge n_0$ together imply 39. Fix an $n \ge n_0$ and a $\delta \le \delta_0$, and let (enlarge the probability space for X^n) θ be a random variable independent of $\mathcal{F}^n = \sigma\big(X^s_n : s \ge 0\big)$ and uniformly distributed over $J = [0, 2\delta]$. For the moment, fix an x in D_∞ and points t_1 and t_2 satisfying $0 \le t_1 \le t_2$. Let μ be the uniform distribution over J, and let $I = [0, \delta]$, $M_i = \{s \in J : |x(t_i + s) - x(t_i)| < \epsilon\}$, and $d = t_2 - t_1$.
Suppose that

$$t_2 - t_1 \le \delta \tag{41}$$

and

$$\mu(M_i) = P\big(\theta \in M_i\big) > \frac{3}{4}, \qquad \text{for } i = 1, 2 \tag{42}$$

If $\mu(M_2 \cap I) \le \frac{1}{4}$, then $\mu(M_2) \le \frac{3}{4}$, which is a contradiction. Hence, $\mu(M_2 \cap I) > \frac{1}{4}$, and for $d(0 \le d \le \delta)$, $\mu((M_2 + d) \cap J) \le \mu((M_2 \cap I) + d) = \mu((M_2 \cap I))\frac{1}{4}$. Thus $\mu(M_1) + \mu((M_2 + d) \cap J) > 1$, which implies $\mu(M_1 \cap (M_2 + d)) > 0$. There is therefore an s such that $s \in M_1$ and $s - d \in M_2$, from which follows

$$|x(t_1) - x(t_2)| < 2\epsilon. \tag{43}$$

Thus, (41) and (42) together implies (43). To put it another way, if (41) holds but (43) does not, then either $P\big(\theta \in M_1^c\big) \geq \frac{1}{4}$ or $P\big(\theta \in M_2^c\big) \geq \frac{1}{4}$. Therefore,

$$\begin{aligned} P\Big(\big|X^n_{\tau_2} - X^n_{\tau_1}\big| \geq 2\epsilon, \tau_2 - \tau_1 \leq \delta\Big) &\leq \sum_{i=1}^{2} P\left[P\Big(\big|X^n_{\tau_i+\theta} - X^n_{\tau_i}\big| \geq \epsilon|\mathcal{F}^n\Big) \geq \frac{1}{4}\right] \\ &\leq 4\sum_{i=1}^{2} P\Big(\big|X^n_{\tau_i+\theta} - X^n_{\tau_i}\big| \geq \epsilon\Big). \end{aligned}$$

Since $0 \leq \theta \leq 2\delta \leq 2\delta_0$, and since θ and $\mathcal{F}^n$ are independent, it follows by (39) that the final term here is at most 8η. Therefore, condition 1 implies condition 2.

This is Aldous's theorem:

Theorem 3.15 (Aldous) *If* 38 *and Condtion* 1^o *hold, then* $\{X^n\}$ *is tight.*

PROOF. By theorem 16.8, it is enough to prove that

$$\lim_{a\to\infty} \limsup_{n} P\Big(w'_m(X^n, \delta) \geq \epsilon\Big) = 0. \tag{44}$$

Let Δ_k be the set of nonnegative dyadic rationals[1] $j/2^k$ of order k. Define random variables $\tau^n_0, \tau^n_1, ...$ by $\tau^n_0 = 0$ and

$$\tau^n_i = \min\{t \in \Delta_k : \tau^n_{i-1} < t \leq m, |X^n_t - X^n_{\tau^n_{i-1}}| \geq \epsilon\},$$

with $\tau^n_i = m$ if there is no such t. The τ^n_i depend on ϵ, m, and k as well as on i and n, although the notation does not show this. It is easy to prove by induction that the τ^n_i are all stopping times.
Because of (3.14), we can assume that condition 2 holds. For given ϵ, η, m, choose δ' and n_0 so that

$$P\Big(\big|X^n_{\tau_i} - X^n_{\tau_{i-1}}\big| \geq \epsilon, \tau^n_i - \tau^n_{i-1} \leq \delta'\Big) \leq \eta$$

for $i \geq 1$ and $n \geq n_0$. Since $\tau^n_i < m$ implies that $\big|X^n_{\tau_i} - X^n_{\tau_{i-1}}\big| \geq \epsilon$. we have

$$P\Big(\tau^n_i < m, \tau^n_i - \tau^n_{i-1} \leq \delta'\Big) \leq \eta, \quad i \geq 1, n \geq n_0 \tag{45}$$

Now choose an integer q such that $q\delta \geq 2m$. There is also a δ such that

$$P\Big(\tau^n_i < m, \tau^n_i - \tau^n_{i-1} \leq \delta\Big) \leq \frac{\eta}{q}, \quad i \geq 1, n \geq n_0. \tag{46}$$

[1]dyadic rational is a rational number whose denominator is a power of two, i.e., a number of the form a/2b where a is an integer and b is a natural number; for example, 1/2 or 3/8, but not 1/3. These are precisely the numbers whose binary expansion is finite.

But then

$$P\left(\bigcup_{i-1}^{q}\{\tau_i^n < m, \tau_i^n - \tau_{i-1}^n \leq \delta\}\right) \leq \eta, \quad n \geq n_0 \tag{47}$$

Although the τ_i^n depend on k, 45 and 47 hold for all k simultaneously. By 45,

$$\begin{aligned} E(\tau_i^n - \tau_{i-1}^n | \tau_q^n < m) &\geq \delta' P(\tau_i^n - \tau_{i-1}^n \geq \delta' | \tau_q^n < m) \\ &\geq \delta'\Big(1 - \frac{\eta}{P(\tau_q^n < m)}\Big), \end{aligned}$$

and therefore,

$$\begin{aligned} m &\geq E(\tau_q^n | \tau_q^n < m) \\ &= \sum_{i=1}^{q} E(\tau_i^n - \tau_{i-1}^n | \tau_q^n < m) \\ &\geq q\delta'\Big(1 - \frac{\eta}{P(\tau_q^n < m)}\Big). \end{aligned}$$

Since $q\delta' \geq 2m$ by the choice of q, this leads to $P(\tau_q^n < m) \geq 2\eta$. By this and 47,

$$P\left(\{\tau_q^n < m\} \cup \bigcup_{i-1}^{q}\{\tau_i^n < m, \tau_i^n - \tau_{i-1}^n \leq \delta\}\right) \leq 3\eta, \quad k \geq 1, n \geq n_0 \tag{48}$$

Let A_{n_k} be the complement of the set in 48. On this set, let v be the first index for which $\tau_v^n = m$. Fix an n beyond n_0. There are points $t_i^k(\tau_i^n)$ such that $0 = t_0^k < \cdots < t_v^k = m$ and $t_i^k - t_{i-1}^k > \delta$ for $1 \leq i < v$. And $|X_t^n - X_s^n| < \epsilon$ if s, t lie in the same $[t_{i-1}^k, t_i^k)$ as well as in Δ_k. If $A_n = \limsup_k A_{n_k}$, then $P(A_n) \geq 1 - 3\eta$, and on A_n there is a sequence of values of k along which v is constant($v \leq q$) and, for each $i \leq v$, t_i^k converges to some t_i. But then, $0 = t_0 < \cdots < t_v = m, t_i - t_{i-1} \geq \delta$ for $i < v$, and by right continuity, $|X_t^n - X_s^n| \leq \epsilon$ if s, t lie in the same $[t_{i-1}, t_i)$. It follows that $w'(X^n, \delta) \leq \epsilon$ on a set of probability at least $1 - 3\eta$ and hence 44.

From the corollary to Theorem follows this one:

Corollary 3.2 *If for each m, the sequences $\{X_0^n\}$ and $\{j_m(X^n)\}$ are tight on the line, and if Condition 1^o holds, then $\{X^n\}$ is tight.*

4 Central Limit Theorem for Semimartingales and Applications

4.1 Local characteristics of semimartingale

In this chapter we study the central limit theorem by Lipster and Shiryayev. We begin by giving some preliminaries. We considered $(\Omega, \{\mathcal{F}_t\}, \mathcal{F}, P)$ a filtered probability space, where $\mathbf{F} = \{\mathcal{F}_t, t \geq 0\}$ is a non-decreasing family of sub $\sigma-$field of $\mathcal{F}$, satisfying $\bigcap_{t \geq s} \mathcal{F}_t = \mathcal{F}_s$. We say that $\{\underline{X}, \mathbf{F}\}$ is a martingale, if for each t, $X_t \in \underline{X} = \{X_t\} \subset L_1(\Omega, \mathcal{F}_t, P)$ and $E\Big(X(t)|\mathcal{F}_s\Big) = X_s$ a.e. P. WLOG, we assume $\{X_t, t \geq 0\}$ is $D[0, \infty)$ valued(or a.s. it is cadlag) as we can always find a version. A martingale X is said to be square-integrable if $\sup_t EX_t^2 < \infty$. We say that $\{X_t, t \geq 0\}$ is locally square integrable martingale if there exists an increasing sequence σ_n of $(\mathcal{F}_t)$-stopping times such that $0 \leq \sigma_n < \infty$ a.e., $\lim_n \sigma_n = \infty$, and $\{X(t \wedge \sigma_n)1_{\{\sigma_n > 0\}}\}$ is a square integrable martingale. A process $(\underline{X}, \mathbf{F})$ is called a semi-martingale if it has the decomposition

$$X_t = X_0 + M_t + A_t$$

where $\{M_t\}$ is local martingale, $M_0 = 0$, A is right continuous process with A_0, A_t, $\mathcal{F}_t$-measurable and has sample paths of finite variation. We now state condition of A to make this decomposition unique (called canonical decomposition). For this we need the following. We say that a sub σ-field of $[0, \infty) \times \Omega$ generated by sets of the form $(s, t] \times A$, $0 \leq s < t < \infty$ with $A \in \mathcal{F}_s$ and $\{0\} \times B (B \in \mathcal{F}_0)$ is the σ-field of predictable sets from now on called predictable σ-algebra $\mathcal{P}$. In the above decomposition A is measurable $\mathcal{P}$ then it is canonical.
We rework that if the jumps of the semi-martingale are bounded then the decomposition is canonical.
We now introduce the concept of local characteristics.
Let $(\underline{X}, \mathbf{F})$ be a semi-martingale. Set

$$\tilde{X}(t) = \sum_{s \leq t} \Delta X(s) 1(|\Delta X(s)| \geq \epsilon)$$

Then $\underline{X}(t) = X(t) - \tilde{X}(t)$ is a semi-martingale with unique canonical decomposition

$$\underline{X}(t) = X(0) + M(t) + A(t)$$

where $(M, \mathbf{F})$ is a local martingale and A is predictable process of finite variation. Thus,

$$X(t) = X(0) + M(t) + A(t) + \tilde{X}(t).$$

Let

$$\mu\Big((0, t]; A \cap \{|X| \geq \epsilon\}\Big) = \sum 1\Big(\Delta X(s) \in A, |\Delta X(s)| > \epsilon\Big)$$

and $\nu\Big((0,t];\cdot\cap\{|X|\geq\epsilon\}\Big)$ its predictable projection (Jacod, (1979), p.18). Then for each $\epsilon>0$, we can write

$$X(t)=X(0)+M'(t)+A(t)+\int_0^t\int_{|x|>\epsilon}x\nu(ds,dx)$$

where $(M',\mathbf{F})$ is a local martingale. Now the last two terms are predictable. Thus the semi-martingale is described by M',A, and ν. We thus have the following:

Definition 4.1 *The local characteristic of a semi-martingale X is defined by the triplet (A,C,ν), where*

1. *A is the predictable process of finite variation appearing in the above decomposition of $\underline{X}(t)$.*

2. *C is a continuous process defined by C_t*

$$C_t=[X,X]_t^c=<M^c>_t$$

3. *ν is the predictable random measure on $R_+\times R$, the dual predictable projection of the measure μ associated to the jumps of X given on $(\{0\}^c)$ by*

$$\mu(w,dt,dx)=\sum_{s>0}1(\Delta X(s,w)\neq 0)\delta_{(s,\Delta X(s,w))}(dt,dx)$$

with δ being the dirac delta measure.

4.2 Lenglart Inequality

Lemma 4.1 (McLeish Lemma(1974)) *Let $F_n(t), n = 1, 2, ...$ and $F(t)$ be in $D[0, \infty)$ such that $F_n(t)$ is increasing in t for each n, and $F(t)$ is a.s. continuous. Then*

$$\sup_t |F_n(t) - F(t)| \to 0$$

where sup is taken over compact set, and there exists $\{t_n\}$ such that

$$F_n(t) \to_P F(t)$$

PROOF. For $\epsilon > 0$ choose $\{t_{n_i}, i = 0, 1, ..., k\}$ for fixed $k \geq 1/\epsilon$ such that $t_{n_i} \to i\epsilon$ and $F_n(t_{n_i}) \to_p F(i\epsilon)$ as $n \to \infty$. Then

$$\begin{aligned} \sup_t |F_n(t) - F(t)| &\leq \sup_i |F_n(t_{n_{i+1}}) - F_n(t_{n_i})| + \sup_i |F_n(t_{n_i}) - F(t_{n_i})| \\ &+ \sup_i |F_n(t_{n_{i+1}}) - F(t_{n_i})| + \epsilon \end{aligned}$$

As $n \to \infty$, choose ϵ such that $|F((i+1)\epsilon) - F(i\epsilon)|$ is small.

We assume that A_t is an increasing process for each t and is $\mathcal{F}_t$−measurable.

Definition 4.2 *An adopted positive right continuous process is said to be dominated by an increasing predictable process A if for all finite stopping times T we have*

$$EX_T \leq EA_T$$

Example. Let M_t^2 is square martingale. Consider $X_t = M_t^2$. Then, we know $X_t - < M >_t$ is a martingale, and hence, $X_T - < M >_T$ is a martingale. Thus,

$$\begin{aligned} & E(X_T - < M >_T) = EX_0 = 0 \\ \Rightarrow \quad & EX_T = E < M >_T \end{aligned}$$

Let

$$X_t^* = \sup_{s \leq t} |X_s|$$

Lemma 4.2 (M1.) *Let T be a stopping time and X be dominated by increasing process A(as above). Then,*

$$P(X_T^* \geq c) \leq \frac{E(A_T)}{c}$$

for any positive c.

PROOF. Let $S = \inf\{s \leq T \wedge n : X_s \geq c\}$. Clearly, $s \leq T \wedge n$. Thus,

$$\begin{aligned} EA_T &\geq EA_S \quad (\text{ since } A \text{ is an increasing process}) \\ &\geq EX_S \quad (\text{ since } X \text{ is dominated by } A) \\ &\geq \int_{\{X^*_{T\wedge n} > c\}} X_S dP \quad (\text{ since } X_S > 0 \text{ on } \{X^*_{T\wedge n} > c\}) \\ &\geq c \cdot P(X^*_{T\wedge n} > c) \end{aligned}$$

Therefore, we let n go to ∞, then we get

$$EA_T \geq c \cdot P(X^*_T > c).$$

Theorem 4.1 (Lenglart Inequality) *If X is dominated by a predictable increasing process, then for every positive c and d*

$$P(X^*_T > c) \leq \frac{E(A_t \wedge d)}{c} + P(A_T > d).$$

PROOF. It is enough to prove for predictable stopping time $T > 0$,

$$P(X^*_{T-} \geq c) \leq \frac{1}{c} E(A_{T-} \wedge d) + P(A_{T-} \geq d). \tag{49}$$

We choose $T' = \infty$. Then T' is predictable, $\sigma_n = n$ and apply to $X^T_t = X_{t\wedge T}$ for T finite stopping time $X^{T^*}_{T-} = X^*_T$.

To prove (49)

$$\begin{aligned} P(X^*_{T-} \geq c) &= P\big(X^*_{T-} \geq c, A_{T-} < d\big) + P\big(X^*_{T-} \geq c, A_{T-} \geq d\big) \\ &= E\Big(1[\{X^*_{T-} \geq c\} \cap \{A_{T-} < d\}]\Big) + E\Big(1[\{X^*_{T-} \geq c\} \cap \{A_{T-} \geq d\}]\Big) \\ &= E\Big(1[\{X^*_{T-} \geq c\} \cap \{A_{T-} < d\}]\Big) + P(A_{T-} \geq d) \\ &\leq P(X^*_{T-} \geq c) + P(A_{T-} \geq d) \end{aligned} \tag{50}$$

Let $S = \{t : A_t \geq d\}$. It is easy to show that S is a stopping time. Also, S is predictable. On $\{\omega : A_{T-} < d\}$, $S(\omega) \geq T(\omega)$, and hence

$$1(A_{t-} < d) X^*_{T-} \leq X^*_{(T\wedge S)-}.$$

By (50), we have

$$\begin{aligned} P(X^*_{T-} \geq c) &\leq P(X^*_{T-} \geq c) + P(A_{T-} \geq d) \\ &\leq P(X^*_{T\wedge S} \geq c) + P(A_{T-} \geq d) \end{aligned}$$

Let $\epsilon > 0$, $\epsilon < c$ and $S_n \nearrow S \wedge T$. Then,

$$\begin{aligned} P(X^*_{(T\wedge S)-} \geq c) &\leq \liminf_n P(X^*_{S_n} \geq c - \epsilon) \text{ (by Fatou's Lemma)} \\ &\leq \frac{1}{c-\epsilon} \lim_{n\to\infty} EA_{S_n} \\ &= \frac{1}{c-\epsilon} EA_{(S\wedge T)} \text{ (by Monotone Convergence Theorem)} \end{aligned}$$

Since ϵ is arbitrary,

$$\begin{aligned} P(X^*_{(T\wedge S)-} \geq c) &\leq \frac{1}{c} EA_{S_n} \\ &\leq \frac{1}{c} E(A_{(T-\wedge d)}) \end{aligned}$$

This completes the proof.

Corollary 4.1 *Let $M \in \mathcal{M}^2_{LOC}((\mathcal{F}_t), P)$ (class of locally square integrable martingale). Then,*

$$P(\sup_{t\leq T} |M_t| > a) \leq \frac{1}{a^2} E(< M >_T \wedge b) + P(< M >_T \geq b)$$

Proof. Use $X_t = |M_t|^2$, $c = a^2$, $b = d$, $A_t =< M >_t$.

Lemma 4.3 *Consider $\{\mathcal{F}^n_t, P\}$. Let $\{M^n\}$ be locally square martingale. Assume that*

$$< M^n >_t \longrightarrow_p f(t),$$

where f is continuous and deterministic function(Hence, f will be increasing function). Then, $\{P \circ (M^n)^{-1}\}$ on $D[0,\infty)$ is relatively compact on $D[0,\infty)$.

PROOF. It suffices to show for $T < \infty$, and any $\eta > 0$ there exists $a > 0$ such that

$$\sup_n P\big(\sup_{t\leq T} |M^n_t| > a\big) < \eta. \tag{51}$$

For each $T < \infty$ and $\eta, \epsilon > 0$, there exist n_0, δ such that for any stopping time τ(w.r.t $\mathcal{F}$, $\tau \leq T, \tau + \delta < T$).

$$\sup_{n\geq n_0} P\big(\sup_{0\leq t\leq \delta} |M^n_{\tau+t} - M^n_t| \geq \epsilon\big) < \eta \tag{52}$$

Observe that by corollary to Lenglast Inequality,

$$P(\sup_{t\leq T}|M^n_t| > a) \leq \frac{1}{a^2}E(< M^n >_T \wedge b) + P(< M^n >_T \geq b)$$

Let $b = f(T) + 1$, then under the hypothesis there exists n_1 such that for all $n \geq n_1$

$$P(< M^n >_T \geq b) > \frac{\eta}{2}.$$

Thus,

$$\sup_n P(\sup_{t\leq T}|M^n_t| > a) \leq \frac{b}{a^2} + \frac{\eta}{2} + \sum_{k=1}^{n_1} P(\sup_{t\leq T}|M^k_t| > a)$$

Choose a large to obtain (51).

We again note that $M^n_{\tau+t} - M^n_\tau$ is a locally square integrable martingale. Hence by (4.1)

$$\begin{aligned}
P\Big(\sup_{0\leq t\leq\delta}|M^n_{\tau+t} - M^n_t| \geq \epsilon\Big) \quad &\leq \quad \frac{1}{\epsilon^2}E\Bigg(\Big(< M^n >_{\tau+\delta} - < M^n >_\tau \Big) \wedge b\Bigg) + P\Big(< M^n >_{\tau+\delta} - < M^n >_\tau \geq \\
&\leq \quad \frac{1}{\epsilon^2}E\Bigg(\sup_{t\leq T}\Big(< M^n >_{t+\delta} - < M^n >_t \Big) \wedge b\Bigg) \\
&\quad\quad + P\Big(\sup_{t\leq T}\Big| < M^n >_{t+\delta} - < M^n >_t \Big| \geq b\Big) \\
&\leq \quad \frac{1}{\epsilon^2}E\Bigg(\sup_{t\leq T}\Big|M^n_{t+\delta} - f(t+\delta)\Big| \wedge b\Bigg) + \frac{1}{\epsilon^2}E\Bigg(\sup_{t\leq T}\Big|M^n_t - f(t+\delta)\Big| \wedge b\Bigg) \\
&\quad\quad + \frac{1}{\epsilon^2}\sup_{t\leq T}|f(t+\delta) - f(t)| + P\Big(\sup_{t\leq T}\Big| < M^n >_{t+\delta} - f(t+\delta)\Big| \geq \frac{b}{3} \geq b\Big) \\
&\quad\quad + P\Big(\sup_{t\leq T}\Big| < M^n >_t - f(t)\Big| \geq \frac{b}{3} \geq b\Big) + 1\Big(\sup_{t\leq T}|f(t+\delta) - f(t)| \geq \frac{b}{3}\Big) \\
&\quad\quad + P\Big(\sup_{t\leq T}\Big| < M^n >_{t+\delta} - < M^n >_t \Big| \geq b\Big)
\end{aligned}$$

Using McLeish Lemma each term goes to 0. This completes the proof.

If our conditions guarantee that for a locally square integrable martingale the associated increasing process converges then the problem reduces to the convergence of finite-dimensional distributions.

4.3 Central Limit Theorem for Semi-Martingale

Theorem 4.2 *Let $\{X^n\}$ be a sequence of semi-martingale with characteristics $(B^n, < X^{nc} >, \nu^n)$ and M be a continuous Gaussian martingale with increasing process $< M >$.*

(i) For any $t > 0$ and $\epsilon \in (0,1)$ let the following conditions be satisfied:
(A)

$$\int_0^t \int_{|x|>\epsilon} \nu^n(ds, dx) \to_p 0$$

(B)

$$B_t^{nc} + \sum_{0\le s\le t} \int_{|x|\le\epsilon} x\nu^n(\{s\}, dx) \to_p 0$$

(C)

$$< X^{nc} >_t + \int_0^t \int_{|x|\le\epsilon} x^2\nu^n(ds, dx) - \sum_{0\le s\le t} \left(\int_{|x|\le\epsilon} x\nu^n(\{s\}, dx) \right)^2 \to_p < M >_t$$

Then $X^n \Rightarrow M$ for finite dimension.

(ii) If (A) and (C) are satisfied as well as the condition

$$\sup_{0<s\le t} \left| B_s^{nc} + \sum_{0\le u\le s} \int_{|x|\le\epsilon} x\nu^n(\{u\}, dx) \right| \to_p 0 \tag{53}$$

for any t and $\epsilon \in (0,1]$ then $X \Rightarrow M$ in $D[0,T]$.

Proof. For $\epsilon \in (0,1)$

$$\begin{aligned} X_t^n &= \left(\sum_{0\le s\le t} \int_{\epsilon<|x|\le 1} x\nu^n(\{s\}, dx) \right) + \left(B_t^{nc} + \sum_{0\le s\le t} \int_{|x|\le\epsilon} x\nu^n(\{s\}, dx) \right) \\ &\quad + \left(\int_0^t \int_{|x|>1} x\mu^n(ds, dx) + \int_0^t \int_{\epsilon<|x|\le 1} (\mu^n - \nu^n)(ds, dx) \right) \\ &\quad + \left(X_t^{nc} + \int_0^t \int_{|x|\le\epsilon} (\mu^n - \nu^n)(ds, dx) \right) \\ &= \alpha_t^n(\epsilon) + \beta_t^n(\epsilon) + \gamma_t^n(\epsilon) + \Delta_t^n(\epsilon) \end{aligned}$$

where

$$\alpha_t^n(\epsilon) = \sum_{0\le s\le t} \int_{\epsilon<|x|\le 1} x\nu^n(\{s\}, dx)$$

$$\begin{aligned}
\beta_t^n(\epsilon) &= B_t^{nc} + \sum_{0\le s\le t} \int_{|x|\le\epsilon} x\nu^n(\{s\}, dx) \\
\gamma_t^n(\epsilon) &= \int_0^t \int_{|x|>1} x\mu^n(ds, dx) + \int_0^t \int_{\epsilon<|x|\le 1} (\mu^n - \nu^n)(ds, dx) \\
\Delta_t^n(\epsilon) &= X_t^{nc} + \int_0^t \int_{|x|\le\epsilon} (\mu^n - \nu^n)(ds, dx)
\end{aligned}$$

By (A) we have

$$\sup_{s\le t} \alpha_s^n(\epsilon) \to_p 0$$

By (B) we have

$$\beta_t^n(\epsilon) \to_p 0$$

By (53) we have

$$\sup_{s\le t} \beta_s^n(\epsilon) \to_p 0$$

Let

$$Y_t^n = \gamma_t^n(\epsilon) + \Delta_t^n(\epsilon).$$

It suffices to prove $Y^n \to M$ on $D[0,T]$ for each T(by the decomposition Y^n does not depend on ϵ). Next, we have

$$\sup_{0<t\le T} |\gamma_t^n(\epsilon)| \le \int_0^T \int_{|x|>1} |x|\mu^n(ds, dx) + \underbrace{\int_0^T \int_{|x|>\epsilon} \mu^n(ds, dx)}_{\downarrow 0 \text{ by Lenglast Inequality}} + \underbrace{\int_0^T \int_{|x|>\epsilon} \nu^n(ds, dx)}_{\downarrow 0 \text{ by (A)}} \tag{54}$$

Therefore, if we can show that the first term of RHS goes to 0, then $\sup_{0<t\le T} |\gamma_t^n(\epsilon)| \to 0$. We have

$$\int_0^T \int_{|x|>1} |x|\mu^n(ds, dx) = \sum_{0<s\le T} |\Delta X_s^n| 1_{(|\Delta X_s^n|>1)}$$

For $\delta \in (0,1)$,

$$\Big\{ \sum_{0<s\le T} |\Delta X_s^n| 1_{(|\Delta X_s^n|>1)} > \delta \Big\} \subset \Big\{ \sum_{0<s\le T} 1_{(|\Delta X_s^n|>1)} > \delta \Big\}$$

and

$$\sum_{0<s\le T} 1_{(|\Delta X_s^n|>1)} = \int_0^T \int_{|x|>1} \mu^n(ds, dx) \to_p 0$$

by Lenglast Inequality. Therefore, by (54), we have

$$\sup_{0<t\le T} |\gamma_t^n(\epsilon)| \to 0.$$

Now, only thing left is to show that

$$\Delta_t^n(\epsilon) \to 0$$

$$\Delta_t^n(\epsilon) = X_t^{nc} + \int_0^T \int_{|x|\le\epsilon} x(\mu^n - \nu^n)(ds, dx)$$

Since $(\mu^n - \nu^n)$ is martingale, and X^{nc} is martingale, Δ^n is martingale. Since $\Delta^n(\epsilon) \in \mathcal{M}_{LOC}\Big((\mathcal{F}^n)_t, P\Big)$,

$$\begin{aligned} < \Delta^n(\epsilon) >_t \quad &= \quad < X^{nc} >_t + \int_0^t \int_{|x|\le\epsilon} x^n \nu^n(ds, dx) \\ &\quad - \sum_{0<s\le t} \Big(\int_{|x|\le\epsilon} x\nu^n(\{s\}, dx) \Big)^2 \\ &\quad \longrightarrow < M >_t \end{aligned}$$

by condition (C).

By McLeish Lemma,

$$\sup_{t\le T} | < \Delta^n(\epsilon) >_t - < M >_t | \to_p 0 \tag{55}$$

We showed $\sup_{t\le T} |\gamma_t^n(\epsilon)| \to 0$. Combining this with (55), we have

$$\max\left(\sup_{t\le T} | < \Delta^n(\epsilon) >_t - < M >_t |, \sup_{t\le T} |\gamma_t^n(\epsilon)| \right) \to 0$$

Then, there exists $\{\epsilon_n\}$ such that

$$\sup_{t\le T} | < \Delta^n(\epsilon_n) >_t - < M >_t | \to 0, \quad \sup_{t\le T} |\gamma_t^n(\epsilon_n)| \to 0 \tag{56}$$

$$M_t^n = \Delta_t^n(\epsilon_n), \quad Y_t^n = \Delta_t^n(\epsilon_n) + \gamma_t^n(\epsilon_n)$$

It suffices to prove that $M^n \Rightarrow M$. $\{M_t^n = \Delta_t^n(\epsilon_n)\}$ is compact by (4.3) and (55). It suffices to prove finite-dimensional convergence.

Let $H(t), 0 \le t \le T$ be a piecewise constant left-continuous function assuming finitely many values. Let

$$N_t^n = \int_0^t H(s) dM_s^n, \quad N_t = \int_0^t H(s) dM_s.$$

Since M is Gaussian, N is also Gaussian.
Remark. Cramer-Wold Criterion for $\mathcal{D}_f-$ convergence

$$Ee^{iN_T^n} \to Ee^{iN_T} = e^{-\frac{1}{2}\int_0^T H^2(s) d<M>_s}$$

Let A be predictable, $A \in \mathcal{A}_{LOC}(\mathcal{F}, P)$

$$e(A)_t = e^{A_t} \prod_{0 \le s \le t} (1 + \Delta A_s) e^{-A_s}$$

Then e^{A_t} will be a solution of $dZ_t = Z_{t-} dA_t$. If $m \in \mathcal{M}_{LOC}$ then

$$A_t = -\frac{1}{2} < m^c >_t + \int_0^t \int_{R-\{0\}} (e^{isx} - 1 - ix) \nu_m(ds, dx)$$

Lemma 4.4 *For some $a > 0$ and $c \in (0,1)$, let $< m >_\infty \le a$, $\sup_t |\Delta m_t| \le c$. Then $(e(A_t), \mathcal{F}_t)$ is such that*

$$|e(A)_t| \ge \exp\Big(- \frac{2a}{1 - c^2} \Big)$$

and the process $(Z_t, \mathcal{F}_t)$ with $Z_t = e^{imt}\Big(e(A)_t\Big)^{-1}$ is a uniformly integrable martingale.

We will use the lemma to prove

$$Ee^{iN_T^n} \to Ee^{iN_T} = e^{-\frac{1}{2}\int_0^T H^2(s) d<M>_s}$$

Case 1. Let us first assume $< M^n >_T \le a$ and $a > 1+ < M >_T$. Observe that

1. By (56), $< N^n >_t \to < N >_t$
2. $|\Delta_t^n| \le 2\epsilon_n$
3. $|\Delta N_t^n| \le 2\lambda\epsilon_n = d_n$ where $\lambda = \max_{t \le T} |H(s)|$

We want to prove

$$E \exp\Big(iN_T^n + \frac{1}{2} < N >_T \Big) \to 1 \tag{57}$$

Let A_t^n be increasing process associated with N_t^n. Let $Z_t = e^{iN_t^n}\Big(e(A^n)_t\Big)^{-1}$. Choose n_0 such that $d_{n_0} = 2\lambda\epsilon_{n_0} \le 1/2$. By (4.4), Z^n is a martingale with $EZ_T^n = 1$. To prove (57) is equivalent to proving

$$\lim_{n\to\infty} \left(E \exp\Big(iN_T^n + \frac{1}{2} < N >_T \Big) - \underbrace{Ee^{iN_T^n}\Big(e(A^n)_T\Big)^{-1}}_{=EZ_T^n = 1} \right) = 0 E \exp\Big(iN_T^n + \frac{1}{2} < N >_T \Big) \to 1 \tag{58}$$

So it is sufficient to prove

$$e(A^n)_T \to e^{-\frac{1}{2}<N>_T}$$

Recall
$$A_t^n = -\frac{1}{2} < N >_t + \int_0^t \int_{|x| \le d_n} (e^{ix} - 1 - ix) \tilde{\nu}^n(ds, dx)$$
Let
$$\alpha_t^n = \int_{|x| \le d_n} (e^{ix} - 1 - ix) \tilde{\nu}^n(\{t\}, dx)$$
Since $(e^{ix} - 1 - ix) \le x^2/2$, we have $\alpha_t^n \le d_n^2/2$. Therefore,
$$\begin{aligned} \sum_{0 \le t \le T} |\alpha_t^n| &= \frac{1}{2} \int_0^T \int_{|x| \le d_n} x^2 \tilde{\nu}^n(dt, dx) \\ &= \frac{1}{2} < N^n >_T \\ &= \frac{1}{2} \frac{\lambda^2 a}{2} \end{aligned}$$
Then,
$$\prod_{0 < t \le T} (1 + \alpha_t^n) e^{-\alpha_t^n} \to 1$$
By definition of $e(A)_t$, it remains to prove
$$\frac{1}{2} < N^{nc} >_T - \int_0^T \int_{|x| \le d_n} (e^{isx} - 1 - isx) \tilde{\nu}^n(ds, dx) \to_p \frac{1}{2} < N >_T$$
By observation (a) and the form of $< N^n >_T$, it suffices to prove
$$\int_0^T \int_{|x| \le d_n} (e^{isx} - 1 - isx) \tilde{\nu}^n(ds, dx) \to_p 0$$
We have
$$\begin{aligned} \int_0^T \int_{|x| \le d_n} \underbrace{\left((e^{isx} - 1 - isx) + \frac{x^2}{2} \right)}_{\le \frac{|x|^3}{6}} \tilde{\nu}^n(ds, dx) &\le \frac{d_n}{6} \int_0^T \int_{|x| \le d_n} x^2 \tilde{\nu}^n(ds, dx) \\ &\le \frac{d_n}{6} < N^n >_T \\ &\le \frac{d_n}{6} \lambda^2 a \\ &\longrightarrow 0 \end{aligned}$$
To dispose of assumption define
$$\tau_n = \min\{t \le T : < M^n >_t \ge < M >_T + 1\}$$
Then τ_n is stopping time. We have $\tau_n = T$ if $< M^n >_t << M >_T +1$. Let $\tilde{M}^n = M^n_{t \wedge \tau_n}$. Then
$$< \tilde{M}^n >_T \le 1 + < M >_T + \epsilon_n^2 \le 1 + < M >_T + \epsilon_1^2$$

and

$$\lim_n P\Big(| < \tilde{M}^n >_t - < \tilde{M} >_t | > \epsilon\Big) \leq \lim_n P(\tau_n > T) = 0$$

Next,

$$\begin{aligned}
\lim_{n\to\infty} Ee^{iN_T^n} &= \lim_{n\to\infty} E\left(e^{iN_t^n} - e^{iN_{t\wedge\tau_n}^n}\right) + \lim_{n\to\infty} Ee^{iN_{t\wedge\tau_n}^n} \\
&= \lim_{n\to\infty} E\left(e^{iN_T^n} - e^{iN_{T\wedge\tau_n}^n}\right) + Ee^{iN_T} \\
&= Ee^{iN_T}
\end{aligned}$$

The last equality follows from

$$\lim_{n\to\infty} \left| E\Big(e^{iN_T^n} - e^{iN_{T\wedge\tau_n}^n}\Big)\right| \leq 2 \lim_{n\to\infty} P(\tau_n > T) = 0.$$

This completes the proof.

4.4 Application to a Survival Analysis

Let X be a positive random variable. Let F and f be cumulative distribution function and probability density function of X. Then, **survival function** $\bar{F}$ is defined as

$$\bar{F}(t) = P(X > t) = 1 - F(t)$$

Then, we have

$$\begin{aligned} P\big(t < X \leq t + \triangle t | X > t\big) &= \frac{P\big(t < X \leq t + \triangle t\big)}{\bar{F}(t)} \\ &= \frac{\int_t^{t+\triangle t} dF(s)}{\bar{F}(t)} \end{aligned}$$

Since we know

$$\frac{1}{\triangle t} \int_t^{t+\triangle t} f(x)ds \longrightarrow f(t)$$

as $\triangle t \to 0$, **hazard rate**, is defined as

$$\begin{aligned} h(t) &= \frac{f(t)}{\bar{F}(t)} \\ &= -\frac{d}{dt} \log \bar{F}(t) \end{aligned}$$

Therefore, survival function can be written as

$$\bar{F}(t) = \exp\Big(- \int_0^t h(s)ds\Big)$$

If integrated hazard rate is given, then it determines uniquely life distribution. For example, think about the following:

$$\tau_F = \sup\{s : F(s) < 1\}$$

Consider now the following problem arising in clinical trials.

Let

1. $X_1, ..., X_n$ be i.i.d F(life time distribution)
2. $U_1, ..., U_n$ be i.i.d measurable function with distribution function G with $G(\infty) < 1$ which means U_i are not random variable in some sense(censoring times).

Now, consider

1. indicator for "alive or not at time s": $1(X_i \le U_i, X_i \wedge U_i \le s)$
2. indicator for "alive and leave or not at time s": $U_i 1(X_i \wedge U_i \le s)$
3. indicator for "leave or not at time s": $1(X_i \wedge U_i \ge s)$

and $\sigma-$field

$$\mathcal{F}_t^n = \sigma\Bigg(\big\{1(X_i \le U_i, X_i \wedge U_i \le s), U_i 1(X_i \wedge U_i \le s), 1(X_i \wedge U_i \ge s), s \le t, i = 1, 2, ..., n\big\}\Bigg)$$

$\mathcal{F}_t^n$ is called **information contained in censored data**.

Let

$$\beta(t) = \int_0^t h(s)ds.$$

It $\hat{\beta}(t)$ is an estimate of $\beta(t)$, then we can estimate survival function. Estimator of survival function will be

$$\hat{\bar{F}}(t) = e^{-\hat{\beta}(t)}$$

which will be approximately be

$$\prod_{s \le t} \Big(1 - d\big(\hat{\beta}(s)\big)\Big).$$

This is alternate estimate of survival function, which is called, **Nelson Estimate**.

Let

$$\begin{aligned} N_n(t) &= \sum_{i=1}^n 1(X_i \le U_i, X_i \wedge U_i \le t) \\ Y_n(t) &= \sum_{i=1}^n 1(X_i \wedge U_i \ge t) \end{aligned}$$

Then, $\hat{\beta}(t)$, which is called **Breslow estimator**, will be

$$\hat{\beta}(t) = \int_0^t \frac{dN_n(s)}{Y_n(s)} \approx \frac{\int_t^{t+\triangle t} dF(s)}{\bar{F}(t)}$$

Now, we consider another estimator of survival function, which is called **Kaplan-Meier estimator**. It will be

$$\prod_{s \le t} \Big(1 - \frac{\triangle N_n(s)}{Y_n(s)}\Big)$$

Richard Gill showed asymptotic properties of Kaplan-Meier estimator.

We can show that

$$e^{-\hat{\beta}(t)} - \prod_{s \leq t} \Big(1 - \frac{\triangle N_n(s)}{Y_n(s)}\Big) = O\Big(\frac{1}{n}\Big) \tag{59}$$

by using following lemma.

Lemma 4.5 ***C.1*** *Let $\{\alpha^n(s), 0 \leq s \leq T, n \geq 1\}$ be real-valued function such that*

1. *$\{s \in (0,u] : \alpha^n(s) \neq 0\}$ is $P-$a.e at most countable for each n*
2. *$\sum_{0<s\leq u} |\alpha^n(s)| \leq C$ with C constant*
3. *$\sup_{s \leq u}\{|\alpha^n(s)|\} = O(a_n)$ where $a_n \searrow 0$ as n goes ∞.*

Then,

$$\sup_{t \leq u} \Bigg| \prod_{0<s\leq t} (1 - \alpha^n(s)) - \prod_{0<s\leq t} e^{-\alpha^n(s)} \Bigg| = O(a_n)$$

PROOF. We choose n_0 large such that for $n \geq n_0$ $O(a_n) < \frac{1}{2}$. Since

$$\begin{aligned}
\prod_{0<s\leq t} (1-\alpha^n(s))e^{\alpha^n(s)} &= \exp\Big(\sum_{0<s\leq t} \log(1-\alpha^n(s)) + \alpha^n(s)\Big) \\
&= \exp\Big(\sum_{0<s\leq t}\sum_{j=2}^{\infty} \frac{(-1)^{j+2}}{j}(\alpha^n(s))^j\Big) \text{ (by Taylor expansion.)},
\end{aligned}$$

for $n \geq n_0$ we have

$$\begin{aligned}
\Bigg| \prod_{0<s\leq t} (1-\alpha^n(s))e^{\alpha^n(s)} - 1 \Bigg| &\leq \Bigg| \exp\Big(\sum_{0<s\leq t}\sum_{j=2}^{\infty} \frac{(-1)^{j+2}}{j}(\alpha^n(s)^j - 1)\Big)\Bigg| \\
&\leq e^{\eta} \Bigg| \sum_{0<s\leq t}\sum_{j=2}^{\infty} \frac{(-1)^{j+2}}{j}(\alpha^n(s))^j \Bigg|
\end{aligned}$$

where

$$0 \wedge \sum_{0<s\leq t}\sum_{j=2}^{\infty} \frac{(-1)^{j+2}}{j}(\alpha^n(s))^j < \eta < 0 \vee \sum_{0<s\leq t}\sum_{j=2}^{\infty} \frac{(-1)^{j+2}}{j}(\alpha^n(s))^j$$

For large n

$$\begin{aligned}
\left| \sum_{0<s\leq t} \sum_{j=2}^{\infty} \frac{(-1)^{j+2}}{j} (\alpha^n(s))^j \right| &\leq \sum_{0<s\leq t} \sum_{j=2}^{\infty} \frac{\left|\alpha^n(s)\right|^j}{j} \\
&\leq \underbrace{\sup_{s\leq u} \left|\alpha^n(s)\right|}_{=O(a_n)} \underbrace{\sum_{0<s\leq t} \left|\alpha^n(s)\right|}_{\leq t\cdot M} \underbrace{\sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^{j-2} \frac{1}{j}}_{<\infty} \\
&\longrightarrow 0
\end{aligned}$$

$\sum_{0<s\leq t} \left|\alpha^n(s)\right| \leq t \cdot M$ holds since $\left|\alpha^n(s)\right|$ will be bounded by M.

In order to prove (59), we let

$$\alpha^n(s) = \frac{1}{Y_n(s)}, \quad s \leq T.$$

and

$$\triangle N(s) = 0, \quad s > T.$$

We get $a_n = 1/n$ by using **Glivenko-Cantelli theorem**.

4.5 Asymptotic Distributions $\hat{\beta}(t)$ and Kaplan-Meier Estimates

X_i and U_i are defined as previously. Again, define $N_n(t), Y_n(t)$ as

$$\begin{aligned} N_n(t) &= \sum_{i=1}^{n} 1(X_i \le U_i, X_i \wedge U_i \le t) \\ Y_n(t) &= \sum_{i=1}^{n} 1(X_i \wedge U_i \ge t) \end{aligned}$$

Then,

$$\begin{aligned} \beta(t) &= \int_0^t h(s)ds \\ &= \int_0^t \frac{f(s)}{\bar{F}(s)}ds \\ &= \int_0^t \frac{dF(s)}{1-F(s)} \\ \hat{\beta}(t) &= \int_0^t \frac{dN_n(s)}{Y_n(s)} \end{aligned}$$

Using above lemma, we can show that

$$\begin{aligned} & \sup_{t\le u} \left| \prod_{s \le t} \Big(1 - \frac{\triangle N_n(s)}{Y_n(s)}\Big) - e^{-\hat{\beta}(t)} \right| \\ = & \sup_{t\le u} \left| \underbrace{\prod_{s \le t} \Big(1 - \frac{\triangle N_n(s)}{Y_n(s)}\Big)}_{\text{K.M estimator}} - \underbrace{\exp\Big(- \int_0^t \frac{dN_n(s)}{Y_n(s)} Big)}_{\text{Nelson estimator}} \right| \\ = & O\Big(\frac{1}{n}\Big) \end{aligned}$$

Let

$$\begin{aligned} Q_1(s) &= P(X_1 \wedge U_1 \le s, X_1 \le U_1) \\ H(s) &= P(X_1 \wedge U_1 \le s) \\ \beta_1(t) &= \int_0^t \frac{dQ_1(s)}{(1 - H(s-))} \end{aligned}$$

Assume that $X($ with $F)$ and U (with G) are independent. Then,

$$\begin{aligned} \triangle Q_1(s) &= P(s \le X_1 \wedge U_1 \le s + \triangle s, X_1 \le U_1) \\ &= P(s \le X_1 \le s + \triangle s, X_1 \le U_1) \end{aligned}$$

$$
\begin{aligned}
&= P(s \leq X_1 \leq s + \triangle s, U_1 \geq s + \triangle s) \\
&= P(s \leq X_1 \leq s + \triangle s) \cdot P(U_1 \geq s + \triangle s) \\
dQ_1(s) &= (1 - G(s-))dF(s)
\end{aligned}
$$

Similarly,

$$
(1 - H(s-)) = (1 - G(s-))(1 - F(s)). \tag{60}
$$

Then,

$$
\begin{aligned}
\beta_1(t) &= \int_0^t \frac{dQ_1(s)}{(1 - H(s-))} \\
&= \int_0^t \frac{(1 - G(s-))dF(s)}{(1 - G(s-))(1 - F(s))} \\
&= \int_0^t \frac{dF(s)}{(1 - F(s))}
\end{aligned}
$$

O. Aalen (Annals of Statistics, 1978) considered this technique. He considered $P(U_i < \infty) = 0$. So, data is not censored and we figure out true distribution of X_i.

Lemma 4.6 ***(PL 1.)*** *Let*

$$
\mathcal{F}_t^n = \sigma\left(\Big\{1(X_i \leq U_i, X_i \wedge U_i \leq s), U_i 1(X_i \wedge U_i \leq s), 1(X_i \wedge U_i \geq s), s \leq t, i = 1, 2, ..., n\Big\}\right)
$$

Suppose we have

1. $(N_n(t), \mathcal{F}_t^n)$ *is Poisson process and*

$$
\Big\{N_n(t) - \int_0^t Y_n(s) \frac{dQ_1(s)}{(1 - H(s-))}\Big\}
$$

is martingale.

2. $\hat{\beta}_n(t) = \int_0^t \frac{dN_n(t)}{Y_n(t)}$

Then, $m_n(t) = \hat{\beta}_n(t) - \beta_1(t)$ *is locally square integrable martingale, and increasing process* $< m_n >_t$ *will be*

$$
< m_n >_t = \int_0^t \frac{1}{Y_n(s)} \frac{dQ_1(s)}{(1 - H(s-))}
$$

Remark. If X and U are independent, from (60) and (60), we have

$$\frac{dQ_1(s)}{(1-H(s-))} = \frac{dF(s)}{1-F(s)}$$

Let us assume that $\{X_i\}$ and $\{U_i\}$ are independent and

$$A_n(t) = \int_0^t Y_n(s)\frac{dQ_1(s)}{(1-H(s-))}$$

From the previous theorem we know that $m_n(t) = (\hat{\beta}_n(t) - \beta_1(t))$ is a locally square integrable martingale with

$$< m_n >_t = \int_0^t \frac{1}{Y_n(s)}\frac{dQ_1(s)}{(1-H(s-))}$$

Hence

$$\sqrt{n}(\hat{\beta}_n(t) - \beta_1(t)) \Rightarrow_{a.s.} \gamma_t$$

and

$$\begin{aligned} < \sqrt{n} m_n >_t &= \int_0^t \frac{n}{Y_n(s)}\frac{dQ_1(s)}{(1-H(s-))} \to C_1(t) \\ C_1(t) &= \int_0^t \frac{dQ_1(s)}{(1-H(s-))^2} \\ < \gamma >_t &= C_1(t) \end{aligned}$$

Also $< m_n >_t = A_n(t)$ which gives by Glivenko-Cantelli Lemma $< m_n >_t = O\left(\frac{1}{n}\right)$. Using Lenglart inequality we get

$$\sup_{s\leq t}\left|\hat{\beta}_n(s) - \hat{\beta}_1(s)\right| \to_p 0 \tag{61}$$

Hence $\hat{\beta}_n(s)$ is consistent estimate of integrated hazard rate under the independence assumption above. We note that under this assumption

$$\beta_1(t) = \int_0^t \frac{dF(s)}{(1-F(s))}$$

and

$$C_1(t) = \int_0^t \frac{dF(s)}{(1-F(s))^2(1-G(s-))}$$

With $\tau_H = \inf\{s : H(s-) < 1\}$, the above results hold for all $t < \tau_H$ only.

Lemma 4.7 *1. For $t < \tau_H$*

$$\sup_{s\leq t}\left|\exp\Big(-\int_0^t \frac{dF(s)}{(1-F(s))}\Big) - \exp\Big(-\int_0^t \frac{dN_n(s)}{Y_n(s)}\Big)\right| \to_p 0$$

2.

$$\sqrt{n}\left[\exp\Big(-\int_0^{\cdot}\frac{dF(s)}{(1-F(s))}\Big)-\exp\Big(-\int_0^{\cdot}\frac{dN_n(s)}{Y_n(s)}\Big)\right]\to_{\mathcal{D}}\gamma.\exp\Big(-\int_0^{\cdot}\frac{dF(s)}{(1-F(s))}\Big)$$

in $D[0,t]$ *for* $t<\tau_H$ *with* γ *as above.*

Proof. Using Taylor expansion we get

$$\exp\Big(-\int_0^t\frac{dF(s)}{(1-F(s))}\Big)-\exp(-\hat\beta_n(t))=\exp\Big(-\int_0^t\frac{dF(s)}{(1-F(s))}\Big)\Big\{(\beta_1(t)-\hat\beta_n(t))+\frac{(\beta_1(t)-\hat\beta_n(t))^2}{2}\exp(-h_n)$$

with h_n is a random variable satisfying $\beta_1(t)\wedge\hat\beta_n(t)\le h_n\le\beta_1(t)\vee\hat\beta_n(t)$. Since for $t<\tau_H$, $\exp(-h_n)$ is bounded by convergence of $\sup_{s\le t}(\hat\beta_n(s)-\beta_1(s))\to_p 0$, the result follows. To prove the second part, note that

$$\sqrt{n}(\beta_1(\cdot)-\hat\beta_n(\cdot))^2=\sqrt{n}(\beta_1(\cdot)-\hat\beta_n(\cdot))(\beta_1(\cdot)-\hat\beta_n(\cdot))\Rightarrow_{\mathcal{D}}\gamma.\cdot 0=0$$

by Slutsky theorem and the first term converges in distribution to $\gamma.$.

Theorem 4.3 (R.Gill) *Let* $\hat F_n(t)=\prod_{s\le t}\Big(1-\frac{\Delta N_n(s)}{Y_n(s)}\Big)$. *Then under independence of* $\{X_i\}$,$\{U_i\}$, *we get that*

$$\frac{n(\hat F_n(\cdot)-F(\cdot))}{1-F(\cdot)}\Rightarrow\gamma.$$

in $D[0,t]$ *for* $t<\tau_H$.

Proof. We have $\exp(-\beta_1(t))=1-F(t)$ for $t<\tau_H$. Hence by previous Lemma

$$\frac{\sqrt{n}}{(1-F(\cdot))}\big(\exp(-\hat\beta_n(\cdot))-(1-F(\cdot))\big)\Rightarrow_{\mathcal{D}}\gamma.$$

Using (3.11) we get the result.

5 Central Limit Theorems for dependent random variables.

In this chapter we study central limit theorems for dependent random variable using Skorokhod embedding theorem.

Theorem 5.1 *Martingale Central Limit Theorem (discrete) Let $\{S_n\}$ be a martingale. Let $S_0 = 0$ and $\{W(t), 0 \leq t < \infty\}$ be Brownian motion. Then there exists a sequence of stopping time, $0 = T_0 \leq T_1 \leq T_2 \cdots \leq T_n$ with respect to $\mathcal{F}_t^W$ such that*

$$(S_0, ..., S_n) =_d (W(T_0), ..., W(T_n))$$

Proof. We use induction.
$T_0 = 0$
Assume there exists $(T_0, ..., T_{k-1})$ such that

$$(S_0, ..., S_{k-1}) =_d (W(T_0), ..., W(T_{k-1}))$$

Note that the strong Markov property implies $\{W(T_{k-1} + t) - W(T_{k-1}), t \geq 0\}$ is a Brownian motion, independent of $\mathcal{F}_t^W$. Look at conditional distribution of $S_k - S_{k-1}$ given $S_0 = s_0, ..., S_{k-1} = s_{k-1}$, if it is regular. Denote it by

$$\mu(S_0, ..., S_{k-1}; B \in \mathcal{B}(R)) = P\Big(S_k - S_{k-1} \in B \big| S_0 = s_0, ..., S_{k-1} = s_{k-1}\Big)$$

Since S_k is a martingale, we have

$$0 = E\Big(S_k - S_{k-1} \big| S_0, ..., S_{k-1}\Big) = \int x \mu_k(S_0, ..., S_{k-1} l; dx)$$

By Skorokhod's representation theorem, we see that for a.e. $S \equiv (S_0, ..., S_{k-1})$, there exists a stopping time $\tilde{\tau}_S$(exist time from (U_k, V_k)) such that

$$W(T_{k-1} + \tilde{\tau}_S) - W(T_{k-1}) = \tilde{W}(\tau_k) =_d \mu_k(S_0, ..., S_{k-1}; \cdot)$$

We let $T_k = T_{k-1} + \tilde{\tau}_S$ then

$$(S_0, S_1, ..., S_k) =_d (W(T_0), ..., W(T_k))$$

and the result follows by induction.

Remark. If $E(S_k - S_{k-1})^2 < \infty$, then

$$E\Big(\tilde{\tau}_S \big| S_0, ..., S_{k-1}\Big) = \int x^2 \mu_k(S_0, ..., S_{k-1}; dx)$$

since $W_t^2 - t$ is a martingale and $\tilde{\tau}_S$ is the exit time from a randomly chosen interval $(S_{k-1} + U_k, S_{k-1} + V_k)$.

Definition 5.1 *We say that $X_{n,m}, \mathcal{F}_{n,m}, 1 \le m \le n$, is a* ***martingale difference array*** *if $X_{n,m}$ is $\mathcal{F}_{n,m}-$measurable and $E(X_{n,m}|\mathcal{F}_{n,m-1}) = 0$ for $1 \le m \le n$, where $\mathcal{F}_{n,0} = \{\emptyset, \Omega\}$.*

Notation. Let

$$S_{(u)} = \begin{cases} S_k, & \text{if } u = k \in \mathbf{N}; \\ \text{linear on } u, & \text{if } u \in [k, k+1] \text{ for } k \in \mathbf{N}. \end{cases}$$

and

$$S_{n,(u)} = \begin{cases} S_{n,k}, & \text{if } u = k \in N; \\ \text{linear on } u, & \text{if } u \in [k, k+1]. \end{cases}$$

Consider $X_{n,m}(1 \le m \le n)$ be triangular arrays of random variables with

$$\begin{aligned} EX_{n,m} &= 0 \\ S_{n,m} &= X_{n,1} + \cdots + X_{n,m} \end{aligned}$$

Lemma 5.1 *If*

$$S_{n,m} = W(\tau_m^n)$$

and

$$\tau_{[ns]}^n \to_P s \textit{ for } s \in [0,1]$$

then $||S_{n,(n\cdot)} - W(\cdot)||_\infty \to_P 0$.

Proof is given before.

Theorem 5.2 *Let $\{X_{n,m}, \mathcal{F}_{n,m}\}$ be a martingale difference array and $S_{n,m} = X_{n,1} + \cdots + X_{n,m}$. Assume that*

1. *$|X_{n,m}| \leq \epsilon_n$ for all m, and $\epsilon_n \to 0$ as $n \to \infty$*
2. *with $V_{n,m} = \sum_{k=1}^{m} E\Big(X_{n,k}^2 \big| \mathcal{F}_{n,k-1}\Big)$, $V_{n,[nt]} \to t$ for all t.*

Then $S_{n,(n\cdot)} \Rightarrow W(\cdot)$.

Proof. We stop $V_{n,k}$ first time if it exceeds 2(call it k_0) and set $X_{n,m} = 0, m > k_0$. We can assume without loss of generality

$$V_{n,n} \leq 2 + \epsilon_n^2$$

for all n. By theorem (5.1) we can find stopping times $T_{n,1}, ..., T_{n,n}$ so that

$$(0, S_{n,1}, ..., S_{n,n}) =_d (W(0), W(T_{n,1}), ..., W(T_{n,n}))$$

By lemma (5.1), it suffices to show that $T_{n,[nt]} \to_P t$ for each t. Let

$$t_{n,m} = T_{n,m} - T_{n,m-1} \text{ with } (T_{n,0} = 0)$$

By Skohorod embedding theorem, we have

$$E\Big(t_{n,m} \big| \mathcal{F}_{n,m-1}\Big) = E\Big(X_{n,m}^2 \big| \mathcal{F}_{n,m-1}\Big)$$

The last observation with hypothesis (2) imply

$$\sum_{m=0}^{[nt]} E\Big(t_{n,m} \big| \mathcal{F}_{n,m-1}\Big) \to_P t$$

Observe that

$$\begin{aligned}
E\Big(T_{n,[nt]} - V_{n,[nt]}\Big)^2 &= E\Bigg(\sum_{m=1}^{[nt]} \Big(\underbrace{t_{n,m} - E\Big(t_{n,m} \big| \mathcal{F}_{n,m-1}\Big)}_{\text{these two terms are orthogonal}} \Big)\Bigg)^2 \\
&= \sum_{m=1}^{[nt]} E\Bigg(t_{n,m} - E\Big(t_{n,m} \big| \mathcal{F}_{n,m-1}\Big)\Bigg)^2 \\
&\leq \sum_{m=1}^{[nt]} E\Big(t_{n,m}^2 \big| \mathcal{F}_{n,m-1}\Big) \\
&\leq \sum_{m=1}^{[nt]} C \cdot E\Big(X_{n,m}^4 \big| \mathcal{F}_{n,m-1}\Big) \text{ (we will show}
\end{aligned}$$

$$
\begin{aligned}
&\quad \text{that } C = 4.) \qquad (62)\\
\le &\quad \sum_{m=1}^{[nt]} C\epsilon_n^2 E\Big(X_{n,m}^2 \Big| \mathcal{F}_{n,m-1}\Big) \text{ (by assumption (1))}\\
= &\quad C\epsilon_n^2 V_{n,n}\\
\le &\quad C\epsilon_n^2(2+\epsilon_n^2) \to 0
\end{aligned}
$$

Since L^2 convergence implies convergence in probability,

$$
E\Bigg(T_{n,[nt]} - V_{n,[nt]}\Bigg)^2 \longrightarrow 0
$$

and

$$
V_{n,[nt]} \longrightarrow_P 0
$$

together implies

$$
T_{n,[nt]} \longrightarrow_P t
$$

Proof of (62) If θ is real, then

$$
E\Bigg(\exp\Big(\theta\big(W(t) - W(s)\big) - \frac{1}{2}\theta^2(t-s)\Big)\Big|\mathcal{F}_S^W\Bigg) = 1
$$

Since

$$
E\Bigg(\exp\Big(\theta W(t) - \frac{1}{2}\theta^2 t\Big)\Big|\mathcal{F}_s^W\Bigg) = \exp\Big(\theta W(s) - \frac{1}{2}\theta^2 s\Big),
$$

we know that $\Big\{\exp\Big(\theta W(t) - \frac{1}{2}\theta^2 t\Big), \mathcal{F}_t^W\Big\}$ is a Martingale. Then, for all $A \in \mathcal{F}_s^W$,

$$
\begin{aligned}
E1_A\Bigg(\exp\Big(\theta W(s) - \frac{1}{2}\theta^2 s\Big)\Bigg) &= \int_A \Bigg(\exp\Big(\theta W(s) - \frac{1}{2}\theta^2 s\Big)\Bigg) dP\\
&= \int_A \exp\Big(\theta W(t) - \frac{1}{2}\theta^2 t\Big) dP \text{ (by definition of conditional expectation)}\\
&= E1_A\Bigg(\exp\Big(\theta W(t) - \frac{1}{2}\theta^2 t\Big)\Bigg)
\end{aligned}
$$

Take a derivative in θ and find a value at $\theta = 0$.

number of derivative	
1	$W(t)$ is MG
2	$W^2(t) - t$ is MG
3	$W^3(t) - 3tW(t)$ is MG
4	$W^4(t) - 6tW^2(t) + 3t^2$ is MG

For any stopping time τ,

$$E(W^4(\tau) - 6\tau W^2(\tau) + 3\tau^2) = 0$$

Therefore,

$$\begin{aligned} &\underbrace{EW_\tau^4}_{\geq 0} - 6E(\tau W_\tau^2) = -3EW_\tau^2 \\ \Rightarrow \quad & EW_\tau^2 \leq 2E(\tau W_\tau^2) \end{aligned}$$

Since

$$E(\tau W_\tau^2) \leq \Big(E\tau^2\Big)^{1/2} \cdot \Big(EW_\tau^4\Big)^{1/2}$$

by Schwartz Inequality, we have

$$\Big(E\tau^2\Big)^{1/2} \leq 2\Big(EW_\tau^4\Big)^{1/2}$$

Therefore,

$$E\Big(t_{n,m}^2 \big| \mathcal{F}_{n,m-1}\Big) \leq 4E\Big(X_{n,m}^4 \big| \mathcal{F}_{n,m-1}\Big)$$

Theorem 5.3 ***(Generalization of Lindberg-Feller Theorem.)*** *Let $\{X_{n,m}, \mathcal{F}_{n,m}\}$ be a martingale difference array and $S_{n,m} = X_{n,1} + \cdots + X_{n,m}$. Assume that*

1. $V_{n,[nt]} = \sum_{k=1}^{[nt]} E\Big(X_{n,k}^2 \big| \mathcal{F}_{n,k-1}\Big) \to_P t$
2. $\widehat{V}(\epsilon) = \sum_{m \leq n} E\Big(X_{n,m}^2 1\big(|X_{n,m}| > \epsilon\big) \big| \mathcal{F}_{n,m-1}\Big) \to_P 0$, *for all* $\epsilon > 0$.

Then $S_{n,(n\cdot)} \Rightarrow W(\cdot)$. For i.i.d. $X_{n,m}$, and $t = 1$, we get Lindberg-Feller Theorem.

Lemma 5.2 *There exists $\epsilon_n \to 0$ such that $\epsilon_n^2 \hat{V}(\epsilon_n) \to_P 0$.*

Proof. Since $\hat{V}(\epsilon) \to_P 0$, we choose large N_m such that

$$P\left(\hat{V}\Big(\frac{1}{m}\Big) > \frac{1}{m^3}\right) < \frac{1}{m}$$

for $m \geq N_m$. Here we choose $\epsilon_n = \frac{1}{m}$ with $n \in [N_m, N_{m+1})$. For $\delta > \frac{1}{m}$, we have

$$P\left(\epsilon_n^{-2}\widehat{V}(\epsilon_n) > \delta\right) \leq P\left(m^2\widehat{V}\Big(\frac{1}{m}\Big) > \frac{1}{m}\right) < \frac{1}{m}.$$

This completes the proof of lemma.

Let

$$\begin{aligned}
\overline{X}_{n,m} &= X_{n,m}1(|X_{n,m}| \le \epsilon_n) \\
\widehat{X}_{n,m} &= X_{n,m}1(|X_{n,m}| > \epsilon_n) \\
\widetilde{X}_{n,m} &= \overline{X}_{n,m} - E\Big(\overline{X}_{n,m}\big|\mathcal{F}_{n,m-1}\Big)
\end{aligned}$$

Lemma 5.3 $\widetilde{S}_{n,[n\cdot]} \Rightarrow W(\cdot)$

Proof. We will show that $\widetilde{X}_{n,m}$ satisfies Theorem (5.2).

$$\begin{aligned}
|\widetilde{X}_{n,m}| &= \Big|\overline{X}_{n,m} - E\Big(\overline{X}_{n,m}\big|\mathcal{F}_{n,m-1}\Big)\Big| \\
&\le \Big|\overline{X}_{n,m}\Big| + \Big|E\Big(\overline{X}_{n,m}\big|\mathcal{F}_{n,m-1}\Big)\Big| \\
&\le 2\epsilon_n \to 0
\end{aligned}$$

and hence, the first condition is satisfied. Since

$$X_{n,m} = \overline{X}_{n,m} + \widetilde{X}_{n,m},$$

we have

$$\begin{aligned}
E\Big(\overline{X}^2_{n,m}\big|\mathcal{F}_{n,m-1}\Big) &= E\Big((X_{n,m} - \widehat{X}_{n,m})^2\big|\mathcal{F}_{n,m-1}\Big) \\
&= E\Big(X^2_{n,m} - 2X_{n,m}\widehat{X}_{n,m} + \widehat{X}_{n,m})^2\big|\mathcal{F}_{n,m-1}\Big) \\
&= E\Big(X^2_{n,m}\big|\mathcal{F}_{n,m-1}\Big) - E\Big(\widehat{X}^2_{n,m}\big|\mathcal{F}_{n,m-1}\Big). \qquad (63)
\end{aligned}$$

Last equality follows from $E\Big(X_{n,m}\widehat{X}_{n,m}\big|\mathcal{F}_{n,m-1}\Big) = E\Big(\widehat{X}^2_{n,m}\big|\mathcal{F}_{n,m-1}\Big)$. Since $X_{n,m}$ is a martingale difference array, and hence, $E(X_{n,m}|\mathcal{F}_{n,m-1}) = 0$. The last observation implies $E\Big(\overline{X}_{n,m}\big|\mathcal{F}_{n,m-1}\Big) = -E\Big(\widehat{X}_{n,m}\big|\mathcal{F}_{n,m-1}\Big)$, and hence,

$$\begin{aligned}
\Big[E\Big(\overline{X}_{n,m}\big|\mathcal{F}_{n,m-1}\Big)\Big]^2 &= \Big[E\Big(\widehat{X}_{n,m}\big|\mathcal{F}_{n,m-1}\Big)\Big]^2 \\
&\le E\Big(\widehat{X}^2_{n,m}\big|\mathcal{F}_{n,m-1}\Big) \text{ (by Jensen's inequality)}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{m=1}^{n}\Big[E\Big(\overline{X}_{n,m}\big|\mathcal{F}_{n,m-1}\Big)\Big]^2 &\le \sum_{m=1}^{n} E\Big(\widehat{X}^2_{n,m}\big|\mathcal{F}_{n,m-1}\Big) \\
&= \widehat{V}(\epsilon_n) \\
&\to_P 0
\end{aligned}$$

by given condition. Finally,

$$
\begin{aligned}
\sum_{m=1}^{n} E\Big(\widetilde{X}_{n,m}^2 \big| \mathcal{F}_{n,m-1}\Big) &= \sum_{m=1}^{n} E\Big(\overline{X}_{n,m}^2 \big| \mathcal{F}_{n,m-1}\Big) - \sum_{m=1}^{n} E\Big(\overline{X}_{n,m} \big| \mathcal{F}_{n,m-1}\Big)^2 \\
&\quad \text{(by the conditional variance formula)} \\
&= \sum_{m=1}^{n} \bigg(E\Big(X_{n,m}^2 \big| \mathcal{F}_{n,m-1}\Big) - E\Big(\widehat{X}_{n,m}^2 \big| \mathcal{F}_{n,m-1}\Big) \bigg) - \sum_{m=1}^{n} E\Big(\overline{X}_{n,m} \big| \mathcal{F}_{n,m-1}\Big)^2 \\
&\quad \text{(from equation (63))} \\
&= \sum_{m=1}^{n} E\Big(X_{n,m}^2 \big| \mathcal{F}_{n,m-1}\Big) - \underbrace{\sum_{m=1}^{n} E\Big(\widehat{X}_{n,m}^2 \big| \mathcal{F}_{n,m-1}\Big)}_{=\widehat{V}(\epsilon_n) \to_P 0} \\
&\quad - \underbrace{\sum_{m=1}^{n} \Big[E\Big(\overline{X}_{n,m} \big| \mathcal{F}_{n,m-1}\Big) \Big]^2}_{\to_P 0},
\end{aligned}
$$

and hence, we get

$$\lim_{n \to \infty} \sum_{m=1}^{n} E\Big(\widetilde{X}_{n,m}^2 \big| \mathcal{F}_{n,m-1}\Big) = \lim_{n \to \infty} \sum_{m=1}^{n} E\Big(X_{n,m}^2 \big| \mathcal{F}_{n,m-1}\Big)$$

Since

$$V_{n,[nt]} = \sum_{m=1}^{[nt]} E\Big(X_{n,m}^2 \big| \mathcal{F}_{n,m-1}\Big) \to_P t,$$

we conclude that

$$\sum_{m=1}^{[nt]} E\Big(\widetilde{X}_{n,m}^2 \big| \mathcal{F}_{n,m-1}\Big) \to_P t,$$

This show that the second condition is satisfied, and hence, completes the proof.

Lemma 5.4

$$||S_{n,(n\cdot)} - \widetilde{S}_{n,(n\cdot)}||_\infty \leq \sum_{m=1}^{n} \Big| E\Big(\overline{X}_{n,m} \big| \mathcal{F}_{n,m-1}\Big) \Big|$$

Proof. Note that if we prove this lemma, then, since $\sum_{m=1}^{n} \Big| E\Big(\overline{X}_{n,m} \big| \mathcal{F}_{n,m-1}\Big) \Big| \to_P 0$(we will show this), and we construct a Brownian motion with $||\widetilde{S}_{n,(n\cdot)} - W(\cdot)||_\infty \to 0$, the desired result follows from the triangle inequality.

Since $X_{n,m}$ is martingale difference array, we know that $E\Big(\overline{X}_{n,m}|\mathcal{F}_{n,m-1}\Big) = -E\Big(\widehat{X}_{n,m}|\mathcal{F}_{n,m-1}\Big)$, and hence,

$$\begin{aligned}
\sum_{m=1}^{n}\left|E\Big(\overline{X}_{n,m}|\mathcal{F}_{n,m-1}\Big)\right| &= \sum_{m=1}^{n}\left|E\Big(\widehat{X}_{n,m}|\mathcal{F}_{n,m-1}\Big)\right| \\
&\leq \sum_{m=1}^{n}E\Big(|\widehat{X}_{n,m}||\mathcal{F}_{n,m-1}\Big) \text{ by Jensen} \\
&\leq \frac{1}{\epsilon_n}\sum_{m=1}^{n}E\Big(\widehat{X}^2_{n,m}|\mathcal{F}_{n,m-1}\Big) \\
&\quad (\text{if } |X_{n,m}| > \epsilon_n,\ \widehat{X}_{n,m} \leq \tfrac{X^2_{n,m}}{\epsilon_n} = \tfrac{\widehat{X}^2_{n,m}}{\epsilon_n}) \\
&= \frac{\widehat{V}(\epsilon_n)}{\epsilon_n} \rightarrow_p 0 \text{ (by lemma (5.2))}
\end{aligned}$$

On $\{|X_{n,m}| \leq \epsilon_n, 1 \leq m \leq n\}$, we have $\overline{X}_{n,m} = X_{n,m}$, and hence, $S_{n,(n\cdot)} = \overline{S}_{n,(n\cdot)}$. Thus,

$$\begin{aligned}
||S_{n,(n\cdot)} - \widetilde{S}_{n,(n\cdot)}||_\infty &= ||S_{n,(n\cdot)} - \overline{S}_{n,(n\cdot)} + \sum_{m=1}^{[n\cdot]}E\Big(\overline{X}_{n,m}|\mathcal{F}_{n,m-1}\Big)||_\infty \\
&\leq \sum_{m=1}^{[n\cdot]}\left|E\Big(\overline{X}_{n,m}|\mathcal{F}_{n,m-1}\Big)\right| \rightarrow_P 0
\end{aligned}$$

Now, complete the proof, we have to show that lemma (5.4) holds on $\{|X_{n,m}| > \epsilon_n, 1 \leq m \leq n\}$. It suffices to show that

Lemma 5.5

$$P\Big(|X_{n,m}| > \epsilon_n,\ \textit{for some } m, 1 \leq m \leq n\Big) \rightarrow 0$$

To prove lemma (5.5), we use Dvoretsky's proposition.

Proposition 5.1 *(Dvoretsky) Let $\{\mathcal{G}_n\}$ be a sequence of $\sigma-$fields with $\mathcal{G}_n \subset \mathcal{G}_{n+1}$. If $A_n \in \mathcal{G}_n$ for each n, then for each $\delta \geq 0$, measurable with respect to $\mathcal{G}_0$,*

$$P\Big(\bigcup_{m=1}^{n}A_m|\mathcal{G}_0\Big) \leq \delta + P\Big(\sum_{m=1}^{n}P(A_m|\mathcal{G}_{m-1}) > \delta|\mathcal{G}_0\Big) \tag{64}$$

Proof. We use induction.

i) $n=1$
We want to show

$$P\Big(A_1|\mathcal{G}_0\Big) \leq \delta + P\Big(P(A_1|\mathcal{G}_0) > \delta|\mathcal{G}_0\Big) \tag{65}$$

Consider $\Omega_\ominus = \{\omega : P(A_1|\mathcal{G}_0) \leq \delta\}$. Then (65) holds. Also, consider $\Omega_\oplus = \{\omega : P(A_1|\mathcal{G}_0) > \delta\}$. Then

$$\begin{aligned}
P\Big(P(A_1|\mathcal{G}_0) > \delta|\mathcal{G}_0\Big) &= E\Big(\mathbf{1}\Big(P(A_1|\mathcal{G}_0) > \delta\Big)|\mathcal{G}_0\Big) \\
&= \mathbf{1}\Big(P(A_1|\mathcal{G}_0) > \delta\Big) \\
&= 1
\end{aligned}$$

and hence (65) also holds.

ii) $n>1$
Consider $\omega \in \Omega_\oplus$. Then

$$\begin{aligned}
P\Big(\sum_{m=1}^{n} P(A_m|\mathcal{G}_{m-1}) > \delta|\mathcal{G}_0\Big) &\geq P\Big(P(A_1|\mathcal{G}_0) > \delta|\mathcal{G}_0\Big) \\
&= 1_{\Omega_\oplus}(\omega) \\
&= 1
\end{aligned}$$

Then, (64) holds. Consider $\omega \in \Omega_\ominus$. Let $B_m = A_m \cap \Omega_\ominus$. Then, for $m \geq 1$,

$$\begin{aligned}
P(B_m|\mathcal{G}_{m-1}) &= P(A_m \cap \Omega_\ominus|\mathcal{G}_{m-1}) \\
&= P(A_m|\mathcal{G}_{m-1}) \cdot P(\Omega_\ominus|\mathcal{G}_{m-1}) \\
&= P(A_m|\mathcal{G}_{m-1}) \cdot 1_{\Omega_\ominus}(\omega) \\
&= P(A_m|\mathcal{G}_{m-1})
\end{aligned}$$

Now suppose $\gamma = \delta - P(B_1|\mathcal{G}_0) \geq 0$, and apply the last result for $n-1$ sets(induction hypothesis).

$$P\Big(\bigcup_{m=2}^{n} B_m|\mathcal{G}_1\Big) \leq \gamma + P\Big(\sum_{m=2}^{n} P(B_m|\mathcal{G}_{m-1}) > \gamma|\mathcal{G}_1\Big).$$

Recall $E\Big(E(X|\mathcal{G}_0)\big|\mathcal{G}_1\Big) = E(X|\mathcal{G}_0)$ if $\mathcal{G}_0 \subset \mathcal{G}_1$. Taking conditional expectation w.r.t $\mathcal{G}_0$ and noting $\gamma \in \mathcal{G}_0$,

$$\begin{aligned}
P\Big(\bigcup_{m=2}^{n} B_m|\mathcal{G}_0\Big) &\leq P\left(\gamma + P\Big(\sum_{m=2}^{n} P(B_m|\mathcal{G}_{m-1}) > \gamma|\mathcal{G}_1\Big)|\mathcal{G}_0\right) \\
&= \gamma + P\Big(\sum_{m=2}^{n} P(B_m|\mathcal{G}_{m-1}) > \gamma|\mathcal{G}_0\Big) \\
&= \gamma + P\Big(\sum_{m=1}^{n} P(B_m|\mathcal{G}_{m-1}) > \delta|\mathcal{G}_0\Big)
\end{aligned}$$

Since $\cup B_m = (\cup A_m) \cap \Omega_\ominus$, on $\Omega_\ominus$ we have

$$\sum_{m=1}^{n} P(B_m|\mathcal{G}_{m-1}) = \sum_{m=1}^{n} P(A_m|\mathcal{G}_{m-1}).$$

Thus, on $\Omega_\ominus$,

$$P\Big(\bigcup_{m=2}^{n} A_m|\mathcal{G}_0\Big) \le \delta - P(A_1|\mathcal{G}_0) + P\Big(\sum_{m=1}^{n} P(A_m|\mathcal{G}_{m-1}) > \delta|\mathcal{G}_0\Big).$$

The result follows from

$$P\Big(\bigcup_{m=1}^{n} A_m|\mathcal{G}_0\Big) \le P(A_1|\mathcal{G}_0) + P\Big(\bigcup_{m=2}^{n} A_m|\mathcal{G}_0\Big)$$

by using monotonicity of conditional expectation and $1_{A\cup B} \le 1_A + 1_B$.

Proof of lemma (5.5). Let $A_m = \{|X_{n,m}| > \epsilon_n\}$, $\mathcal{G}_m = \mathcal{F}_{n,m}$, and let δ be a positive number. Then, lemma (4.3.3) implies

$$P(|X_{n,m}| > \epsilon_n \text{ for some } m \le n) \le \delta + P\Big(\sum_{m=1}^{n} P(|X_{n,m}| > \epsilon_n|\mathcal{F}_{n,m-1}) > \delta\Big)$$

To estimate the right-hand side, we observe that "Chebyshev's inequality" implies

$$\sum_{m=1}^{n} P(|X_{n,m}| > \epsilon_n|\mathcal{F}_{n,m-1}) \le \epsilon_n^{-2} \sum_{m=1}^{n} E(\widehat{X}_{n,m}^2|\mathcal{F}_{n,m-1}) \to 0$$

so $\limsup P(|X_{n,m}| > \epsilon_n$ for some $m \le n) \le \delta$. Since δ is arbitrary, the proof of lemma and theorem is complete.

Theorem 5.4 ***(Martingale cental limit theorem)*** *Let $\{X_n, \mathcal{F}_n\}$ be Martingale difference sequence, $X_{n,m} = X_m/\sqrt{n}$, and $V_k = \sum_{n=1}^{k} E\Big(X_n^2\Big|\mathcal{F}_{n-1}\Big)$. Assume that*

1. *$V_k/k \to_P \sigma^2$*
2. *$n^{-1}\sum_{m\le n} E\Big(X_m^2 1\Big(|X_m| > \epsilon\sqrt{n}\Big)\Big) \to 0$*

Then,

$$\frac{S_{(n\cdot)}}{\sqrt{n}} \Rightarrow \sigma W(\cdot)$$

Definition 5.2 *A process $\{X_n, n \ge 0\}$ is called* ***stationary*** *if*

$$\{X_m, X_{m+1}, ..., X_{m+k}\} =_{\mathcal{D}} \{X_0, X_1, ..., X_k\}$$

for any m, k.

Definition 5.3 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A measurable map $\varphi : \Omega \to \Omega$ is said to be* ***measure preserving*** *if $P(\varphi^{-1}A) = P(A)$ for all $A \in \mathcal{F}$.*

Theorem 5.5 *If $X_0, X_1, ...$ is a stationary sequence and $g : \boldsymbol{R}^{\boldsymbol{N}} \to \boldsymbol{R}$ is measurable then $Y_k = g(X_k, X_{k+1}, ...)$ is a stationary sequence.*

Proof. If $\mathbf{x} \in \mathbf{R}^{\mathbf{N}}$, let $g_k(\mathbf{x}) = g(x_k, x_{k+1}, ...)$, and if $B \in \mathcal{R}^{\mathbf{N}}$ let

$$A = \{\mathbf{x} : (g_0(\mathbf{x}), g_1(\mathbf{x}), ...) \in B\}$$

To check stationarity now, we observe:

$$\begin{aligned} P\Big(\{\omega : (Y_0, Y_1, ...) \in B\}\Big) &= P\Big(\{\omega : (g(X_0, X_1, ...), g(X_1, X_2, ...), ...) \in B\}\Big) \\ &= P\Big(\{\omega : (g_0(\mathbf{X}), g_1(\mathbf{X}), ...) \in B\}\Big) \\ &= P\Big(\{\omega : (X_0, X_1, ...) \in A\}\Big) \\ &= P\Big(\{\omega : (X_k, X_{k+1}, ...) \in A\}\Big) \text{ (since } X_0, X_1, ... \text{ is a stationary sequence)} \\ &= P\Big(\{\omega : (Y_k, Y_{k+1}, ...) \in B\}\Big) \text{ (Check this!)} \end{aligned}$$

Definition 5.4 *Assume that θ is measure preserving. A set $A \in \mathcal{F}$ is said to be* ***invariant*** *if $\theta^{-1}A = A$.*

Definition 5.5 *A measure preserving transformation on $(\Omega, \mathcal{F}, P)$ is said to be ergodic if $\mathcal{I}$ is trivial, i.e., for every $A \in \mathcal{I}$, $P(A) \in \{0, 1\}$.*

Example. Let $X_0, X_1, ...$ be i.i.d. sequence. If $\Omega = \mathbf{R}^{\mathbf{N}}$ and θ is the shift operator, then an invariant set A has $\{\omega : \omega \in A\} = \{\omega : \theta\omega \in A\} \in \sigma(X_1, X_2, ...)$. Iterating gives

$$A \in \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, ...) = \mathcal{T}, \text{the tail } \sigma\text{−field}$$

so $\mathcal{I} \subset \mathcal{T}$. For an i.i.d. sequence, Kolmogorov's 0-1 law implies $\mathcal{T}$ is trivial, so $\mathcal{I}$ is trivial and the sequence is ergodic. We call θ is ergodic transformation.

Theorem 5.6 *Let $g : \boldsymbol{R}^{\boldsymbol{N}} \to \boldsymbol{R}$ be measurable. If $X_0, X_1, ...$ is an ergodic stationary sequence, then $Y_k = g(X_k, X_{k+1}, ...)$ is ergodic.*

Proof. Suppose $X_0, X_1, ...$ is defined on sequence space with $X_n(\omega) = \omega_n$. If B has $\{\omega : (Y_0, Y_1, ...) \in B\} = \{\omega : (Y_1, Y_2, ...) \in B\}$, then $A = \{\omega : (Y_0, Y_1, ...) \in B\}$ is shift invariant.

Theorem 5.7 *Birkhoff's Ergodic Theorem. For any* $f \in L_1(P)$,

$$\frac{1}{n}\sum_{m=0}^{n-1} f(\theta^m \omega) \to E(f|\mathcal{G}) \ \textit{a.s. and in } L_1(P)$$

where θ *is measure preserving transformation on* $(\Omega, \mathcal{F}, P)$ *and* $\mathcal{G} = \{A \in \mathcal{F} : \theta^{-1}A = A\}$.

The proof is based on an odd integration inequality due to Yosida and Kakutani(1939).

Theorem 5.8 *Suppose* $\{X_n, n \in \boldsymbol{Z}\}$ *is an ergodic stationary sequence of martingale differences, i.e.,* $\sigma^2 = EX_n^2 < \infty$ *and* $E(X_n|\mathcal{F}_{n-1}) = 0$ *with respect to* $\mathcal{F}_n = \sigma(X_m, m \le n)$. *Let* $S_n = X_1 + \cdots + X_n$. *Then,*

$$\frac{S_{(n\cdot)}}{\sqrt{n}} \Rightarrow \sigma W(\cdot)$$

Proof. Let $u_n = E(X_n^2|\mathcal{F}_{n-1})$. Then u_n can be written as $\theta(X_{n-1}, X_{n-2}, ...)$, and hence by theorem (5.6), u_n is stationary and ergodic. By Birkhoff's ergodic theorem($\mathcal{G} = \{\emptyset, \Omega\}$),

$$\frac{1}{n}\sum_{m=1}^{n} u_m \to Eu_0 = EX_0^2 = \sigma^2 \text{ a.s.}$$

The last conclusion shows that (i) of theorem (5.4) holds. To show (ii), we observe

$$\begin{aligned}
\frac{1}{n}\sum_{m=1}^{n} E\Big(X_m^2 1\Big(|X_m| > \epsilon\sqrt{n}\Big)\Big) &= \frac{1}{n}\sum_{m=1}^{n} E\Big(X_0^2 1\Big(|X_0| > \epsilon\sqrt{n}\Big)\Big) \text{ (because of stationarity)} \\
&= E\Big(X_0^2 1\Big(|X_0| > \epsilon\sqrt{n}\Big)\Big) \to 0
\end{aligned}$$

by the dominated convergence theorem. This completes the proof.

Let's consider stationary process

$$X_n = \sum_{k=0}^{\infty} c_k \xi_{n-k} \quad \xi_i \text{are i.i.d.}$$

If ξ_i are i.i.d., X_n is not definitely stationary but it is not martingale difference process. This is called **Moving Average Process**. What we will do is we start with stationary ergodic process, and then we will show that limit of this process is the limit of martingale difference sequence. Then this satisfies the conditions of martingale central limit theorem.

We can separate phenomena into two parts: new information(non-deterministic) and non-new information(deterministic).

$$EX_nX_0 = \int_0^{2\pi} e^{in\lambda} dF(\lambda)$$

where F is spectral measure. In case $X_n = \sum_{k=0}^{\infty} c_k \xi_{n-k}$, then $F \ll$ Lebesgue measure.

Theorem 5.9 *There exist $\bar{c}_k$ and ϕ such that*

- $f(e^{i\lambda}) = \left|\phi(e^{i\lambda})\right|^2$
- $\phi(e^{i\lambda}) = \sum_{k=0}^{\infty} \bar{c}_k e^{ik\lambda}$

if and only if

$$\int_0^{2\pi} \log f(\lambda) dF(\lambda) > -\infty \qquad \left(or \int_{-\infty}^{\infty} \frac{\log f(\lambda)}{1+\lambda^2} dF(\lambda) > -\infty \right)$$

Now we start with $\{X_n : n \in \mathbf{Z}\}$ ergodic stationary sequence such that

- $EX_n = 0, EX_n^2 < \infty$
- $\sum_{n=1}^{\infty} ||E(X_0|\mathcal{F}_{-n})||_2 \infty$

Idea is if we go back then X_n will be independent X_0.

Let

$$\begin{aligned} H_n &= \{Y \in \mathcal{F}_n \text{ with } EY^2 < \infty\} = L^2(\Omega, \mathcal{F}_n, P) \\ K_n &= \{Y \in H_n \text{ with } EYZ = 0 \text{ for all } Z \in H_{n-1}\} = H_n \ominus H_{n-1} \end{aligned}$$

Geometrically, $H_0 \supset H_{-1} \supset H_{-2} \ldots$ is a sequence of subspaces of L^2 and K_n is the orthogonal complement of H_{n-1}. If Y is a random variable, let

$$(\theta^n Y)(\omega) = Y(\theta^n \omega),$$

i.e., θ is isometry(measure-preserving) on L^2. Generalizing from the example $Y = f(X_{-j}, ..., X_k)$, which has $\theta^n Y = f(X_{n-j}, ..., X_{n+k})$, it is easy to see that if $Y \in H_k$, then $\theta^n Y \in H_{k+n}$, and hence $Y \in K_j$ then $\theta^n Y \in K_{n+j}$.

Lemma 5.6 *Let P be a projection such that $X_j \in H_{-j}$ implies $P_{H_{-j}} X_j = X_j$. Then,*

$$\begin{aligned} \theta P_{H_{-1}} X_j &= P_{H_0} X_{j+1} \\ &= P_{H_0} \theta X_j \end{aligned}$$

Proof. For $j \leq -1$,

$$\theta \underbrace{P_{H_{-j}} X_j}_{X_j} = \theta X_j = X_{j+1}$$

We will use this property. For $Y \in H_{-1}$,

$$\begin{aligned}
& X_j - P_{H_{-1}} X_j \perp Y \\
\Rightarrow \quad & \big(X_j - P_{H_{-1}} X_j, Y\big)_2 = 0 \\
\Rightarrow \quad & \Big(\theta\big(X_j - P_{H_{-1}} X_j\big), \theta Y\Big)_2 = 0 \text{ (since } \theta \text{ is isometry on } L^2)
\end{aligned}$$

Since $Y \in H_{-1}$, θY generates H_0. Therefore, for all $Z \in H_0$, we have

$$\begin{aligned}
& \Big(\big(\theta X_j - \theta P_{H_{-1}} X_j\big), Z\Big)_2 = 0 \\
\Rightarrow \quad & \big(\theta X_j - \theta P_{H_{-1}} X_j\big) \perp Z \\
\Rightarrow \quad & \theta P_{H_{-1}} X_j = P_{H_0} \theta X_j = P_{H_0} X_{j+1}
\end{aligned}$$

We come now to the central limit theorem for stationary sequences.

Theorem 5.10 *Suppose $\{X_n, n \in \boldsymbol{Z}\}$is an ergodic stationary sequence with $EX_n = 0$ and $EX_n^2 < \infty$. Assume*

$$\sum_{n=1}^{\infty} ||E(X_0|\mathcal{F}_{-n})||_2 < \infty$$

Let $S_n = X_1 + ... + X_n$. Then

$$\frac{S_{(n\cdot)}}{\sqrt{n}} \Rightarrow \sigma W(\cdot)$$

where we do not know what σ is.

Proof. If X_0 happened to be in K_0 since $X_n = \theta^n X_0 \in K_n$ for all n, and taking $Z = 1_A \in H_{n-1}$ we would have $E(X_n 1_A) = 0$ for all $A \in \mathcal{F}_{n-1}$ and hence $E(X_n|\mathcal{F}_{n-1}) = 0$. The next best thing to having $X_n \in K_0$ is to have

$$X_0 = Y_0 + Z_0 - \theta Z_0 \qquad (*)$$

with $Y_0 \in K_0$ and $Z_0 \in L^2$. Let

$$\begin{aligned}
Z_0 &= \sum_{j=0}^{\infty} E(X_j|\mathcal{F}_{-1}) \\
\theta Z_0 &= \sum_{j=0}^{\infty} E(X_{j+1}|\mathcal{F}_0) \\
Y_0 &= \sum_{j=0}^{\infty} \Big(E(X_j|\mathcal{F}_0) - E(X_j|\mathcal{F}_{-1})\Big)
\end{aligned}$$

Then we can solve $(*)$ formally

$$Y_0 + Z_0 - \theta Z_0 = E(X_0|\mathcal{F}_0) = X_0. \tag{66}$$

We let

$$S_n = \sum_{m=1}^{n} X_m = \sum_{m=1}^{n} \theta^m X_0 \quad \text{and } T_n = \sum_{m=1}^{n} \theta^m Y_0.$$

We want to show that T_n is martingale difference sequence. We have $S_n = T_n + \theta Z_0 - \theta^{n+1} Z_0$. The $\theta^m Y_0$ are a stationary ergodic martingale difference sequence (ergodicity follows from (5.6)), so (5.8) implies

$$\frac{T_{(n\cdot)}}{\sqrt{n}} \Rightarrow \sigma W(\cdot) \quad \text{where } \sigma^2 = EY_0^2.$$

To get rid of the other term, we observe

$$\frac{\theta Z_0}{\sqrt{n}} \to 0 \text{ a.s.}$$

and

$$\begin{aligned} P\Big(\max_{0\le m\le n-1} \big|\theta^{m+1} Z_0\big| > \epsilon\sqrt{n}\Big) &\le \sum_{m=0}^{n-1} P\Big(\big|\theta^{m+1} Z_0\big| > \epsilon\sqrt{n}\Big) \\ &= nP\Big(\big|Z_0\big| > \epsilon\sqrt{n}\Big) \\ &\le \epsilon^{-2} E\Big(Z_0^2 \mathbf{1}_{\{|Z_0|>\epsilon\sqrt{n}\}}\Big) \to 0 \end{aligned}$$

The last inequality follows from the stronger form of Chevyshev,

$$E\Big(Z_0^2 \mathbf{1}_{\{|Z_0|>\epsilon\sqrt{n}\}}\Big) \ge \epsilon^2 nP\Big(\big|Z_0\big| > \epsilon\sqrt{n}\Big).$$

Therefore,

$$\begin{aligned} \frac{S_n}{\sqrt{n}} &= \frac{T_n}{\sqrt{n}} + \underbrace{\frac{\theta Z_0}{\sqrt{n}}}_{\to_p 0} - \underbrace{\frac{\theta^{n+1} Z_0}{\sqrt{n}}}_{\to_p 0} \\ \Rightarrow \quad \frac{S_n}{\sqrt{n}} &\to_p \frac{T_n}{\sqrt{n}} \\ \Rightarrow \quad \lim_{n\to\infty} \frac{S_{(n\cdot)}}{\sqrt{n}} &= \lim_{n\to\infty} \frac{T_{(n\cdot)}}{\sqrt{n}} = \sigma W(\cdot). \end{aligned}$$

Theorem 5.11 *Suppose* $\{X_n, n \in \boldsymbol{Z}\}$*is an ergodic stationary sequence with* $EX_n = 0$ *and* $EX_n^2 < \infty$. *Assume*

$$\sum_{n=1}^{\infty} ||E(X_0|\mathcal{F}_{-n})||_2 < \infty$$

Let $S_n = X_1 + ... + X_n$. *Then*

$$\frac{S_{(n\cdot)}}{\sqrt{n}} \Rightarrow \sigma W(\cdot)$$

where

$$\sigma^2 = EX_0^2 + 2\sum_{n=1}^{\infty} EX_0 X_n.$$

If $\sum_{n=1}^{\infty} EX_0X_n$ diverges, theorem will not be true. We will show that $\sum_{n=1}^{\infty}\Big|EX_0X_n\Big| < \infty$. This theorem is different from previous theorem since we now specify σ^2.

Proof. First,

$$\begin{aligned}
\Big|EX_0X_m\Big| &= \Bigg|E\Big(E\big(X_0X_m|\mathcal{F}_0\big)\Big)\Bigg| \\
&\leq E\Big|X_0E\big(X_m|\mathcal{F}_0\big)\Big| \\
&\leq ||X_0||_2 \cdot \Big|\Big|E\big(X_m|\mathcal{F}_0\big)\Big|\Big|_2 \text{ (by Cauchy Schwarz inequality)} \\
&= ||X_0||_2 \cdot \Big|\Big|E\big(X_0|\mathcal{F}_{-m}\big)\Big|\Big|_2 \text{ (by shift invariance)}
\end{aligned}$$

Therefore, by assumption,

$$\sum_{n=1}^{\infty}\big|EX_0X_n\big| \leq ||X_0||_2 \sum_{n=1}^{\infty}\big|\big|E\big(X_0|\mathcal{F}_{-m}\big)\big|\big|_2 < \infty.$$

Next,

$$\begin{aligned}
ES_n^2 &= \sum_{j=1}^{n}\sum_{k=1}^{n} EX_jX_k \\
&= nEX_0^2 + 2\sum_{m=1}^{n-1}(n-m)E_0X_m
\end{aligned}$$

From this, it follows easily that

$$\frac{ES_n^2}{n} \to EX_0^2 + 2\sum_{m=1}^{\infty} E_0X_m.$$

To finish the proof, let $T_n = \sum_{m=1}^{n} \theta^m Y_0$, observe $\sigma^2 = EY_0^2$(we proved), and

$$\begin{aligned}
n^{-1}E(S_n - T_n)^2 &= n^{-1}E(\theta Z_0 - \theta^{n+1}Z_0)^2 \\
&\leq \frac{3EZ_0^2}{n} \to 0
\end{aligned}$$

since $(a-b)^2 \leq (2a)^2 + (2b)^2$.

We proved central limit theorem of ergodic stationary process. We will discuss examples: M-dependence and Moving Average.

Example 1. M-dependent sequences. Let X_n, $n \in \mathbf{Z}$ be a stationary sequence with $EX_n = 0, EX_n^2 < \infty$. Assume that $\sigma(\{X_j, j \le 0\})$ and $\sigma(\{X_j, j \ge M\})$ are independent. In this case, $E(X_0|\mathcal{F}_{-n}) = 0$ for $n > M$, and $\sum_{n=0}^{\infty} ||E(X_0|\mathcal{F}_{-n})||_2 < \infty$. Let $\mathcal{F}_{-\infty} = \cap_m \sigma(\{X_j, j \ge M\})$ and $\mathcal{F}_k = \sigma(\{X_j, j \le k\})$. If $m - k > M$, then $\mathcal{F}_{-\infty} \perp \mathcal{F}_k$. Recall Kolmogorov 0-1 law. If $A \in \mathcal{F}_k$ and $B \in \mathcal{F}_{-\infty}$, then $P(A \cap B) = P(A) \cdot P(B)$. For all $A \in \cup_k \mathcal{F}_k$, $A \in \sigma\big(\cup_k \mathcal{F}_k\big)$. Also, $A \in \mathcal{F}_{-\infty}$ where $\mathcal{F}_{-\infty} \subset \cup_k \mathcal{F}_{-k}$. Therefore, by Kolmogorov 0-1 law, $P(A \cap A) = P(A) \cdot P(A)$, and hence, $\{X_n\}$ is stationary. So, (5.8) implies

$$\frac{S_{n,(n\cdot)}}{\sqrt{n}} \Rightarrow \sigma W(\cdot)$$

where

$$\sigma^2 = E_0^2 + 2\sum_{m=1}^{M} EX_0 X_m.$$

Example 2. Moving Average. Suppose

$$X_m = \sum_{k\ge 0} c_k \xi_{m-k} \quad \text{where } \sum_{k \ge 0} c_k^2 < \infty$$

and ξ_i, $i \in \mathbf{Z}$ are i.i.d. with $E\xi_i = 0$ and $E\xi_i^2 = 1$. Clearly $\{X_n\}$ is stationary sequence since series converges. Check whether $\{X_n\}$ is ergodic. We have

$$\underbrace{\bigcap_n \sigma\big(\{X_m, m \le n\}\big)}_{\text{trvivial algebra}} \subset \underbrace{\bigcap_n \sigma\big(\{\xi_k, k \le n\}\big)}_{\text{trvivial algebra}},$$

therefore, by Kolmogorov 0-1 law, $\{X_n\}$ is ergodic. Next, if $\mathcal{F}_{-n} = \sigma\big(\{\xi_m, m \le -n\}\big)$, then

$$\begin{aligned} ||E(X_0|\mathcal{F}_{-n})||_2 &= ||\sum_{k\ge n} c_k \xi_k||_2 \\ &= \Big(\sum_{k \ge n} c_k^2\Big)^{1/2} \end{aligned}$$

If, for example, $c_k = (1+k)^{-p}$, $||E(X_0|\mathcal{F}_{-n})||_2 \sim n^{(1/2-p)}$, and (5.11) applies if $p > 3/2$.
Mixing Properties
Let $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$, and

$$\alpha(\mathcal{G}, \mathcal{H}) = \sup_{A \in \mathcal{G}, B \in \mathcal{H}} \Big\{ \big|P(A \cap B) - P(A)P(B)\big| \Big\}$$

If $\alpha = 0$, $\mathcal{G}$ and $\mathcal{H}$ are independent so α measures the dependence of two $\sigma-$algebras.

Lemma 5.7 *Let $p, q, r \in (1, \infty]$ with $1/p + 1/q + 1/r = 1$, and suppose $X \in \mathcal{G}$, $Y \in \mathcal{H}$ have $E|X|^p, E|Y|^q < \infty$. Then*

$$|EXY - EXEY| \leq 8||X||_p||Y||_q\Big(\alpha(\mathcal{G}, \mathcal{H})\Big)^{1/r}$$

Here, we interpret $x^0 = 1$ for $x > 0$ and $0^0 = 0$.

Proof. If $\alpha = 0$, X and Y are independent and the result is true, so we can suppose $\alpha > 0$. We build up to the result in three steps, starting with the case $r = \infty$.

(a). $r = \infty$

$$|EXY - EXEY| \leq 2||X||_p||Y||_q$$

Proof of (a) Hölder's inequality implies $|EXY| \leq ||X||_p||Y||_q$, and Jensen's inequality implies

$$||X||_p||Y||_q \geq \big|E|X|E|Y|\big| \geq |EXEY|$$

so the result follows from the triangle inequality.

(b). $X, Y \in L^\infty$

$$|EXY - EXEY| \leq 4||X||_\infty||Y||_\infty\alpha(\mathcal{G}, \mathcal{H})$$

Proof of (b) Let $\eta = sgn\Big(E(Y|\mathcal{G}) - EY\Big) \in \mathcal{G}$. $EXY = E(XE(Y|\mathcal{G}))$, so

$$\begin{aligned}
|EXY - EXEY| &= |E(X(E(Y|\mathcal{G}) - EY))| \\
&\leq ||X||_\infty E|E(Y|\mathcal{G}) - EY| \\
&= ||X||_\infty E(\eta E(Y|\mathcal{G}) - EY) \\
&= ||X||_\infty (E(\eta Y) - E\eta EY)
\end{aligned}$$

Applying the last result with $X = Y$ and $Y = \eta$ gives

$$|E(Y\eta) - EYE\eta| \leq ||Y||_\infty|E(\zeta\eta) - E\zeta E\eta|$$

where $\zeta = sgn(E(\eta|\mathcal{H}) - E\eta)$. Now $\eta = 1_A - 1_B$ and $\zeta = 1_C - 1_D$, so

$$\begin{aligned}
|E(\zeta\eta) - E\zeta E\eta| &= |P(A \cap C) - P(B \cap C) - P(A \cap D) + P(B \cap D) \\
&\quad -P(A)P(C) + P(B)P(C) + P(A)P(D) - P(B)P(D)| \\
&\leq 4\alpha(\mathcal{G}, \mathcal{H})
\end{aligned}$$

Combining the last three displays gives the desired result.

(c) $q = \infty$, $1/p + 1/r = 1$

$$|EXY - EXEY| \le 6||X||_p||Y||_\infty\Big(\alpha(\mathcal{G},\mathcal{H})\Big)^{1-1/p}$$

Proof of (c) Let $C = \alpha^{-1/p}||X||_p$, $X_1 = X1_{(|X|\le C)}$, and $X_2 = X - X_1$.

$$\begin{aligned}|EXY - EXEY| &\le |EX_1Y - EX_1EY| + |EX_2Y - EX_2EY| \\ &\le 4\alpha C||Y||_\infty + 2||Y||_\infty E|X_2|\end{aligned}$$

by (a) and (b). Now

$$E|X_2| \le C^{-(p-1)}E(|X|^p1_{(|X|\le C)}) \le C^{-p+1}E|X|^p$$

Combining the last two inequalities and using the definition of C gives

$$|EXY - EXEY| \le 4\alpha^{1-1/p}||X||_p||Y||_\infty + 2||Y||_\infty\alpha^{1-1/p}||X||_p^{-p+1+p}$$

which is the desired result.

Finally, to prove (5.7), let $C = \alpha^{-1/q}||Y||_q$, $Y_1 = Y1_{(|Y|\le C)}$, and $Y_2 = Y - Y_1$.

$$\begin{aligned}|EXY - EXEY| &\le |EXY_1 - EXEY_1| + |EXY_2 - EXEY_2| \\ &\le 6C||X||_p\alpha^{1-1/p} + 2||X||_p||Y_2||_\theta\end{aligned}$$

where $\theta = (1-1/p)^{-1}$ by (c) and (a). Now

$$E|Y|^\theta \le C^{-q+\theta}E(|Y|^q1_{(|Y|\le C)}) \le C^{-1+\theta}E|Y|^q$$

Taking the $1/\theta$ root of each side and recalling the definition of C

$$||Y_2||_\theta \le C^{-(q-\theta)}||Y||_q^{q/\theta} \le \alpha^{(q-\theta)/q\theta}||Y||_q$$

so we have

$$|EXY - EXEY| \le 6\alpha^{-1/q}||Y||_q||X||_p\alpha^{1-1/p} + 2||X||_p\alpha^{1/\theta-1/q}||Y||_q^{1/\theta+1/q}$$

proving (5.7).

Combining (5.11) and (5.7) gives:

Theorem 5.12 *Suppose X_n, $n \in \mathbf{Z}$ is an ergodic stationary sequence with $EX_n = 0$, $E|X_0|^{2+\delta} < \infty$. Let $\alpha(n) = \alpha(\mathcal{F}_{-n}, \sigma(X_0))$, where $\mathcal{F}_{-n} = \sigma(\{X_m, m \le -n\})$, and suppose*

$$\sum_{n=1}^{\infty}\alpha(n)^{\delta/2(2+\delta)} < \infty$$

If $S_n = X_1 + \cdots X_n$, then

$$\frac{S_{(n\cdot)}}{\sqrt{n}} \Rightarrow \sigma W(\cdot),$$

where

$$\sigma^2 = EX_0^2 + 2\sum_{n=1}^{\infty} EX_0 X_n.$$

Proof. To use (5.7) to estimate the quantity in (5.11) we begin with

$$||E(X|\mathcal{F})||_2 = \sup\{E(XY) : Y \in \mathcal{F}, ||Y||_2 = 1\} \qquad (*)$$

Proof of (*) If $Y \in \mathcal{F}$ with $||Y||_2 = 1$, then using a by now familiar property of conditional expectation and the Cauchy-Schwarz inequality

$$EXY = E(E(XY|\mathcal{F})) = E(YE(X|\mathcal{F})) \le ||E(X|\mathcal{F})||_2 ||Y||_2$$

Equality holds when $Y = E(X|\mathcal{F})/||E(X|\mathcal{F})||_2$.

Letting $p = 2 + \delta$ and $q = 2$ in (5.7), noticing

$$\frac{1}{r} = 1 - \frac{1}{p} - \frac{1}{q} = \frac{\delta}{2(2+\delta)}$$

and recalling $EX_0 = 0$, showing that if $Y \in \mathcal{F}_{-n}$

$$|EX_0 Y| \le 8||X_0||_{2+\delta} ||Y||_2 \alpha(n)^{\delta/2(2+\delta)}$$

Combining this with (*) gives

$$||E(X_0|\mathcal{F}_{-n})||_2 \le 8||X_0||_{2+\delta} \alpha(n)^{\delta/2(2+\delta)}$$

and it follows that the hypotheses of (5.11) are satisfied.

6 Empirical Process

Let $(\Omega, \mathcal{A}, P)$ be an arbitrary probability space and $T : \Omega \to \bar{\mathbf{R}}$ am arbitrary map. The *outer integral* of T with respect to P is defined as

$$E^*T = \inf\{EU : U \geq T, U : \Omega \to \bar{\mathbf{R}} \text{ measurable and } EU \text{ exists}\}.$$

Here, as usual, EU is understood to exist if at least one of EU^+ or EU^- is finite. The *outer probability* of an arbitrary subset B of Ω is

$$P^*(B) = \inf\{P(A) : A \supset B, A \in \mathcal{A}\}.$$

Note that the functions U in the definition of outer integral are allowed to take the value ∞, so that the infimum is never empty.

Inner integral and inner probability can be defined in a similar fashion. Equivalently, they can be defined by $E_*T = -E^*(-T)$ and $P_*(B) = 1 - P^*(B)$, respectively.

In this section, $\mathbf{D}$ is metric space with a metric d. The set of all continuous, bounded functions $f : \mathbf{D} \to \mathbf{R}$ is denoted by $C_b(\mathbf{D})$.

Definition 6.1 *Let $(\Omega_\alpha, \mathcal{A}_\alpha, P_\alpha)$, $\alpha \in I$ be a net of probability space, $X_\alpha : \Omega_\alpha \to \mathbf{D}$, and $P_\alpha = P \circ X_\alpha^{-1}$. Then we say that the net X_α converges weakly to a Borel measure L, i.e., $P_\alpha \Rightarrow L$ if*

$$E^*f(X_\alpha) \to \int f dL, \quad \text{for every } f \in C_b(\mathbf{D})$$

Theorem 6.1 ***(Portmanteau)*** *The following statements are equivalent:*
(i) $P_\alpha \Rightarrow L$;
(ii) $\liminf P_*(X_\alpha \in G) \geq L(G)$ for every open G;
(iii) $\limsup P^*(X_\alpha \in F) \leq L(F)$ for every open F;
(iv) $\liminf E_*f(X_\alpha) \geq \int f dL$ for every lower semicontinuous f that is bounded below;
(v) $\limsup E^*f(X_\alpha) \leq \int f dL$ for everyupper semicontinuous f that is bounded above;
(vi) $\lim P^*(X_\alpha \in B) = \lim P_*(X_\alpha \in B) = L(B)$ for every Borel set B with $L(\partial B) = 0$.
(vii) $\liminf E_*f(X_\alpha) \geq \int f dL$ for every bounded, Lipschitz continuous, nonnegative f.

Definition 6.2 *The net of maps X_α is asymptotically measurable if and only if*

$$E^*f(X_\alpha) - E_*f(X_\alpha) \to 0, \quad \text{for every } f \in C_b(\mathbf{D}).$$

The net X_α is *asymptotically tight* if for every $\epsilon > 0$ there exists a compact set K such that

$$\liminf P_*(X_\alpha \in K^\delta) \geq 1 - \epsilon, \qquad \text{for every } \delta > 0.$$

Here $K^\delta = \{y \in \mathbf{D} : d(y, K) < \delta\}$ is the "$\delta-$enlargement" around K. A collection of Borel measurable maps X_α is called uniformly tight if, for every $\epsilon > 0$, there is a compact K with $P(X_\alpha \in K) \geq 1 - \epsilon$ for every α.

The δ in the definition of tightness may seem a bit overdone. It is not-asymptotic tightness as defined is essentially weaker than the same condition but with K instead of K^δ. This is caused by a second difference with the classical concept of uniform tightness: the enlarged compacts need to contain mass $1 - \epsilon$ only in the limit.

On the other hand, nothing is gained in simple cases: for Borel measurable maps in a Polish space, asymptotic tightness and uniform tightness are the same. It may also be noted that, although K^δ is dependent on the metric, the property of asymptotic tightness depends on the topology only. One nice consequence of the present tightness concept is that weak convergence usually implies asymptotic measurability and tightness.

Lemma 6.1 *(i) $X_\alpha \Rightarrow X$, then X_α is asymptotically measurable.*
(ii) If $X_\alpha \Rightarrow X$, then X_α is asymptotically tight if and only if X is tight.

Proof. (i). This follows upon applying the definition of weak convergence to both f and $-f$.

(ii). Fix $\epsilon > 0$. If X is tight, then there is a compact K with $P(X \in K) > 1-\epsilon$. By the portmanteau theorem, $\liminf P_*(X_\alpha \in K^\delta) \geq P(X \in K^\delta)$, which is larger than $1 - \epsilon$ for every $\delta > 0$. Conversely, if X_α is tight, then there is a compact K with $\liminf P_*(X_\alpha \in K^\delta) \geq 1 - \epsilon$. By the portmanteau theorem, $P(X \in \overline{K^\delta}) \geq 1 - \epsilon$. Let $\delta \downarrow 0$.

The next version of Prohorov's theorem may be considered a converse of the previous lemma. It comes in two parts, one for nets and one for sequences, neither one of which implies the other. The sequence case is the deepest of the two.

Theorem 6.2 ***(Prohorov's theorem.)*** *(i) If the net X_α is asymptotically tight and asymptotically measurable, then it has a subnet $X_{\alpha(\beta)}$ that converges in law to a tight Borel law.*
(ii) If the sequence X_n is asymptotically tight and asymptotically measurable, then it has a subsequence X_{n_j} that converges weakly to a tight Borel law.

6.1 Spaces of Bounded Functions

A vector lattice $\mathcal{F} \subset C_b(\mathbf{D})$ is a vector space that is closed under taking positive parts: if $f \in \mathcal{F}$, then $f = f \vee 0 \in \mathcal{F}$. Then automatically $f \vee g \in \mathcal{F}$ and $f \wedge g \in \mathcal{F}$ for every $f, g \in \mathcal{F}$. A set of functions on $\mathbf{D}$ separates points of $\mathbf{D}$ if, for every pair $x \neq y \in \mathbf{D}$, there is $f \in \mathcal{F}$ with $f(x) \neq f(y)$.

Lemma 6.2 *Let L_1 and L_2 be finite Borel measures on $\mathbf{D}$.*

(i) If $\int f dL_1 = \int f dL_2$ for every $f \in C_b(\mathbf{D})$, then $L_1 = L_2$.
Let L_1 and L_2 be tight Borel probability measures on $\mathbf{D}$.

(ii) If $\int f dL_1 = \int f dL_2$ for every f in a vector lattice $\mathcal{F} \subset C_b(\mathbf{D})$ that contains the constant functions and separates points of $\mathbf{D}$, then $L_1 = L_2$.

Lemma 6.3 *Let the net X_α be asymptotically tight, and suppose $E^* f(X_\alpha) - E_* f(X_\alpha) \to 0$ for every f in a subalgebra $\mathcal{F}$ of $C_b(\mathbf{D})$ that separates points of $\mathbf{D}$. Then the net X_α is asymptotically measurable.*

Let T be an arbitrary set. The space $l^\infty(T)$ is defined as the set of all uniformly bounded, real functions on T: all functions $z : T \to \mathbf{R}$ such that

$$||z||_T := \sup_{t \in T} |z(t)| < \infty$$

It is a metric space with respect to the *uniform distance* $d(z_1, z_2) = ||z_1 - z_2||_T$.

The space $l^\infty(T)$, or a suitable subspace of it, is a natural space for stochastic processes with bounded sample paths. A stochastic process is simply an indexed collection $\{X(t) : t \in T\}$ of random variables defined on the same probability space: every $X(t) : \Omega \to \mathbf{R}$ is a measurable map. If every sample path $t \mapsto X(t, \omega)$ is bounded, then a stochastic process yields a map $X : \Omega \to l^\infty(T)$. Sometimes the sample paths have additional properties, such as measurability or continuity, and it may be fruitful to consider X as a map into a subspace of $l^\infty(T)$. If in either case the uniform metric is used, this does not make a difference for weak convergence of a net; but for measurability it can.

In most cases a map $X : \Omega \to l^\infty(T)$ is a stochastic process. The small amount of measurability this gives may already be enough for asymptotic measurability. The special role played by the marginals $(X(t_1), ..., X(t_k))$, which are considered as maps into $\mathbf{R}^k$, is underlined by the following three results. Weak convergence in $l^\infty(T)$ can be characterized as asymptotic tightness plus convergence of marginals.

Lemma 6.4 *Let $X_\alpha : \Omega_\alpha \to l^\infty(T)$ be asymptotically tight. Then it is asymptotically measurable if and only if $X_\alpha(t)$ is asymptotically measurable for every $t \in T$.*

Lemma 6.5 *Let X and Y be tight Borel measurable maps into $l^\infty(T)$. Then X and Y are equal in Borel law if and only if all corresponding marginals of X and Y are equal in law.*

Theorem 6.3 *Let $X_\alpha : \Omega_\alpha \to l^\infty(T)$ be arbitrary. Then X_α converges weakly to a tight limit if and only if X_α is asymptotically tight and the marginals $(X(t_1), ..., X(t_k))$ converge weakly to a limit for every finite subset $t_1, ..., t_k$ of T. If X_α is asymptotically tight and its marginals converge weakly to the marginals $(X(t_1), ..., X(t_k))$ of a stochastic process X, then there is a version of X with uniformly bounded sample paths and $X_\alpha \Rightarrow X$.*

Proof. For the proof of both lemmas, consider the collection $\mathcal{F}$ of all functions $f : l^\infty(T) \to \mathbf{R}$ of the form

$$f(z) = g(z(t_1), ..., z(t_k)), \quad g \in C_b(\mathbf{R}^k), t_1, ..., t_k \in T, k \in \mathbf{N}.$$

This forms an algebra and a vector lattice, contains the constant functions, and separates points of $l^\infty(T)$. Therefore, the lemmas are corollaries of (6.2) and (6.3), respectively. If X_α is asymptotically tight and marginals converge, then X_α is asymptotically measurable by the first lemma. By Prohorov's theorem, X_α is relatively compact. To prove weak convergence, it suffices to show that all limit points are the same. This follows from marginal convergence and the second lemma.

Marginal convergence can be established by any of the well-known methods for proving weak convergence on Euclidean space. Tightness can be given a more concrete form, either through finite approximation or with the help of the Arzelà-Ascoli theorem. Finite approximation leads to the simpler of the two characterizations, but the second approach is perhaps of more interest, because it connects tightness to continuity of the sample paths $t \mapsto X_\alpha(t)$.

The idea of finite approximation is that for any $\epsilon > 0$ the index set T can be partitioned into finitely many subsets T_i such that the variation of the sample paths $t \mapsto X_\alpha(t)$ is less than ϵ on every one of the sets T_i. More precisely, it is assumed that for every $\epsilon, \eta > 0$, there exists a partition $T = \cup_{i=1}^k T_i$ such that

$$\limsup_\alpha P^*\Big(\sup_i \sup_{s,t \in T_i} \big| X_\alpha(s) - X_\alpha(t) \big| > \epsilon \Big) < \eta.$$

Clearly, under this condition the asymptotic behavior of the process can be described within error margin ϵ, η by the behavior of the marginal $(X_\alpha(t_1), ..., X_\alpha(t_k))$ for arbitrary fixed points $t_i \in T_i$. If the process can thus be reduced to a finite set of coordinates for any $\epsilon, \eta > 0$ and the nets or marginal distributions are tight, then the net X_α is asymptotically tight.

Theorem 6.4 *A net $X_\alpha : \Omega_\alpha \to l^\infty(T)$ is asymptotically tight if and only if $X_\alpha(t)$ is asymptotically tight in* $\mathbf{R}$ *for every t and, for all $\epsilon, \eta > 0$, there exists a finite partition $T = \cup_{i=1}^k T_i$ such that* (6.1) *holds.*

Proof. The necessity of the conditions follows easily from the next theorem. For instance, take the partition equal to disjointified balls of radius δ for a semimetric on T as in the next theorem. We prove sufficiency.

For any partition, as in the condition of the theorem, the norm $||X_\alpha||_T$ is bounded by $max_i|X_\alpha(t_i)| + \epsilon$, with inner probability at least $1 - \eta$, if $t_i \in T_i$ for each i. Since a maximum of finitely many tight nets of real variables is tight, it follows that the net $||X_\alpha||_T$ is asymptotically tight in R.

Fix $\zeta > 0$ and a sequence $\epsilon_m \downarrow 0$. Take a constant M such that $\limsup P^*(||X_\alpha||_T > M) < \zeta$, and for each $\epsilon = \epsilon_m$ and $\eta = 2^{-m}\zeta$, take a partition $T = \cup_{i=1}^k T_i$ as in (6.1). For the moment m is fixed and we do not let it appear in the notation. Let $z_1, ..., z_p$ be the set of all functions in $l^\infty(T)$ that are constant on each T_i and take on only the values $0, \pm\epsilon_m, ..., \pm[M/\epsilon_m]\epsilon_m$. Let K_m be the union of the p closed balls of radius ϵ_m around the z_i. Then, by construction, the two conditions

$$||X_\alpha||_T \leq M \quad \text{and} \quad \sup_i \sup_{s,t \in T_i} \big|X_\alpha(s) - X_\alpha(t)\big| \leq \epsilon_m$$

imply that $X_\alpha \in K_m$. This is true for each fixed m.

Let $K = \cap_{m=1}^\infty K_m$. Then K is closed and totally bounded(by construction of the K_m and because $\epsilon \downarrow 0$) and hence compact. Furthermore for every $\delta > 0$, there is an m with $K^\delta \supset \cap_{m=1}^m K_i$. If not, then there would be a sequence z_m not in K^δ, but with $z_m \in \cap_{m=1}^m K_i$ for every m. This would have a subsequence contained in one of the balls making up K_1, a further subsequence eventually contained in one of the balls making up K_2, and so on. The "diagonal" sequence, formed by taking the first of the first subsequence, the second of the second subsequence and so on, would eventually be contained in a ball of radius ϵ_m for every m; hence Cauchy. Its limit would be in K, contradicting the fact that $d(z_m, K) \geq \delta$ for every m.

Conclude that if X_α is not in K^δ, then it is not in $\cap_{m=1}^m K_i$ for some fixed m. Then

$$\limsup P^*(X_\alpha \notin K^\delta) \leq \limsup P^*\Big(X_\alpha \notin \cap_{m=1}^m K_i\Big) \leq \zeta + \sum_{i=1}^m \zeta 2^{-m} < 2\zeta$$

This concludes the proof of the theorem.

The second type of characterization of asymptotic tightness is deeper and relates the concept to asymptotic continuity of the sample paths. Suppose ρ

is a semimetric on T A net $X_\alpha : \Omega_\alpha \to l^\infty(T)$ is asymptotically uniformly $\rho-$equicontinuous in probability if for every $\epsilon, \eta > 0$ there exists a $\delta > 0$ such that

$$\limsup_\alpha P^*\Big(\sup_{\rho(s,t)<\delta} \big|X_\alpha(s) - X_\alpha(t)\big| > \epsilon\Big) < \eta.$$

Theorem 6.5 *A net $X_\alpha : \Omega_\alpha \to l^\infty(T)$ is asymptotically tight if and only if $X_\alpha(t)$ is asymptotically tight in R for every t and there exists a semimetric ρ on T such that (T,ρ) is totally bounded and X_α is asymptotically uniformly $\rho-$equicontinuous in probability. If, moreover, $X_\alpha \Rightarrow X$, then almost all paths $t \mapsto X(t,\omega)$ are uniformly $\rho-$continuous; and the semimetric ρ can without loss of generality be taken equal to any semimetric ρ for which this is true and (T,ρ) is totally bounded.*

Proof. $(\Leftarrow)$. The sufficiency follows from the previous theorem. First take $\delta > 0$ sufficiently small so that the last displayed inequality is valid. Since T is totally bounded, it can be covered with finitely many balls of radius δ. Construct a partition of T by disjointifying these balls.

$(\Rightarrow)$. If X_α is asymptotically tight, then $g(X_\alpha)$ is asymptotically tight for every continuous map g; in particular, for each coordinate projection. Let $K_1 \subset K_2 \subset \ldots$ be compacts with $\liminf P_*(X_\alpha \in K_m^\epsilon) 1 - 1/m$ for every $\epsilon > 0$. For every fixed m, define a semimetric ρ_m on T by

$$\rho_m(s,t) = \sup_{z\in K_m} |z(s) - z(t)|, \quad s,t \in T$$

Then (T,ρ_m) is totally bounded. Indeed, cover K_m by finitely many balls of radius η, centered at $z_1,\ldots,z_k$. Partition R^k into cubes of edge η, and for every cube pick at most one $t \in T$ such that $(z_1(t),\ldots,z_k(t))$ is in the cube. Since $z_1,\ldots,z_k$ are uniformly bounded, this gives finitely many points $t_1,\ldots,t_p$. Now the balls $\{t : \rho(t,t_i) < 3\eta\}$ cover T: t is in the ball around t_i for which $(z_1(t),\ldots,z_k(t))$ and $(z_1(t_i),\ldots,z_k(t_i))$ fall in the same cube. This follows because $\rho_m(t,t_i)$ can be bounded by $2\sup_{z\in K_m}\inf_i ||z-z_i||_T + \sup_j |z_j(t_i) - z_j(t)|$. Next set

$$\rho(s,t) = \sum_{m=1}^\infty 2^{-m}(\rho_m(s,t)\wedge 1).$$

Fix $\eta > 0$. Take a natural number m with $2^{-m} < \eta$. Cover T with finitely many ρ_m-balls of radius η. Let $t_1,\ldots,t_p$ be their centers. Since $\rho_1 \le \rho_2 \le \ldots$, there is for every t a t_i with $\rho(t,t_i) \le \sum_{k=1}^m 2^{-k}\rho_k(t,t_i) + 2^{-m} < 2\eta$. Thus (T,ρ) is totally bounded for ρ, too. It is clear from the definitions that $|z(s) - z(t)| \le \rho_m(s,t)$ for every $z \in K_m$ and that $\rho_m(s,t)\wedge 1 \le 2^m\rho(s,t)$. Also, if $||z_0 - z||_T < \epsilon$ for $z \in K_m$ then $|z_0(s) - z_0(t)| < 2\epsilon + |z(s) - z(t)|$ for any pair s,t. Deduce that

$$K_m^\epsilon \subset \Big\{z : \sup_{\rho(s,t)<2^{-m}\epsilon} |z(s) - z(t)| \le 3\epsilon\Big\}.$$

Thus for given ϵ and m, and for $\delta < 2^{-m}\epsilon$,

$$\liminf_{\alpha} P_*\Big(\sup_{\rho(s,t)<\delta} \big|X_\alpha(s) - X_\alpha(t)\big| \leq 3\epsilon\Big) \geq 1 - \frac{1}{m}.$$

Finally, if $X_\alpha \Rightarrow X$, then with notation as in the second part of the proof, $P(X \in K_m) \geq 1 - 1/m$; hence X concentrates on $\cup_{m=1}^{\infty} K_m$. The elements of K_m are uniformly ρ_m-equicontinuous and hence also uniformly ρ-continuous. This yields the first statement. The set of uniformly continuous functions on a totally bounded, semimetric space is complete and separable, so a map X that takes its values in this set is tight. Next if $X_\alpha \Rightarrow X$ and X is tight, the X_α is asymptotically tight and the compacts for asymptotical tightness can be shosen equal to the compacts for tightness of X If X has uniformly continuous paths, then the latter compacts can be chosen within the space of uniformly continuous functions. Since a compact is totally bounded, every one of the compacts is necesaarily uniformly equicontinuous. Combination of these facts proves the second statement.

6.2 Maximal Inequalities and Covering Numbers

We derive a class of maximal inequalities that can be used to establish the asymptotic equicontinuity of the empirical process. Since the ineualities have much wider applicability, we temporarily leave the empirical framework.

Let ψ be a nondecreasing, convex function with $\psi(0) = 0$ and X a random variable. Then the Orlicz norm $||X||_\psi$ is defined as

$$||X||_\psi = \inf \left\{ C > 0 : E\psi\Big(\frac{|X|}{C}\Big) \leq 1 \right\}.$$

Here the infimum over the empty set is ∞. Using Jensen's inequality, it is not difficult to check that this indeed defines a norm. The best-known examples of Orlicz norms are those corresponding to the functions $x \mapsto x^p$ for $p \geq 1$: the corresponding Orlicz norm is simply the L_p-norm

$$||X||_p = \Big(E|X|^p\Big)^{1/p}$$

For our purpose, Orlicz norms of more interest are the ones given by $\psi_p(x) = e^{x^p} - 1$ for $p \geq 1$, which give much more weight to the tails of X. The bound $x^p \leq \psi_p(x)$ for all nonnegative x implies that $||X||_p \leq ||X||_{\psi_p}$ for each p. It is not true that the exponential Orlicz norms are all bigger than all L_p-norms. However, we have the inequalities

$$\begin{aligned} ||X||_{\psi_p} &\leq ||X||_{\psi_p} (\log 2)^{1/q-1/p}, \quad p \leq q \\ ||X||_p &\leq p! ||X||_{\psi_1} \end{aligned}$$

Since for the present purposes fixed constants in inequalities are irrelevant, this means that a bound on an exponential Orlicz norm always gives a better result than a bound on an L_p-norm.

Any Orlicz norm can be used to obtain an estimate of the tail of a distribution. By Markov's inequality,

$$P(|X|>x) \leq P\Big(\psi(|X|/||X||_\psi) \geq \psi(x/||X||_\psi)\Big) \leq \frac{1}{\psi(x/||X||_\psi)}$$

For $\psi_p(x) = e^{x^p} - 1$, this leads to tail estimates $\exp(-Cx^p)$ for any random variable with a finite ψ_p-norm. Conversely, an exponential tail bound of this type shows that $||X||_{\psi_p}$ is finite.

Lemma 6.6 *Let X be a random variable with $P(|X|>x) \leq Ke^{-Cx^p}$ for every x, for constants K and C, and for $p \geq 1$. Then its Orlicz norm satisfies $||X||_{\psi_p} \leq \Big((1+K)/C\Big)^{1/p}$.*

Proof. By Fubini's theorem,

$$E\Big(e^{D|X|^p} - 1\Big) = E\int_0^{|X|^p} De^{Ds}ds = \int_0^\infty P(|X|>s^{1/p})De^{Ds}ds$$

Now insert the inequality on the tails of $|X|$ and obtain the explicit upper bound $KD/(C-D)$. This is less than or equal to 1 for $D^{-1/p}$ greater than or equal to $\Big((1+K)/C\Big)^{1/p}$. This completes the proof.

Next consider the $\psi-$norm of a maximum of finitely many random variables. Using the fact that $\max|X_i|^p \leq \sum|X_i|^p$, one easily obtains for the L_p-norms

$$\Big|\Big|\max_{1\leq i\leq m} X_i\Big|\Big|_p = \left(E\max_{1\leq i\leq m} X_i^p\right)^{1/p} \leq m^{1/p}\max_{1\leq i\leq m}||X_i||_p$$

A similar inequality is valid for many Orlicz norms, in particular the exponential ones. Here, in the general case, the factor $m^{1/p}$ becomes $\psi^{-1}(m)$.

Lemma 6.7 *Let ψ be a convex, nondecreasing, nonzero function with $\psi(0)=0$ and $\limsup_{x,y\infty}\psi(x)\psi(y)/\psi(cxy) < \infty$ for some constant c. Then for any random variables $X_1,...,X_n$,*

$$\Big|\Big|\max_{1\leq i\leq m} X_i\Big|\Big|_\psi = \leq K\psi^{-1}(m)\max_{1\leq i\leq m}||X_i||_p$$

for a constant K depending only on ψ.

Proof. For simplicity of notation assume first that $\psi(x)\psi(y) \leq \psi(cxy)$ for all $x, y \geq 1$. In that case, $\psi(x/y) \leq \psi(cx)/\psi(y)$ for all $x \geq y \geq 1$. Thus, for $y \geq 1$ and any C,

$$\begin{aligned}
\max \psi\left(\frac{|X_i|}{Cy}\right) &\leq \max\left[\frac{\psi(c|X_i|/C)}{\psi(y)} + \psi\Big(\frac{|X_i|}{Cy}\Big)\mathbf{1}\Big\{\frac{|X_i|}{Cy} < 1\Big\}\right] \\
&\leq \sum \frac{\psi(c|X_i|/C)}{\psi(y)} + \psi(1).
\end{aligned}$$

Set $C = c \max ||X_i||_\psi$, and take expectations to get

$$E\psi\left(\frac{\max|X_i|}{Cy}\right) \leq \frac{m}{\psi(y)} + \psi(1)$$

When $\psi(1) \leq 1/2$, this is less than or equal to 1 for $y = \psi^{-1}(2m)$, which is greater than 1 under the same condition. Thus,

$$\Big|\Big| \max_{1\leq i\leq m} X_i \Big|\Big|_\psi \leq \psi^{-1}(2m)c \max ||X_i||_\psi.$$

By the convexity of ψ and the fact that $\psi(0) = 0$, it follows that $\psi^{-1}(2m) \leq 2\psi^{-1}(m)$. The proof is complete for every special ψ that meets the conditions made previously. For a general ψ, there are constants $\sigma \leq 1$ and $\tau > 0$ such that $\phi(x) = \sigma\psi(\tau x)$ satisfies the conditions of the previous paragraph. Apply the inequality to ϕ, and observe that $||X||_\psi \leq ||X||_\phi/(\sigma\tau) \leq ||X||_\psi/\sigma$.

For the present purposes, the value of the constant in the previous lemma is irrelevant. The important conclusion is that the inverse of the $\psi-$function determines the size of the $\psi-$norm of a maximum in comparison to the $\psi-$norms of the individual terms. The $\psi-$norms grows slowest for rapidly increasing ψ. For $\psi(x) = e^{x^p} - 1$, the growth is at most logarithmic, because

$$\psi_p^{-1}(m) = (\log(1+m))^{1/p}$$

The previous lemma is useless in the case of a maximum over infinitely many variables. However, such a case can be handled via repeated application of the lemma via a method known as chaining. Every random variable in the supremum is written as a sum of "little links," and the bound depends on the number and size of the little links needed. For a stochastic process $\{X_t : t \in T\}$, the number of links depends on the entropy of the index set for the semimetric

$$d(s,t) = ||X_s - X_t||_\psi.$$

The general definition of "metric entropy" is as follows.

Definition 6.3 ***(Covering numbers)*** *Let (T, d) be an arbitrary semi-metric space. Then the covering number $N(\epsilon, d)$ is the minimal number of balls of radius ϵ needed to cover T. Call a collection of points $\epsilon-$separated if the distance between each pair of points is strictly larger than ϵ. The packing number $D(\epsilon, d)$ is the maximum number of $\epsilon-$separated points in T. The corresponding entropy numbers are the logarithms of the covering and packing numbers, respectively.*

For the present purposes, both covering and packing numbers can be used. In all arguments one can be replaced by the other through the inequalities

$$N(\epsilon, d) \leq D(\epsilon, d) \leq N\Big(\frac{\epsilon}{2}, d\Big)$$

Clearly, covering and packing numbers become bigger as $\epsilon \downarrow 0$. By definition, the semimetric space T is totally bounded if and only if the covering and packing numbers are finite for every $\epsilon > 0$. The upper bound in the following maximal inequality depends on the rate at which $D(\epsilon, d)$ grows as $\epsilon \downarrow 0$, as measured through an integral criterion.

Theorem 6.6 *Let ψ be a convex, nondecreasing, nonzero function with $\psi(0) = 0$ and $\limsup_{x,y\to\infty} \psi(x)\psi(y)/\psi(cxy) < \infty$, for some constant c. Let $\{X_t : t \in T\}$ be a separable stochastic process with*

$$||X_s - X_t||_\psi \leq Cd(s, t), \quad \textit{for every } s, t$$

for some semimetric d on T and a constant C. Then, for any $\eta, \delta > 0$,

$$\Big|\Big| \sup_{d(s,t)\leq\delta} |X_s - X_t| \Big|\Big|_\psi \leq K\left[\int_0^\eta \psi^{-1}(D(\epsilon, d))d\epsilon + \delta\psi^{-1}(D^2(\eta, d))\right],$$

for a constant K depending on ψ and C only.

Corollary 6.1 *The constant K can be chosen such that*

$$\Big|\Big| \sup_{s,t} |X_s - X_t| \Big|\Big|_\psi \leq K \int_0^{diamT} \psi^{-1}(D(\epsilon, d))d\epsilon,$$

where diam T is the diameter of T.

Proof. Assume without loss of generality that the packing numbers and the associated "covering integral" are finite. Construct nested sets $T_0 \subset T_1 \subset \cdots \subset T$ such that every T_j is a maximal set of points such that $d(s, t) > \eta 2^{-j}$ for every $s, t \in T_j$ where "maximal" means that no point can be added without destroying the validity of the inequality. By the definition of packing numbers, the number of points in T_j is less than or equal to $D(\eta 2^{-j}, d)$.

"Link" every point $t_{j+1} \in T_{j+1}$ to a unique $t_j \in T_j$ such that $d(t_j, t_{j+1}) \leq \eta 2^{-j}$. Thus, obtain for every t_{k+1} a chain $t_{k+1}, t_k, ..., t_0$ that connects it to a point in T_0. For arbitrary points s_{k+1}, t_{k+1} in T_{k+1}, the difference in increments along their chains can be bounded by

$$\begin{aligned} |(X_{s_{k+1}} - X_{s_0}) - (X_{t_{k+1}} - X_{t_0})| &= \Big| \sum_{j=0}^{k} (X_{s_{j+1}} - X_{s_j}) - \sum_{j=0}^{k} (X_{t_{j+1}} - X_{t_j}) \Big| \\ &\leq 2 \sum_{j=0}^{k} \max |X_u - X_v| \end{aligned}$$

where for fixed j the maximum is taken over ll links (u, v) from T_{j+1} to T_j. Thus the jth maximum is taken over at most $\#T_{j+1}$ links, with each link having a $\psi-$norm $||X_u - X_v||_\psi$ bounded by $Cd(u, v) \leq C\eta 2^{-j}$. It follows with the help of lemma (6.7) that , for a constant depending only on ψ and C,

$$\begin{aligned} \Big|\Big| \max_{s,t \in T_{k+1}} |(X_s - X_{s_0}) - (X_t - X_{t_0})| \Big|\Big|_\psi &\leq K \sum_{j=0}^{k} \psi^{-1}(D(\eta 2^{-j-1}, d)) \eta 2^{-j} \\ &\leq 4K \int_0^\eta \psi^{-1}(D(\epsilon, d)) d\epsilon. \end{aligned}$$

In this bound, s_0 and t_0 are the endpoints of the chains starting at s and t, respectively.

The maximum of the increments $|X_{s_{k+1}} - X_{t_{k+1}}|$ can be bounded by the maximum on the left side of (67) plus the maximum of the discrepancies $|X_{s_0} - X_{t_0}|$ at the end of the chains. The maximum of the latter discrepancies will be analyzed by a seemingly circular argument. For every pair of endpoints s_0, t_0 of chains starting at two points in T_{k+1} within distance δ of each other, choose exactly one pair s_{k+1}, t_{k+1} in T_{k+1}, with $d(s_{k+1}, t_{k+1}) < \delta$, whose chains end at s_0, t_0. By definition of T_0, this gives at most $D^2(\eta, d)$ pairs. By the triangle inequality,

$$|X_{s_0} - X_{t_0}| \leq |(X_{s_0} - X_{s_{k+1}}) - (X_{t_0} - X_{t_{k+1}})| + |X_{s_{k+1}} - X_{t_{k+1}}|$$

Take the maximum over all pairs of endpoints s_0, t_0 as above. Then the corresponding maximum over the first term on the right in the last display is bounded by the maximum in the left side of (67). It $\psi-$norm can be bounded by the right side of this equation. Combine this with (67) to find that

$$\Big|\Big| \max_{s,t \in T_{k+1}, d(s,t) < \delta} |(X_s - X_{s_0}) - (X_t - X_{t_0})| \Big|\Big|_\psi \leq 8K \int_0^\eta \psi^{-1}(D(\epsilon, d)) d\epsilon + \Big|\Big| \max |X_{s_{k+1}} - X_{t_{k+1}}| \Big|\Big|_\psi$$

Here the maximum on the right is taken over the pairs s_{k+1}, t_{k+1} in T_{k+1} uniquely attached to the pairs s_0, t_0 as above. Thus the maximum is over at most $D^2(\eta, d)$ terms, each of whose $\psi-$norm is bounded by δ. Its $\psi-$norm is bounded by $K\psi^{-1}(D^2(\eta, d))\delta$.

Thus the upper bound given by the theorem is a bound for the maximum of increments over T_{k+1}. Let k tend to infinity to conclude the proof. The corollary follows immediately from the previous proof, after noting that, for η equal to the diameter of T, the set T_0 consists of exactly one point. In that case $s_0 = t_0$ for every pair s, t, and the increments at the end of the chains are zero. The corollary also follows from the theorem upon taking $\eta = \delta = diamT$ and noting that $D(\eta, d) = 1$, so that the second term in the maximal inequality can also be written $\delta\psi^{-1}(D(\eta, d))$. Since the function $\epsilon \mapsto \psi^{-1}(D(\epsilon, d))$ is decreasing, this term can be absorbed into the integral, perhaps at the cost of increasing the constant K.

Though the theorem gives a bound on the continuity modulus of the process, a bound on the maximum of the process will be needed. Of course, for any t_0,

$$\Big|\Big|\sup_t |X_t|\Big|\Big|_\psi \leq ||X_{t_0}||_\psi + K\int_0^{diamT} \psi^{-1}(D(\epsilon, d))d\epsilon$$

Nevertheless, to state the maximal inequality in terms of the increments appears natural. The increment bound shows that the process X is continuous in $\psi-$norm, whenever the covering integral $\int_0^\eta \psi^{-1}(D(\epsilon, d))d\epsilon$ converges for some $\eta > 0$. It is a small step to deduce the continuity of almost all sample paths from this inequality, but this is not needed at this point.

6.3 Sub-Gaussian Inequalities

A standard normal variable has tails of the order $x^{-1}\exp\Big(-\frac{x^2}{2}\Big)$ and satisfies $P(|X| > x) \leq 2\exp\Big(-\frac{x^2}{2}\Big)$ for every x. By direct calculation one finds a ψ_2-norm of $\sqrt{8/3}$. In this section we study random variables satisfying similar tail bounds.

Hoeffding's inequality asserts a "sub-Gaussian" tail bound for random variables of the form $X = \sum X_i$ with $X_1, ..., X_n$ i.i.d. with zero means and bounded range. The following special case of Hoeffding's inequality will be needed.

Theorem 6.7 ***(Hoeffding's inequality)*** *Let $a_1, ..., a_n$ be constants and $\epsilon_1, ..., \epsilon_n$ be independent Rademacher random variables; i.e., with $P(\epsilon_i = 1) = P(\epsilon_i = -1) = 1/2$. Then*

$$P\Big(|\sum \epsilon_i a_i| > x\Big) \leq 2e^{-\frac{x^2}{2||a||^2}},$$

for the Euclidean norm $||a||$. Consequently, $||\sum \epsilon_i a_i||_{\psi_2} \leq \sqrt{6}||a||$.

Proof. For any λ and Rademacher variable ϵ, one has $Ee^{\lambda\epsilon} = (e^{\lambda} + e^{-\lambda}) \leq e^{\lambda^2/2}$, where the last inequality follows after writing out the power series. Thus by Markov's inequality, for any $\lambda > 0$,

$$P\Big(\sum \epsilon_i a_i > x\Big) \leq e^{-\lambda x} Ee^{\lambda \sum_{i=1}^n a_i \epsilon_i} \leq e^{(\lambda^2/2)||a||^2} - \lambda x.$$

The best upper bound is obtained for $\lambda = x/||a||^2$ and is the exponential in the probability bound of the lemma. Combination with a similar bound for the lower tail yields the probability bound. The bound on the $\psi-$norm is a consequence of the probability bound in view of (6.6).

A stochastic process is called sub-Gaussian with respect to the semi-metric d on its index set if

$$P(|X_s - X_t| > x) \leq 2e^{-\frac{x^2}{2d^2(s,t)}}, \quad \text{for every } s,t \in T, x > 0$$

any Gaussian process is sub-Gaussian for the standard deviation semimetric $d(s,t) = \sigma(X_s - X_t)$. another example is Rademacher process

$$X_a = \sum_{i=1}^n a_i \epsilon_i, \quad a \in R^n$$

for Rademacher variables $\epsilon_1, ..., \epsilon_n$. By Hoeffding's inequality, this is sub-Gaussian for the Euclidean distance $d(a,b) = ||a - b||$.

Sub-Gaussian processes satisfy the increment bound $||X_s - X_t||_{\psi_2} \leq \sqrt{6}d(s,t)$. Since the inverse of the ψ_2-function is essentially the square root of the logarithm, the genereal maximal inequality leads for sub-Gaussian processes to a bound in terms of an entropy integral. Furthermore, because of the special properties of the logarithm, the statement can be slightly simplified.

Corollary 6.2 *Let $\{X_t : t \in T\}$ be a separable sub-Gaussian process. Then for every $\delta > 0$,*

$$E \sup_{d(s,t)\leq\delta} |X_s - X_t| \leq K \int_0^{\delta} \sqrt{\log D(\epsilon, d)}d\epsilon,$$

for a universal constant K. In particular, for any t_0,

$$E \sup_t |X_t| \leq E|X_{t_0}| + K \int_0^{\infty} \sqrt{\log D(\epsilon, d)}d\epsilon.$$

Proof. Apply the general maximal inequality with $\psi_2(x) = e^{x^2} - 1$ and $\eta = \delta$. Since $\psi_2^{-1}(m) = \sqrt{\log(1+m)}$, we have $\psi_2^{-1}(D^2(\delta,d)) \leq \sqrt{2}\psi_2^{-1}(D(\delta,d))$. Thus the second term in the maximal inequality can first be replaced by $\sqrt{2}\delta\psi^{-1}(D(\eta, d))$ and next be incorporated in the first at the cost of increasing the constant. We obtain

$$\Big|\Big| \sup_{d(s,t)\le\delta} |X_s - X_t| \Big|\Big|_{\psi_2} \le K \int_0^\delta \sqrt{\log(1 + D(\epsilon, d))} d\epsilon.$$

Here $D(\epsilon, d) \ge 2$ for every ϵ that is strictly less than the diameter of T. Since $\log(1 + m) \le 2\log m$ for $m \ge 2$, the 1 inside the logarithm can be removed at the cost of increasing K.

6.4 Symmetrization

Let $\epsilon_1, ..., \epsilon_n$ be i.i.d. Rademacher random variables. Instead of the empirical process

$$f \mapsto (P_n - P)f = \frac{1}{n}\sum_{i=1}^n (f(X_i) - Pf),$$

consider the symmetrized process

$$f \mapsto P_n^o f = \frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i),$$

where $\epsilon_1, ..., \epsilon_n$ are independent of $(X_1, ..., X_n)$. Both processes have mean function zero. It turns out that the law of large numbers or the central limit theorem for one of these processes holds if and only if the corresponding result is true for the other process. One main approach to proving empirical limit theorems is to pass from $P_n - P$ to P_n^o and next apply arguments conditionally on the original X's. The idea is that, for fixed $X_1, ..., X_n$, the symmetrized empirical measure is a Rademacher process, hence a sub-Gaussian process, to which (6.2) can be applied.

Thus we need to bound maxima and moduli of the process $P_n - P$ by those of the symmetrized process. To formulate such bounds, we must be careful about the possible nonmeasurability of suprema of the type $||P_n - P||_{\mathcal{F}}$. The result will be formulated in terms of outer expectation, but it does nto hold for every choice of an underlying probability space on which $X_1, ..., X_n$ are defined. Throughout this part, if outer expectations are involved, it is assumed that $X_1, ..., X_n$ are the coordinate projections on the product space $(\mathcal{X}^n, \mathcal{A}^n, P^n)$, and the outer expectations of functions $(X_1, ..., X_n) \mapsto h(X_1, ..., X_n)$ are computed for P^n. thus "independent" is understood in terms of a product probability space. If auxiliary variables, independent of the X's, are involved, as in the next lemma, we use a similar convention. In that case, the underlying probability space is assumed to be of the form $(\mathcal{X}^n, \mathcal{A}^n, P^n) \times (\mathcal{Z}, \mathcal{C}, Q^)$ with $X_1, ..., X_n$ equal to the coordinate projections on the first n coordinates and the additional variables depending only on the $(n + 1)$st coordinate.

The following lemma will be used mostly with the choice $\Phi(x) = x$.

Lemma 6.8 *(**Symmetrization**) For every nondecreasing, convex $\Phi : R \to R$ and class of measurable functions $\mathcal{F}$,*

$$E^*\Phi\Big(||P_n - P||_{\mathcal{F}}\Big) \le E^*\Phi\Big(2||P_n^0||_{\mathcal{F}}\Big),$$

where the outer expectations are computed as indicated in the preceding paragraph.

Proof. Let $Y_1, ..., Y_n$ be independent copies of $X_1, ..., X_n$, defined formally as the coordinate projections on the last n coordinates in the product space $(\mathcal{X}^n, \mathcal{A}^n, P^n) \times (\mathcal{Z}, \mathcal{C}, Q^) \times (\mathcal{X}^n, \mathcal{A}^n, P^n)$. The outer expectations in the statement of the lemma are unaffected by this enlargement of the underlying probability space, because coordinate projections are perfect maps. For fixed values $X_1, ..., X_n$,

$$\begin{aligned} ||P_n - P||_{\mathcal{F}} &= \sup_{f\in\mathcal{F}} \frac{1}{n}\Big|\sum_{i=1}^{n}\big(f(X_i) - Ef(Y_i)\big)\Big| \\ &\le E_Y^* \sup_{f\in\mathcal{F}} \frac{1}{n}\Big|\sum_{i=1}^{n}\big(f(X_i) - f(Y_i)\big)\Big| \end{aligned}$$

where E_Y^* is the outer expectation with respect to $Y_1, ..., Y_n$ computed for P^n for given, fixed values of $X_1, ..., X_n$. Combination with Jensen's inequality yields

$$\Phi\Big(||P_n - P||_{\mathcal{F}}\Big) \le E_Y\Phi\left(\Big|\Big|\frac{1}{n}\sum_{i=1}^{n}\big(f(X_i) - f(Y_i)\big)\Big|\Big|_{\mathcal{F}}^{*Y}\right),$$

where $*Y$ denotes the minimal measurable majorant of the supremum with respect to $Y_1, ..., Y_n$, still with $X_1, ..., X_n$ fixed. Because Φ is nondecreasing and continuous, the $*Y$ inside Φ can be moved to E_Y^*. Next take the expectation with respect to $X_1, ..., X_n$ to get

$$E^*\Phi\Big(||P_n - P||_{\mathcal{F}}\Big) \le E_X^* E_Y^* \Phi\left(\Big|\Big|\frac{1}{n}\sum_{i=1}^{n}\big(f(X_i) - f(Y_i)\big)\Big|\Big|_{\mathcal{F}}\right).$$

Here the repeated outer expectation can be bounded above by the joint outer expectation E^* by Fubini's theorem.

Adding a minus sign in front of a term $\big(f(X_i) - f(Y_i)\big)$ has the effect of exchanging X_i and Y_i. By construction of the underlying probability space as a product space, the outer expectation of any function $f(X_1, ..., X_n, Y_1, ..., Y_n)$ remains unchanged under permutations of its $2n$ arguments. hence the expression

$$E^*\Phi\left(\Big|\Big|\frac{1}{n}\sum_{i=1}^{n} e_i\big(f(X_i) - f(Y_i)\big)\Big|\Big|_{\mathcal{F}}\right)$$

is the same for any $n-$tuple $(e_1, ..., e_n) \in \{-1, 1\}^n$. Deduce that

$$E^*\Phi\Big(||P_n - P||_{\mathcal{F}}\Big) \leq E_\epsilon E^*_{X,Y}\Phi\left(\Big|\Big|\frac{1}{n}\sum_{i=1}^n \epsilon_i\big(f(X_i) - f(Y_i)\big)\Big|\Big|_{\mathcal{F}}\right).$$

Use the triangle inequality to separate the contributions of the X's and the Y's and next use the convexity of Φ to bound the previous expression by

$$\frac{1}{2}E_\epsilon E^*_{X,Y}\Phi\left(2\Big|\Big|\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i)\Big|\Big|_{\mathcal{F}}\right) + \frac{1}{2}E_\epsilon E^*_{X,Y}\Phi\left(2\Big|\Big|\frac{1}{n}\sum_{i=1}^n \epsilon_i f(Y_i)\Big|\Big|_{\mathcal{F}}\right)$$

By perfectness of coordinate projections, the expectation $E^*_{X,Y}$ is the same as E^*_X and E^*_Y in the two terms, respectively. Finally, replace the repeated outer expectations by a joint outer expectation. This completes the proof.

The symmetrization lemma is valid for any class $\mathcal{F}$. In the proofs of Glivenko-Cantelli and Donsker theorems, it will be applied not only to the original set of functions of interest, but also to several classes constructed from such a set $\mathcal{F}$. The next step in these proofs is to apply a maximal inequality to the right side of the lemma, conditionally on $X_1, ..., X_n$. At that point we need to write the joint outer expectation as the repeated expectation $E^*_X E_\epsilon$, where the indices remaining variables. Unfortunately, Fubini's theorem is not valid for outer expectations. To overcome this problem, it is assumed that the integrand in the right side of the lemma is jointly measurable in $(X_1, ..., X_n, \epsilon_1, ..., \epsilon_n)$. Since the Rademacher variables are discrete, this is the case if and only if the maps

$$(X_1, ..., X_n) \mapsto \Big|\Big|\sum_{i=1}^n e_i f(X_i)\Big|\Big|_{\mathcal{F}}$$

are measurable for every $n-$tuple $(e_1, ..., e_n) \in \{-1, 1\}^n$. For the intended application of Fubini's theorem, it suffices that this is the case for the completion of $(\mathcal{X}^n, \mathcal{A}^n, P^n)$.

Definition 6.4 ***(Measurable Class)*** *A class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \to R$ on a probability space $(\mathcal{X}, \mathcal{A}, P)$ is called a $P-$measurable class if the function* (6.4) *is measurable on the completion of $(\mathcal{X}^n, \mathcal{A}^n, P^n)$ for every n and every vector $(e_1, ..., e_n) \in R^n$.*

Glivenko-Cantelli Theorems

In this section we prove two types of Glivenko-Cantelli theorems. The first theorem is the simplest and is based on entropy with bracketing. Its proof relies on finite approximation and the law of large numbers for real variables. The second theorem uses random L_1-entropy numbers and is proved through symmetrization followed by a maximal inequality.

Definition 6.5 ***(Covering numbers)*** *The covering number $N(\epsilon, \mathcal{F}, ||\cdot||)$ is the minimal number of balls $\{g : ||g - f|| < \epsilon\}$ of radius ϵ needed to cover the set $\mathcal{F}$. The centers of the balls need not belong to $\mathcal{F}$, but they should have finite norms. The entropy(without bracketing) is the logarithm of the covering number.*

Definition 6.6 ***(Bracketing numbers)*** *Given two functions l and u, the bracket $[l, u]$ is the set of all functions f with $l \leq f \leq u$. An $\epsilon-$bracket is a bracket $[l, u]$ with $||u - l|| < \epsilon$. The bracketing number $N_{[]}(\epsilon, \mathcal{F}, ||\cdot||)$ is the minimum number of $\epsilon-$brackets needed to cover $\mathcal{F}$. The entropy with bracketing is the logarithm of the bracketing number. In the definition of the bracketing is the logarithm of the bracketing number. In the definition of the bracketing number, the upper and lower bounds u and l of the brackets need not belong to $\mathcal{F}$ themselves but are assumed to have finite norms.*

Theorem 6.8 *Let $\mathcal{F}$ be a class of measurable functions such that $N_{[]}(\epsilon, \mathcal{F}, L_1(P)) < \infty$ for every $\epsilon > 0$. Then $\mathcal{F}$ is Glivenko-Cantelli.*

Proof. Fix $\epsilon > 0$. Choose finitely many $\epsilon-$brackets $[l_i, u_i]$ whose union contains $\mathcal{F}$ and such that $P(u_i - l_i) < \epsilon$ for every i. Then for every $f \in \mathcal{F}$, there is a bracket such that

$$(P_n - P)f \leq (P_n - P)u_i + P(u_i - f) \leq (P_n - P)u_i + \epsilon$$

Consequently,

$$\sup_{f \in \mathcal{F}} (P_n - P)f \leq \max_i (P_n - P)u_i + \epsilon.$$

The right side converges almost surely to ϵ by the strong law of large numbers for real variables. Combination with a similar argument for $\inf_{f \in \mathcal{F}}(P_n - P)f$ yields that $\limsup ||P_n - P||^*_{\mathcal{F}} \leq \epsilon$ almost surely, for every $\epsilon > 0$. Take a sequence $\epsilon_m \downarrow 0$ to see that the limsup must actually be zero almost surely. This completes the proof.

An envelope function of a class $\mathcal{F}$ is any function $x \mapsto F(x)$ such that $|f(x)| \leq F(x)$, for every x and f. The minimal envelope function is $x \mapsto sup_f |f(x)|$. It will usually be assumed that this function is finite for every x.

Theorem 6.9 *Let $\mathcal{F}$ be a P-measurable class of measurable functions with envelope F such that $P^*F < \infty$. Let $\mathcal{F}_M$ be the class of functions $f\mathbf{1}\{F \leq M\}$ when f ranges over $\mathcal{F}$. If $\log N(\epsilon, \mathcal{F}_M, L_1(P_n)) = o^*_P(n)$ for every ϵ and $M > 0$, then $||P_n - P||^*_{\mathcal{F}} \to 0$ both almost surely and in mean. In particular, $\mathcal{F}$ is Glivenko-Cantelli.*

Proof. By the symmetrization lemma, measurability of the class $\mathcal{F}$, and Fubini's theorem,

$$\begin{aligned} E^*||P_n - P||_{\mathcal{F}} &\leq 2E_X E_\epsilon \Big|\Big| \frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i) \Big|\Big|_{\mathcal{F}} \\ &\leq 2E_X E_\epsilon \Big|\Big| \frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i) \Big|\Big|_{\mathcal{F}_M} + 2P^* F\{F > M\} \end{aligned}$$

by the triangle inequality, for every $M > 0$. For sufficiently large M, the last term is arbitrarily small. To prove convergence in mean, it suffices to show that the first term converges to zero for fixed M. Fix $X_1, ..., X_n$. If $\mathcal{G}$ is an $\epsilon-$net in $L_1(P_n)$ over $\mathcal{F}_M$, then

$$E_\epsilon \Big|\Big| \frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i) \Big|\Big|_{\mathcal{F}_M} \leq E_\epsilon \Big|\Big| \frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i) \Big|\Big|_{\mathcal{G}} + \epsilon.$$

The cardinality of $\mathcal{G}$ can be chosen equal to $N(\epsilon, \mathcal{F}_M, L_1(P_n))$. Bound the L_1-norm on the right by the Orlicz-norm for $\psi_2(x) = \exp(x^2) - 1$, and use the maximal inequality (6.7) to find that the last expression does not exceed a multiple of

$$\sqrt{1 + \log N(\epsilon, \mathcal{F}_M, L_1(P_n))} \sup_{f \in \mathcal{G}} \Big|\Big| \frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i) \Big|\Big|_{\psi_2|X} + \epsilon,$$

where the Orlicz norms $|| \cdot ||_{\psi_2|X}$ are taken over $\epsilon_1, ..., \epsilon_n$ with $X_1, ..., X_n$ fixed. By Hoeffding's inequality, they can be bounded by $\sqrt{6/n}(P_n f^2)^{1/2}$, which is less than $\sqrt{6/n}M$. Thus the last displayed expression is bounded by

$$\sqrt{1 + \log N(\epsilon, \mathcal{F}_M, L_1(P_n))}\sqrt{\frac{6}{n}}M + \epsilon \rightarrow_{P^*} \epsilon$$

It has been shown that the left side of (6.4) converges to zero in probability. Since it is bounded by M, its expectation with respect to $X_1, ..., X_n$ converges to zero by the dominated convergence theorem. This concludes the proof that $||P_n - P||^*_{\mathcal{F}}$ in mean. That it also converges almost surely follows from the fact that the sequence $||P_n - P||^*_{\mathcal{F}}$ is a reverse martingale with respect to a suitable filtration.

Donsker Theorems

Uniform Entropy In this section weak convergence of the empirical process will be established under the condition that the envelope function F be squre integrable, combined with the uniform entropy bound

$$\int_0^\infty \sup_Q \sqrt{\log N(\epsilon ||F||_{Q,2}, \mathcal{F}, L_2(Q))} d\epsilon < \infty.$$

Here the supremum is taken over all finitely discrete probability measures Q on $(\mathcal{X}, \mathcal{A})$ with $||F||_{Q,2}^2 = \int F^2 dQ > 0$. These conditions are by no means necessary, but they suffice for many examples. Finiteness of the previous integral will be referred to as the uniform entropy condition.

Theorem 6.10 *Let $\mathcal{F}$ be a class of measurable functions that satisfies the uniform entropy bound* (6.4)*. Let the class $\mathcal{F}_\delta = \{f - g : f, g \in \mathcal{F}, ||f - g||_{P,2} < \delta\}$ and $\mathcal{F}_\infty^2$ be $P-$measurable for every $\delta > 0$. If $P^* F^2 < \infty$, then $\mathcal{F}$ is $P-$Donsker.*

Proof. Let $\delta_n \downarrow 0$ be arbitrary. By Markov's inequality and the symmetrization lemma,

$$P^*(||G_n||_{\mathcal{F}_{\delta_n}} > x) \leq \frac{2}{x} E^* \Big|\Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i f(X_i) \Big|\Big|_{\mathcal{F}_{\delta_n}}.$$

Since the supremum in the right-hand side is measurable by assumption, Fubini's theorem applies and the outer expectation can be calculated as $E_X E_\epsilon$. Fix $X_1, ..., X_n$. By Hoeffding's inequality, the stochastic process $f \mapsto \{n^{-1/2} \sum_{i=1}^n \epsilon_i f(X_i)\}$ is sub-Gaussian for the $L_2(P_n)-$seminorm

$$||f||_n = \sqrt{\frac{1}{n} \sum_{i=1}^n f^2(X_i)}.$$

Use the second part of the maximal inequality (6.2) to find that

$$E_\epsilon \Big|\Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i f(X_i) \Big|\Big|_{\mathcal{F}_{\delta_n}} \leq \int_0^\infty \sqrt{\log N(\epsilon, \mathcal{F}_{\delta_n}, L_2(P_n))} d\epsilon.$$

For large values of ϵ the set $\mathcal{F}_{\delta_n}$ fits in a single ball of radius ϵ around the origin, in which case the integrand is zero. This is certainly the case for values of ϵ larger than θ_n, where

$$\theta_n^2 = \sup_{f \in \mathcal{F}_{\delta_n}} ||f||_n^2 = \Big|\Big| \frac{1}{n} \sum_{i=1}^n f^2(X_i) \Big|\Big|_{\mathcal{F}_{\delta_n}}.$$

Furthermore, covering numbers of the class $\mathcal{F}_\delta$ are bounded by covering numbers of $\mathcal{F}_\infty = \{f - g : f, g \in \mathcal{F}\}$. The latter satisfy $N(\epsilon, \mathcal{F}_\infty, L_2(Q)) \leq N^2(\epsilon/2, \mathcal{F}, L_2(Q))$ for every measure Q.

Limit the integral in (6.4) to the interval $(0, \theta_n)$, make a chagne of variables, and bound the integrand to obtain the bound

$$\int_0^{\theta_n/||F||_n} \sup_Q \sqrt{\log N(\epsilon ||F||_{Q,2}, \mathcal{F}, L_2(Q))} d\epsilon ||F||_n.$$

Here the supremum is taken over all discrete probability measures. The integrand is integrable by assumption. Furthermore, $||F||_n$ is bounded below by $||F_*||_n$, which converges almost surely to its expectation, which may be assumed positive. Use the Cauch-Schwarz inequality and the dominated convergence theorem to see that the expectation of this integral converges to zero provided $\theta_n \to_{P^*} 0$. This would conclude the proof of asymptotic equicontinuity.

Since $\sup\{Pf^2 : f \in \mathcal{F}_{\delta_n}\} \to 0$ and $\mathcal{F}_{\delta_n} \subset \mathcal{F}_\infty$, it is certainly enough to prove that

$$||P_n f^2 - Pf^2||_{\mathcal{F}_\infty} \to_{P^*} 0.$$

This is a uniform law of large numbers for the class $\mathcal{F}_\infty^2$. This class has integrable envelope $(2F)^2$ and is measurable by assumption. For any pair f, g of functions in $\mathcal{F}_\infty$,

$$P_n|f^2 - g^2| \leq P_n|f - g|4F \leq ||f - g||_n ||4F||_n$$

It follows that the covering number $N(\epsilon||2F||_n^2, \mathcal{F}_\infty^2, L_1(P_n))$ is bounded by the covering number $N(\epsilon||F||_n, \mathcal{F}_\infty, L_2(P_n))$. By assumption, the latter number is bounded by a fixed number, so its logarithm is certainly $o_P^*(n)$, as required for the uniform law of large numbers, (6.9). This concludes the proof of asymptotic equicontinuity.

Finally we show that $\mathcal{F}$ is totally bounded in $L_2(P)$. By the result of the last paragraph, there exists a sequence of discrete measures P_n with $||(P_n - P)f^2||_{\mathcal{F}_\infty}$ converging to zero. Take n sufficiently large so that the supremum is bounded by ϵ^2. by assumption, $N(\epsilon, \mathcal{F}, L_2(P_n))$ is finite. Any $\epsilon-$net for $\mathcal{F}$ in $L_2(P_n)$ is a $\sqrt{2}\epsilon-$net in $L_2(P)$. This completes the proof.

References

[1] Billingsley P. *Convergence of Probability Measure*(2nd Edition), John Wiley, New York, 1999.

[2] Chari, R. and Madrekar V. *On Asymptotic Theory in Censored Data Analysis without independent censoring RM-431*, Stat and Prob., MSU, 1983.

[3] Durrett, R. *Probability: Theory and Examples*(3rd Edition), Thomson(Books/Cole) Belmont, 2005.

[4] Gill, R.(1980) *Large Sample behavior of Product Limit Estimator on the whole line*, Ann.Stat. 49-58.

[5] Jacod, J. and Shiryayev, A.N. *Limit Theorems for Stochastic Processes*(2nd Edition), Grundlagen der Math Wiss. 288, Springer-Verlag, Berlin 2003.

[6] Vander Vaart, A. and Wellnes, J.A. *Weak Convergence and Empirical Processes(with applications to statistics)*, Springer, NY, 2000.