Martingale Methods in Statistics

Eric V. Slud
Mathematics Department
University of Maryland, College Park

Eric V. Slud
Statistics Program
Mathematics Department
University of Maryland
College Park, MD 20742

Contents

Chapter 1

Counting Processes, with Statistical Heuristics

This chapter treats the simple counting processs, that is, the class of random right-continuous increasing step-functions $N(t)$ with isolated jumps of unit height which occur at random times $T_1, T_2, \ldots$. Much of the chapter is taken up with examples and applications, and the terminology of 'hazards' and 'compensators' is introduced and interpreted. Although this chapter is less formal than those that follow, two crucial formulas (Theorems 1.1 – 1.3 are proved by calculation, namely the general formulas for the compensator- and variance-process of the simple counting process.

1.1 The Indicator Counting-Process

The simplest nontrivial example of a counting processis

$$N(t) \equiv I_{[T \le t]} = \begin{cases} 1 & \text{if} \quad T \le t \\ 0 & \text{if} \quad T > t \end{cases} \quad , \qquad t \ge 0$$

where T is a nonnegative-valued (waiting time) random variable with distribution function $F(u) \equiv P\{T \le u\} = P\{N(u) = 1\}$. The only randomness in the function $N(\cdot)$ is in the location of T , its single jump. The graph of N looks like

As a matter of notation, let $\bar{F}(u) = 1 - F(u-)$ denote the probability $P\{T \ge u\}$. Then $F(\cdot)$ is right-continuous while $\bar{F}(\cdot)$ is left-continuous with $\bar{F}(0) = 1$. When F has density $f(\cdot)$ with respect to Lebesgue measure on $[0, \infty)$, the *hazard intensity* $h(\cdot)$ of T is defined Lebesgue almost everywhere by

$$h(t) \equiv \lim_{\delta \to 0+} \delta^{-1} P\{T < t + \delta \mid T \ge t\} = f(t)/\bar{F}(t).$$

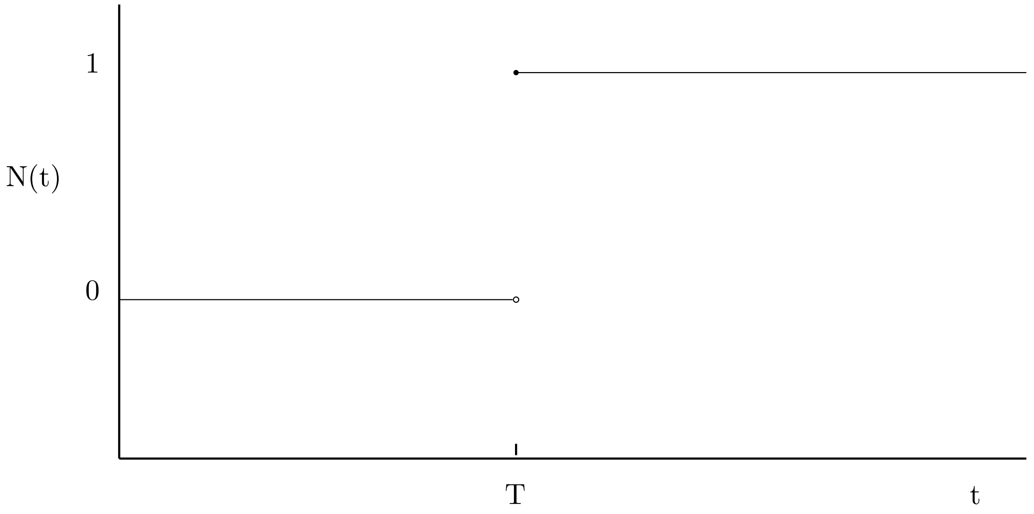

Figure 1: Graph of an indicator counting process $N(\cdot)$.

For small δ, $\delta \cdot h(t)$ can be interpreted as the approximate conditional probability that the random waiting time T is at most $t+\delta$ time units, given that it is least t. More generally, without any restrictions on the distribution function $F(\cdot)$, the *cumulative hazard function* $H(\cdot)$ for T is defined by

$$H(t) \equiv H_T(t) \equiv \int_{0-}^{t} [1/\bar{F}(x)]\, dF(x). \tag{1.1}$$

We give this formula a precise meaning in the following subsection. For justification of all assertions given there, see Apostol (1957), pp. 191-224.

1.1.1 *Digression on Riemann-Stieltjes Integrals.*

Say that a real-valued function r of a real variable x has *isolated discontinuities* if for every x there is a sufficiently small positive δ such that r is continuous on $(x-\delta, x)$ and on $(x, x+\delta)$. Suppose that G is a nondecreasing right-continuous function on $(a, b]$, that q is a left-continuous function for which the right-hand limits $q(x+)$ exist on $(a, b]$, and that both G and q have isolated discontinuities. If both G and q are uniformly bounded on $(a, b]$, then define for $a < t \le b$ the Riemann-Stieltjes integral

$$\int_a^t q(x)\, dG(x) \equiv \lim_{\{x_j:\, 0 \le j \le m\}} \sum_{j=0}^{m-1} q(x_j)\, [G(x_{j+1}) - G(x_j)]$$

where the limit is taken over finite strictly increasing sequences $\{x_j\}_{j=1}^m$ in $(a, t]$, for which as $m \longrightarrow \infty$,

$$\text{mesh}(\{x_j\}) \equiv \max_j\, (x_{j+1} - x_j) \longrightarrow 0$$

The limit does exist, and is equal to

$$\sum_{x\in(a,t]:\,\Delta G(x)>0} q(x)\,\Delta G(x) + \int_a^t q(x)\,dG_c(x)$$

where

$$\Delta G(x) \equiv G(x) - G(x-) \quad , \qquad G_c(x) \equiv G(x) - \sum_{s\in(a,x]:\,\Delta G(x)>0} \Delta G(s)$$

respectively denote the jumps in G and the continuous nondecreasing part of G. If $G(\cdot)$ is piecewise continuously differentiable with $G' = g$, then $G_c = G$, and $\int_a^t q(x)dG(x) = \int_a^t q(x)g(x)dx$. If $G(t) = \int_a^t r(x)dL(x)$ is given by a Riemann-Stieltjes integral as just defined, then G is piecewise continuously differentiable, and $\int_a^t q(x)dG(x) = \int_a^t q(x)r(x)dL(x)$.

The definition of Stieltjes integral is extended to unbounded q and G first for nonnegative q, then in general by decomposing into positive and negative parts of q whenever $\int_a^t |q(x)|\,dG(x) < \infty$. Another extension — to allow right-continuous integrands r with left limits $r(x-)$ and isolated jumps — is given by

$$\int_a^t r(x)\,dG(x) \equiv \int_a^t r(x-)\,dG(x) + \sum_{x\in(a,t]} \Delta r(x)\cdot\Delta G(x)$$

$$= \lim_{\text{mesh}(\{x_i\})\to 0} \;\sum_j\; r(x_{j+1})\cdot[G(x_{j+1}) - G(x_j)]$$

where the limit is taken over partitions of $(a,t]$ as before. These definitions agree precisely with the abstract Lebesgue integral $\int I_{(a,t]}(x)r(x)d\mu(x)$ on the Borel sets of the real line, where μ is the measure defined to satisfy $\mu((a,x]) \equiv G(x) - G(a)$ for each $x \geq a$.

An easy consequence of the foregoing statements is the integration-by-parts formula, valid if r and L are each right-continuous nondecreasing, and therefore also if r and L are each the difference of two such functions.

$$\int_a^t r(x)\,dL(x) + \int_a^t L(x-)\,dr(x) = r(t)L(t) - r(a+)L(a+)$$

$$= \lim_{\{x_j\}} \;\sum_j\; \{\, r(x_{j+1})\,[L(x_{j+1}) - L(x_j)] + L(x_j)\,[r(x_{j+1}) - r(x_j)]\,\}$$

A good exercise for the interested reader is to extend by limiting arguments the foregoing definitions and results to integrators G for which the discontinuities need not be isolated. See the solution to Exercise 1 for an indication of how such limiting arguments are made.

1.1.2 *First Theorems about Compensators.*

Consider again the indicator $N(\cdot) = I_{[T\leq\cdot]}$ and the cumulative hazard function H defined in (1.1). If $F'(t) \equiv f(t)$ exists and is piecewise continuous, then

$$H(t) - \int_0^t (1-F(u))^{-1} f(u)\, du = -\log(1-F(t))$$

Exercise 1 *Show in general that*

$$1 - F(t) = e^{-H_c(t)} \cdot \prod_{x\in(0,t]:\, \Delta H(x)>0} (1 - \Delta H(x))$$

Use the result to conclude that always $1 - F(t) \leq \exp(-H(t))$. □

The importance of the function H derives in large part from:

Theorem 1.1 *Let* c_s *be any bounded left-continuous function on* $[0,\infty)$ *which takes the constant value* c *on* (s,∞)*. Then the process*

$$M(t) \equiv N(t) - \int_{0-}^t I_{[T\geq u]}\, dH(u) = N(t) - H(\min\{t,T\}), \qquad t \geq 0$$

satisfies: $E\{c_s(T)\cdot(M(t)-M(s))\} = 0$ *for all* $t \geq s$.

Remark 1.1 In this Theorem, the r.v. $c_s(T)$ serves as a general way of assigning weights to the successive increments $M(u)-M(v)$ for $u > v$, using only information about $N(\cdot)$ available by watching up to time v. This information consists of the exact value of the r.v. T if $T \leq v$, but otherwise only of the single fact $T > v$. □

Proof. Note first that $E\left\{\int_{0-}^t I_{[T\geq u]} dH(u)\right\} = E\{H(\min\{t,T\})\}$

$$= \int_{0-}^\infty H(\min\{t,s\})\, dF(s) = \int_{0-}^\infty \int_{0-}^{\min(t,s)} dH(x)\, dF(s)$$

Now switch the order of integration in the last double integral by the Fubini-Tonelli Theorem (A.1 in Appendix A), obtaining

$$\int_{0-}^t \int_{x-}^\infty dF(s)\, dH(x) = \int_{0-}^t \bar{F}(x)\,[\bar{F}(x)]^{-1}\, dF(x) = F(t)$$

Now

$$E\{c_s(T)\cdot(M(t)-M(s))\} \;=\; E\{I_{[T\leq s]}\cdot c_s(T)\cdot 0 \;+\; I_{[T>s]}\cdot c\cdot(M(t)-M(s))\}$$

since $M(t) = M(s)$ by definition whenever $T \leq s$, and $c_s(T) = c$ by assumption whenever $T > s$. The last expectation is equal to

$$c \cdot E\{N(t) - N(s) - \int_s^t I_{[T\geq u]}\, dH(u)\} = c \cdot (P\{s < T \leq t\} - F(t) + F(s)) = 0$$

□

Theorem 1.1 says essentially that the increments $N(s+\delta) - N(s)$ and $H(\min\{s+\delta, T\} - H(\min\{s, T\})$, which are 0 unless $T > s$, have the same expected values. This can be interpreted to say that the conditional distribution of $N(s+\delta) - N(s)$ given the value T if $T \leq s$, but otherwise given only that $T > s$, is approximately $Binom(1, I_{[T>s]} \cdot [H(s+\delta) - H(s)])$. Thus the conditional variance of $N(s+\delta) - N(s)$ is

$$I_{[T>s]} \cdot [H(s+\delta) - H(s)] \cdot [1 - H(s+\delta) + H(s)]$$

and we define an accumulated or integrated conditional-variance process

$$V(t) \equiv \int_{0-}^t I_{[T\geq u]} \cdot [1 - \Delta H(u)]\, dH(u).$$

Theorem 1.2 *Let $d_s(\cdot)$ be any bounded left-continuous function with a constant value d on (s, ∞). Then for $t \geq s$, $E\{d_s(T)(M^2(t) - V(t))\} = 0$.*

Proof. Fix $s < t$. Observe that $M^2(t) = M^2(s) - V(t) + V(s) = 0$ by definition whenever $T < s$. Then by Theorem 1.1 with $c_s(T) \equiv M(s)\, d_s(T)\, I_{[T\geq s]}$,

$$\begin{aligned} E\{d_s(T) \quad \cdot \quad & [M^2(t) - M^2(s) - V(t) + V(s)]\} \\ &= d \cdot E\left\{\left([M(t) - M(s)]^2 - V(t) + V(s)\right) I_{[T\geq s]}\right\} \\ &= d \cdot E\left\{[I_{[s<T\leq t]} - \textstyle\int_s^t I_{[T\geq u]} dH(u)]^2 - \int_s^t I_{[T\geq u]}(1 - \Delta(u)) dH(t)\right\} \end{aligned}$$

By expanding the square, recalling $\Delta H(x) = \Delta F(x)/\bar{F}(x)$, and rearranging terms, we find the last line equal to

$$d \cdot \Big(F(t) - F(s) - 2 \cdot \textstyle\int_s^t P\{u \leq T \leq t\}\, dH(u) + \int_s^t \int_s^t P\{T \geq \max(u,v)\} dH(u)\, dH(v)$$

$$- \textstyle\int_s^t \bar{F}(u)\, dH(u) + \int_s^t \bar{F}(u)\, \Delta H(u)\, dH(u)\Big)$$

$$= d \cdot \Big(-2 \cdot \textstyle\int_s^t (F(t) - F(u) + \Delta F(u))\, dH(t) + 2 \cdot \int_s^t \bar{F}(u)\, \Delta H(u)\, dH(u)\Big)$$

$$+ 2 \cdot \textstyle\int_s^t \int_u^t \bar{F}(v)\, \Delta H(v)\, dH(u)\Big)$$

$$= d \cdot \Big(-2 \cdot \textstyle\int_s^t (F(t) - F(u) + \Delta F(u))\, dH(u) + \; 2 \cdot \int_s^t \Delta F(u)\, dH(u)$$

$$+ 2 \cdot \textstyle\int_s^t (F(t) - F(u))\, dH(u)\Big) = 0$$

1.2 Some Formal Definitions

This section provides a mathematical glossary to connect the heuristics of this chapter with the formal mathematical prerequisites sketched in Appendix A. For more detailed mathematical information, see Appendix A and the references cited there.

1. **[Conditional Expectation].** Let $(\Omega, \mathcal{F}, P)$ be a probability space; let $V_1, V_2, \ldots, V_k$ and X be real-valued random variables defined on $(\Omega, \mathcal{F}, P)$ such that $E|X| \equiv \int |X|\, dP < \infty$, and let

$$\mathcal{G} \equiv \sigma(V_1, \ldots, V_k) \equiv \sigma([V_1 \leq x] : \ 1 \leq i \leq k, \ x \in R\})$$

be the σ-algebra generated by the r.v.'s V_i, $i =, \ldots, k$. Then by the Radon-Nikodym Theorem, Section A.3, there is a random variable $E\{X|\mathcal{G}\} \equiv E\{X|V_1, \ldots, V_k\}$ on $(\Omega, \mathcal{F})$ called the *conditional expectation of* X *given* $\mathcal{G}$ and uniquely characterized P-almost surely by the property:

for every bounded continuous function $\gamma : \mathbb{R}^k \longmapsto \mathbb{R}$,
$\int \gamma(V_1, \ldots, V_k) \cdot X\, dP = \int \gamma(V_1, \ldots, V_k) \cdot E\{X \mid \mathcal{G}\}\, dP$

Throughout the chapter and the book, the reader can make the definitions and calculations more concrete by imagining all σ-algebras to be generated by finitely many random variables $V_1, \ldots, V_k$, and by expressing all conditional expectations via 'regular conditional probability densities' (see A.3) as integrals with respect to a conditional probability density given $V_1, \ldots, V_k$. In this spirit, while Theorem 1.1 asserts that $E\{M(t) - M(s) \mid \mathcal{F}_s\} - 0$ whenever $0 \leq s \leq t$, where

$$\mathcal{F}_t \equiv \sigma(N(s) : \ 0 \leq s \leq t) = \sigma(I_{[t \geq t]}, T \cdot I_{[T \leq t]})$$

the calculations in its proof have been carried out in terms of conditional densities given $[T \leq s]$ together with the value of T or given $[T > s]$.

2. **[Stochastic Process].** A *stochastic process* $\{M(t) : \ t \in S\}$ defined on the probability space $(\Omega, \mathcal{F}, P)$ and the set S is simply a family of F-measurable real-valued random variables indexed by elements of S . The index-sets S used in this book are (subsets of) $Z, [0, \infty], \mathbb{R}$, and (in Chapter 12) $\mathbb{R}^2$. Although our notations suppress the dependence of the random variables $M(t) \equiv M(t, \omega)$ on $\omega \in \Omega$, the stochastic process $\{M(t) : t \in S\}$ can also be regarded as a function $M : S \times \Omega \longmapsto \mathbb{R}$ about which some further regularity conditions, such as joint measurability or 'separability', are explicitly imposed (as in Doob 1953). In this book, we adopt the much stronger and intuitively more meaningful restriction (when $S \subset \mathbb{R}$):

$$\text{for all } \omega \text{ in a measurable subset } \omega_1 \text{ of } \Omega \text{ for which } P\{\Omega_1\} = 1, \text{ the function } M(\cdot,\omega) : S \longmapsto \mathbb{R} \text{ is right-continuous and has limits from the left at every point } x \in S. \tag{1.2}$$

This means that not only do each of the events

$$\{\omega \in \Omega : \lim_{s \to x+} M(s,\omega) = M(x,\omega)\} \quad \text{and} \quad \{\omega \in \Omega : \lim_{s \to x-} M(s,\omega) \quad \text{exists}\}$$

for $x \in \text{int}(S)$ have P-probability 1, but also that all these events are contained in a single event Ω_1 of probability 1. Denote by $D(S)$ the set of real-valued functions f on a subset S of $\mathbb{R}$ which are right-continuous and have left-hand limits. The statement (1.2) says just that the *random function* $M(\cdot) \equiv M(\cdot,\omega) : S \longmapsto \mathbb{R}$ is almost surely an element of $D(S)$, i.e., for each ω belonging to some F-measurable set Ω_1 with P-probability 1, the function $M(\cdot,\omega) : S \longmapsto \mathbb{R}$ belongs to $D(S)$. Many books on stochastic processes call the graph $\{(t, M(t,\omega)) : t \in S\}$ of the random function $M(\cdot,\omega)$ its *path* or *time-trajectory*, and we will too.

3. **[Process adapted to a σ-algebra family].** Suppose that the collection $\{\mathcal{F}_t : t \in S\}$ of sub-σ-algebras of $\mathcal{F}$ on Ω, where S is a Borel subset of $\mathbb{R}$, are *increasing* in the sense that whenever $s,t \in S$ and $s < t$, also $\mathcal{F}_s \subset \mathcal{F}_t$. (Two σ-algebras $\mathcal{G} \equiv \sigma(Y_1, \ldots, Y_k)$ and $\mathcal{H} \equiv \sigma(Z_1, \ldots, Z_m)$ generated by finite collection of random variables satisfy $\mathcal{G} \subset \mathcal{H}$ if and only if each of the r.v.'s Y_1 can be expressed as a Borel-measurable function of $Z_1, \ldots, Z_m$.) A stochastic process $\{M(t) : t \in S\}$ on $(\Omega, \mathcal{F}, P)$ is said to be *adapted* to $\{\mathcal{F}_t\}$ if each r.v. $M(t) \equiv M(t,\omega)$ for $t \in S$ is $\mathcal{F}_t$ measurable as a function of $\cdot \in \Omega$.

4. **[Martingale].** A stochastic process $M(t)$ on $(\Omega, \mathcal{F}, P)$ which is adapted to an increasing σ-algebra family $\{\mathcal{F}_t : t \in S\}$ is called a *martingale* with respect to $\{\mathcal{F}_t\}$, or simply an $\mathcal{F}_t$ martingale, if all expectations $E|M(t)|$ are finite and if

$$\text{for all} \quad s,t \in S \quad \text{with} \quad s < t, \qquad E\{M(t) \mid \mathcal{F}_s\} \quad \text{a.s.} \quad (P)$$

This definition applies equally well to the *discrete-time* where the parameter-set S is a discrete subset of R such as Z, as to the *continuous-time* case where S is a subinterval of $\mathbb{R}$.

In sophisticated treatments of continuous-time martingales, it is usually assumed that each $\mathcal{F}_t$ contains all subsets of Ω which are contained in $\mathcal{F}$-measurable sets of P-probability 0, i.e., that each $\mathcal{F}_t$ is *complete* with respect to P, and that $\bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t$, i.e., that $\{\mathcal{F}\}$ is *right-continuous*. These assumptions are used primarily to guarantee that each $\mathcal{F}_t$ martingale $M(t)$ is *equivalent* to some $\mathcal{F}_t$ martingale $M_*(t)$ satisfying (1.2), in the sense that $P\{M(t) \neq M_*(t)\} = 0$ for each t. Since the processes $M(\cdot)$ in this book will always be assumed to satisfy (1.2), we dispense with these technical assumptions on $\mathcal{F}_t$.

In the foregoing terminology, Theorems 1.1 and 1.2 respectively say that the processes $M(t)$ and $M^2(t) - V(t)$ are $\mathcal{F}_t$ martingales. Another calculation which can be similarly summarized is the following.

Exercise 2 *Suppose that* $T_1, T_2, \ldots, T_n$ *are independent and identically distributed random variables with distribution function* F, *and let*

$$\hat{F}_n(t) = n^{-1} \sum_{i=1}^{n} I_{[T_i \le t]} \quad \textit{and} \quad F_t \equiv \sigma(I_{[T_i \le t]}, \quad T_i I_{[T_i \le t]}, \quad i = 1, \ldots, n)$$

Show that $[\hat{F}_n(t) - F(t)/[1 - F(t)]$ *is a* $\{\mathcal{F}_t\}$ *martingale on the interval* $[0, \tau_F)$, *where* $\tau_F \equiv \sup\{s : F(s)\}$, *by fixing* $s < t < \tau_F$ *and calculating*

$$I_{[\hat{F}_n(s)<1]} \cdot E\left(\frac{\hat{F}_n(t) - F(t)}{1 - F(t)} - \frac{\hat{F}_n(s) - F(s)}{1 - F(s)} \,\middle|\, \mathcal{F}_s \right) .$$

□

<u>5.</u> **[Counting Process].** A (*simple*) *counting process* $\{N(t)\}_{t \ge 0}$ on $(\Omega, \mathcal{F}, P)$ is a right-continuous nondecreasing random step-function with isolated jumps of unit height and $N(0) = 0$. It is uniquely determined by the times $T_1, T_2, \ldots$ at which jumps take place. With probability 1,

$$N(t) = n \quad \text{if and only if} \quad T_n \le t < T_{n+1}, \quad n \ge 0 \tag{1.3}$$

where $T_0 \equiv 0$ by convention, and for $n \ge 0$,

$$T_{n+1} \equiv \inf\{t : N(t) > n\} \qquad (\text{possibly} = \infty) \tag{1.4}$$

By assumption $P\{T_{n+1} - T_n > 0 \mid T_1, \ldots, T_n\} = 1$ whenever $T_n < \infty$.

<u>6.</u> **[Compensator and Variance-Process of a Counting Process].** Let $N(t)$ be any counting process such that $E\{N(t)\} < \infty$ for all finite t. As will be proved in Theorem 3.4 and Remark 5.2, there is a uniquely determined stochastic process $\{A(t) : t \in [0, \infty)\}$ adapted to $\mathcal{F}_t \equiv \sigma(N(s) : s \le t) = \sigma(N(t), \{T_i : 1 \le i \le N(t)\})$, called the *compensator* of $N(\cdot)$, such that $A(\cdot) \in D([0, \infty))$ P-a.s., and

(i) $A(0) = 0$, and for $s < t$, $A(s) \le A(t)$

(ii) $A(t-) = A(t)$ whenever $P\{\Delta N(t) = 0 \mid F_{t-}\} = 1$, and $\Delta A(t)$ is measurable with respect to $\mathcal{F}_{t-} \equiv \sigma(\mathcal{F}_s : s < t)$

(iii) $M(t) \equiv N(t) - A(t)$ is a $\mathcal{F}_t$ martingale.

Similarly, if $E(N^2(t)) < \infty$ for all $t < \infty$, there is a unique $\{F_t\}$-adapted stochastic process $V(\cdot)$ called the (*predictable-variance process* for $N(\cdot)$, with paths almost surely in $D([0,\infty))$ and satisfying (i) and (ii) almost surely, for which

(iii′) $(N(t) - A(t))^2 - V(t)$ is a $\{F_t\}$ martingale.

A lot of effort (e.g., in Liptser and Shiryayev 1977, Chapters 4–5) often goes into characterizing $A(\cdot)$ and $V(\cdot)$ uniquely within a much larger class of processes — the class of $\mathcal{F}_t$ 'predictable' processes be described in Section 5.2 — and into supplying conditions under which $A(\cdot)$ essentially determines (the probability law of) $N(\cdot)$. Although such questions are of interest in applications to filtering of the theory we develop, they are irrelevant to us here, since we exhibit $A(\cdot)$ and $V(\cdot)$ explicitly in Theorem 1.3 and then prove results about the martingales of (iii) and (iii′). Nevertheless, we make implicit reference to uniqueness by talking about 'the' compensator and variance-process associated with a counting process.

In the terminology of the present paragraph, Theorems 1.1 and 1.2 say:

if $N(t)$ is a simple counting process with precisely one jump at the random time T, then its compensator $A(t)$ and variance-process $V(T)$ are given in terms of T and H by
$$A(t) = H(\min\{t,T\}) \ , \quad V(t) = \int_0 I_{[T\ge u]} \cdot (1 - \Delta H(u))dH(u) \tag{1.5}$$

1.3 Another Class of Examples — Poisson Processes

Let h be a deterministic nonnegative Borel-measurable function of $[0,\infty)$ such that $H(t) \equiv \int_0^t H(s)ds < \infty$ for each $t < \infty$. The counting process $N(\cdot)$ adapted to an increasing family $\mathcal{F}_t$ of σ-algebras generated by N alone, is called a *Poisson counting process with intensity* h if for all real numbers s,t,u with $0 \le s < t$,

$$E\{e^{iu(N(t)-N(s))} \mid \mathcal{F}_s\} = e^{(e^{iu}-1)(H(t)-H(s))} \text{ a.s.}$$

or equivalently, if all increments $N(t) - N(s)$ with $s < t$ are independent of $\mathcal{F}_s$ and if for each integer $k \ge 0$,

$$P\{N(t) - N(s) = k\} = \frac{1}{k!}\,(H(t) - H(s))^k e^{-(H(t)-H(s))}$$

For such a counting process, the independence of increments implies for any integrable $\mathcal{F}_s$-measurable random variable c_s that when $s < t$,

$$E\{c_s \cdot (N(t) - N(s))\} = E\{N(t) - N(s)\} \cdot E\{c_s\} = E(c_s) \cdot (H(t) - H(s))\}$$

Therefore, in the case $A(t) \equiv H(t) = \int_0^t h(x)$ is a deterministic compensator. Similarly

$$E\{c_1 \cdot [N(t) - N(s) - H(t) + H(s)]^2 = E(c_s) \cdot (H(t) - H(s))\}$$

so that the variance-process $V(\cdot)$ is also nonrandom and equal to $H(\cdot)$.

A theorem due to Watanabe (Brémaud 1981, pp. 21–25) says that if $\mathcal{F}_0$ is the trivial σ-algebra, if h is any nonnegative function with $H(t) \equiv \displaystyle\int_0^t h(x)dx$ finite for finite t, and if a counting process N is restricted only by the assumption that $A(t) \equiv H(t)$ is a compensator for N, i.e., satisfies conditions (i)–(iii) of Section 1.2 above, then N must already be a Poisson counting process with intensity h. See the book of Karlin and Taylor (1975) for extensive discussion and applications of the Poisson processes.

In this subsection, there are no essential changes in statements of results if the *cumulative intensity function* $H(t)$ is continuous but not necessarily absolutely continuous.

1.4 Compensator of a General Univariate Counting-Process

Theorems 1.1 and 1.2 would remain valid if F were a *subdistribution function*, that is, a right-continuous function with $F(0) = 0$ for which $F(\infty)$ might be smaller than 1. This would mean that the counting process N can be allowed with positive probability to make no jumps at all. One important way in which such a zero-or-one-jump process can arise is as an indicator $N(t) \equiv I_{[T \leq \min\{C,t\}]}$, where T and C are nonnegative random variables only the smaller of which can actually be observed. In this case $F(t) \equiv P\{T \leq \min(C,t)\}$, and the appropriate choice of σ-algebra family to describe the information observable up to time t would be

$$\mathcal{F}_t \;\equiv\; \sigma(\, I_{[T \leq \min\langle C,T\rangle]},\, I_{[T \leq \min\langle C,t\rangle]},\, \min\{C < T\} \cdot I_{[\min\langle C,T\rangle \leq t]}\,)$$

Here T might be the waiting-time until failure for an organism or device, and C might represent a random 'censoring' time after which failure could not be directly observed. In this case, we would say that $[T > C]$ is the event that the failure-time T is *censored.* Alternatively T and C might be so-called 'latent' waiting times to failure from different causes, with the process N indicating only failure from the cause associated with time T. Then $[T > C]$ would be described as the event that the device or individual was 'lost to observation' or 'lost to follow-up' at the time C of failure from a 'competing cause'. See the solution of Exercise 3 for further discussion of the latter situation, which is called a *competing risks* survival experiment.

If the (sub-) distribution function $F(t) = P\{T \le t \mid \mathcal{F}_0\}$ is specified conditionally given a nontrivial σ-algebra $\mathcal{F}_0$ of information known or observable at time 0, and if

$$\mathcal{F}_t = \sigma(\mathcal{F}_0, N(s) : 0 \le s \le t) \tag{1.6}$$

then the proofs of Theorems 1.1 and 1.2 given above still show that $M(t)$ and $M^2(t) - V(t)$ are $\mathcal{F}_t$ martingales.

Exercise 3 (a) Show that the compensator given in Theorem 1.1 or equation (1.5) for $N(t) = I_{[T \ge \min(C,t)]}$ is the same as for $I_{[T \ge t]}$ on the interval $[0, C]$ if C is a nonrandom constant.

(b) Find a (necessary and) sufficient condition on the joint distribution of random variables T and C for the compensator of $N(t)$ with respect to the family $\{\mathcal{F}_t\}$ of display (1.6) to be almost surely equal on the random interval $[0, C]$ to the compensator $H(\min\{t, T, C\})$ of $I_{[T \le t]}$ with respect to $\sigma(I_{[T \le t]},\ T \cdot I_{[T \le t]})$.

(c) Interpret the condition you found in (b) in case a joint continuous density for (T, C) exists. □

The following theorem is due to Jacod (1975). For another proof, see Theorem 18.2 and Lemma 18.12 of Lipster and Shiryayev (1977, vol. 2). See Section A.3 for general properties of the *regular conditional probability distributions* (with distribution functions F_k) used here.

Theorem 1.3 *Suppose that* N *is a counting process satisfying (1.3) and (1.4); that* $EN(t) < \infty$ *for each* $t < \infty$*; that equation (1.6) holds and that*

$$P\{T_{k+1} - T_k > 0 \mid T_1, \ldots, T_k, F_0\} = 1 \qquad \textit{a.s. for all} \quad k \ge 0$$

Let

$$F_k(x) \equiv P\{T_{k1} - T_k \le x \mid F_0, T_1, \ldots, T_k\}$$

$$\bar{F}_k(x) \equiv 1 - F_k(x-), \qquad H_k(x) \equiv \textstyle\int_0^x [\bar{F}_k(u)]^{-1}\, dF_k(u)$$

and define nondecreasing right-continuous $\mathcal{F}_k$ *adapted processes* $A(\cdot)$ *and* $V(\cdot)$ *by* $A(0) \equiv A(T_0) \equiv 0$, $V(0) \equiv V(T_0) \equiv 0$, *and for* $T_k < t \le T_{k+1}$

$$A(t) = A(T_k) + \int_0^{t-T} dH_k(u), \qquad V(t) = V(T_k) + \int_0^{t-T} [1 - \Delta H_k(u)]\, dH_k(u)$$

Then $N(t) - A(t)$ *is a* $\mathcal{F}_t$ *martingale, and under the additional assumption that* $EN^2(t) < \infty$ *for each finite* t*, so is* $(N(t) - A(t))^2 - V(t)$*.*

Proof. The idea is to verify the compensator properties separately on each of the intervals $(T_{k-1}, T_k]$ using Theorems 1.1 and 1.2 by considering only counting processes with 0 or 1 jumps. For this purpose, define for $k = 1, 2, \ldots,$ and $0 \le t < \infty$,

$$N_k(t) \equiv N(\min\{t, T_k\}) - N(\min\{t, T_{K-1}\}) \equiv I_{[T_{k-1} < t \le T_k]}$$

and note that the compensator (respectively, variance-process) of N_k with respect to $\mathcal{F}_t$ is obviously the same as the compensator (variance process) of $N_k(\cdot)$ with respect to $\mathcal{G}_{k,t} \equiv \sigma(T_1, \ldots, T_k, \mathcal{F}_t)$.

Then Theorem 1.1, together with the opening remarks of this subsection, say that a compensator for the 0-or-1 jump counting process $N_k(t)$ is given by

$$A_k(t) \equiv A(\max\{T_{k-1}, \min\{t, T_k\}\}) - A(T_{k-1})$$

and (the proof of) Theorem 1.2 says that $(N_k(\cdot) - A_k(\cdot))^2$ is compensated by the process $V_k(T) \equiv V(\max\{T_{k-1}, \min\{t, T_k\}\}) - V(t, T_{k-1})$.

Observe that to apply the proofs of the Theorems 1.1 and 1.2, what is needed is actually not quite equation (1.6) but the weaker condition that for all $0 \le s \le t \le u < \infty$, almost surely on the event $[T_n \le s]$ $(= [N(s) \ge n]\,)$,

$$P\left\{T_{n+1} \ge u \,\middle|\, \{T_i\}_{i-1}^n, \mathcal{F}_t\right\} \;=\; P\left\{T_{n+1} \ge u \,\middle|\, \{T_i\}_{i=1}, I_{[T_{n+1} > t]}, \mathcal{F}_s\right\}$$

This observation will be exploited in Chapter 3, specifically in Theorem 3.5. An important application where the σ-algebras $\mathcal{F}_t$ satisfy this more general condition but not (1.6) will be given in Section 5.4.

Now for each finite integer L define the processes

$$m_L(t) \;\equiv\; \sum_{k=1}^{L} (N_k(\cdot) - A_k(\cdot))\,, \qquad v_L(t) \;\equiv\; \sum_{k=1}^{L} \{(N_k(t) - A_k(t))^2 - V_k(t)\}$$

Each of these processes is a finite sum of $\mathcal{F}_t$ martingales and is therefore itself a martingale. Moreover, by inspection

$$N(\min\{t, T_L\} = \sum_{k=0}^{L} N_k(t), \qquad A(\min\{t, T_L\}) = \sum_{k=0}^{L} A_k(t) \tag{1.7}$$

By monotone convergence these two processes respectively converge to $N(t)$ and $A(t)$ as $L \to \infty$. For fixed $s < t > \infty$, dominate the random variables $m_L(t)$ by the integrable random variables $N(t) + A(t)$ to conclude by the Dominated Convergence Theorem from the martingale property of $m_L(\cdot)$ that

$$N(s) - A(s) = \lim_{L\to\infty} m_L(s) \;=\; \lim_{L\to\infty} E\{m_L(t) \mid \mathcal{F}_s\}$$

and

$$E\{\lim_{L\to\infty} m_L(t) \mid \mathcal{F}_s\} \;=\; E\{N(t) - A(t) \mid \mathcal{F}_s\}$$

i.e., that $N(\cdot) - A(\cdot)$ is also a $\mathcal{F}_t$ martingale.

For each L, and fixed $s < t$,

$$\begin{aligned} &E\{[m_L(t)]^2 - [m_L(s)]^2 \mid \mathcal{F}_s\} \\ &\qquad = E\{m_L(t) - m_L(s)]^2 \mid \mathcal{F}_s\} \qquad \text{(by martingale property, c.f. (3.4))} \\ &\qquad = E\left\{ \left(\textstyle\sum_{k=0}^{L}(N_k(t) - A_k(t) - N_k(s) - A_k(s))\right)^2 \mid \mathcal{F}_s \right\} \qquad \text{(by (1.7))} \\ &\qquad = \textstyle\sum_{k=0}^{L} E\{[N_k(t) - A_k(t) - N_k(s) + A_k(s)]^2 \mid \mathcal{F}_s\} \end{aligned}$$

The last equality follows by the identity for all j:

$$E\{N_j(t) - N_j(s) - A_j(t) + A_j(s) \mid T_1, \ldots, T_{j-1}, \mathcal{F}_0\} = 0$$

Then, by the martingale property for $N_k - A_k$ and $(N_k - A_k)^2 - V_k$ for each k, we have the last sum of conditional expectation

$$= \sum_{k=0}^{L} E\{(V_k(t) - V_k(t) \mid \mathcal{F}_s\}) = E\{V_k(t) - V(s) \mid \mathcal{F}_l\}$$

Thus $M(\min\{t, T_L\}) - A(\min\{t, T_L\}))^2 - V(\min\{t, T_L\})$ is a $\{\mathcal{F}_t\}$ martingale for each finite integer L.

Under the assumption that $EN^2(t) < \infty$, we check that $EA^2(T) < \infty$. Indeed, for each finite L and t,

$$\begin{aligned} EN(\min\{t, T_L\}) &\geq EV(\min\{t, T_L\}) \\ &= E(N(\min\{t, T_L\}) - A(\min\{t, T_L\}))^2 \\ &\geq \left([EA^2(\min\{t, T_L\}))^{1/2} - [EN^2((\min\{t^2, T_L\})]I^{1/2}\right)^2 \end{aligned}$$

where the last inequality comes from the Cauchy-Schwarz inequality. By the Monotone Convergence Theorem, if $EN^2(t) < \infty$, then

$$EA^2(t) \;=\; \lim_{L\to\infty} EA^2(\min\{t, T_L\}) \leq\; 2\{EN(t) + EN^2(t)\}$$

For $t < \infty$, $v_L(t) \equiv N(\min\{t, T_L\}) - A(\min\{t, T_L\})^2 - V(\min\{s, T_L\})$ is dominated for all L by the integrable variable $N^2(t) + A^2(t)$. By equation (1.7), the almost surely pointwise limit of $v_L(\cdot)$ is $(N(\cdot) - A(\cdot))^2 - V(\cdot)$ as $L \to \infty$. It follows as before by the Dominated Convergence Theorem that $(N(t) - A(t))^2 - V(t)$ is a $\{\mathcal{F}_t\}$ martingale. $\square$

1.5 Mantel's Successive Contingency-Table Method for Two-Sample Survival Data.

This section shows in the context of censored survival data the striking notational simplicity achieved by expressing relevant statistics as Stieltjes integrals with respect to compensated counting processes.

Suppose that we can observe

$$\begin{array}{lll} t_i \equiv \{X_i, C_i\}, & \delta_i \equiv I_{[x_i \le c_i]}, & i = 1, \ldots, n \\ s_j \equiv \min\{Y_j, D_j\}, & \epsilon_j \equiv I_{[Y_i, D_i]}, & j = 1, \ldots, m \end{array}$$

where $\{X_i\}_{i=1}^n$ is a sequence of independent and identically distributed random variables independent of the independent and identically distributed sequence $\{Y_i\}_{j=1}^m$, and $\{C_1\}$ and $\{D_j\}$ are known censoring times. Here the X_i and Y_j are regarded as the waiting times for two differently treated groups A and B of medical patients, from entry into a clinical trial until death. If it is known in advance exactly when all the data from the trial will be collected and analyzed, then the *administrative censoring time* is simply the duration from the entry into the trial of the i'th patient of group A until the end of the trial, and D_j can be understood similarly. Since it may make sense to think of the times of entry as random variables, upon which any statistical analysis of the clinical trial should be made conditional, we define $\mathcal{F}_0$ as the σ-algebra generated by the C_i and D_j. A statistical null hypothesis in this experiment might be that all X_i and Y_j have the same distribution function F.

Mantel's (1966) idea was to analyze data of the form described above by considering a series of 2×2 contingency tables summarizing the survival experience of all patients stillunder observation at various amounts of time after entry into the trial. Define

$$N_i^A(t) \equiv \delta_i I_{[t_i \le t]}, \quad i = 1, \ldots, n; \qquad N_j^B(t) \equiv \epsilon_j I_{[s_j \le t]}, \quad j = 1, \ldots, m;$$

$$\begin{array}{lll} N^A(t) \equiv \sum_{i=1} N^A(t), & N^B(t) \equiv \sum_{j=1} N^B(T), & N(t) \equiv N^A(t) + N^B(t) \\ R^A(t) \equiv \sum_{i=1}^n I_{[t_i \ge t]}, & R^B(t) \equiv \sum_{j=1}^m I_{[s_j \ge t]}, & R(t) \equiv R^A(t) + R^B(t) \end{array}$$

The processes $N^A(t)$ and $N^B(t)$ count deaths, respectively in patient groups A and B, which can be observed before the end of the trial among patients who have been in the trial for time t or less. The processes $R^A(t)$ and $R^B(t)$, called the groupwise *numbers at risk* at time-on-test t, count the numbers of patients in groups A and B who had been in the trial for at least time t and who had not died by time t. The very natural idea of summarizing the survival experience of all those in the trial with respect to duration t under study by means of the counting processs N^A, R^A, N^B, R^B is called the *life-table method* and is standard in analyzing survival data. Mantel (1966) formed the following type of contingency-table, one at each of the distinct times t of jumps in $N(\cdot)$, i.e., at the distinct times t_i for which $\delta = 1$ or the times s for which $\epsilon = 1$.

	#Deaths at T	#Survivors past T	Totals at risk
Group A	$x_A(t)$	$R^A(t) - x_A(t)$	$R^A(t)$
Group B	$X_B(t)$	$R^B(t) - x_B(t)$	$R^B(t)$
Totals over 2 groups	$\Delta N(t)$	$R(t) - \Delta N(t)$	$R(t)$

In this table, the number of group A deaths observed at time t is $x_A(t) \equiv \sum_{i=1}^n \Delta N_i^A(t)$. Under the null hypothesis that all X_i and Y_i are independent and identically distributed with distribution function F, the lifetimes of the $R(t)$ individuals 'at risk' at time-on-test t are invariant under permutation of labels, or *exchangeable* (Feller 1971, p. 228). Thus conditionally given the marginal table-totals $\Delta N(t), R^A(t)$, and $R(t)$, as well as the other data (t_i, δ_i) and s_j, ϵ_j observable before time t, i.e. for which t_i or s_j is less than t, the random variable $x_A(t)$ is *hypergeometrically distributed* (Feller 1957, pp. 43 *ff.*) with parameters $R^A(t), \Delta N(t)$, and $R(t)$. Therefore, $x_A(T)$ has conditional expectation $R^A(t)\,\Delta N(t)/R(t)$ and conditional covariance

$$R^A(t)\,R^B(t)\,\Delta N(t) \cdot (R(t) - \Delta N(t))/[R^2(t) \cdot R(t) - 1)]$$

With a motivation which is retrospect is very similar to that of the running conditional expectation $A(T)$ and accumulated conditional variance $V(t)$ of the previous sections, Mantel formed what is now called the *Mantel-Haenszel* or *logrank statistic* by summing $x_A(t)$ over all distinct jump-times t for N, centering the sum by the sum of null-hypothetical conditional expectations, and scaling it by the square root of the sum of conditional variance, obtaining

$$\sum_{t \geq 0:\, \Delta N(t) > 0} \left[x_A(t) - \frac{R^A(t)\Delta N(t)}{R(t)} \right] \Big/ \left(\sum_t \frac{\Delta N(t)\{R(t) - \Delta N(t)\} R^A(t) R^B(t)}{R^2(t)\{R(t) - 1\}} \right)^{1/2}$$

We focus now on re-expressing the numerator of Mantel's statisticin terms of the N and R processes, using the identities

$$N^A(t) \equiv \sum_{0 \geq s \leq t:\, \Delta N(s) > 0} x_A(s)$$

and

$$x_A(t) - \frac{\Delta N(t) \cdot R^A(t)}{R(t)} = x_A(t) - \frac{R^A(t)}{R(t)}(x_A(t) + x_B(t) = \frac{R^B(t)}{R(t)} x_A(t) - \frac{R^A(t)}{R(t)} x_B(t)$$

It is not hard to show from Theorem 1.1, although we will not do it until Section 5.4, that under our null hypothesis the processes

$$M_A(t) \equiv N^A(t) - \int_0^t R^A(u)\, dH(u)\,, \qquad M_B(t) \equiv N^B(t) - \int_0^t R^B(u)\, dH(u)$$

are martingales with respect to the σ-algebra family $\mathcal{F}_t$ generated by all r.v.'s which would be observable by time t. Here H is the cumulative-hazard function associated with F, and the integrals are defined as Stieltjes integrals. This fact lends added interest and statistical importance to the following expressions for Mantel's numerator.

$$\sum_{\substack{s:\, 0 \le s \le t \\ \Delta N(s) > 0}} \left[x_A(s) - \Delta N(s)\, \frac{R^A(s)}{R(s)} \right] \;=\; \int_{0-}^t \left[dN^A(u) - \frac{R^A(u)}{R(u)}\, dN(u) \right] \tag{1.8}$$

$$= \int_{0-}^t \left(\frac{R^B(u)}{R(u)}\, dN^A(u) - \frac{R^A(u)}{R(u)}\, dN(u) \right) - \int_{0-}^t \left(\frac{R^B}{R}\, R^A(u) - \frac{R^A}{R}\, R^B(u) \right) dH(u)$$

$$= \int_{0-}^t \left(\frac{R^B(u)}{R(u)}\, dM_A(u) \right) - \int_{0-}^t \left(\frac{R^A(u)}{R(u)}\, dM_B(u) \right)$$

The martingale behavior of expression (1.8) turns out to have a special relationship to its large-sample behavior. The standard techniques to be developed in Chapters 4 and 5 apply naturally to prove that formula (1.8) has asymptotically normal distribution for large m and n, under simple regularity conditions on $\{C_i\}$ and $\{D_j\}$. The same techniques show quite generally that Mantel's denominator squared is a consistent asymptotic-variance estimator for expression (1.8) under the null hypothesis.

1.6 Further Examples of Martingale-Related Test Statistics

While the previous example shows how martingale and compensator formalism relates to an important test statistic originally introduced on more intuitive grounds, there are many more recent examples of statistics, conceived for use in special hypothesis-testing situations, which completely owe their existence to thinking based on martingales. Two such examples will be presented in this section, one due to Diaconis and Graham (1981) in the context of ESP experimentation with feedback to the experimental subjects, and the other a new nonparametric test for trend in time-sequence data.

1.6.1 *'Skill-Scoring' in Card-Guessing Experiments with Feedback [following Diaconis and Graham 1981]*

Consider an experiment in which a human subject is to be tested for possible ESP by being asked to guess successively the colors of a deck of n shuffled cards in order, where the deck is composed of cards r colors: c_1 of the first color, $\ldots, c_r$ of the r'th color. The card-guesser will successively, for $1 = 1, 2, \ldots, n$, announce his guess g_i of the color of the i'th card, based on 'partial feedback' information concerning earlier cards and guesses. Let the shuffled order of the cards be $\pi(1), \ldots, \pi(n)$, where the random permutation $\pi(\cdot)$ of the symbols $\{1, 2, \ldots, n\}$ is assumed to be distributed uniformly over all $n!$ possible permutations. This corresponds to an assumption of perfect shuffling. The guesser may use auxiliary randomization such as flipping a coin. The feedback is assumed to be of such a form that after each guess g_i, the guesser is told whether his guess is right or wrong, i.e., the value of $\delta_i \equiv I_{g_i = \text{color of } \pi(i)]}$, as well as the value f_i of some function of $\pi(1), \ldots, \pi(i)$ and $g_1, \ldots, g_i$ which is specified as part of the experimental design.

For example, if there are two colors, red and black, then these rules dictate that the guesser know after guess g_i exactly which of the cards $\pi(1), \ldots, \pi(i)$ were red ('complete feedback'). If the colors are understood as the four suits of ordinary playing-cards, then the guesser's information as of just after the i'th guess might be nothing more than which of the first i guesses are correct, or might, for example, include also either the exact values $\pi(1), \ldots, \pi(i-4)$ (full information after a delay of four guesses) or the exact values $\pi(1) \bmod 2, \ldots, \pi(i) \bmod 2$.

The point of this formulation, as indeed of the article of Diaconis and Graham (1981), is the great flexibility of experimental designs which allow a rigorous and intuitively sensible analysis. Diaconis and Graham propose to base the hypothesis test of whether the guesses $\{g_i\}$ are no better than 'purely random' upon their *skill-scoring statistic*

$$S_n = \left\{ n \cdot \frac{r-1}{r^2} \right\}^{-1/2} \sum_{i=1}^{n} \left(\delta_i - E\{\delta_i \mid (f_j, g_j, \delta_j : 1 \le j < 1)\} \right) \tag{1.9}$$

where the conditional expectations are calculated as though the next guess g_i is *purely random* given the present state of knowledge. That is, expectations are calculated under the null hypothesis that g_i is a possible guess which is conditionally independent of $\pi(i)$ given $\mathcal{F}_{i-1} \equiv \sigma((f_j, g_j, \delta_j) : 1 \le j < i)$. The statistic(1.9), which evidently has expectation 0 under the null hypothesis, makes sense because its increments adjust the number of correct guesses by the expected number under purely random guessing given the available information.

Our methodological interest in this and the next example arises from the fact that δ_i are event-indicators whose values are naturally referred under a null hypothesis to a changing conditional-information base. Formally, this is expressed through the observation that

$$\{ (\delta_i - E[\delta_i \mid (f_j, g_j, \delta_j : 1 \le j < i)]) \}_{i=1}^n$$

forms a *martingale difference* sequence, i.e., that the sequences S_n of its partial sums over i form a martingale. Using the martingale central limit theorem of McLeish (1974), to be proved in chapter 4, one concludes that under the null hypothesis, S_n is asymptotically for large n distributed as a standard normal random variable. To apply McLeish's Theorem, one needs to know

$$n^{-1} \sum_{i=1}^n E\{\delta_i \mid \mathcal{F}_{i-1}\} \cdot (1 - E\{\delta_i \mid \mathcal{F}_{i-1}\}) \xrightarrow{P} \frac{r-1}{r} \quad \text{as} \quad n \to \infty$$

which Diaconis and Graham prove cleverly (pp. 12–20 of their paper) by comparing the best and worst guessing-strategies under the null hypothesis.

1.6.2 *A General Hypothesis Test for Trend*

In many settings where one observes time-sequence data $\{X_i\}_{i=1}^n$ but cannot a priori assume independence or identical distribution, it is important to be able to distinguish the nonparametric null hypothesis

$$(\mathbf{H_0}) : \{X_i\} \quad \text{satisfies} \quad P(r_i = k \mid X_i, \ldots, X_{i-1}) = \frac{1}{i} \quad \text{for} \quad 1 \le k \le i$$

against *trend* alternatives of the form

$$(\mathbf{T}) : P\{r_i = k \mid X_1, \ldots, X_{i-1}\} \begin{cases} \text{increase} \\ \text{decrease} \end{cases} \quad \text{in} \quad k \quad \text{for each} \quad i$$

where $r_i \equiv 1 + \sum_{j=1}^{i-1} I_{[x_j \le x_i]}$ denotes the *rank* of X_i among $\{X_j : j \le i\}$. In particular, suppose that $X_i^0 = \sum_{k=1}^p a_k y_{ki} + e_i$ for $i = 1, \ldots, n$, where $Y = (y_{ki} : 1 \le k \le p, \quad 1 \le i \le n)$ is a known $p \times n$ design matrix and where $\{e\}_{i=1}^n$ is an independent sequence of identically and continuously distributed 'errors'. Then for fixed p-vectors $(a_1^j, \ldots, a_p^j)$, $j = 0, 1$, the problem of testing

$$(\mathbf{H_0})^* : (a_1, \ldots, a_p) = (a_1^0, \ldots, a_p^0)$$

versus

$$(\mathbf{H_A})^* : \text{for all} \quad t, \quad \{P(X_i^0 \le t)\}_{i=1}^n \quad \text{has the same rank-order as} \quad \left\{ \sum_{k=1}^p a_k^1 y_{ki} \right\}_{i=1}^n$$

is a special instance of testing (H_0) against (T) for

$$X_i = X_{in} = Z_{\pi(i),n} \qquad \text{where} \qquad Z_{j,n} \equiv X_j^0 - \sum_{k=1}^p a_k^0 y_{kj}$$

and $\pi(\cdot)$ is a permutation of $\{1,\ldots,n\}$ such that $\sum_{k=1}^{p}(a_k^1-a_k^0)y_{k,\pi(i)}$ is monotonically increasing or decreasing in $i=1,2,\ldots,n$.

A general approach to testing (H_0) versus (T) is to fix a sequence $h_1(\cdot),h_2(\cdot),\ldots$ of increasing functions from $[0,1]$ to R and to form the statistic

$$M \;\equiv\; \frac{\sum_{i=1}^{n}\{h_i(r_i/i)-i^{-1}\sum_{j=1}^{i}h_i(j/i)}{\left(\sum_{i=1}^{n}[i^{-1}\sum_{j=1}^{i}h^2(j/i)-(i^{-1}\sum_{j=1}^{i}h_i(j/i))^2]\right)}$$

An especially simple choice for the functions $h_i(\cdot)$, namely $h_i(x)\equiv i\,x$ for all i, gives M_n a form identical of the *two-simple Wilcoxon* statistic(Rao 1973, p. 500). Intuitively, one can view this M_n as a nonparametric statisticfor testing the 'equality in distribution' of the two 'samples' $\{X_i\}_{i=1}^{n}$ and $\{1,\ldots,n\}$ by testing the correspondence between their rank-orderings.

The generality of the statistics M_n derives from the fact that

$$\left\{\sum_{j=1}^{m}[h_i(r_i/i)-i^{-1}\sum_{j=1}^{i}h_i(j/i)]\right\}_{m=1}^{n}$$

is a martingale sequence under (H_0), and by the Martingale Central Limit Theorem of McLeish (Theorem 4.1), the condition

$$\max_{i\le n}\frac{\left(\sum_{j=1}^{n}[h_i(r_i/i)-i^{-1}\sum_{j=1}^{i}h_i(j/1)\right)^2}{\sum_{m=1}^{n}\left\{m^{-1}\sum_{j=1}^{m}h^2(j/m)-(m^{-1}\sum_{j=1}^{m}h_m(j/m))^2]\right\}^2}\longrightarrow 0 \quad\text{as}\qquad n\longrightarrow\infty$$

implies that $M_n \xrightarrow{D} N(),1)$ as $n\to\infty$. This condition is not at all restrictive, and makes the untried statistics M_n seem natural in time-series tests for trend.

1.7 References

1. Aalen, O. (1978) Nonparametric inference for a family of counting processs. *Ann. Statist.* **6**, 701-726.

2. Apostol, T. (1957) *Mathematical Analysis.* Addison-Wesley: Reading, Mass. (pp. 190–225)

3. Diaconis, P. and Graham, R. (1981) The analysis of sequential experiments with feedback to subjects. *Ann. Statist.* **9**, 3–23.

4. Jacod, J. (1975) Multivariate point processes: predictable projection, Radon-Nikodym derivative, representation of martingales. *Zeitschr. f. Wahrsch. verw. Geb.* **31**, 235–253.

5. Lipster, R. and Shiryayev, A. (1977) *Statistics of Random Processes*, vol. 2. (Chapter 18)

6. Mantel, N. (1966) Evaluations of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Reports* **50**, 163–170.

Chapter 2

Weak Convergence of Probability Laws of Random Functions

2.1 Definitions and Portmanteau Theorem

This chapter provides the probabilistic machinery or talking about asymptotic distribution theory as $n \longrightarrow \infty$ for sequences $\{M_n(T) : 0 \leq t \leq T,\ n \geq 1\}$ of stochastic processs related to compensated counting process martingales. As will appear in the examples concluding this chapter, and in many others throughout the book, it turns out not to be enough for statistical purposes to understand simply the behavior for large n of the *finite-dimensional distributions* of M_n, that is, of the joint distributions of finite collections $(M_n(t_1), \ldots, M_n(t_k))$ for fixed $t_1, \ldots, t_k$. One must instead treat each $M_n(\cdot) \equiv M_n(\cdot, \omega)$ as a random function in $D[0,T]$, with a view to understanding the convergence of the sequence of probability laws to M_n defined as probability measures on suitable classes of subsets of $D[0,T]$. A theory of this kind can be built up within the more tractable space $C[0,T]$ continuous functions on $[0,T]$ because the $M_n(\cdot)$ will contain normalizing constants such as $\sqrt{n}$ which will tend tomake the heights of its jumps uniformly small in probability.

Now fix $T = 1$, and write $C \equiv C[0,1]$ as the metric space of real-valued continuous functions on $[0,1]$, with distance between f and g in C given by $\|f - g\| \equiv \sup\{|f(t) - g(t)| : 0 \leq t \leq 1\}$. Then C is complete and separable. That is, Cauchy sequences converge in C, and by the Weierstrass Approximation Theorem (A.)

$$P \equiv \left\{ \sum_{k=0}^{N} a_k t^k : N \geq 0, \quad a_k \quad \text{rational} \right\}$$

is a countable set of functions dense in C. Throughout this chapter, probability measures are defined on

$$B \equiv \sigma(\{f \in C : \|f - f_0\| < r, \quad r \quad \text{rational} \quad f_0 \in P\})$$

which is the same as the Borel σ-algebra generated by the open subsets of C in the supremum-norm topology.

Let $\{\mu_n\}_{n\geq 1}$ and μ be probability measures on the Borel subsets $\mathcal{B}(S)$ of a metric space S which may be either C or $\mathbb{R}^k$ for some $k \geq 1$. We say that μ_n *converges weakly* to μ as $n \longrightarrow \infty$, and write $\mu_n \xrightarrow{w} \mu$ on S, if for all bounded continuous functionals $\gamma : S \longmapsto \mathbb{R}$

$$\int \gamma(f)\, \mu_n(df) \longrightarrow \int \gamma(f)\, \mu(df) \quad \text{as} \quad n \longrightarrow \infty$$

If X_n and X are random elements of S defined on (Ω, F, P) such that for all $A \in \mathcal{B}(S)$

$$\mu_n(A) = P\{\omega \in \Omega : X_n(\cdot, \omega) \in A\}, \quad \mu(A) = P\{X(\cdot) \in A\}$$

that is, μ_n and ν are the *probability laws* or *distributions* of X_n and X, then we say X_n *converges in distribution* and write $X_n \xrightarrow{D} X$ in S if $\mu_n \xrightarrow{w} \mu$ when $n \longrightarrow \infty$.

For $A \in \mathcal{B}(S)$, let A^0, $\bar{A}$, and ∂A, respectively be defined by $A^0 = \mathit{int}(A) =$ interior of A, $\bar{A} = (A^0)^c =$ closure of A, where ^{c} denotes the complement of a set, and $\partial A = \bar{A}/A^0 =$ boundary of A.

Theorem 2.1 (Portmanteau Theorem, Billingsley 1968.) *If S is a metric space, equal either to $C[0,T]$ or $\mathbb{R}^k$, and if μ and μ_n for $n \geq 1$ are probability measures on $B(S)$, then the following are equivalent.*

(a) $\mu_n \xrightarrow{w} \mu$ on S as $n \longrightarrow \infty$;

(b) for all closed $A \subset S$, $\limsup_{n\longrightarrow\infty} \mu_n(A) \longrightarrow \mu(A)$;

(c) for all $A \in \mathcal{B}(S)$ with $\partial A = \phi$, $\mu_n(A) \longrightarrow \mu(A)$ as $n \longrightarrow \infty$.

Proof. $((a) \Longrightarrow (b))$ First suppose $\mu_n \xrightarrow{w} \mu$. For each $M \geq 1$ and closed $A \subset S$, define $k_{A,M}(g) \equiv 1 - \min\{1, M \cdot d(g, A)\}$ for $g \in S$, where

$$d(g, K) \equiv \inf\{\|g - f\| : \ f \in K\} \quad \text{for} \quad K \in \mathcal{B}(S).$$

Then $k_{A,M}(g) - I_A(g) \leq I_{[g\in A^c,\, d(g,A)\leq M^{-1}]}$. Thus $k_{A,M}$ is bounded and converges pointwise to I_A, so by the definition of weak convergence and then the bounded convergence theorem

$$\lim_{n\longrightarrow\infty} \int k_{A,M}\, d\mu_n \;=\; \lim_{M\longrightarrow\infty} \int k_{A,M}\, d\mu \;=\; \mu(A)$$

Therefore,

$$\limsup_{n \longrightarrow \infty} \mu_n(A) \;\le\; \limsup_{n \longrightarrow \infty} \int k_{A,M}\, d\mu \;=\; \mu(A)$$

$((b) \Longrightarrow (c))$ For any $A \in \mathcal{B}(S)$, apply (b) to both $\bar{A}$ and $(A^0)^c$ to find

$$\limsup_{n \longrightarrow \infty} \mu_n(\bar{A}) \;\le\; \mu(\bar{A}), \quad \limsup_{n \longrightarrow \infty} \mu_n((A^0)^c) \;\le\; \mu((A^0)^c)$$

Thus, since $\mu(A^0) = 1 - \mu((A^0)^c$ and $A^0 \subset A \subset \bar{A}$,

$$\mu\,(A^0) \;\le\; 1 - \limsup_{n \longrightarrow \infty} \mu_n(A^0)^c) \;\le\; \limsup_{n \longrightarrow \infty} \mu_n(A)$$

$$\le\; \limsup_{n \longrightarrow \infty} \mu_n(A) \;\le\; \limsup_{n \longrightarrow \infty} \mu_n(\bar{A}) \;\le\; \mu(\bar{A})$$

from which (a) follows immediately.

$((c) \Longrightarrow (a))$ Let $\gamma : S \longmapsto \mathbb{R}$ be any bounded and continuous function(al) on S, and fix $\delta > 0$. Choose a finite sequence of real numbers r_i so that

$$r_0 \;\equiv\; \inf_{f \in S} \gamma(f) - \frac{1}{2}\,\delta \;<\; r_1 \;<\; \cdots \;<\; r_m \;\equiv\; \sup_{f \in S} \gamma(f)$$

with $|r_i - r_{i-1}| \le \delta$ and $\mu(\partial\{f \in S : \gamma(f) \le r_i\}) = 0$ for $i = 1, \ldots, m$. This can be done because only countably many numbers r_i can violate the last condition. Putting $A_i \equiv \{f \in S : r_{i-1} < \gamma(f) \le r_i\}$, we have $\mu(\partial A_i) = 0$ and $\sup_{f \in S} |\gamma(f) - \sum_{i=1}^m r_i I_{A_i}(f)| \le \delta$. Now (c) implies that

$$\int \sum_{i=1}^m r_i\, I_{A_i}\, d\mu_n \;\longrightarrow\; \int \sum_{i=1}^m r_i\, I_{A_i}\, d\mu_n \quad \text{as} \quad n \longrightarrow \infty$$

where $\left|\int \gamma d\mu_n - \int \gamma d\mu\right| \le 3\delta$ for all large n. Since δ was arbitrary, we have proved the assertion of weak convergence in (a). □

The next Theorem and Corollary, often referred to as the *Continuous Mapping Theorem*, show how the weak convergence of probability measures on C or D is ordinarily used. See Section 2.4 for some statistical applications.

Theorem 2.2 *If* $X_n \xrightarrow{D} X$ *in* C *as* $n \longrightarrow \infty$, *and if* $\gamma : C \longmapsto \mathbb{R}$ *is a continuous functional, then* $\gamma(X_n) \xrightarrow{D} \gamma(X)$ *in* $\mathbb{R}$.

Proof. If $f : \mathbb{R} \longmapsto \mathbb{R}$ is bounded and continuous, then $f(\gamma(\cdot))$ is a bounded continuous functional on C. By the distributional convergence of X_n,

$$Ef(\gamma(X_n)) = \int_C f(\gamma(g))d\mu_n(g) \longrightarrow \int f(\gamma(g))d\mu(g) = Ef(\gamma(X))$$

as $n \longrightarrow \infty$. The general definition of weak convergence applied to the real random variables $\gamma(X_n)$ now implies that $\gamma(X_n)$ converges in distribution to $\gamma(X)$ as $n \longrightarrow \infty$.

Now and throughout the rest of the book, we extend the definition of distributional convergence so that if X_n for $n \geq 1$ are random elements of $D[0,T]$ and X is a random element of $C[0,T]$, then we say that $X_n \xrightarrow{D} X$ in D, if there exist random functions $\widetilde{X}_n(\cdot,\omega)$ in C defined on the same probability space (Ω, F, P) such that

$$\|\widetilde{X}_n(\cdot) - X(\cdot)\| \xrightarrow{P} 0, \quad \text{and} \quad \widetilde{X}_n \xrightarrow{D} X \quad \text{in} \quad C \quad \text{as} \quad n \longrightarrow \infty.$$

Corollary 2.3 *If $\{X_n\}$ is a sequence of random functions in $D[0,T]$ which converges in distribution according to our extended definition to a continuous random function X, then for any functional γ from $D[0,T]$ to R which is continuous with respect to the sup-norm.*

Proof. Let $\widetilde{X}_n$ be a sequence of continuous random functions assumed to exist satisfying $\|\widetilde{X}_n - X_n\| \xrightarrow{P} 0$ as $n \longrightarrow \infty$. Since γ is a continuous functional on C, Theorem 2.2) implies

$$\gamma(\widetilde{X}_n) \xrightarrow{P} \gamma(X) \quad \text{in} \quad \mathbb{R} \quad \text{as} \quad n \longrightarrow \infty$$

and norm continuity of γ on D implies $|\gamma(\widetilde{X}_n) - \gamma(LX_n)| \xrightarrow{P} 0$. The Corollary now follows from Slutsky's theorem. □

2.2 Criteria for Distributional Convergence

The purpose of this section is to develop criteria for distributional convergence as $n \longrightarrow \infty$ of stochastic processes X_n as far as possible in terms of probabilities concerning random vectors $(X_n(t_1), \ldots, X_n(t_k))$ formed by evaluating these processes at arbitrary finite sets of times $t_1, t_2, \ldots, t_k$. Recall that for a continuous process X on $[0,T]$, i.e., a stochastic process $X(t,\omega)$ such that each path-function $X(\cdot,\omega)$ is continuous on $[0,T]$, the law of X is the probability measure $\mu \equiv \mu_X$ on $C = C[0,t]$ defined by

$$\mu(A) \equiv \mu_X((\{f \in C : \ f \in A\}) \equiv P(\{\omega \in \Omega : \ X(\cdot,\omega) \in A\}), \quad A \in \mathcal{B}(C)$$

Where a sequence $\{X_n\}$ of stochastic processes is under consideration, we shall write μ_n in place of μ_{X_n} for their laws. Let $\mathbf{t} \equiv \mathbf{t}(\mathbf{k}) = (t_1, \ldots, t_k)$ denote an arbitrary k-tuple of elements of $[0,T]$, where $k \geq 1$. For a particular process X with law μ, define the law of the random k-vector $(X(t_1), \ldots, X(t_k)$ on Borel-measurable subsets B of $\mathbb{R}^k$ by

$$\begin{aligned} \mu_{\mathbf{t}(\mathbf{k})}(B) \;&\equiv\; \mu_{t_1,\ldots,t_k}(B) \;\equiv\; \mu(\{f \in C : (f(t_1), \ldots, f(t_k)) \in B\}) \\ &= \; P(\{\omega \in \Omega : (X(t_1,\omega), \ldots, X(t_k,\omega)) \in B\}) \end{aligned}$$

The measures $\mu_{\mathbf{t}(\mathbf{k})}$ for all k and $\mathbf{t}(\mathbf{k})$ are called the *finite-dimensional distributions* either for μ or for X.

For the rest of this section, let $\{t_j\}_{j=1}^{\infty}$ be a fixed enumeration (Appendix A.) of the set $[0,T]\cap\mathbb{Q}$, that is, of all rational numbers in $[0,t]$. Since for each $f_0 \in C$

$$\{f \in C : \|f - f_0\| \le r\} = \bigcap_{j\ge 1}\{f : |f(t_j) - f(t_j)| \le r\}$$

it is easy to see that $\mathcal{B}(C) = \sigma(\{f \in C : f(t_j) \le r\},\ j \ge 1,\ r \in \mathbb{Q})$. Since the collection of $\mu_{\mathbf{t}(\mathbf{k})}$ determines all μ-probabilities of finite intersections of the generating class of sets just given for $\mathcal{B}$, the Extension Theorem of Caratheodory (Appendix B.1) implies that the finite-dimensional distributions $\mu_{\mathbf{t}(\mathbf{k})}$ of μ determine the collection of μ probabilities on C. Conversely, the Kolmogorov-Daniell Extension Theorem (Appendix B.5) says that any mutually consistent family of $\mu_{t(k)}$ does uniquely determine a probability measure ν on $\mathbb{R}^\infty$, the space of real infinite sequences. The latter fact will be used to characterize the distributions of infinite random sequences $\underline{\xi} = (\xi_j, j \ge 1)$ obtained as distributional limits of sequences $XX_n(T_j),\ j \ge 1$ as $n \longrightarrow \infty$. At that point, a further hypothesis ("tightness", defined in the next paragraph) on the sequence of probability laws of X_n is needed to prove that there exists a continuous stochastic process X for which such a limiting random sequence ξ has the same distribution on $\mathbb{R}^\infty$ as $(X(t_j),\ j \ge 1)$. These steps are carried out in Theorem 2.4) below.

A family $\{\nu_\alpha : \alpha \in I\}$ of probability measures on C is called *tight* if for each $\delta > 0$ there is a compact subset $K_\delta \subset C$ such that for all $\alpha \in I,\ \mu_\alpha(K_\delta) \ge 1 - \delta$. Here K is (sequentially) *compact* if every countable sequence $\{f_n\} \subset K$ has some subsequence converging uniformly to some $f \in K$. The classical and very tractable criterion of Arzelà-Ascoli for a subset of C to be compact, will be given later in this section.

Theorem 2.4 (Prohorov 1956) *If* $\{\mu_\alpha\}_{\alpha\in I}$ *is tight on* C*, then every infinite subsequence* $\{\mu_{\alpha(n)}\}_{n\ge 1}$ *has a further subsequence converging weakly on* C*.*

Proof. The Helly Selection Theorem on $\mathbb{R}^k$ (Appendix B.6) asserts that if $\{F_n\}$ is a sequence of joint distribution functions on $\mathbb{R}^k$, then there exists a subsequence converging pointwise, at all continuity points of the limit, to a function G which has all the properties of a joint distribution function on $\mathbb{R}^k$ except that its range may be contained in a subinterval of $[0,1]$. If $(-N,N)^k$ is an open rectangle with measure $\ge 1-\delta$ according to all the F_n for $n \ge 1$, then it follows as in Theorem 2.1)((a) $\Longrightarrow$ (b)) that

$$\int_{(-N,N^k)} dG(x) \;\ge\; 1 - \int_{(-N,N^k)^c} dF_n(x) \;\ge\; 1-\delta$$

If such a rectangle exists for each $\delta > 0$, then G is a proper joint distribution function.

If K is a compact subset of C, then it is easy to check that

$$K_{\mathbf{t}} \equiv K_{\mathbf{t(k)}} \equiv K_{t_1,\ldots,t_k} \equiv \{(f(t_1),\ldots,f(t_k)) : f \in K\}$$

is a closed and bounded, and therefore compact, subset of $\mathbb{R}^k$. Thus if $\{\mu_\alpha : \alpha \in I\}$ is tight, and if for each $\delta > 0$, K_δ is a compact subset of C such that $\mu_\alpha(K_\delta) \geq 1-\delta$ for all α, then for each $t(k)) \equiv (t_1,\ldots,t_k)$, $f \in k\}$,

$$\mu_{t(k)}(K_{\delta t(k)}) \;=\; \mu_\alpha(\{f : (f(t_1),\ldots,f(t_k)) \in K_{\delta t_1,\ldots,t_k}\}) \;\geq\; 1-\delta$$

By the Helly Selection Theorem, for each $\{\mu_{\delta(n)}\}_{n\geq 1}$ and k, there exists a subsequence $\{mu_{\alpha\prime(n)-n\geq 1}\}$ such that as $n \longrightarrow \infty$,

$$\mu_{\alpha\prime(n)t_1,\ldots,t_k} \xrightarrow{w} \quad \text{some probability measure} \quad \nu_{t(k)} \equiv \nu_{t_1,\ldots,t_k}$$

on $\mathbb{R}_k$. Apply this argument to successive subsequences for $k = 1, 2, \ldots$. Then the diagonal argument (Appendix A.), just as in the proof of the Helly Theorem, yields a subsequence $\{\alpha^0(n)\}$ of $\{\alpha(n)\}$ such that simultaneously for all k and $\mathbf{t} \equiv \mathbf{t(k)} \equiv (t_1,\ldots,t_k)$,

$$\mu_{\alpha^0(n)t(k)} \xrightarrow{w} \nu_{t(k)} \quad \text{as} \quad n \longrightarrow \infty.$$

The laws $\nu_{t(k)}$ on $\mathbb{R}^k$ are *mutually consistent*, that is, for each $(t_1,\ldots,t_k)$ and measurable $A \subset \mathbb{R}^{k-1}$,

$$\nu_{t_1,\ldots,t_k}(A \times \mathbb{R}) = \nu_{(t_1,\ldots,t_{k-1})}(A).$$

By the Kolmogorov Extension Theorem (Appendix B.5), there is a probability measure ν on $\mathbb{R}^\infty$ with finite-dimensional distributions

$$\nu(\{\underline{a} = (a_1, a_2, \ldots) \in \mathbb{R}^\infty : (a_1,\ldots,a_k) \in A\}) = \nu_{t(k)}(A) \in B(\mathbb{R}^k).$$

We will next use the key fact that for compact subsets $K \subset C$,

$$\{(f(t_1, f(t_2),\ldots) \in R^\infty : f \in K\} = \bigcap_{k\geq 1}\{\underline{a} \in \mathbb{R}^\infty : (a_1,\ldots,a_k) \in K_{\mathbf{t(k)}}\} \qquad (2.1)$$

To verify (refcompact), observe first that for every k, by definition

$$K \subset \{g \in C : \quad \text{for some} \quad f \in K, \;\; g(t_i) = f(t_i) \quad \text{for} \quad i = 1,\ldots,k\}$$

which says precisely that the left-hand side of (2.1) is contained in the right-hand side. However, if the sequence $\underline{a}$ is an element of the right-hand side of (2.1), then for every $k \geq 1$, there is an element g_k of K such that $a_j = g_k(t_j)$ for every $j \leq k$. Since D is compact the sequence $\{g_k\}$ must have a subsequence converging (in the supremum-norm topology) to an element $f \in K$. Since $a_j = g_k(t_j)$ for all $k \geq j$, also $f(t_j) = a_j$. In other words, $\underline{a}$ belongs to the left-hand side of (2.1).

By Theorem 2.1 and the weak convergence of the finite-dimensional distributions of $\mu_{\alpha^0(n)}$ as $n \longrightarrow \infty$ for each $k \geq 1$ and each $\delta > 0$,

$$\limsup_{n\longrightarrow\infty} \mu_{\alpha^0(n)}(A) \leq \limsup_{n\longrightarrow\infty} \mu_{\alpha^0(n)}(A \cap K_\delta) + \mu_{\alpha^0(n)}(K_\delta^c))$$

$$\leq \limsup_{n\longrightarrow\infty} \left\{\mu_{\alpha^0(n)}(\{f \in C : (f((t_1), \ldots, f(t_k)) \in (A \cap K_\delta)_{t(k)}\}) + \mu_{\alpha^0(n)}(K_\delta^c)\right\}$$

$$\leq \nu(\{\underline{a} \in \mathbb{R}^\infty : (a_1, \ldots, a_k) \in (A \cap K_\delta)_{t(k)}\}) + \delta.$$

Letting $k \longrightarrow \infty$ and using countable additivity of ν on $\mathbb{R}^\infty$, along with (2.1) applied to the compact set $A \cap K$, we find for each closed A ,

$$\limsup_{n\longrightarrow\infty} \mu_{\alpha^0(n)}(A) \leq \nu(\{\underline{a} \in \mathbb{R}^\infty : \quad \text{for} \quad k \geq 1, \ (a_1, \ldots, a_k) \in (A \cap K_\delta)_{t(k)}\}) + \delta$$

$$= \nu(\{(f(t_1), f(t_2), \ldots) : f \in A \cap K_\delta\} + \delta. \tag{2.2}$$

Now let $\delta \longrightarrow 0$, and conclude

$$\limsup_{n\longrightarrow\infty} \mu_{\alpha^0(n)}(A) \leq \nu(\{(f(t_1), f(t_2), \ldots) : f \in A\}).$$

Take A to be all of C shows that $\nu(\{(f(t_1), f(t_2), \ldots) : f \in C\}) = 1$. Finally, define a probability measure μ on C by

$$\mu(A) \equiv \nu(\{(f(t_1), f(t_2), \ldots) : f \in A\}), \quad A \in B(C).$$

Then (2.2) and Theorem 2.1 immediately imply $\mu_{\alpha^0(n)} \xrightarrow{w} \mu$ on C and $n \longrightarrow \infty$, and we have found a subsequence $\{\mu_{\alpha^0(n)}\}$ of $\{\mu_{\alpha(n)}\}$ converging weakly on C. □

The classical criterion for compactness of $K \subset C$ is the *Arzelà-Ascoli Theorem.* A set $K \subset C$ has compact closure in C if and only if it is **bounded** and **uniformly equicontinuous**, i.e., if and only if

$$\sup_{f\in K} |f(0)| < \infty \quad \text{and} \quad \lim_{\delta\downarrow 0} \sup_{f\in K} \sup_{\substack{0\leq s,t\leq 1 \\ |s-t|<\delta}} |f(s) - f(t)| = 0.$$

A discussion and proof of this result can be found in Billingsley (1968, p. 221) or in Coddington and Levinson (?). The idea of the next theorem is to re-express information about compactness from Arzelà-Ascoli in a form useful in applying Theorem 2.4.

Theorem 2.5 *$\{\mu_n : n \geq 1\}$ is tight on C if and only if for all $\alpha, \beta > 0$ there exist $M < \infty$ and $\delta > 0$ such that for all $n \geq 1$*

$$\mu_n(\{f : |f(0)| > M\}) < \beta, \quad \mu_n(\{f : \sup_{|s-t|\leq\delta} |f(s) - f(t)| > \alpha\}) < \beta \tag{2.3}$$

Proof. If (2.3) holds, then for given $\beta > 0$ put

$$K_\beta \equiv \{f : |f(0)| \leq M\} \cap \bigcap_{k=1}^{\infty} \{f : \sup_{|s-t|<\delta(k)} |f(s) - f(t)| \leq k^{-1}\}$$

where $\delta(k)$ and M are chosen by (2.3) so that

$$\mu_n(\{f \in C : \sup_{|s-t|\leq\delta(k)} |f(s) - f(t)| > k^{-1}\}) \leq \beta/s^{k+1}, \ \mu_n(\{f : |f(0)| \leq M\}) \leq \frac{1}{2}\beta.$$

Then $\mu_n(K_\beta) \geq 1 - \beta$ for $n \geq 1$, and K_β is compact by Arzelà-Ascoli. Conversely, if for given $\beta > 0$, K is compact subset of C with $\mu_n(K^c) \geq \beta$ for all $n \geq 1$, and if $\alpha > 0$ is given, then Arzelà-Ascoli Theorem says that there exist M and δ for which

$$K \subset \{f \in C : |f(0)| \leq M, \ \sup_{|s-t|<\delta} |f(s) - f(t)| \leq \alpha\}.$$

This inclusion immediately implies (2.3). □

Remark 2.1 If $\{\mu_n\}$ is a family of probability laws on C for which

(α) $((\mu_n)\mathbf{t} \xrightarrow{w} \mu_{\mathbf{t}}$ on $\mathbb{R}^k$ as $n \longrightarrow \infty$ for each $\mathbf{t} = (t_1, \cdots, t_k)$, and

(β) $\{\mu_n\}$ is tight on C,

then $\mu_n \xrightarrow{w} \mu$ for some probability measure μ on C with finite-dimensional distributions $\mu_{\mathbf{t(k)}}$. This is so because, by Prohorov's Theorem and (β), each subsequence of $\{\mu_n\}$ has a weakly convergent subsequence and the weak limits must all be μ due to (α). Condition (α) is usually the conclusion of some Central Limit Theorem, and (β) is verified through (2.3) in particular situations. □

In applications of weak convergence, one often encounters stochastic processes which are not continuous but which almost surely belong to some $D[a, b]$. If the jumps of these processes become small as the sequence-index goes to ∞, then condition (2.3) is enough to ensure weak convergence in our extended sense in D, to a probability law on C.

Theorem 2.6 *Suppose $\{X_n : n \geq 1\}$ is a sequence of random functions in $D[a, b]$ such that*

(i) for each α and $\beta > 0$, there exists $\delta > 0$ and a finite integer n_0 such that for all $n \geq n_0$,

$$P\{\omega : \sup_{|s-t|<\delta} |X_n(t, \omega) - X_n(t, \omega)| \geq \alpha\} < \beta;$$

(ii) $\limsup_{M\to\infty} P\{\omega : |X_n(a,\omega)| \geq M\} \;=\; 0$*; and*

(iii) for all finite subsets $(t_1, \ldots, t_k)$ *of rational numbers in* $[a,b]$,

$$(X_n(t_1), \ldots, X_n(t_k)) \xrightarrow{D} (X(t_1), \ldots, X(t_k)) \quad \textit{in} \quad \mathbb{R}^k \quad \textit{as} \quad n \longrightarrow \infty,$$

where $(X(r) : r \in [a,b] \cap \mathbb{Q})$ *is some random sequence in* $\mathbb{R}^\infty$.

Then in our extended sense, $X_n \xrightarrow{D} X$ *in* $D[a,b]$ *as* $n \longrightarrow \infty$, *and with probability* 1 *the limiting process* X *is continuous on* $[a,b]$. *Conversely, if* $X_n \xrightarrow{D} X$ *in* D, *with* X *an element of* $C[a,b]$, *then conditions (i)–(iii) hold.*

Proof. The idea is to approximate the random elements X_n of D by linearly-interpolated continuous random functions $\widetilde{X}_n(\cdot)$. For each $n \geq 1$ and $0 \leq t \leq 1$, let

$$\widetilde{X}_n \;\equiv\; (1 - \{nt\})\, X_n([nt]/n) \;+\; \{nt\}\, X_n(([nt]+1)/n)$$

where $[x]$ denotes the greatest integer less than or equal to x, and $\{x\} \equiv x - [x]$ is the fractional part of x. Then $X(\cdot)$ is continuous by definition, and by (i)

$$\sup_{t\in[a,b]} |\widetilde{X}_n(t) - X_n(T)| \xrightarrow{P} 0 \quad \text{as} \quad n \longrightarrow \infty$$

Now (i)–(iii) for $\{X_n\}$ evidently imply that (i)–(iii) hold with X_n replaced by $\widetilde{X}_n$. Note that by continuity of the random functions $\widetilde{X}_1, \ldots, \widetilde{X}_{n_0}$, for fixed α and β a still smaller δ' can be chosen so that the probability-inequality (i) holds with $\widetilde{X}_n$ replacing X_n and with δ' replacing δ, for all $n \geq 1$. Theorem 2.5 and Remark 2.1 say that $\widetilde{X}_n \xrightarrow{D} X$ as $n \longrightarrow \infty$, and that $\{X(r) : r \in [a,b] \cap \mathbb{Q}\}$ is almost surely the sequence of values at rational points of a continuous function on $[a,b]$. But this is exactly what the extended definition of convergence in distribution in D for X_n requires.

For the converse direction, observe that distributional convergence in extended sense in D implies the existence of continuous stochastic processes $X_n^\ddagger$ satisfying

$$X_n^\ddagger \xrightarrow{D} X \quad \text{in} \quad C, \quad \text{and} \quad \|X_n^\ddagger - X_n\| \xrightarrow{P} 0 \quad \text{as} \quad n \longrightarrow \infty \tag{2.4}$$

Theorem 2.5 implies (i)–(iii) hold with X_n replaced by $X_n^\ddagger$, and the second part of (2.4) then implies that (i)–(iii) hold for X_n. □

Remark 2.2 In order for (i) of Theorem 2.6 to hold, the jumps of $X_n(\cdot)$ must become uniformly small in probability in the sense that

$$\sup\{|X_n(t) - X_n(r-)| : t \in [a,b]\} \xrightarrow{P} 0 \quad \textit{as} \quad n \longrightarrow \infty$$

□

Remark 2.3 It is easy to check from the definitions that if $X_n(\cdot)$ for $n \geq 1$ are random functions in $D[0,1]$ which converge in distribution as $n \longrightarrow \infty$ to an almost surely continuous random function X, and if $\gamma : [0,1] \longmapsto [a,b]$ is a strictly monotone continuous nonrandom function, then as $n \longrightarrow \infty$, $X_n^0 \gamma^{-1} \longrightarrow X^0 \gamma^{-1}$ in $D[a,b]$. This comment is helpful in interpreting weak convergence on $D[a,b]$ when $[a,b]$ is (semi-) infinite. □

An important extension of Corollary 2.3 and Remark 2.3 is

Corollary 2.7 *Suppose* $\{X_n\}_{n\geq 1}$ *is a sequence of* $D[a,b]$ *stochastic processes which converge in distribution as* $n \longrightarrow \infty$ *to the continuous* $D[a,b]$ *random function* X*, and suppose* $\gamma : D[a,b] \longrightarrow D[c,d]$ *is a supremum-norm-continuous functional which sends continuous functions to continuous functions. Then*

$$\gamma(X_n) \xrightarrow{D} \gamma(X) \qquad \textit{in} \quad D[c,d] \quad \textit{as} \quad n \longrightarrow \infty.$$

Proof. Observe first that by Corollary 2.3, for fixed $c \leq t_1 \leq \cdots \leq t_k \leq d$ and real $\alpha_1, \ldots, \alpha_k$, as $n \longrightarrow \infty$

$$\sum_{i=1}^{k} \alpha_i \, \gamma(X_n)(t_i) \xrightarrow{D} \sum_{i=1}^{k} \alpha_i \, \gamma(X)(t_i) \qquad \text{in} \quad \mathbb{R}$$

By the Cramèr-Wold device (Appendix B.6) the finite-dimensional distributions of $\gamma(X_n)$ or $\gamma(X_n^{\ddagger})$ converge weakly to those of $\gamma(X)$, where $\{X_n^{\ddagger}\}$ is any sequence of stochastic processes in $C[a,b]$ for which (2.4) holds. From Theorems 2.5 and 2.6, it follows that the probability laws μ_n of X_n form a tight family of probability measures on C and therefore that for each $\beta > 0$, there exists a compact subset $K \equiv K(\beta)$ of C such that $P\{X_n \in K(\beta)\} \geq 1 - \beta$ for all $n \geq 1$. Then, since $\gamma(\cdot)$ is continuous on $D[a,b]$ and sends continuous functions to continuous functions, the set $\gamma(K(\beta)) \equiv \{g \circ f : \, f \in K(\beta)\}$ of $D[c,d]$ functions is actually a compact subset of $C[c,d]$. At the same time,

$$P\left\{\gamma(X_n^{\ddagger}) \in \gamma(K(\beta))\right\} \; \geq \; 1 - \beta \qquad \text{for all} \quad n \geq 1$$

so that the family of probability laws $\mu_n \circ \gamma^{-1}$ of $\gamma(X_n)$ are also tight on $C[c,d]$, and the Corollary will follow from Remark 2.1 if we show

$$\|\gamma(X_n^{\ddagger}) - \gamma(X_n)\| \xrightarrow{P} 0 \qquad \text{as} \quad n \longrightarrow \infty \tag{2.5}$$

Now continuity of γ on $D[a,b]$ in supremum norm together with sup-norm compactness of $K(\beta)$ in $D[a,b]$ imply as $\delta \longrightarrow 0$,

$$\sup\{\|\gamma(f) - \gamma(g)\| : \; f \in K(\beta), \quad g \in D[a,b], \quad \|f - g\|\} \; < \; \delta \longrightarrow 0 \tag{2.6}$$

If (2.6) were false, there would exist sequences $f_n \in K(\beta)$ and $g_n \in D[a,b]$, with $\|f_n - g_n\| \le \delta_n$ and $\delta_n \longrightarrow 0$, such that $\|\gamma(f_n) - \gamma(g_n)\|$ is bounded away from 0 as $n \longrightarrow \infty$. Since $K(\beta)$ is compact, there exists $f \in K(\beta)$ such that some subsequence $\{f_{n'}\}$ of $\{f_n\}$ converges in sup-norm to f. Thus $g_{n'}$ also converges to f. By continuity of γ, $\gamma(g_{n'})$ and $\gamma(f_{n'})$ converge to $\gamma(f)$ in sup-norm, contradicting the boundedness away from 0 of $\|\gamma(f_{n'}) - \gamma(g)_{n'})\|$. This contradiction proves (2.6).

From (2.6) and the observation that $\|X_n^\ddagger - X_n\| \xrightarrow{P} 0$ as $n \longrightarrow \infty$, we conclude that for all α, $\beta > 0$ and sufficiently small $\delta = \delta(\alpha, \beta)$,

$$P\{\|(X_n^\ddagger) - \gamma(X_n\| > \alpha\} \le P\{X_n^\ddagger \in K(\beta)^c\} + P\{\|X_n^\ddagger - X_n\| \ge \delta\}$$

Therefore (2.5) and the Corollary have been proved. □

Remark 2.4 If the **modulus of continuity** $\omega_f(\cdot)$ of each function f in $D[a,b]$ is defined by

$$\omega_f(r) \equiv \sup\{|f(s) - f(t)| : s \le s, t \le b, |s-t| < r\}$$

then the tightness conditions (2.3) of Theorem 2.5 and (i) of Theorem 2.6 can be understood as uniform in-probability bounds $\omega_x(\delta) = o_p(1)$ as $\delta \longrightarrow 0$ on the moduli of continuity of the random functions X_n. From this point of view, it is clear why the tightness-conditions (2.3) and (i) ensure that any possible limiting distribution of X_n must assign probability 1 to C. □

For a unified and general method of deriving tightness criteria for laws of processes C_n based on moments of $|X_n(t) - X_n(s)|$, see §3 of the Stroock- and Varadhan-edited 1973 NYU Seminar Notes.

2.3 Construction of the Wiener Process

The limiting continuous random functions arising in problems of statistics can in most cases be understood as simple transformations of a single continuous random function with Gaussian finite-dimensional distributions, namely the **Wiener process** or **standard Brownian motion**. A straightforward construction of Wiener process using Theorem 2.6 is given in this section.

For each $n \ge 1$, let $\{Z_{j,n} : j = 1, \ldots, 2^n\}$ be an independent and identically distributed sequence of $N(0,1)$ random variables and define a random process $W_n(\cdot)$ in $D[0,1]$ by

$$W_n(T) \equiv 2^{-n/2} \sum_{j=1}^{[2^n t]} Z_{j,n} \qquad \text{for} \quad 0 \le t \le 1$$

where $[x]$ again denotes the greatest integer less than or equal to x. It is easy to see that W_n has Gaussian, or multivariate-normal, finite-dimensional distributions with

$$EW_n(t) = 0 \qquad \text{and} \qquad E = \{W_n(s)W_n(t)\} = [2^{-n} \cdot \min(s,t)]$$

for $0 \leq s,t \leq 1$. As $n \longrightarrow \infty$, the finite-dimensional distributions of W_n converge weakly to those of a family $\{W(t) : 0 \leq t \leq 1\}$ of jointly normal random variables with $EW(t) = 0$ and $E\{W(s)W(t)\} - \min(s,t)$. This follows because a sequence of normal random vectors is easily seen to converge in distribution if the corresponding sequences of means and covariances converge. The limiting process W is called the **Wiener process**, and Theorem 2.6 will be used to show that it is almost surely a continuous random function on $[0,1]$. Indeed, (iii) of Theorem 2.6 has just been observed to hold, and since $W_n(0) = 0$ for all n, (ii) is also obvious. Since the maximum jump-size $|W_n(t) - W_n(t-)|$ over $t \in [0,1]$ is by definition equal to $2^{-n/2} \max\{|Z_{j,n}| : 1 \leq j \leq 2^n\}$, the convergence in Remark 2.2 can be verified directly. To check (i) of Theorem 2.6, calculate for each $m \geq 0$ and $n \geq m$,

$$P\left\{\sup_{|s-t|<2^{-m}} |W_n(t) - W_n(s)| > 3\alpha\right\}$$

$$\leq \; P\left\{\max_{0 \leq j < 2^{-m}} \; \sup_{0 \leq t \leq 2^{-m}} \; |W_n(t) - W_n(2^{-m}j)| > \alpha\right\}$$

since one of $|W_n(s) - W_n(2^{-m}[2^m s])|$, $|W_n(2^{-m}[2^m t]) - W_n(2^m[2^m(s)])|$, or $|W_n(T) - W_n(2^{-m}[2^m t])|$ must be $> \alpha$ if $|W_n(t) - W_n(s)| > 3\alpha$. The last displayed probability is equal to

$$P\left\{2^{-n/2} \max_{0 \leq j < 2^m} \max_{1 \leq k \leq 2^{n-m}} \; |\textstyle\sum_{i=1}^k \; Z_{j+2^{n-m}+i,n}| > \alpha\right\}$$
$$\leq \; 2^M \left\{\max_{1 \leq k \leq 2^{n-m}} |Z_{1,n} + \cdots + Z_{k,n}| > \; \alpha \, 2^{n/2}\right\} \tag{2.7}$$

where the last step has of course used the fact that $Z_{j,n}$ are independent and identically distributed $N(0,1)$ random variables. When $n \geq m$, independence and joint symmetry of $Z_{j,n}$ imply for $\lambda \equiv \alpha \, 2^{n/2}$, $L \equiv 2^{n-m}$, and $S_k \equiv \sum_{i=1}^k Z_{i,n}^k$ for $k \geq 1$, that

$$P\{\max_{1 \leq k \leq L} |S_k| \geq \lambda\} = 2P\{\max_{1 \leq k \leq L} S_k \geq \lambda\} = 2\textstyle\sum_{k=1}^L P\{\max_{1 \leq j < k} S_j < \lambda \leq S_k\}$$

$$\leq \; 4 \textstyle\sum_{k=1}^L P\{\max_{1 \leq j < k} \; S_j < \lambda \leq S_k \leq S_L\} \; \leq \; 4\,P\{S_L \geq \lambda\}$$

Thus, when $n \geq m$, (2.7) is $\leq \; 2^{m+2}\,[1 - \Phi(\alpha\, 2^{n/2}/2^{(n-m)/2})]$. Now the well-known (Feller 1957, vol. 1, p. 166) tail-inequality

$$1 - \Phi(x) \; \leq \; (2\pi)^{-1/2}\, x^{-1}\, e^{-x^2/2} \qquad \text{for} \quad x > 0$$

implies that the probabilities (2.7) for $n \geq m$ are bounded by

$$C(\alpha)\; 2^{m/2}\; e^{-\alpha^2\, 2^{m-1}}$$

where $C(\alpha) > 0$ does not depend upon m. This last expression can be made arbitrarily small by choosing m large.

Thus (i)–(iii) of Theorem 2.6 hold for the processes W_n, and that theorem implies

$$W_n \xrightarrow{D} W \quad \text{in} \quad D[0,1] \quad \text{as} \quad n \longrightarrow \infty, \quad \text{and} \quad W \in C[0,1] \quad \text{a.s.}$$

The definition of Wiener process $W(\cdot)$ is summarized by

$$\left.\begin{aligned} W(0) \;\equiv\; 0 \text{ a.s., and each increment} \quad & W(t) - W(s) \quad \text{for} \quad 0 \le s \le t \le 1 \\ \text{is a} \qquad N(0, t-s) \quad & \text{r.v. independent of} \quad \{W(u) : 0 \le u \le s\}. \end{aligned}\right\} \tag{2.8}$$

As has just been proved, W is an almost surely continuous process. In addition, the random variables $W(t)$ have finite moment-generating functions, and

Proposition 2.8 *The Wiener process* $W(t)$ *for* $t \in [0,1]$ *as well as the processes* $W^2(t) - t$ *and* $e^{W(t) - t/2}$ *are martingales.*

Proof. The σ-algebra family implicit in the Proposition is defined by $F_t^W \equiv \sigma(W(u) : 0 \le u \le t)$. The independence in (2.8) says for $0 \le s \le t \le 1$ that for any continuous real-valued function γ of $W(t) - W(s)$,

$$E\{\gamma(W(t) - W(s)) \mid F_t^W\} \;=\; E\{\gamma(W(t) - W(s))\} \quad \text{a.s.}$$

and the normal distribution part of (2.8) says

$$E\gamma(W(t) - W(s)) = \begin{cases} 0 & \text{if } \gamma(x) \equiv x \\ t - s & \text{if } \gamma(x) \equiv x^2 \\ \exp\{\frac{1}{2}(t-s)\} & \text{if } \gamma(x) \equiv \exp(s) \end{cases}$$

Therefore, almost surely

$$E\{W(t) - W(s) \mid F_t^W\} = 0$$

$$E\{W^2(t) - W^2(s) \mid F_t^W\} = E\{W(t) - W(s))^2 + 2(W(s)W(t - W(s)) \mid F_t^w\} = t - s$$

$$E\{e^{W(t) - \frac{1}{2}t} - e^{W(s) - \frac{1}{2}s} \mid F_t^W\} = e^{W(s) - \frac{1}{2}t} - E\{e^{W(t) - W(s)} - e^{\frac{1}{2}(t-s)} \mid F_t^W\} = 0.$$

□

Exercise 4 *Let* $N_n(t)$, $0 \le t \le 1$, *be a Poisson counting process with cumulative rate* $n\wedge(t)$, *where* $\wedge(\cdot)$ *is a fixed continuous increasing function, and let* $X_n(t) \equiv n^{-\frac{1}{2}}(N_n(t) - n \wedge (t))$. *Prove that* X_n *converges in distribution in* D *as* $n \longrightarrow \infty$, *and describe the* *limit.* **Hint:** *imitate the tightness argument of this section, this time using the special properties of the Poisson distribution through the inequality*

$$P\{|X-\lambda| \geq a\} \leq \exp[\frac{1}{2}a^2\lambda], \quad \textit{where} \quad X \sim \textit{Poisson}\,(\lambda).$$

□

Note that our definition of Wiener process so far applies only to random functions on $[0,1]$. Here are two ways to extend to $[0,T]$, with $1 < T \leq \infty$. The simplest is to regard the random function W_0 in $C[0,1]$ as having been constructed and to define $W(t) \equiv T^{\frac{1}{2}} W_0(t/T)$ for $0 \leq t \leq T$.

Alternatively, one can regard independent Wiener processes $W_1, W_2, \ldots$ as having been constructed on $[0,1]$, and define

$$W(t) \equiv \sum_{j=1}^{n} W_j(1) + W_{n+1}(t-n) \quad \text{for} \quad 0 \leq n \leq t < n+1.$$

The reader should verify that these two definitions of Wiener process on $[0,\infty)$ both yield processes W with jointly normal finite-dimensional distributions, $EW(t) \equiv 0$, and $EW(s)W(t) = s,t$ for $0 \leq s,t < T$.

2.4 Examples of Statistical Uses for Weak Convergence

2.4.1 *Sequential Hypothesis Test or Test Based on Boundary-Crossing.*

Suppose that a test statistic $U_n(t)$ is defined for each time $t \in [0,T]$ in terms of the information observed up to time t based on an experimental sample of potential size n. Suppose also that the convergence $U_n \longrightarrow U$ in $D[0,T]$ for some random continuous function U can be established under a null hypothesis H_0 as the sample-size parameter n goes to ∞. Then for any fixed continuous function $b(\cdot) \geq \beta > 0$ on $[0,T]$, the hypothesis test of H_0 with

$$\text{rejection region } \equiv \left\{ \sup_{0 \leq t \leq T} \mid U_n(t)/b(t) \geq 1 \right\}.$$

will have type-I error-probability α approximately equal to

$$P(\sup_{0 \leq t \leq T]} |U(t)/b(t)| \geq 1)$$

This follows by Theorem 2.2) and continuity of the functional γ which sends $f \in D$ to $\|f(\cdot)/b(\cdot)\|$, as long as the distribution of the random variable $\gamma(U)$ has no probability atoms.

Suppose that $U_n(t)$ is given by $n^{\frac{1}{2}}[F_n(t)]$, where $F_n(\cdot)$ is the empirical distribution function (defined in Exercise 2 of Chapter 1) for independent and identically distributed random variables $T_1, \ldots, T_n$ which have the continuous null-hypothetical distribution function F. If $b(\cdot) \equiv b$ is a positive constant, then the boundary-crossing test just described is the well-known one-sample Kolmogorov-Smirnov test of goodness fit of $\{T_1 : 1 \le i \le n\}$ to F.

2.4.2 *Sampling Up to a Random Time.*

Consider a survival experiment such as the two-sample experiment described in Section 1.5. We may wish to terminate the experiment not at a fixed time but, either for ethical or economic reasons, at time τ depending on observed data. Suppose, for example, that we intend to terminate at the time of the $[cn]$'th observed death, where $0 < c \le 1$ is fixed, or more generally at a random time τ_n about which we know that $\tau_n \xrightarrow{P} t_0$ for some nonrandom constant t_0 under a null-hypothetical model as the sample-size parameter n goes to ∞. Even if the asymptotic distribution as $n \longrightarrow \infty$ of $U_n(t_0)$ were known, one ordinarily needs to know that U_n converges in D, to a continuous limiting random function U, in order to conclude that $U_n(\tau_n) - U_n(t_0) \xrightarrow{P} 0$ as $n \longrightarrow \infty$. In such a case both $U_n(\tau_0)$ and $U_n(t_0)$ have limiting distributions equal to the distribution of $U(t_0)$. To prove this, note that if $U_n \xrightarrow{D} U$ in $D[0,T]$ with $0 < t_0 < T$, then the converse direction of Theorem 2.6 yields property (i) for $\{U_n\}$. For arbitrary $\alpha > 0$, first choose $\delta > 0$ from (i) so small that

$$P\left\{\sup_{|s-t|<\delta} | U_n(s) - U_n(t) > \alpha\right\} < \frac{1}{2}\alpha \quad \text{for all} \quad n.$$

Next choose n so large that $P\{|T_n - t_0| \ge \delta\} < \frac{1}{2}\alpha$. For all larger n, $P\{|U_n(t_0) - U_n(\tau_n)| > \alpha\}$, so that $U_n(\tau_n) - U_n(t_0) \xrightarrow{P} 0$ as $n \longrightarrow \infty$.

2.4.3 *Asymptotic Distribution of Level-Crossing Times. [adapted from Brookmeyer and Crowley, Appendix 1]*

Suppose that $S_n(\cdot)$ estimates the function $S(t) \equiv P\{X > t\}$ of a survival-time random variable X on $[0, \tau_0]$, based on a data-sample of size n. Assume that as $n \longrightarrow \infty$, it is known that

(a) $m \equiv S^{-1}(\frac{1}{2}) < \tau_0$ is a point of left and right decrease for $S(\cdot)$, i.e., for each small enough $\delta > 0$, $S(m - \delta) > \frac{1}{2} > S(m + \delta)$;

(b) $n^{\frac{1}{2}}(S_n - S) \xrightarrow{D}$ some random function Y in $C[0, \tau_0]$;

(c) if $m_n \equiv S_n^{-1}(\frac{1}{2} \equiv \inf\{x : S_n(x) \leq \frac{1}{2}\}$, then $n^{\frac{1}{2}}\left(S(m_n) - \frac{1}{2}\right) \longrightarrow 0$.

Then it is not hard to check that $m_n \xrightarrow{P} m$, and by (b) and the result proved in 2.2.2 above,

$$n^{\frac{1}{2}}\left(S(m_n) - S(m_n) - S_n(m) + S(m_n)\right) \xrightarrow{P} \infty$$

from which it follows that $n^{\frac{1}{2}}(\frac{1}{2} - S(m_n)) \longrightarrow -Y(m)$ as real random variables.

If S is differentiable at m, with $S\prime(m) < 0$, this gives the asymptotic distribution of $n^{\frac{1}{2}}(m_n - m)$ by the "delta method" (Rao 1973, pp. 385–388). There are two points of view from which such a result is interesting: one can think of m either as the crossing time by the curve S of the level $\frac{1}{2}$, or of m_n as a generalized median-estimator.

2.5 References

1. Billingsley, P. (1968) *Convergence of Probability Measures* (Chapter 1).

2. Brookmeyer, R. and Crowley, J. (1982) A K-sample median test for censored data". *J.A.S.A.* **77**, 433–440.

3. Prohorov, Yu. (1956) Convergence of random processes and limit theorems in probability theory. *Theory Probability Appl.* **1**, 157–214.

4. Stroock, D. and Varadhan, S. (1973) *Topics in Probability Theory.* NYU Seminar Notes (§3).

Chapter 3

Tools From the Theory of Martingales, I

This chapter supplies definitions, basic properties, and distribution inequalities for (sub-) martingale stochastic processes $\{M(t) : a \leq t \leq b\}$. It has been shown in Chapter 1 that counting process $\{N(t), t \geq 0\}$ have naturally associated "compensated" martingales, and in Chapter 2 that the Wiener process $W(t)$ as well as $W^2(t) - t$ and $exp\{W(t) - \frac{1}{2}t\}$ are martingales for $t \geq 0$. The class of statistically interesting martingales will be dramatically enlarged with the introduction in Chapter 5 of stochastic integrals. The primary focus of this chapter is to construct via pointwise limiting operations the compensators of increasing adapted stochastic processes adapted to a filtration, including the predictable-variation and quadratic-variation processes associated with a large class of locally square-integrable martingales. As a result, it is shown for the key examples which arise in this book, how compensators and variation processes are calculated and what the formulas mean intuitively.

3.1 Basic Properties and Inequalities

Recall that for a random variable $X : (\Omega, F, P) \mapsto \mathbb{R}$ and a sub-σ-algebra $G \subset F$, the conditional expectation $E(X|G)$ is characterized almost surely by:

$$\int Y E\{X|G\} dP = \int Y\, X\, dP$$

for all bounded $G-$measurable random variables Y. The following two basic properties of conditional expectation will be used frequently and without further comment, but see Appendix I for references:

Repeated conditioning: if $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ are sub-σ-algebras and Z is a bounded $\mathcal{G}$-measurable random variable, then

$$E(Z \cdot E(X|\mathcal{G}) \,|\, \mathcal{H}) \;=, E(ZX \,|\, \mathcal{H}) \tag{3.1}$$

Conditional Jensen inequality: if $\gamma : \mathbb{R} \longmapsto \mathbb{R}$ is convex and $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$, then

$$E(\gamma(X) \,|\, \mathcal{G}) \;\geq\; \gamma(E(X|\mathcal{G}) \tag{3.2}$$

A particular consequence of (3.2) is that $E(E(X|\mathcal{G}))^2 \leq E(X^2)$, while if $E(X^2) < \infty$, then (3.1) readily implies

$$E(\, E(X|\mathcal{G}) \,\cdot\, \{X - E(X|\mathcal{G})\}) \;=\; 0$$

These two comments yield the very useful corollary:

$$\text{if} \quad EX^2 < \infty, \qquad \text{then} \quad EX^2 \;=\; E(E(X|\mathcal{G}))^2 \,+\, E(X - E(X|\mathcal{G}))^2 \tag{3.3}$$

A stochastic process $\{M(t) :\, a \leq t \leq b\}$, which is almost surely right-continuous, with limits from the left (i.e., is a random function almost surely in $D[a, b]$, is said to be **adapted** to an **increasing family** or **filtration** $\{\mathcal{F}_t :\, a \leq t \leq b\}$ of sub-σ- algebras of $\mathcal{F}$ if for each t, $M(t)$ is $\mathcal{F}_t$ measurable, and whenever $s \leq t$, $\mathcal{F}_s \subset \mathcal{F}_t$. Such a process $M(\cdot)$ is called a **martingale** [respectively **submartingale**] if each $M(t)$ is integrable and for each $s < t$, $E(M(t)|\mathcal{F}_s) = M(s)$ almost surely [respectively, $E(M(t)|\mathcal{F}_s) \geq M(s)$ a.s.] Whenever the increasing family $\mathcal{F}_t$ is not explicitly mentioned, the σ- algebras $\mathcal{F}_t \,\equiv\, \sigma(M(s) :\, s \leq t)$ will be understood.

Remark 3.1 Any right-continuous process $M(\cdot)$ of integrable random variables which is almost surely nondecreasing is a submartingale. Note also that if either $M(\cdot)$ is a martingale and γ any convex function on $\mathbb{R}$, or if M is a submartingale and γ a nondecreasing convex function, then (3.2) immediately implies that $\{\gamma \circ M(t) :\, a \leq t \leq b\}$ is a submartingale.

The following two fundamental theorems are stated without proof, but all further developments to be based upon them are self-contained. [See Appendix I for discussion of uniform integrability.]

Submartingale Convergence Theorem [Doob 1953, pp. 324-5,358]. If M is a uniformly integrable submartingale adapted to $\{\mathcal{F}_t\}$ on $(a, b) \subset \mathbb{R}$, then there exist random variables

$M(b-)$ and $M(a+)$ to which $M(t)$ converges a.s. and in the mean, respectively as $t \uparrow b$ and as $t \downarrow a$. In addition,

$$E\{M(b-)|\mathcal{F}_t\} \geq M(t) \quad \text{a.s.foreach} \quad t \in (a,b)$$

with equality if M is a martingale.

Burkholder Inequalities [Burkholder 1973, pp. 21-3]. For each $p > 1$, there are constants $C_p,\, c_p$ such that, if $\{M(t) : t \in [a,b]\}$ is a martingale and $\{\alpha_j,\, \beta_j)\}_{j=1}^m$ is a finite system of disjoint subintervals of $[a,b]$, then

$$\begin{aligned} \frac{1}{C_p}\, E\left|\sum_{j=1}^m (M(\beta_j) - M(\alpha_j))\right|^p &\leq E\left[\sum_{j=1}^m (M(\beta_j) - M(\alpha_j))\right]^{\frac{1}{2}p} \\ &\leq c_p\, E\left|\sum_{j=1}^m (M(\beta_j) - M(\alpha_j))\right|^p . \end{aligned} \tag{3.4}$$

These Burkholder inequalities are essentially discrete-time results. An extension due to Millar for continuous-time martingales, in which the p^{th} absolute moment of $M(t) - M(a)$ for martingale $M(\cdot)$ is bounded in terms of the $\frac{1}{2}p^{th}$ moment for the quadratic-variation process $[M](t) - [M](a)$, will be proved in Chapter 4.

One of the important uses of the (sub-) martingale property is the maximal inequality of Doob, which generalized the famous Kolmogorov Inequality (Loève 1955, p. 235) for tail probabilities of the largest partial sum of a sequence of independent random variables $\{X_i : 1 \leq i \leq n\}$.

Development of Doob's inequality requires the concept of *stopping time.* A **stopping time** τ with respect to an increasing family of σ-algebras $\{\mathcal{F}_t : a \leq t \leq b\}$ is a random variable with values in $[a,b]$ such that the event $[\tau \leq t]$ is $\mathcal{F}_t$-measurable for each t. Then

$$[\tau < t] = \bigcup_{\{(s<t,\, s\in\mathbb{Q}:\, \tau\leq s)\}}$$

and $[\tau = t]$ are also $\mathcal{F}_t$ measurable. It is an easy and worthwhile exercise to show that if τ and σ are each stopping times with respect to $\mathcal{F}_t$, then so are

$$\tau \wedge \sigma \equiv \min(\tau,\sigma) \qquad \text{and} \qquad \tau v \sigma \equiv \max(\tau,\sigma)$$

Theorem 3.1 (Submartingale Maximal Inequality, Doob 1953, pp. 317-8) . *If* $\{M(t) : a \leq t \leq b\}$ *is either a martingale or a nonnegative submartingale, and if* $p > 1$, *then*

$$E\left[\sup_{a\leq t\leq b} |M(t)|^p\right] \leq (p/(p-1))^p E|M(t)|^p .$$

Proof. In the martingale case, $|M|$ is a nonnegative submartingale by Remark 3.1. Hence, we treat only the submartingale case and drop $|\ \ |$. It will be enough to show, for all $\{t_1, \cdots, t_n\} \subset (a, b)$ and $L > 0$, that

$$E[min\{L, \max_{1\le i\le n} M(T_i)\}]^p \le (p/(p-1)))^p EM^p(b) \tag{3.5}$$

since the assertion of the Theorem then follows by the Monotone Convergence Theorem upon letting $L \longrightarrow \infty$ and $\{t_i\}$ increase to a dense set in $[a, b]$. Here the max tends to the sup because $M \in D[a, b]$ almost surely.

Fix L and $\{t_i : i = 1, \cdots, n\}$ and $\beta > 0$, and let $0 < a \le L$. Define

$$Y \equiv \min\{L, \max_{1\le i\le n} M(t_i)\} \ \ and \ \ \tau \ \equiv \begin{cases} \min\{t_i \le b : M(t_i) \ge \beta\} \\ \\ 0 \ \text{ if no such } t_i \text{ exists.} \end{cases}$$

Then by definition and (3.1), $E(M(b) - M(t_i)) =$

$$E\sum_i I_{[\tau=t_i]}(M(b) - M(t_i)) = \sum_i E\left[I_{[\tau=t_i]}(M(b) - M(t_i))|\mathcal{F}_{t_i}\}\right]$$

each summand of which is ≥ 0 by the submartingale property. Therefore,

$$EM(b) \ge EM(\tau) = E(I_{[Y\ge\beta]}(M(\tau)) + E(I_{[Y<\beta]}M(b)).$$

Since $E(I_{[Y\ge\beta]} \cdot (M(\tau)) \ge BI_{[Y\ge\beta]}$, it follows that

$$\int_{[Y\ge B]} M(b)dP = E(M(b)(1 - I_{[Y<\beta]})) \ge BP\{Y \ge B\}.$$

Next, integrating by parts and substituting the last inequality gives

$$EY^p = \int_0^L P\{Y > B)d(\beta^p) \le \int \beta^{-1} \int_{[Y\ge\beta]}$$

which by the Fubini-Tonelli Theorem is equal to

$$\begin{aligned} \textstyle\int M(b) \int_0^Y p\beta^{p-2} d\beta dP &= \tfrac{1}{p-1} E(Y^{p-1}M(b)) \\ &\le (p/(p-1))(E(M(b)^p)^{1/p} \\ &\quad\ (EY^p)^{(p-1)/p} \text{ Hölder's inequality, (A.)} \end{aligned}$$

Divide through by $(EY^p)^{(p-1)/p}$ and raise both sides of the resulting inequality to the power p to complete the proof. □

Remark 3.2 *The first part of the proof of Theorem 3.1 shows that if* M *on* $[a, b]$ *is a* $\{\mathcal{F}_t\}$ *submartingale and* τ *is a* $\{\mathcal{F}_t\}$ *stopping timetaking values almost surely in a discrete (nonrandom) subset of* $[a, t]$, *where* $t < b$, *then* $EM(t) \ge EM(\tau)$. *The restriction to discrete-valued* τ *will be removed in Lemma 3.1.*

A useful corollary of Theorem 3.1 is that martingales M on $[a,b]$ with $EM^2(t)$ uniformly bounded, have $M^2(t)$ uniformly integrable.

Corollary 3.2 *(i) If* $\{M(t) : a \le y \le n\}$ *is either a martingale or nonnegative submartingale, then it is a uniformly integrable family of random variables with* $\{\sup_t |M(t)| \ge \beta\} \le \beta^{-1}EM(b)$. *(ii) If* $\{M(t) : a \le y \le n\}$ *is uniformly integrable, where* $M(B) \equiv M(b-)$.

Proof. The inequality of (i) was proved as part of the proof of Theorem 3.1. Then respectively with $k=1$ or 2 in (i), (ii), the submartingale property of $|M(t)|^k$ implies

$$E|M(t)|^k I_{[|M(t)| \ge \beta]} \le E|M(b)|^k I_{[|M(t)| \ge \beta]} \le E|M(b)|^k I_{[\sup_t |M(t)| \ge \beta]}$$

which by integrability of $|M(t)|^k$ becomes arbitrarily small as β becomes large since the probability of the last event becomes small. (See Appendix (B.4) if further justification is needed.) In (ii), the Martingale Convergence Theorem says that $M(b) \equiv M(b-)$ exists a.s. Theorem 3.1 and the Dominated Convergence Theorem tell that as t increases to b, $[M(t) - M(b-)]^2$ being dominated must converge in mean square to 0, and $EM^2(t)$ converges to $EM^2(b-)$.

Because of this corollary, any martingale M defined initially on $[a,b]$ for which $\sup\{EM^2(t) : t \in (a,b)\} < \infty$ is called a **square-integrable martingale**. The conventional definitions $M(a) \equiv M(a+)$ and $M(b) \equiv M(b-)$, allowed by the Martingale Convergence Theorem in case M had not been defined at a or b, will apply throughout this book. From now on, the notations $x \wedge y \equiv \min(x,y)$ and $xvy \equiv \max(x,y)$ will also be used freely.

Example. (Compound Renewal Process martingales) Since martingale theory was in large part invented to generalize the fruitful theory of partial sums of independent random variables, it is instructive for the reader not versed in probability theory to specialize each martingale related theorem to the following case.

Let $X \equiv \{X_i : i = 1, 2, \cdots\}$ be a sequence of independent integrable random variables; let $Y \equiv \{Y_j : j = 1, 2, \cdots\}$ be an independent sequence, independent of X, of positive random variables; and define for $t \ge 0$,

$$\begin{aligned} M(t) &\equiv \textstyle\sum_{i=1}^{\infty} I_{[Y_1 + \cdots + Y_j \le t]} X_j \\ \mathcal{F}_t &= \sigma(Y_j, X_j : j \ge 1, Y_1 + \cdots + Y_j \le t). \end{aligned} \tag{3.6}$$

If one imagines the increment X_j to be included in a cumulative sum at the observation-time $Y_1 + \cdots + Y_j$, then $M(t)$ is the accumulated value of the sum up to time t. The times of successive increments form a renewal point-process on the half-line (Karlin and Taylor 1975, Chapter 5), and the integrability for each finite t of

$$N(t) \equiv \max\{j \geq 0 : \; Y_1 + \cdots + Y_j \leq t$$

(Karlin and Taylor 1975, pp. 181-2) ensures that $N(t)$ is almost surely finite and thus that $M(t)$ is almost surely well-defined. If the random variables X_i were all equal to 1, then $M \equiv N$ would be a **renewal counting process**. If the Y_j were all 1, then M would be a random walk. The class of M defined in (3.6) is that of **compound renewal processes**, which has many important applications (Karlin and Taylor 1975, Feller 1971, pp. 180-90). For example, in actuarial (i.e., insurance) problems X_i represents the amount of an insurance claim at the death- or accident-occurrence time $Y_1 + \cdots + Y_i$. Actuarial applications are discussed further in Sections 4.4.2 and 5.8. In Remark 5.2, $M(t)$ is interpreted as a stochastic integral.

Now assume that all the expectations $EX_i = 0$. By independence of X_j from $\{X_i : 1 \leq i < j\} \cup \{Y_i : 1 \leq i \leq j\}$, it is easy to see that $E(X_j \, I_{[s<Y_1+\cdots+Y_j \leq t]} \,|\, \mathcal{F}_s) = 0$. By integrability of $N(t)$, M is a martingale on $[0, \infty)$. The inequalities of the present section specialize to say about M that for each $p > 1$, with $C'_p \equiv (p/(p-1))^p \, C_p$

$$\begin{array}{lcl} E \max_{1 \leq m \leq n} |\sum_{k=1}^m X_k|^p & \leq & \left(\frac{p}{p-1}\right)^p E \, |\sum_{k=1}^m X_k|^p \leq C'{}_p E \left(\sum_{k=1}^n X_k X_k^2\right)^{\frac{1}{2}p} \\ E \sup_{0 \leq t \leq T} |M(t)^p & \leq & (p/p-1))^p E|M(T)|^p \leq C'_p E \left(\sum^{N(T)_{k=1}} X_k X_k^2\right)^{\frac{1}{2}p} \end{array} \tag{3.7}$$

The last inequality is obtained by applying Burkholder's inequality to $M(T)$ decomposed into the sum of increments $M(Y_1 \wedge T)$, $M(Y_2 \wedge T) - M(Y_1 \wedge T)$, $\cdots$, $M(Y_m \wedge T) - M(Y_{m-1} \wedge T)$, $M(T) - M(Y_m \wedge T)$, each of which has conditional expectation 0 given the previous ones, and then letting $m \longrightarrow \infty$. Interesting information is obtained from (3.7) when X_i takes the values ± 1 with equal probabilities $\frac{1}{2}$. These inequalities then tell that the p^{th} moments of $\max_{1 \leq m \leq n} |X_1 + \cdots + X_M|$ and $\sup_{0 \leq t \leq T} |M(t)|$ are bounded by the constant C'_p respectively multiplying $n\frac{1}{2}p$ and $E\{N(T)\}^{\frac{1}{2}p}$.

Exercise 5 *Fix an integer* n, *and suppose* N *is a Poisson*(n) *random variable, independent of an independent and identically distributed sequence of random variables* $\{X_i\}_{i=1}^\infty$ *with distribution function* F *on* $\mathbb{R}$. *Define* $M(t) \equiv \sum_{j=1}^N I_{[X_j \leq t]} - nF(t)$ *and* $\mathcal{F}_t \equiv \sigma(M(s) : s \leq t)$. *Show that* M *is a* $\{\mathcal{F}_t\}$ *martingale.*

Hint: $\sum_{j=1}^N I_{[X_j \leq t]}$ *is a homogeneous Poisson process on* $\mathbb{R}$.

3.2 Local Martingales

A local (sub-) martingale is a stochastic process $\{M(t), t \in [a,b])\}$ adapted to a $\sigma-$algebra family $\{\mathcal{F}_t\}$ with respect to which there exists a sequence $\{\tau_n\}_{n\geq 1}$ of stopping times such that $\tau_n \wedge b$ increases almost surely to b as $n \longrightarrow \infty$, and such that for each $n \geq 1$ the process $M(\cdot \wedge \tau_n)$ is a (sub-) martingaleadapted to $\{\mathcal{F}_t\}$ on $[a,b)$. Then M is said to be locally square-integrable if each of the (sub-) martingales $M(\cdot \wedge \tau_n)$ is square-integrable, that is, if for each n, $\sup\{EM^2(t \wedge t_n) : t \in [a,b)\} < \infty$. It is useful to know that (sub-) martingales M on $[a,b)$ are local (sub-) martingales with respect to any sequence $\{\tau_n\}$ of stopping times increasing to b . That is the central assertion of the following Lemma.

Lemma 3.1 *If* $\{X(t), t \in [a,b)\}$ *is an almost surely right-continuous* $\{\mathcal{F}_t\}$ *adapted submartingale, then for each* $a \leq s \leq t < b$ *and stopping time* τ*,*

$$E\{X(t \wedge \tau\}) \,|\, \mathcal{F}_s \geq X(s \wedge \tau) \qquad a.s.$$

and for each constant c, $X_{\tau,c}(\cdot) \equiv \max\{c, X(\cdot \wedge \tau)\}$ *is a uniformly integrable submartingaleon* $[a,t]$.

Proof. For uniform integrability, we follow Chow, Robbins, and Siegmund (1971, pp. 14*ff*). For $s < t$ and $n \geq 1$, let $x_j \equiv s + 2^{-n}(t-s)j$ for $j = 0, 1, 2, \cdots$ and let $\tau(n) \equiv \inf(x_j : j \geq 0, \tau \geq x\}$. Then $\tau(n)$ is a stopping time, and as $n \longrightarrow \infty$, $\tau(n)$ decreases almost surely to τ. If $a \leq u < w \leq b$, then (3.1) and the $\mathcal{F}_{x_j}$ measurability of the event $[x_j < w \wedge \tau]$ imply that

$$\begin{aligned} & E\{X(w \wedge \tau(n)) - X(u \wedge \tau(n))\})|\mathcal{F}_u\} \\ = \;& E\left\{\textstyle\sum_j (X(w \wedge x_{j+1} \wedge \tau(n)) - X(\max\{u, x_j\} \wedge \tau(n))|\mathcal{F}_u\right\} \\ = \;& E\left(\textstyle\sum_j I_{[x_{j+1} \geq u,\, x_j < w \wedge \tau]} E\{X(w \wedge x_{j+1}) - X(uvx_j)|\mathcal{F}_{uvx_j}\}|\mathcal{F}_u\right) \end{aligned}$$

which is ≥ 0 by the submartingaleproperty of X. Thus $X(\cdot \wedge \tau(n))$ is a submartingale for each n, and by Remark 3.1 applied to the increasing convex function $\gamma(x) = \max\{c, x\}$,

$$X_{\tau(n),c}(\cdot) \equiv \max\{c, X(\min\{\cdot, \tau(n)\})\}$$

is a $\{\mathcal{F}_t\}$ submartingale, $t \in [a,b)$. For each real β and $u \in [a,t]$, if $B \equiv B(n,c,u,\beta)$ denotes the $\mathcal{F}_u$ measurable event $[X_{\tau(n),c}(u) > \beta]$, then

$$\begin{aligned} \beta P[X_{\tau(n),c}(u) > \beta] &= P\{B\} < \textstyle\int_B X_{\tau(n),c}(u)dP \leq \int_B X_{\tau(n),c}(t)dP \\ &= \textstyle\int_B \max\{c, X(t)\}dP. \end{aligned}$$

The two displayed inequalities respectively express the submartingale property of $X_{\tau(n),c}$ and of $\max\{c, X(\cdot)\}$.

Therefore, when β is large, $P\{X_{\tau_n,c}(u) > \beta\}$ is uniformly small for all n and all $u \in [a,t]$. The family $X_{\tau_n,c}(u) : n \geq 1, u \in [a,t]\}$ has now been shown to be uniformly integrable. The decrease of $\tau(n)$ to τ and the right-continuity of X imply $X_{\tau_n,c}(u) \to X_{\tau,c}(u)$ as $n \longrightarrow \infty$. It follows for each c (cf. Appendix (B.4)) that for $u \in [s,t]$

$$E\{X_{\tau,c}(u) \,|\, \mathcal{F}_s\} = \lim_{n\longrightarrow\infty} E\{X_{\tau_n,c}(u) \,|\, \mathcal{F}_s\} \geq \lim_{n\longrightarrow\infty} X_{\tau_n,c}(s) = X_{\tau,c}(s).$$

When $c \longrightarrow \infty$, the Monotone convergence Theorem yields $E\{X_t \wedge \tau \,|\, \mathcal{F}_s\} \geq X(s \wedge \tau)$ almost surely. Thus $X(\cdot \wedge \tau)$ and $X_{\tau,c}$ are submartingales. The uniform integrability of $\{X_{\tau,c}(s) : s \in [a,t]\}$ follows because, as $n \longrightarrow \infty$,

$$E\{I_{[|X_{tau_n,c}(u)| \geq \beta]} \,|X_{\tau(n)}(u)|\} \to E\{I_{|X_{\tau_n,c}(u)| \geq \beta]} \,|X_{\tau,c}(u)|\}$$

The simplest example of a local submartingale which is not already a submartingale arises as a counting process. Suppose that, conditionally given the value of a random variable Y, N is a Poisson counting process(Sec. 1.3) with cumulative intensity function $E\{N(t)|Y\} = Y\Lambda_0(t)$, where $\Lambda_0(\cdot)$ is a nonrandom increasing continuous function with $\Lambda_0(0) = 0$ and $\Lambda_0(\infty) = \infty$. Then $P\{N(t) < \infty|Y\} = 1$ for each $t < \infty$, almost surely with respect to Y, which implies also $P\{N(t) < \infty\} = 1$, and N is a simple counting processas described in Section 1.4. Counting processes like this, which are conditionally Poisson given a random cumulative intensity function Λ, are called **doubly stochastic Poisson processes** and have been advocated by Cox (195?) as models for the clustering of random occurrences. In our example, $EN(t) < \infty$ for finite t with intensity $A(t) < \infty$ if and only if $EY < \infty$. Whether expectations are finite or not, the process N is a locally square-integrable submartingale, with the 'localizing' sequence of stopping times $\tau_n = \inf\{t > 0 : N(t) = n\}$. The processes $N(\cdot \wedge \tau_n)$ are submartingales because they are increasing, and square-integrable because they are uniformly bounded by n.

In the preceding paragraph, the technical device of introducing a localizing sequence τ_n had the effect of restricting attention to only the first n occurrences in an unfolding random experiment. This idea suggests the possibility of measuring time for a local martingale by means of some increasing process associated with it. The predictable-variance or cumulative-conditional-variance process to be constructed in the next section serves as such an 'operational time' for a martingale. The main benefit of treating martingales through their operational time- scales is to relate their behavior to a standard form, as expressed for example by the following heuristic principle:

If M is a locally square-integrable martingale, continuous or with small jumps, and $V(\cdot)$ is its predictable-variance, then the large-scale distributional behavior of the graph of $(V(t), M(t))$ is like that of the graph of $(s, W(s))$ for a Wiener process W.

Important general theorems justifying and applying (3.2) are proved in Section 3.4 and Chapter 4.

3.3 Constructive Doob-Meyer Decomposition

We develop in this section for a large class of submartingales a more general analog of the compensator A and variance-process V associated with counting processes N as introduced in Chapter 1. The main result is the **Doob-Meyer Decomposition** (Lipster and Shiryayev 1977, vol. 1, Chap. 3). The approach here is to obtain the compensator as a limit in probability of a sequence of processes defined directly in terms of a given submartingale via conditional-expectation operations. Although this approach restricts slightly the class of submartingales for which we prove the Decomposition, the class is still ample for statistical applications. Since the need to pass to equivalent versions of processes will be avoided entirely, the proofs require less complicated measure theory than in presentations done in full generality.

The key idea of the following theorem is to treat continuous-time submartingales by discretizing time finely but nonrandomly. See Helland (1982, pp. 86-7) for a good survey of techniques and counterexamples related to this idea. For later reference, we define the concept of discretizing time by sequence of stopping times. A **partition sequence** $\{Q(k)_{k\geq q} \equiv \{t_{jk} : j \geq 0\}_{k\geq q}$ of subdivisions of $[0, T)$ adapted to $\mathcal{F}_t$, for $T \leq \infty$, is defined to be a doubly indexed set of $\mathcal{F}_t$ stopping times t_{jk} such that $t_{0k} \equiv 0$ almost surely, and

$$\left.\begin{array}{rl} (i) & t_{j+1,k} \leq t_{jk} \text{ a.s., and for each } k, \;\; t_{jk} \uparrow T \;\; as \;\; j \longrightarrow \infty \\ (ii) & t_{jk} \in Q(k+1) \equiv \{t_{i,k+1} : i = 0, 1, 2, \cdots\}, \;\; all \;\; j \geq 0, k \geq 1 \\ (iii) & as \;\; \text{k} \longrightarrow \infty, \\ & \text{mesh} \;\; Q(k) \equiv \max\{t_{j+j1,k} - t_{jk} : j \geq 0, t_{jk} < T\} \xrightarrow{P} \;\; 0 \end{array}\right\} \tag{3.8}$$

We regard $Q(k) \equiv \{t_{jk}\}_j$ as partitioning $[0, T)$ into the system of disjoint random intervals $[t_{jk}, t_{j+1,k})$. Condition (i) says for each $t > T$ that at most finitely many such intervals intersect $[0, T]$; (ii) says that the partitions $Q(k)$ are *nested* in the sense that $Q(k) \subset Q(k+1)$; and (iii) says the partitions become arbitrarily fine as the index k increases.

Theorem 3.3 *For fixed* $T \leq \infty$, *let* $(X(t), t \in [0, T])$ *be a* $\mathcal{F}_t$ *submartingale, with*

$X(0) = 0$ and $E\sup_{t\le T} X^2(t) < \infty$, such that

There is a nondecreasing right-continuous $\mathcal{F}_0$ measurable random function $\Lambda(\cdot)$ on $[0,T)$ with $\Lambda(0)=0$, and a positive adapted process h with $E\int_0^t h\,d\Lambda < \infty$ for $t<T$, such that for all nonrandom times $0 \le s < t < T$,

$$E\left\{\sum_{u\in(s,t)} \Delta X(u)\,|\,\mathcal{F}_s\right\} \le \int_s^T E\{h(u)|\mathcal{F}_s\}\,d\Lambda(u). \tag{3.9}$$

Then the random variables

$$A_k(t) \;=\; \sum_{j\ge 0} I_{[t_{jk}\le t]}\, E\{X(t\wedge t_{j+1,k}) - X(t_{jk})\,|\,\mathcal{F}_{t_{jk}}\}$$

converge in probability for each $t<T$ to a random variable $A(t)$ measurable with respect to the increasing family $\mathcal{F}_t$ of σ-algebras. The stochastic process $(A(t), t\in[0,T))$, called the **compensator** *of X, is nondecreasing and right-continuous almost surely, and $X-A$ is a $\{\mathcal{F}_t\}$ martingale.*

Remark 3.3 *In the terminology of Brown (1978), the compensator $A(\cdot)$ is called calculable if for each t, $A(t)$ is the limit in probability of $A_k(t)$. Accordingly, what we show here is that submartingales with square-integrable suprema, with sums of absolute jumps square-integrable up to each finite t, and which satisfy (3.9), have* **calculable compensators**. *Observe that for submartingales X which are nonnegative, the hypothesis $E\sup_s X^2(s) < \infty$ is no more restrictive than $EX^2(T) < \infty$, by Remark 3.1 and Theorem 3.1.*

Remark 3.4 *The class of processes with absolutely summable jumps can also be understood as the right-continuous processes with left limits which can be obtained as the sum of a continuous process with one of locally bounded variation over any nonrandom partition-sequence $\{Q(k)\}$, i.e., such that for $t<T$,*

$$\sup_k \sum_j I_{[t_{jk}<t]}\,|X(t_{j+1,k}) - X(t_{j+1})| < \infty \quad a.s. \tag{3.10}$$

For any right-continuous process X with left limits, recall the notation $\Delta X(s) \equiv X(s) - X(s-)$. If $\sum_{s\le t} |\Delta X(s)| < \infty$ almost surely at each $t<T$, then the process $U(t) \equiv \sum_{s\le t} \Delta X(s)$ is by definition right-continuous with locally bounded variation, and $X-U$ is almost surely continuous. The continuous processes arising in applications are typically derived from either the Wiener process or the continuous compensators of counting processes.

Proof of Theorem 3.3. The proof steps are numbered for easy reference.

(1) (*Doob Decomposition*). The sequence $\{A_k(t_{jk})\}_j$ is characterized uniquely for each k by the properties that $A_k(0) \equiv 0$, that $A_k(t_{j+1,k})$ is $\mathcal{F}_{t_{jk}}$ measurable for each $j \geq 0$, and that for $t_{jk} < T$, $X(t_{jk}) - A_k(t_{jk})$ is a discrete-time martingale sequence with respect to $\{\mathcal{F}_{t_{jk}}\}_j$. This is easy to check through the formula

$$X(t) - A_k(t) \;=\; \sum_j I_{[t_{jk}<t]}\,\big(X(t \wedge t_{j+1,k}) - E\{X(t \wedge t_{j+1,k})\,|\,\mathcal{F}_{t_{jk}}\}\big) \tag{3.11}$$

The definition of A_k and the submartingale property for X immediately imply that $A_k(t) \geq A_k(t_{jk})$ almost surely for each j, k, and $t_{jk} \leq t \leq t_{j+1,k}$. By definition, A_k is almost surely right-continuous and adapted to $\{\mathcal{F}_{t-}, t \in \{0,T]\}$.

(2) With the object of examining the convergence of $A_k(t)$ as $k \longrightarrow \infty$, we find in this step an upper bound for all expectations $E((A_k(t_{jk}) - A_m(t_{jk}))^2)$ when integers $k < m$ and j are fixed. Throughout this proof-step, we denote for $r = k$ or m and for all i,

$$z_{ir} \;\equiv\; E\{X(t_{i+1,r}) - X(t_{ir})\,|\,\mathcal{F}_{t_{ir}}\}. \tag{3.12}$$

Now (3.8)(i) implies both that $Q(k) \subset Q(m)$ and that there are only finitely many t_{ir} which are less than t_{jk}. Thus the variables $A_k(t_{jk}) - A_m(t_{jk})$ are well defined, and by their definition,

$$\begin{aligned} E(A_k(t_{Jk}) - A_m(t_{Jk}))^2 \;&=\; E\left\{\sum_{j<J}\sum_l (E\{z_{lm}\,|\,\mathcal{F}_{t_{jk}}\} - z_{lm})^2\right\} \\ &=\; \sum_{j<J} E\left\{z_{jk}\sum_l z_{lm}\right\}^2 \end{aligned} \tag{3.13}$$

where for each fixed j, the summations over l for which $t_{jk} \leq t_{lm} < t_{j+1,k}$ are finite, with $E\left\{\sum_l z_{lm}\,|\,\mathcal{F}_{t_{jk}}\right\} = z_{jk}$. In the second line of (3.13), the cross-terms involving pairs (j',l') and (j,l) for $j' > j$ have been dropped because they are mutually orthogonal by virtue of (3.3). By (3.1) and $\mathcal{F}_{t_{ir}}$ measurability of z_{ir},

$$\begin{aligned} E(z_{jk}\textstyle\sum_l z_{lm}) \;&=\; E\textstyle\sum_l z_{jk}\,E\left\{X(t_{l+1,m}) - X(t_{lm})\,|\,\mathcal{F}_{t_{lm}}\right\} \\ &=\; E\{z_{jk}\,E\left[X(t_{j+1,k}) - X(t_{jk})\,|\,\mathcal{F}_{t_{jk}}\right]\} \;=\; E\{Z_{jk}\}^2. \end{aligned}$$

Substitute the last equalities into (3.13) to obtain

$$\begin{aligned} E(A_k(t_{Jk}) - A_m(t_{Jk}))^2 \;&=\; \textstyle\sum_{j\leq J} E\{z_{jk}\}^2 \\ &=\; 2\textstyle\sum_{j<J}\sum_{l,l'}\sum_{:l<l'} E\{[X(t_{l'+1,m}) - X(t_{l'm})]\,z_{lm}\}. \end{aligned}$$

where the double (l, l') summation is taken first over indices l' such that $t_{lm} < t_{l'm} < t_{j+1,k}$, completing the proof that

$$E(A_k(t_{Jk}) - A_m(t_{Jk}))^2 = \sum_{j \le J} \left\{ \sum_l Ez_{lm}^2 - Ez_{jk}^2 \right\} + 2 \sum_{j \le J} \sum_l E\{z_{lm}^2 \, (X(t_{j+1,k}) - X(t_{l+1,m})\} \tag{3.14}$$

where z_{lm} and z_{jk} are as defined in (3.12) for fixed k, m, and J.

In applying (3.14), it is helpful to remember that $A_k - A_m$ is a discrete-time martingale with respect to $(\mathcal{F}_t : t \in Q(k)\}$, so that the SubmartingaleMaximal Inequality (Theorem 3.1) implies

$$E\left(\max_{j \le J} (A_k(t_{jk}) - A_m(t_{jk}))^2 \right) \le 4\, E\{A_k(t_{jk}) - A_m(t_{Jk})\}^2. \tag{3.15}$$

(3) Some further bounds on terms of (3.14) will be useful. To obtain them, we appeal to the following Lemma, proved as Exercise 6.

Lemma 3.2 *Suppose that* $\{Y_n, V_n : n \ge 1\}$ *are arbitrary random variables on* $\{\Omega, \mathcal{F}, P\}$ *for which* $\{\mathcal{H}_n : n \ge 1\}$ *is an increasing family of sum-*σ*-algebras of* $\Omega, \mathcal{F}, P)$ *for which* $E\{Y_n | \mathcal{H}_n\} \ge 0\}$ *and* $E\{V_n | \mathcal{H}_0\} \ge 0$. *Suppose moreover that* $E\{V_n | \mathcal{H}_n\}$ *is square-integrable for all* n *and that there exists a square-integrable dominating random variable* Y *such that* $\left| \sum_{n \ge r} Y_n \right| \le Y$ *almost surely. Then*

$$E\left\{ \sum_n E\{Y_n | \mathcal{H}_n\} \, E\{V_n | \mathcal{H}_n\} \right\} \le E\{Y \sup_n E\{V_n | \mathcal{H}_n\}\}.$$

Exercise 6 *Prove Lemma 3.2.*

In the context of the previous step, let $\mathcal{J}$ denote any set of indices j bounded above by $J - 1$. For fixed k, apply Lemma 3.2 with $n = (j, l)$, $\mathcal{H}_n = \mathcal{F}_{t_{lm}}$, $Y_n = X(t_{l+1,m}) - X(t_{lm})$, $V_n \le I_{[j \in \mathcal{J}]} \, (X(t_{j+1,k}) - X(t_{lm}))$, and $Y \le 2 \sup_s |X_s|$ to obtain

$$\begin{aligned} & \sum_{j \in \mathcal{J}} \sum_l E\{z_{lm} \, (z_{lm} + 2\, E[X(t_{j+1,k}) - X(t_{l+1,m}) \,|\, \mathcal{F}_{t_{lm}}])\} \\ \le \;& 2 \sum_{j \in \mathcal{J}} \sum_l E\{z_{lm} \, E[X(t_{j+1,k}) - X(t_{lm}) \,|\, \mathcal{F}_{t_{lm}}])\} \\ + \;& 2\, E\{\sup_{s \le T} |X(s)| \, \max_{j \in \mathcal{J}, l} E[X(t_{j+1,k}) - X(t_{lm}) | \mathcal{F}_{t_{lm}}])\} \end{aligned}$$

Combining these inequalities with (3.14), we summarize the results of the last step and this one in the assertion that for all J and all sets $\mathcal{J}$ of integer indices less than or equal to $J - 1$,

$$\begin{aligned} E\{A_k(t_{jk}) \;-\;& A_m(t_{jk})^2 \le \sum_{j \in \mathcal{J}^c} \left[\sum_l z_{lm}^2 - z_{jk}^2 \right] \\ +\;& 2\, E \sum_{j < c,\, j \in \mathcal{J}} \sum_l z_{lm} \, E\{X(t_{jk}) - X(t_{lm}) \,|\, \mathcal{F}_{t_{jk}}\} \\ +\;& 4\, E\{\sup_{s \le T} |X(s)| \, \max_{j \in \mathcal{J}, l} E[X(t_{jk}) - X(t_{lm}) | \mathcal{F}_{t_{lm}}])\}. \end{aligned}$$

(4) In this and the next step, we prove that, as $m, k \longrightarrow \infty$ in such a way that $m > k > i$,

$$E\{\sup\{A_k(t_{jk}) - A_m(t_{jk}) : j \ge 0\}\}^2 \to 0. \tag{3.16}$$

This step will accomplish several preliminary reductions, based on a fixed, arbitrarily small $\epsilon > 0$. First choose $\delta > 0$ so small that

$$E\left\{[\sup_{s \le T} |X^2(s) + \sup_{s \le T} E\{X(T)|\mathcal{F}_s\}]^2 I_A\right\} \le \frac{\epsilon}{10} \quad if \quad P(A) \le \delta \tag{3.17}$$

which can be done by Appendix (B.4), since both sup's are integrable, by square-integrability of $\sup_s X^2(s)$ together with Theorem 3.1 applied to the submartingale $[E\{X(T)|\mathcal{F}_s\}]^2$.

Next, we use (3.9) to find a finite set $K \equiv \{x_1, \cdots, x_p\}$ of atoms of Λ and an integer k_0 such that for all $k \ge k_0$

$$E \sup_{j,u,s} E[X(s) - X(u)|\mathcal{F}_u] : t_{jk} \le u \le s \le t_{j+1,k},\, K \cap (u, s] = \phi\} \le \delta^2 \tag{3.18}$$

and

$$E \sum_{i=1}^{p} \sup\{|X(s) - X(u)| : \max\{t_{jk} : t_{jk} < x_i\} \le s \le u < x_i\} \le \delta^2. \tag{3.19}$$

To see that this is possible, observe that X is the sum of a pure-jump process $U(s) \equiv \sum_{u \le s} \Delta X(u)$ and a continuous process $Z \equiv X(s) - U(s)$, and that both $\sup_s U^2(s)$ and $\sup_s Z^2(s)$ are integrable. Recall by (3.8)(iii) that as $k \longrightarrow \infty$,

$$\delta_k \equiv \;\; mesh(Q(k)) \; = \; \sup\{t_{j+1,k} - t_{jk} : j = 0, 1, \cdots\} \to 0.$$

Therefore, continuity of Z and integrability of $\sup_s Z^2(s)$ imply

$$\omega_k \equiv \sup\{\sup\{|Z(s) - Z(u)| : 0 \le s \le u < T, u - s \le \delta_k\}\} \to 0 \;\text{ in }\; L^2$$
$$E \sup\{E|Z(s) - Z(u)|\mathcal{F}_u\} : 0 \le s \le u < T, u - s \le \delta_k\}^2 \le 4\, E\{\omega_k^2\}0$$

the last inequality following from Theorem 3.1 for the submartingale $[E\{\omega_k|\mathcal{F}_s\}]^2$. Then (3.9) implies

$$E\{U(s)|\mathcal{F}_u\} \le E\left\{\int_s^u h\, d\Lambda \,|\, \mathcal{F}_s\right\}$$

and the integrands of (3.18) are dominated and converge to 0 as K increases to $\{s : \Delta\Lambda(x) > 0\}$ and at the same time $k \longrightarrow \infty$. Thus the existence of K and k_0 in (3.18) follows from the Dominated Convergence Theorem. Now, for the p-element set K just proved to exist, each sup in the sum under the expectation of (3.19) converges almost surely to 0 as $k \longrightarrow \infty$, since the left-hand limits of X at each x_i exist. Also, the sum in (3.19) is dominated by $2p\,\sup_s |X(s)|$. Therefore, replacing k_0 in (3.18) by a sufficiently large integer, by Dominated convergence (3.19) will also be satisfied.

From now on, for fixed J and $m > k > k_0$, with k chosen so large that at most one element of K lies in any single interval $(t_{jk}, t_{j+1,k}]$, let

$$\mathcal{J}^c \equiv \{j \geq 0 : \, j < J, \, K \cap (t_{jk}, t_{j+1,k}] = \phi\}$$

(5) Let $t \in Q(i)$ and $\epsilon > 0$ be arbitrary, and fix δ, $k_0 \geq i$, and K satisfying (3.17)-(3.19). By (3.16),

$$\begin{aligned} E(A_k(t) - A_m(t))^2 \;&\leq\; E \textstyle\sum_{j \in \mathcal{J}^c} \left\{ \sum_l z_{lm}^2 - z_{jk}^2 \right\} \\ &+\; 2E \textstyle\sum_{j \in \mathcal{J}^c} \sum_l z_{lm} \, E\{X(t_{j+1,k}) - X(t_{l+1,m}) \,|\, \mathcal{F}_{t_{jk}}\} \\ &+\; 4\,E\{\sup_{s \leq T} |X(s)| \, \max_{j \in \mathcal{J}, l} E[X(t_{j+1,k}) - X(t_{lm}) \,|\, \mathcal{F}_{t_{lm}}])\}. \end{aligned} \tag{3.20}$$

The integrand in the third line of (3.20) is dominated by $2\,\sup_s\, X^2(s) + 2\,\sup_{s \leq T}\, [E\{X(T)|\mathcal{F}_s\}]^2$, while by (3.19) and the definition of J,

$$E \max_{j \in \mathcal{J}, l} \left\{ E\left[X(t_{j+1,k}) - X(t_{lm}) \,|\, \mathcal{F}_{t_{jk}}\right] \right\} \leq \delta^2$$

Now (3.17) shows the third line of (3.20) is $\leq \epsilon$, since

$$P\left\{ \omega : \max_{j \in \mathcal{J}, l} E\left[X(t_{j+1,k}) - X(t_{lm}) \,|\, \mathcal{F}_{t_{jk}}\right] \geq \delta \right\} \leq \delta^{-1}\delta^2 = \delta.$$

Consider the second line of (3.20). For each $j \leq c$ in $\mathcal{J}$, fix $\xi(j) \in K \cap (t_{jk}, t_{j+1,k}]$. We partition the inner sum into those l for which $\xi(j) \leq t_{l+1,m}$ and those for which $\xi(j) > t_{l+1,m}$. By Lemma 3.2 with $n = (j,l)$, $\mathcal{H}_n \geq \mathcal{F}_{t_{jk}}$, $V_n = (X(t_{j+1,k}) - X(t_{l+1,m}))\, I_{[j \in \mathcal{J}^c, \, \xi(j) \leq t_{l+1,m}]}\, X(t_{lm})$, and $Y \leq 2 \sup_s |X(s)|$, we have

$$\begin{aligned} &\textstyle\sum_{j \in \mathcal{J}^c} \sum_{l :\, \xi(j) \leq t_{l+1,m}} E\left\{ z_{lm}\, E\left[X(t_{j+1,k}) - X(t_{l+1,m}) \,|\, \mathcal{F}_{t_{jk}}\right] \right\} \\ &\leq 2\,E\{\sup_{s \leq T} |X(s)| \, \max_{j \in \mathcal{J}^c, \, \xi(j) \leq t_{l+1,m}} E\left[X(t_{j+1,k} - X(t_{l+1,m}) \,|\, \mathcal{F}_{t_{lm}}\right])\}. \end{aligned}$$

As in the previous paragraph, by (3.18) and (3.19) the last expression is shown to be $\leq \epsilon$. Next apply Lemma 3.2 with $\mathcal{H}_n = \mathcal{F}_{t_{jk}}$, $V_n \leq X(t_{j+1,k}) - X(t_{l+1,m})$, $Y_n \equiv (X(t_{j+1,k}) - X(t_{lm}))\, I_{[j \in \mathcal{J}^c, \, t_{l+1,m} < \xi(j)]}$, and

$$Y \equiv \sup_{j \in \mathcal{J}^c} \sup\{|X(u) - X(s)| : \, \{t_{jk} \leq s \leq u < \xi(j)\}\}.$$

To check that Y does dominate $\sum_{n \geq r} Y_n$, recall that $\mathcal{J}^c$ consists of all $j < J$ such that $(t_{jk}, t_{j+1,k}]$ contains one of the elements $x_i = \xi(j)$ of K. The result, from Lemma 3.2, is

$$\begin{aligned} &\textstyle\sum_{j \in \mathcal{J}^c} \sum_{l :\, \xi(j) \leq t_{l+1,m}} E\left\{ z_{lm}\, E\left[X(t_{j+1,k}) - X(t_{l+1,m}) \,|\, \mathcal{F}_{t_{jk}}\right] \right\} \\ \leq\;& E\{Y \max_{j \in \mathcal{J}^c, \, \xi(j) \leq t_{l+1,m}} E\left[X(t_{j+1,k}) - X(t_{l+1,m}) \,|\, \mathcal{F}_{t_{lm}}\right])\}. \\ \leq\;& 2\,E\{Y \textstyle\sum_{s \leq T} E[X(T)|\mathcal{F}_s]\}. \end{aligned}$$

By (3.19), $P\{Y \geq \delta\} \leq \delta^{-1} EY \leq \delta$. Hence (3.17) shows that the last expectation is $\leq \epsilon$.

Taken together, the estimates of this step have so far proved that

$$E\{A_k(t) - A_m(t)\}^2 \leq 3\epsilon + \sum_{j \in \mathcal{J}^c,\, t_{lm} \leq \xi(j)} E\left(z_{lm}^2 - z_{jk}^2\right) \tag{3.21}$$

Recall that the cardinality p of $\mathcal{J}^c$ is finite and depends only on ϵ and not on k or m. As $k, m \longrightarrow \infty$, for each $x \in K$ and the unique values $j = j(x)$ and $l = l(x)$ for which $t_{jk} < x \leq t_{j+1,k}$ and $t_{lm} < x \leq t_{l+1,m}$

$$\begin{aligned} z_{lm} &\equiv E\left\{X(t_{l+1,m}) - X(t_{lm}) \,|\, \mathcal{F}_{t_{jk}}\right\} \longrightarrow E\{\Delta X(x)|\mathcal{F}_{x-}\} \\ z_{jk} &\equiv E\left\{X(t_{j+1,k}) - X(t_{jk}) \,|\, \mathcal{F}_{t_{jk}}\right\} \longrightarrow E\{\Delta X(x)|\mathcal{F}_{x-}\} \end{aligned} \tag{3.22}$$

by the Martingale Convergence Theorem and the right-continuity of X. The convergence takes place both almost surely and in mean-square. We conclude immediately from (3.21) and (3.22) that $\{A_k(t) - A_m(t)\}^2$ converges to 0 as $k, m \longrightarrow \infty$. Since $\epsilon > 0$ and $t \in Q(i)$ were arbitrary, and since none of the upper bounds developed for $E\{A_k(t) - A_m(t)\}^2$ depend on t, we appeal to (3.15) to conclude for all sufficiently large $m > k$,

$$E\{\sup_{j:\, t_{jk} \leq t} (A_k(t_{jk}) - A_m(t_{jk}))^2\} \leq 16\epsilon$$

uniformly in $t < T$. Let $t \uparrow T$ and apply the Monotone Convergence Theorem to complete the proof of (3.16).

(6) The assertion in (3.16) can be strengthened to

$$E\{\sup_{s<T} (A_k(s) - A_m(s))\} \longrightarrow 0 \quad \text{as} \quad k, m \longrightarrow \infty. \tag{3.23}$$

To prove this, fix arbitrary $\epsilon > 0$ and $t \in Q(i)$, and let $\delta, K, k_0,$ and J be as in step (4). Then

$$E\left\{\max_{j \in J} (A_k(t_{j+1,k}) - A_k(t_{jk}))^2\right\} \leq E \sum_{j \in J} \left\{E\left[X(t_{j+1,k}) - X(t_{jk}) \,|\, \mathcal{F}_{t_{jk}}\right]\right\}^2.$$

By Lemma 3.2 with $\mathcal{H}_n = \mathcal{F}_{t_{j+1,k}}$, $Y_n le X(t_{j+1,k}) - X(t_{jk})$, and $Y \equiv 2\sup_{s \leq T} |X(s)|$, the last displayed expression is

$$\leq 2E\left\{\sup_{s \leq T} |X(s)| \max_{j \in J} E\left[X(t_{j+1,k}) - X(t_{jk}) | \mathcal{F}_{t_{jk}}\right]\right\} \tag{3.24}$$

Together, (3.18) and (3.19) imply that (3.24) is $\leq \epsilon$.

For each of the finitely many elements x of K, let $j \equiv j(x)$ be defined as at the end of step (5). Reasoning as for (3.22),

$$\left.\begin{array}{l} |A_k(x-) - A_k(t_{jk})| + |A_k(t_{j+1,k}) - A_k(x)| \;\to\; 0 \\ \\ \Delta A_k(x) \;\to\; E\{\Delta X(x)\, E\{\Delta X(x)\,|\,\mathcal{F}_{x-}\} \end{array}\right\} \quad \text{a.s.} \quad \text{and} \quad \text{in} \quad L^2$$

when $k \longrightarrow \infty$, by the Martingale Convergence Theorem and Dominated Convergence. Since (3.24) is $\le \epsilon$, as k becomes large and $m > k$,

$$E\left\{\max_x \max_{t_{jk}<s\le t_{j+1,k}} (A_k(s) - A_m(s) - A_k(t_{jk}) + A_m(t_{jk}))^2\right\} \to 0$$

which together with (3.16) proves (3.23).

(7). By (3.23) the family of nondecreasing right-continuous functions A_k converges uniformly in the mean. For each $t \in [0,T]$, $\{A_k(t)\}_k$ is a Cauchy sequence in $L^1(\Omega, F, P)$, by (3.23) and the Cauchy-Schwarz inequality. Hence, there is a limiting random variable $A(t)$. For any infinite sequence of integers k, there is a nonrandom infinite sequence of integers $k(r)$ such that

$$\sup_{i\ge r}\sup_{t<T} |A_{k(i)}(t) - A_{k(r)}(t)| \;\to\; 0 \quad \text{a.s.}$$

Then the a.s. limit of the random variables $A_{k(i)}(t)$ exists for each t (*cf.* Appendix A.2), and must agree with the limit-in-the-mean $A(t)$. In particular, since $A(t)$ is an a.s. limit of nondecreasing $\mathcal{F}_{t-}$ adapted random variables, it is nondecreasing and $\mathcal{F}_{t-}$ measurable, and $A(0) = 0$. Monotonicity implies the existence of all left limits $A(t-), t \le T$.

Letting $i \longrightarrow \infty$ in $\sup_{t<T} |A_{k(i)}(t) - A_{k(r)}(t)|$, we have

$$\sup_{t<T} |A(t) - A_{k(r)}(t)| \;\to\; 0 \quad \text{a.s.} \tag{3.25}$$

Since A is the uniform limit of a subsequence of an arbitrary subsequence of $\{A_k\}$, it follows that $\sup_{t<T} |A(t) - A_k(t)| \to 0$ in probability (Appendix A.3).

(8) We show finally that the process A is almost surely right-continuous, and that $X - A$ is a martingale. Indeed, (3.25) and the bound of ϵ of (3.24) show, with $\epsilon > 0$ fixed arbitrarily and K as in step (4), that

$$E\left\{\sup\{[A(t_{j+1,k}) - A(t_{jk})]^2 \,:\, K \cap (t_{jk}, t_{j+1,k}] = \phi\}\right\} \le 2\epsilon$$

for all sufficiently large k. For each $x \in K$, (3.25) implies that $A(t_{j(x)+1,k}) - A(x) \to 0$ in probability as $k \longrightarrow \infty$. Since ϵ was arbitrary, A is continuous at every t for which $\Delta\Lambda(t) = 0$, and A is right-continuous at each x for which $\Delta\Lambda(x) > 0$.

To check that $X - A$ is a martingale, observe first that for every $s, t \in \cup_i Q(i)$ with $s < t$, by step (1) together with (3.25),

$$\begin{aligned} E\{X(t) - A(t)|\mathcal{F}_s\} &= \lim_{k\longrightarrow\infty} E\{X(t) - A_k(t)|\mathcal{F}_s\} \\ &= \lim_{k\longrightarrow\infty} [X(s) - A_k(s)] = X(s) - A(s) \quad \text{a.s.} \end{aligned}$$

Now let $0 \le s < t < T$ be arbitrary. Find sequences $\{s_j\}, \{t_j\} \subset \cup_i Q(i)$ with $s_j < t_j < T$ for all j, and s_j and t_j respectively decreasing to s and to t. By the right-continuity of X and A, together with (3.1 and the martingaleconvergence theorem, we conclude

$$\begin{aligned} E\{X(t) &- A(t)|\mathcal{F}_s\} = \lim_{j\longrightarrow\infty} E\{X(t_j) - A(t_j)|\mathcal{F}_s\} \\ &= \lim_{j\longrightarrow\infty} E\{X(t_j) - A(t_j)|\mathcal{F}_s\} = \lim_{j\longrightarrow\infty} E\{X(s_j) - A(s_j)|\mathcal{F}_s\} \end{aligned}$$

which is equal to $X(s) - A(s)$. The martingaleproperty and the Theorem are proved.

Remark 3.5 The simple counting processs N with finitely many jumps clearly satisfy (3.10). If $\mathcal{F}_t$ is generated by $\mathcal{F}_0$ together with $(N(s) : s \le t)$ and possibly some random variables independent of N, then by Theorem 1.3 there is an increasing $\{\mathcal{F}_{t-}\}$ adapted process A for which $N - A$ is a $\mathcal{F}_t$ martingale. If this A is assumed to be absolutely continuous with respect to to the nonrandom $\mathcal{F}_0$ measurable function Λ in the sense that $A(t) = \int^t h(s)d\Lambda(s)$ for some necessarily $\mathcal{F}_t$-adapted process h, then the martingaleproperty of $N - A$ implies (3.9), and Theorem 3.3 shows that A is a calculable compensator. For a single-jump counting process N, we can by Theorem 1.1 dispense with the assumption of absolute continuity and take $h \equiv 1$ and Λ to be the regular conditional sub-distribution function of the jump-time of N given $\mathcal{F}_0$. Piecing together the general simple counting process by means of single-jump processes as in the proof of Theorem 1.3, one proves that the compensator-processes derived in Theorem 1.3 are calculable.

One clear benefit of the slightly restricted class of submartingales treated in Theorem 3.3 is that the convergence in the mean of the approximate (Doob) compensators A_k to A is uniform. Further, the same result holds if (3.9) is replaced by a requirement ensuring only that the compensatorbe continuous, rather than absolutely continuous with respect to a nonrandom increasing function Λ.

Corollary 3.4 *(i) Under the hypotheses of Theorem 3.3,*

$$E \sup_{t<T} |A_k(t) - A(t)| \to 0 \quad \textit{as} \qquad k \longrightarrow \infty \tag{3.26}$$

and the compensator limit A does not depend on the choice of partition-sequence $\{Q(k)\}_k$ satisfying (3.9).

(ii) Assume the hypotheses of Theorem 3.3, with (3.9) replaced by the assumption that there exist a continuous increasing $\mathcal{F}_t$ *adapted process* B *with* $B(0) = 0$ *and* $EB(T-) < \infty$ *such that for* $0 \le s < t < T$

$$E\left\{ \sum_{s<u\le t} \Delta X(u) \,|\, \mathcal{F}_s \right\} \le E\{B(t) - B(s)|\mathcal{F}_s\} \tag{3.27}$$

Then (3.26) and the assertion of Theorem 3.3 hold, with A *continuous.*

Proof. Steps (6) and (7) of the proof of Theorem 3.3 establish (3.26). To see that A does not depend on the choice of $\{Q(k)\}_k$, let $\{Q(k)\}$ and $\{R(k)\}$ be any two partition-sequences satisfying (3.8), and define another partition-sequence $\{S(k)\}$ by $S(k) \equiv Q(k) \cup R(k)$. Denote by superscripts (as in A^Q the partition being used to form an approximate compensator for X. For each $\epsilon > 0$, steps (4)-(6) show that k can be found so large that for any $m \ge k$, $E \sup_t (A^Q(t) - A^S(t))^2 \le \epsilon$. Letting $m \longrightarrow \infty$ in the last inequality and recalling that ϵ is arbitrary proves that the limiting compensator for the $\{S(m)\}$ partition sequence must coincide with that for the partition sequence $\{Q(k)\}$.

For assertion (ii), the proof proceeds exactly as in the Theorem, except that K in step (3) should now be taken to be the empty set (for all ϵ), so that $\mathcal{J} = \{0, 1, 2, \cdots, J - 1\}$. Then (3.18) is proved as before, by Dominated Convergence and (3.27), and (3.19) becomes vacuous. The remainder of steps (3)-(8) yield (3.26) and the compensator properties of A , and (7) proves continuity of A. □

The Theorem and Corollary can be generalized in two directions: $Q(k)$ can be allowed to consist of stopping times, and submartingales can be replaced by local submartingales.

Theorem 3.5 *Let* $(X(t), t \in [0, T))$ *be a* $\mathcal{F}_t$ *adapted local submartingalesuch that*

i for each m, $E \sup\{|\Delta X(t)|^2 : t \in [0, T), |X(t)| \le m\} < \infty$*;*

ii $\sum_{s<T} |\Delta X(s)| < \infty$ *a.s.; and*

iii for some localizing sequence σ_n *of stopping times, (3.9) holds for each of the submartingales* $X((\cdot \wedge \sigma_n)$*.*

Let $\{Q(k)\}_k$ *denote an arbitrary partition sequence of stopping times satisfying (3.8). Then there exists a localizing sequence* $\{\tau_n : n \ge 1\}$ *of stopping times increasing almost* *surely to T as $n \longrightarrow \infty$, and a nondecreasing right-continuous $\mathcal{F}_t$ adapted process A with $A(0) = 0$ and $X - A$ a local martingale, such that for each n,*

$$\lim_{k \longrightarrow \infty} E \sup_{0 \le t \le \tau_n} |A_k^{(n)}(t) - A(t \wedge \tau_n)| \; = \; 0 \tag{3.28}$$

where $A_k^{(n)}$ denotes the $Q(k)$ Doob compensator corresponding to the submartingale $X(\cdot \wedge \tau_n)$.

Proof. The numbering of steps in this Theorem continues that of Theorem 3.3.

(9) First let all hypotheses be exactly as in Theorem 3.3, except that the partition sequence $Q(k)$ satisfying (3.8) is now allowed to consist of stopping times. In this case, we need a new definition to make sense of the σ-algebra $\mathcal{F}_\tau$ representing the information generated by values of all $\mathcal{F}_t$ adapted processes up to the stopping time τ. The definition is

$$\mathcal{F}_\tau \equiv \sigma(B : \text{ for } t \in [0,T], \; B \cap [\tau \le t] \in \mathcal{F}_t)$$

and the repeated-conditioning property continues to hold:

if Y is an integrable random variable and Z is a bounded $\mathcal{F}_\tau$ measurable random variable, where $\tau \ge 0$ is a $\mathcal{F}_t$ stopping time, then

$$E\{E\{YZ|\mathcal{F}_\tau\}\} = E\{Z \, E\{Y|\mathcal{F}_\tau\}\} \tag{3.29}$$

In addition, if $X(t)$ is a right-continuous $\mathcal{F}_t$ submartingalefor $t \in [0, T)$, then for any stopping times $\sigma \le \tau$ (a.s.),

$$E(X(\tau)|\mathcal{F}_\sigma) \; \ge \; X(\sigma) \qquad \text{almost surely} \tag{3.30}$$

This is easy to check via right continuity and the calculation, for arbitrary $B \in \mathcal{F}_\sigma$ and positive integer n,

$$E\left\{ \sum_j E\{[X(\tau \wedge (j+1)/2^n) - X(\tau \wedge j/2^n)] \, I_{B \cap [2^n \sigma \le j]} \right\} \ge 0$$

which relies on $\mathcal{F}_{j/2^n}$ measurability of $B \cap [2^n \sigma \le j]$ together with the submartingale property of $X(\cdot \wedge \tau)$. As a result, if $\{\tau(r)\}$ is any family of $\mathcal{F}_t$ stopping times indexed by the real parameter r, with $\tau(r) \le \tau(s)$ a.s. whenever $r < s$, then $X(\tau(r))$ is a submartingale with respect to the σ-algebra family $\{\mathcal{F}_{\tau(r)}\}$. For further background on the σ-algebras $\mathcal{F}_\tau$, see Lipster and Shiryayev (1977, vol. 1, pp. 25-29).

Careful inspection of the steps (1)–(8), together with repeated applications of (3.29) and (3.30, using $\mathcal{F}_{t_{jk}}$ measurability of events $[t_{jk} < t]$ and $[t_{jk} \le t]$ for all t, and

$\mathcal{F}_{lm}$ measurability of $[t_{jk} \le t_{lm} < t_{j+1,k}]$ for all j and l, shows that all steps remain valid. For this reason, we shall not mention the random or nonrandom character of $Q(k)$ except in Theorem 4.3 and in Chapter 9?, where it becomes crucial to introduce partitions by stopping times.

(10). Now assume only the hypotheses of the present Theorem. Let ρ_n be any sequence of $\mathcal{F}_t$ stopping times increasing to T, for which $X(\cdot \wedge \rho_n)$ is a submartingale. Define $\tau_0 = 0$, and for each $n \ge 1$,

$$\tau_n \equiv \inf\{t \ge \tau_{n-1} : |X(T)| \ge n \quad \text{or} \quad \sum_{s \le t} |\Delta X(s)| \ge n\} \cup \{\rho_n \wedge \sigma_n\}.$$

Denote the submartingale $X(\cdot \wedge \tau_n)$ by $X^{(n)}(\cdot)$, and for each k let $A_k^{(n)}$ denote the Doob compensator A_k with X replaced by $X^{(n)}$ for the partition $Q(k)$. We show in this step that $X^{(n)}$ satisfies all the hypotheses of Theorem 3.3. First, by (i), (ii), and the definition of τ_n,

$$\begin{aligned} & E\{\sup_{t<T} (|X^{(n)}(t)| + \textstyle\sum_{s \le t} |\Delta X^{(n)}(s)|)^2\} \\ \le \quad & 2EE\{\sup_{t<\tau_n} (|X^{(n)}(t)| + \textstyle\sum_{s \le t} |\Delta X^{(n)}(s)|)^2\} + 8\,E|\Delta X(\tau_n)|^2 \\ < \quad & 8n^2 + 8\,E \sup_{t<T} |\Delta X(t)|^2 < \infty. \end{aligned}$$

By Lemma 3.1, each process $X^{(n)}$ is a submartingale, and (3.9) continues to hold for $X^{(n)}$ because, for each n, (iii) implies that

$$\int_0^t h_n(u)\, d\Lambda(u) \; - \sum_{u \le t \wedge \sigma_n} \Delta X(u)$$

is a $\mathcal{F}_t$ is a submartingale and by Lemma 3.1, for $0 \le s < t < T$

$$\begin{aligned} E\left\{\textstyle\sum_{s<u \le t} \Delta X^{(n)}(u) \,|\, \mathcal{F}_t\right\} \quad &= \quad E\left\{\textstyle\sum_{s<u \le t} \Delta X(u)\, I_{[u_n \le \sigma \wedge_n \tau]} \,|\, \mathcal{F}_s\right\} \\ \le E\left\{\textstyle\int_s^t I_{[u \le \tau_n]}\, h_n(u)\, d\Lambda(u) \,|\, \mathcal{F}_s\right\} \quad &\le \quad E\left\{\textstyle\int_s^t h_n(u)\, d\Lambda(u) \,|\, \mathcal{F}_s\right\}. \end{aligned}$$

(11) According to the previous step, Theorem 3.3 applies to each of the processes $X^{(n)}$ with approximate compensators $A_k^{(n)}$. That Theorem together with step (9) says, for each n and partition-sequence $\{Q(k)\}$ of stopping times, that $X^{(n)}$ has a compensator $A^{(n)}$ for which

$$E \sup_{t \le \tau_n} |A_k^{(n)}(t) - A^{(n)}| \to 0 \quad \text{as} \quad k \longrightarrow \infty$$

However, Corollary 3.4(i) implies that when the partition-sequences $Q(k)$ are replaced by $Q_n(k)$ in calculating approximate compensators $A_k^{(n)}$, the limit is $A^{(n)}$. If $n' > n$, then by definition first of $X_k^{(n')}$ with respect to $Q_{n'}(k)$ and then by definition of $X^{(n)}$

and $A_k^{(n)}$ with respect to $Q_n(k)$, we have

$$A_k^{(n)}(t \wedge \tau_n) = \sum_j I_{[t_{jk}<t\wedge\tau_n]} E\left\{X^{(n')}(t_{j+1,k} \wedge t\ \tau_n) - X^{(n')}(t_{jk})|\mathcal{F}_{t_{jk}}\right\}$$

$$= \sum_j I_{[t_{jk}<t\wedge\tau_n]} E\left\{X^{(n)}(t_{j+1,k} \wedge t \wedge \tau_n) - X^{(n)}(t_{jk})|\mathcal{F}_{t_{jk}}\right\} = A_k^{(n)}(t \wedge \tau_n)$$

for all $t \in [0,T)$. Upon taking limits as $k \longrightarrow \infty$, it follows that the compensators $A^{(n)}$ and $A^{(n')}$ are identical processes on the interval $[0,\tau_n]$. Thus the right-continuous nondecreasing process A defined by

$$A(t) \le A^{(n)}(t) \qquad \text{for} \qquad 0 \le t \le \tau_n, \quad \text{all } n$$

exists and satisfies (3.28). By the compensator property of $A^{(n)}$ for $X^{(n)}$, $X - A$ is a local $\mathcal{F}_t$ martingalewith localizing sequence τ_n.

Whenever a submartingale X is defined as the square of a local martingale M with respect to a σ-algebra family $\{\mathcal{F}_t, t \in [0,T)\}$, the compensator A guaranteed to exist under the hypotheses of Theorem 3.5 is known as the variance process $< M >$ associated with M. We have seen in Section 1.3 the explicit calculation both of the compensator A for a simple counting process N with respect to the σ-algebra family $\mathcal{F}_t = \sigma(\mathcal{F}_0,\ N(s) : 0 \le s \le t)$, and of the variance-process $V \equiv < N - A >$ for the compensated local martingale $M \equiv N - A$. In that setting, $V(t)$ was given explicitly by $\int^t (1 - \Delta A(s))\, dA(s)$. More generally, we evaluate or approximate varianceprocesses for square-integrable martingales M, under the hypotheses of Theorem 3.5 on M^2, via the Doob compensators defined for each nonrandom $t \in [0,T)$ by

$$\begin{aligned} V_k(t) &\le \sum_j I_{[t_{jk}<t]} E\left\{M^2(t_{j+1,k} \wedge t) - M^2(t_{jk})|\mathcal{F}_{t_{jk}}\right\} \\ &= \sum_j I_{[t_{jk}<t]} E\left\{M(t_{j+1,k} \wedge t) - M(t_{jk})|\mathcal{F}_{t_{jk}}\right\} \end{aligned} \tag{3.31}$$

The last line of (3.31) follows immediately from the martingale property of M, and justifies our regarding variance processes as cumulative partial sums of conditional variances.

Examples. (a) The continuous submartingales which appear most often in applications are of the form $W(G(\cdot))$ or $W^2(G(\cdot))$ where G is a nondecreasing nonrandom continuous $[0,1]$-valued function on $[0,\infty)$, and W is standard Wiener process. For any such G, if for each $k \ge 1$, $Q_0(k) \equiv \{t_{jk}\}_j$ denotes a nonrandom partition sequence satisfying (3.8), then define $Q(k) \equiv \{G(t_{jk}\}_j$. For $X(\cdot) \equiv W(G(\cdot))$, the Doob compensator corresponding to $Q(k)$ is

$$A_k(t) \equiv \sum_{j:t_{jk}\le t} E\{W(G(t \wedge t_{j+1,k})) - W(G(t_{jk}))\,|\,\mathcal{F}^W_{G(t_{jk})}\} \equiv 0$$

where $\mathcal{F}_s^W$ denotes $\sigma(W(u) : 0 \le u \le s)$, and for the submartingale $Y(\cdot) \equiv W^2(G(\cdot))$, the corresponding discrete-time compensators are

$$A_k^Y(t) \equiv \sum_{j:t_{jk} \le t} [W(G(t \wedge t_{j+1,k})) - G(t_{jk})].$$

These assertions follow immediately from the Gaussian-distribution properties (2.8) of Wiener process. By taking in-probability limits as $k \longrightarrow \infty$ as in Theorem 3.3, the compensatorof $W(G(\cdot))$ is 0, and the compensatorof $W^2(G(\cdot))$ is $G(\cdot) - G(0)$.

(b) The nonhomogeneous Poisson counting process also has a nonrandom compensator and variance process. Let N be such a process with cumulative rate-function Λ, so that for $0 \le s < t$, $N(t) - N(s)$ is a *Poisson*$(\Lambda(t) - \Lambda(s))$ random variable independent of $\mathcal{F}_u \equiv \sigma(N(u) : 0 \le u \le s)$. Then obviously, for any nonrandom partition sequence $Q(k)$,

$$\sum_{j:t_{jk} \le t} E\{N((t \wedge t_{j+1,k})) - N(t_{jk})) \,|\, \mathcal{F}_{t_{jk}}\} = \sum_{j:\, t_{jk} \le t} [\Lambda(t \wedge t_{j+1,k}) - \Lambda(t_{jk})] = \Lambda(t)$$

and

$$\sum_{j:\, t_{jk} \le t} E\{N((t \wedge t_{j+1,k})) - \Lambda(t \wedge t_{j+1,k}) + \Lambda(t_{jk}))^2| = \Lambda(t)$$

The compensator and the variance process are both equal to Λ. □

Exercise 7 . *Let* $U_1, U_2, \cdots$ *be an independent and identically distributed sequence of Exponential random variables with mean* 1, *and define*

$$S_k = \sum_{i=1}^{k} U_i, \qquad N_k(t) \equiv I_{[S_k \le t]}, \qquad 1 \le k \le \infty$$

and

$$\mathcal{F}_t \equiv \sigma(N_k(s) : 0 \le s \le t, \ k \ge 1)$$

Find the compensator of N_∞.

3.4 References

- 1. Barlow, M., Jacka, S., and Yor, M. (1986). Inequalities for a pair of processes stopped at a random time. *Proc. London Math. Soc.* 52, 142-172.

- 2. Brown, T. (1978). A martingaleapproach to the Poisson convergence of simple point processes. *Ann. Prob.* 6, 615-629 (§3).

- 3. Burkholder, D. (1973). Distribution function inequalities for martingales. Ann. Prob. 1, 19-42.
- 4. Chow, Y., Robbins, H., and Siegmund, D. (1971). Great Expectations: *Theory of Optimal Stopping*. Houghton-Mifflin: Boston (Chapter 1).
- 5. Freedman, D. (1975). Tail inequalities for martingales. *Ann. Prob.*, 3, 100-118.
- 6. Helland, I. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* 9, 79-94.
- 7. Liptser, R. and Shiryaev, A. (1977). *Statistics of Random Processes*, vol. 1 (Chapters 1-3).
- 8. Millar, P.W. (1968). Martingale integrals. *Trans. Amer. Math. Soc.* 133, 145-166.
- 9. Rebolledo, R. (1980). Central limit theorems for local martingales. *Zeitschr. f. Wahrsch. u. verw. Geb.* 51, 269-286.
- 10. Steiger, W. (1969). A best-possible Kolomogorov-type inequality for martingales and a characteristic property. *Ann. Math. Statist.* 40, 764-769.

Chapter 4

Martingale Functional Central Limit Theorems

The focus of this Chapter is a Functional Central Limit Theorem (**FCLT**) of Rebolledo (1977, 1980) for continuous-time martingales. This theorem provides readily applicable conditions on a sequence of martingales $\{M_n(t) : t \in [0,T]\}$ — expressed in terms of the magnitude of the largest jumps and in terms of the convergence of variance-processes $< M_n > (\cdot)$ — to converge in distribution in $D[0,T]$ to a process of the form $\sigma \cdot W(G(t))$, where $W(\cdot)$ is a standard Wiener process, σ is a nonrandom constant, and $G(\cdot)$ is a nonrandom function. We will continue to prove theorems on continuous-time processes in somewhat less than their full generality, restricting ourselves throughout to the case of locally square-integrable random functions in $D[0,T]$ with calculable compensators and variance-processes. [The Rebolledo references give statements and proofs in full generality.]

4.1 Discrete time: the McLeish Theorem

Although the goal is the FCLT for continuous-time martingales, the "master theorem" from which it is derived in this chapter applies to discrete-time [or equivalently, piecewise-constant] processes which are sufficiently close to being martingales.

Theorem 4.1 (McLeish, 1974) *Let* $\{X_{ni} : n \geq 1,\ 1 \leq i \leq k_n\}$ *be an array of square-integrable random variables on* $(\Omega, \mathcal{F}, P)$, *and* $\{\mathcal{F}_{ni}\}_{i=0}^{k_n}$ *for each* n *an increasing family of* σ*-algebras with* X_{ni} *measurable with respect to* $\mathcal{F}_{ni}$. *Suppose* $\{k_n(\cdot)\}_n$ *is a sequence of nonrandom* **time-scales**, *i.e., nondecreasing right-continuous integer-valued functions on* $[0,1]$ *with* $k_n(1) = k_n$ *and* $k_n(0) = 0$. *For each fixed* n, *let* $E_i\{\cdot\}$ *denote*

$E\{\cdot|\mathcal{F}_{ni}\}$, and let $\sum_i$ denote summation over all $i = 1, \ldots, k_n$. If as $n \longrightarrow \infty$

$$\sum_i E_{i-1}\{X_{ni}^2 \, I_{[|X_{ni}>\delta]}\} \xrightarrow{P} 0 \quad \textit{for each} \quad \delta > 0 \tag{4.1}$$

$$\sum_{i=1}^{k_n(t)} E_{i-1}\{X_{ni}^2\} \xrightarrow{P} t \qquad \textit{for each} \qquad t \in [0,1] \tag{4.2}$$

$$\sum_i |E_{i-1}\{X_{ni}\}| \xrightarrow{P} 0 \tag{4.3}$$

then

$$W_n(\cdot) \equiv \sum_i X_{ni} \, I_{[i \le k_n(\cdot)]} \xrightarrow{D} W(\cdot) \qquad \textit{in} \;\; D[0,1] \quad \textit{as} \quad n \longrightarrow \infty$$

where W is a Wiener process. If (4.1) and (4.3) hold but the convergence in (4.2) is assumed to hold only for a single $t \in (0,1]$, then

$$W_n(t) \xrightarrow{D} W(t) \sim N(0,t) \qquad \textit{in} \quad \mathbb{R}.$$

Remark 4.1 Condition (4.1) above is known as the *conditional Lindeberg condition* and reduces to the usual Lindeberg condition [see Loève 1955, p. 295; Feller 1971, p. 518] in case the X_{ni} for $i = 1, \cdots, k_n$ form an independent sequence for each n.

Assumption (4.3) says that the random variables form an *approximate martingale difference array* [or m.d.a.]: $\{X_{ni}\}$ is a $\{\mathcal{F}_{ni}\}$ m.d.a. if $E_{i-1}\{X_{ni}\} \equiv 0$ almost surely for $1 \le i \le k_n$, that is, $W_n(\cdot)$ is approximately a martingale. If $\{X_{ni}\}$ is a m.d.a., then $\sum_i E_{i-1}\{X_{ni}^2\} I_{[i \le k_n(\cdot)]}$ is the discrete-time compensator [from the Doob Decomposition, step (1) of the proof of Theorem 3.3] for the submartingale $\sum_i X_{ni}^2 \, I_{[i \le k_n(\cdot)]}$. □

Throughout the proof and the Section, discrete-time partial-sum processes will be viewed as piecewise-constant right-continuous processes in continuous time adapted to the increasing σ-algebra family $\mathcal{G}_n(t) \equiv \mathcal{F}_{n,k_n(t)}$. When $k_n(t)$ is later assumed to be a stopping time, the events $[k_n(t) > i]$ are simply being assumed to be $\mathcal{F}_{ni}$-measurable for all i. When $k_n(\cdot)$ is a nondecreasing process of stopping times, the definition $\mathcal{G}_n(t) \equiv \mathcal{F}_{n,k_n(t)}$ still makes sense [*cf.* Remark 3.3] and can be reinterpreted as

$$\mathcal{G}_n(t) \equiv \sigma(B \cap [k_n(t) \le i] : \; B \in \mathcal{F}_{ni}, \; i \ge 1)$$

The first proof-step is a reduction: the Theorem will be proved with the following assumptions replacing (4.1) and (4.2):

$$\max\{|X_{ni}| : \; 1 \le i \le k_n\} \xrightarrow{P} 0 \qquad \text{as} \quad n \longrightarrow \infty \tag{4.1 $'$}$$

$$\sum_{i:\, i\le k_n(t)} X_{ni}^2 \xrightarrow{P} t \qquad \text{as} \quad n \longrightarrow \infty, \quad \text{for each } \ t\in[0,1] \qquad (4.2\,')$$

Proof that (4.1) and (4.2) imply (4.1 ′) and (4.2 ′). First,(4.1) evidently implies

$$\sum_i P_{i-1}\{|X_{ni}|>\delta\} \xrightarrow{P} 0 \qquad \text{as} \quad n\longrightarrow\infty$$

where $P_{i-1}\{\cdot\} \equiv E_{i-1}\{I_{[\cdot]}$. For arbitrary $\alpha\in(0,1)$, observe for each i that $\sum_{j=1}^i P\{|X_{ni}|>\delta\}$ is $\mathcal{F}_{ni}$-measurable, and consider the martingale

$$M_n(t) \equiv \sum_{i=1}^{k_n(t)} (I_{[|X_{ni}|>\delta]} - P\{|X_{ni}|>\delta\})\, I_{[\sum_{i\le j} P_{j-1}\{|X_{nj}|>\delta\}\,\le\alpha]}$$

for which it is easy to verify that $EM_n^2(t)\le\alpha$ for all $t\in[0,1]$, using (3.3) and the fact that

$$E_{i-1}(I_{[|X_{ni}|>\delta]} - P\{|X_{ni}|>\delta\})^2 \;\le\; P\{|X_{ni}|>\delta\}$$

Now

$$P(\sum_i I_{[|X_{ni}|>\delta]}\ge 1) \le P(|M_n(1)|\ge 1-\alpha) + P(\sum_i P_{i-1}\{|X_{ni}|>\delta\}>\alpha)$$

and since $P(|M_n(1)|\ge 1-\alpha)\le \alpha(1-\alpha)^{-2}$ [by Chebychev's inequality], where α can be chosen arbitrarily small, (4.1 ′) follows.

It remains to show that $\sum_i X_{ni}^2\, I_{[i\le k_n(t)]} \xrightarrow{P} t$ as $n\longrightarrow\infty$ for each $0<t<1$. Observe that if U_{ni} denotes the truncated random variable $X_{ni}I_{[|X_{ni}|\le\delta]}$, then $[i-1< k_n(t),\ \sum_i E_{i-1}U_{ni}^2\le t+\alpha]\in\mathcal{F}_{n,i-1}$, so that

$$\overline{M}_n(t) \equiv \sum_i (U_{ni}^2 - E_{i-1}\{U_{ni}^2\})\, I_{[i\le k_n(t),\, \sum_{j\le i} E_{j-1}U_{nj}^2\le t+\alpha]}$$

is a $\{\mathcal{G}_n(t)\}$ martingale. Again, repeated use of (3.3) shows easily that the variance of $\overline{M}_n(t)$ is

$$\le\; E\left\{\sum_i E_{i-1}(U_{ni}^4)\, I_{[i\le k_n(t),\, \sum_{j\le i} E_{j-1}(U_{nj}^2)\le(t+\alpha)]}\right\}\delta^2$$

where almost surely the conditional variance U_{ni}^2 given $\mathcal{F}_{n,i-1}$ has been bounded above by $E_{i-1}U_{ni}^4$, and where U_{ni}^4 has been bounded above by $\delta^2 U_{ni}^2$. Thus for each $t\le 1$ and arbitrary $\alpha>0$, one can choose $\delta>0$ small enough and then $n_0\equiv n_0(\delta)$ large enough so that by (4.1 ′), (4.2) and (4.1), for all $n\ge n_0$ the probability is $\ge 1-\alpha$ that simultaneously

$$t-\alpha \le \sum_i E_{i-1}\{X_{ni}^2\, I_{[|X_{ni}|\le\delta]}\}\, I_{[i\le k_n(t]} \;\le t+\alpha$$

and

$$\sum_{i\le k_n(t)} (X_{ni}^2\, I_{[|X_{ni}|\le\delta]} - E_{i-1}\{X_{ni}^2\, I_{[|X_{ni}|\le\delta]}\})$$

Hence for each $n \geq n_0$,

$$P\{|\sum_i X_{ni}^2 \, I_{[i \leq k_n(t)]} - t| \geq 2\alpha\} \leq \alpha$$

Exercise 8 *Suppose (4.1)–(4.3) hold for* $\{X_{ni}\}$, *and define*

$$Y_{ni} \;=\; X_{ni} \, I_{[|X_{ni}| \leq \frac{1}{2}]} - E_{i-1}\{X_{ni} \, I_{[|X_{ni}| \leq \frac{1}{2}]}\}$$

Show that (4.1 ′) and (4.2 ′) hold for the array $\{Y_{ni}\}$.

This exercise implies that there is no loss in generality in proving Theorem 4.1 under the auxiliary assumption

$$E_{i-1}(X_{ni}) \;=\; 0 \qquad \text{and} \qquad |X_{ni}| \leq 1 \quad \text{a.s.} \quad \text{for} \quad 1 \leq i \leq k, \; n \geq 1 \qquad\qquad (4.3')$$

since (4.3 ′) does hold for the array $\{Y_{ni}\}$, and if (4.1)-(4.3) hold and $s \leq t \leq 1$, then $|\sum_{i \leq k_n(s)} (Y_{ni} - X_{ni})| \leq$

$$\sum_i |X_{ni}| \, I_{[|X_{ni}| > \frac{1}{2}]} + \sum_i |E_{i-1} X_{ni}| \, + \, 2 \sum_i E_{i-1}\{X_{ni}^2 \, I_{|X_{ni}| > \frac{1}{2}}\}$$

the three terms of which $\xrightarrow{P} 0$ as $n \longrightarrow \infty$ by (4.1 ′), (4.3), and (4.1) respectively.

Proof of Theorem from (4.1 ′)-(4.3 ′). Weak convergence of the finite-dimensional distributions of $W_n(\cdot)$ will be proved by first establishing a Central Limit Theorem (**CLT**) for each sequence of random variables $W_n(s) = \sum_{i \leq k_n(s)} X_{ni}$ for fixed $s \in [0, 1]$. Suppose it has been proved for each s that $W_n(s) \longrightarrow N(0, s)$ in distribution. Then fix any $0 < s_1 < s_2 < \cdots < s_m \leq 1$ and any $\alpha_1, \alpha_2, \ldots \alpha_m \in \mathbb{R}$ with $\sum_{i=1}^m |\alpha_j| \leq 1$, and define

$$U_{ni}(\alpha, \mathbf{s}) \;\equiv\; X_{ni} \,\cdot\, \sum_{j=1}^m \alpha_j \, I_{[i \leq k_n(s)]}$$

Then the variables $U_{ni}(\alpha, \mathbf{s})$ are measurable with respect to $\mathcal{F}_{ni}$, and the array $\{U_{ni}(\alpha, \mathbf{s})\}$ satisfies (4.1 ′) and (4.3 ′) by inspection. Also, by (4.2 ′) for $\{X_{ni}\}$, when $n \longrightarrow \infty$

$$\begin{aligned} &\textstyle\sum_i U_{ni}^2(\alpha, \mathbf{s}) \;=\; \sum X_{ni}^2 \, (\sum_{j=1}^m \alpha_i \, I_{[i \leq k_n(s_j)]})^2 \\ &= \textstyle\sum_{i \leq k_n(s_1)} X_{ni}^2 \, (\alpha_1 + \cdots + \alpha_m)^2 \, + \, \sum_i X_{ni}^2 \, I_{[k_n(s_1) < i \leq k_n s(2)]} \, (\alpha_2 + \cdots + \alpha_m)^2 \\ &+ \cdots \textstyle\sum_i X_{ni}^2 \, I_{[k_n(s_{m-1}) < i \leq k_n(s_m)]} \, \alpha_m^2 \;\xrightarrow{P}\; \sigma_m^2(\alpha, \mathbf{s}) \;\equiv\; \sum_{i=1}^m (s_i - s_{i-1})(\sum_{j=i}^m \alpha_j)^2 \end{aligned}$$

where by convention $s_0 = 0$. The Central Limit Theorem for arrays satisfying (4.1 ′)–(4.3 ′) would then imply

$$\sum_{j=1}^m \alpha_j \, W_n(s_j) \;\xrightarrow{D}\; N(0, \sigma^2(\alpha, \mathbf{s})) \qquad \text{as} n \longrightarrow \infty$$

and the limiting distribution is precisely the same as that of $\sum_i \alpha_i W(s_i)$ [compare (2.8)]. Therefore, by the Cramèr-Wold trick of taking joint characteristic functions and applying the Lévy Continuity Theorem], it follows that

$$(W_n(s_1), \ldots, W_n(s_m)) \stackrel{D}{\longrightarrow} (W(s_1), \ldots, W(s_m)) \qquad \text{in} \quad \mathbb{R}^m \quad \text{as} \quad n \longrightarrow \infty$$

Now if $\{X_{ni}\}$ satisfies (4.1 ′), (4.3 ′), and (4.2 ′) for the single fixed value $t = s \in [0,1]$, then $\{Z_{ni}\}$ defined by

$$Z_{ni} \equiv X_{ni} I_{[i \le k_n(s), \sum_{j<i} X_{nj}^2 \le 2]}$$

does also; and $\sum_i Z_{ni}^2 \le 3$ almost surely and $P\{Z_{ni} \ne X_{ni}$ for some $i = 1, \ldots, k_n(s)\} \to 0$ as $n \longrightarrow \infty$. Thus a CLT for $\sum_{i \le k_n(s)} Z_{ni}$, which is what we shall prove, implies a CLT for $W_n(s) = \sum_{i \le k_n(s)} X_{ni}$. Define

$$I_n \equiv \exp\left(it \sum_j Z_{nj} \right) \quad , \qquad T_n \equiv \prod_j (1 + it\, Z_{nj})$$

where both the summation and product range over all $j = 1, \ldots, k_n(s)$, and where i now denotes a complex square root of -1. Observe that $E|I_n|^2 = 1$ and, by conditioning repeatedly, that $ET_n = 1$. The following fact noted by McLeish (1974) will also be used below:

$$\text{if} \quad e^{ix} \equiv (1 + ix)\, e^{-x^2/2 + r(x)} , \qquad \text{then} \quad |r(x) \le |x|^3 \quad \text{for} \quad |x| \le 1 \tag{4.4}$$

where x is real. Next, using $1 + x \le e^x$, calculate

$$E|T_n|^2 = E \prod_j \left(1 + t^2 Z_{nj}^2\right) \le E \exp(t^2 \sum_j Z_{nj}^2) \le e^{3t^2} < \infty$$

On the other hand,

$$I_n = T_n\, e^{-t^2/2} + T_n \left(\exp\{-\frac{t^2}{2} \sum_j Z_{nj}^2 + \sum_j r(tZ_{nj})\} - e^{-t^2/2} \right)$$

and by the inequality $|x + y|^2 \le |x|^2 + |y|^2$,

$$E|I_n - T_n e^{-t^2/2}|^2 \le 2 \left(E|I_n|^2 + E|T_n|^2\, e^{-t^2} \right) \le K(t) < \infty$$

where $K(\cdot)$ does not depend on n. By (4.4), (4.1 ′), and (4.2 ′), for $|t| < 1$

$$|\sum_j r(tZ_{nj})| \le |t|^3 (\sum_j Z_{nj}^2) \max_j |Z_{nj}| \stackrel{P}{\longrightarrow} 0$$

Thus for large n,

$$P\{|\exp(-\frac{t^2}{2} \sum_j Z_{nj}^2 + \sum_j r(tZ_{nj})) - e^{-t^2/2}| > \delta\} < \delta$$

and

$$P\{|T_n| > \delta^{1/2}\} \,<\, \delta^{-1}\, E|T_n|^2 \,\le\, \delta^{-1}\, e^{3\,t^2} ????$$

so that $|I_n - T_n e^{-t^2/2}|$ converges in probability to 0 as $n \longrightarrow \infty$ and has second moment uniformly bounded by $K(t)$, and hence converges to 0 in the mean. But $E\{T_n e^{-t^2/2}\} = e^{-t^2/2}$ then implies $EI_n \to e^{-t^2/2}$ as $n \longrightarrow \infty$ for each $t \in (-1,1)$. By the Lévy Continuity Theorem, our CLT for $\{Z_{nj}\}$, and therefore the weak convergence of the finite-dimensional distributions of $W_n(\cdot)$ to those of $W(\cdot)$, is proved.

By the weak-convergence theory of Chapter 2, all that remains in proving $W_n(\cdot) \stackrel{D}{\longrightarrow} W(\cdot)$ in $D[0,1]$ as $n \longrightarrow \infty$ is to verify condition (2.3) of Theorem 2.5, that is,that for each α, $\beta > 0$, there exist $\delta > 0$ and $n_1 < \infty$ such that

$$\text{for} \quad n \ge n_1, \qquad P\{ \sup_{0<s,t<1,\, |s-t|<\delta} |W_n(s) - W_n(t)| \,\} \,>\beta\} \,<\alpha \tag{4.5}$$

Condition (4.5), which we shall prove below, is apparently weaker than (2.3) of Theorem 2.5. But it implies first that there exists a sequence $\{\tilde{W}_n\}$ of linearly-interpolated continuous processes for which (4.5) also holds and $\|W_n - \tilde{W}_n\| \stackrel{P}{\longrightarrow} 0$ as $n \longrightarrow \infty$. Then, by choosing $\delta_0 = \delta(\alpha,\beta)$ still smaller, (4.5) can be seen to hold for $\tilde{W}_n$ for all $n \ge 1$. Finally, the convergence of the finite-dimensional distributions of $\tilde{W}_n$ together with Theorem 2.5 imply that $\tilde{W}_n$ and therefore W_n converge in distribution to W in D as $n \longrightarrow \infty$. To prove (4.5) in the current context, again pass from $\{X_{ni}\}$ to the equivalent [by (4.2 $'$)] sequence $Z_{ni} \equiv X_{ni}\, I_{[\sum_{j<i} X_{nj} \le 2]}$. Since $P\{\max_i |Z_{ni} - X_{ni}| > 0\} \to 0$ as $n \longrightarrow \infty$, it suffices to check (4.5) with $W_n(\cdot)$ replaced by the martingale $\overline{W}_n(\cdot) \equiv \sum_{i \le k_n(\cdot)} Z_{ni}$. Now $P\{\sup_{|s-t|} |\overline{W}_n(s) - \overline{W}_n(t)| > \alpha\} \le$

$$\sum_{j:\, j\delta \le 1} P\{ \sup_{j\delta < t \le (j+1)\delta} |\overline{W}_n(t) - \overline{W}_n(j\delta)| > \beta/3\} \le C(\beta) \sum_{j:\, j\delta \le 1} E|\overline{W}_n((j+1)\delta) - \overline{W}_n(j\delta)|^3$$

by the Submartingale Maximal Inequality applied to the martingales $W_n(\cdot) - W_n(j\delta)$ on $[j\delta, (j+1)\delta]$. Next, by the Burkholder Inequality applied to the same martingales, the last expression is

$$\le \overline{C}(\beta) \sum_{j:\, j\delta \le 1} E|\sum_k Z_{nk}^2\, I_{[k_n(j\delta < k \le k_n((j+1)\delta)]}|^{3/2}$$

where $C(\beta)$ and $\overline{C}(\beta)$ do not depend upon n or δ. But as $n \longrightarrow \infty$, the random variables

$$\sum_i Z_{ni}^2\, I_{[k_n(j\delta) < k \le k_n((j+1)\delta)]}$$

are uniformly bounded and converge in probability for each j with $(j+1)\delta \le 1$ to δ as $n \longrightarrow \infty$. [This is simply (4.2 $'$) for Z_{ni}.] Therefore

$$\limsup_{n \longrightarrow \infty} P\{ \sup_{|s-t|<\delta} |\overline{W}_n(s) - \overline{W}_n(t)| > \beta\} \le \overline{C}(\beta)(1 + \frac{1}{\delta})\delta^{3/2}$$

which can be made as small as desired by choosing small $\delta > 0$. The proofs of (4.5) and Theorem 4.1 are complete. □

The Central Limit Theorem used in showing weak convergence of finite-dimensional distributions is due essentially to Brown and Dvoretzky (both 1971, cited in McLeish 1974], although the proof given here is McLeish's. An immediate corollary of Theorem 4.1 is the famous Donsker Invariance Principle:

Corollary 4.2 *If* $\{X_i\}_{i=1}^{\infty}$ *is an independent and identically distributed sequence with mean* μ *and finite variance* σ^2, *then as* $n \longrightarrow \infty$,

$$\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{[nt]} (X_i - \mu) \xrightarrow{D} W(t) \qquad \text{in} \quad D[0,1]$$

Proof. In Theorem 4.1, take $X_{ni} \equiv (X_i - \mu)/(\sigma\sqrt{n})$, $\mathcal{F}_{ni} \equiv \sigma(X_1, \ldots, X_i)$, and $k_n(t) \equiv [nt]$, where $[\cdot]$ denotes the greatest-integer function. Then (4.3) is obvious, (4.2) follows from $E_{i-1}(X_{ni}^2) = 1/n$, and (4.1) follows from

$$\sum_{i=1}^{[nt]} E_{i-1}\{X_{ni}^2 \, I_{[|X_{ni}|>\delta]}\} \le \sigma^{-2} \, E\{(X_1 - \mu)^2 \, I_{[|X_1-\mu|>\sigma\delta\sqrt{n}]}\}$$

4.2 Discrete-time Theory : Extensions

There are several complements to the McLeish Theorem 4.1 which will make the later extension to continuous-time martingale sequences easier. They concern the same setting as Theorem 4.1, but the time-scales $k_n(\cdot)$ will now be allowed to take infinite values or to be random right-continuous integer-valued processes such that each $k_n(t)$ is a stopping time. First, if k_n is allowed to be $+\infty$, there is no change in the statement or proof of Theorem 4.1. However, to make sense of the case $k_n = \infty$, one must observe that the condition (4.2) or (4.2 ′) is effectively ensuring that the sum $\sum_i X_{ni}^2$ is almost surely finite. Therefore the Theorem is also valid if $k_n(\cdot)$ is a right-continuous nondecreasing integer-valued process such that for each $s \in [0,1]$, $k_n(s)$ is a stopping time. That is, if (4.1) and (4.3) are assumed for the variables X_{ni}, and if (4.2) is assumed for the random variables $X'_{ni} \equiv X_{ni} \, I_{[k_n(1)\ge i]}$, then measurability of $[k_n(1) \ge i]$ with respect to $\mathcal{F}_{n,i-1}$ implies by (3.1) that

$$E\{(X'_{ni})^r\} = E\{(X_{ni})^r\} \cdot I_{[k_n(1)\ge i]} \quad \text{a.s. for} \quad r = 1, 2, \ldots$$

and assumptions (4.1)–(4.3) hold for $\{X'_{ni}\}$. The reader should check that the steps in the proof of Theorem 4.1, applied to the partial sums of the X'_{ni}, remain valid without change when the $k_n(t)$ are stopping times, so that as $n \longrightarrow \infty$,

$$\sum_i X'_{ni}\, I_{[k_n(\cdot)]} \xrightarrow{D} W(\cdot) \quad \text{in} \quad D[0,1]$$

Suppose next, as in the previous paragraph, that (4.1) and (4.3) hold for $\{X_{ni}\}$, and that $\{k_n(t)\}$ is a right-continuous non-decreasing integer-valued family of stopping times for $t \in [0,T]$, but now assume

$$\sum_i E_{i-1}\{X_{ni}^2\}\, I_{[k_n(t) \geq i]} \xrightarrow{P} F(t) \quad \text{as } n \longrightarrow \infty \quad \text{for each } t \in [0,T] \tag{4.6}$$

where $F(\cdot)$ is a nonrandom nondecreasing continuous function with $F(0) = 0$ and $F(T) > 1$. Then define $\bar{k}_n(t) \equiv k_n(F^{-1}(t))$ for $0 \leq t \leq F(1)$, where $F^{-1}(t) \equiv \inf\{x : F(x) > t\}$. Again $\{k_n(t)\}$ is a nondecreasing right-continuous family of stopping times. Since $F(F^{-1}(t)) = t$ for all t, the previous paragraph implies that

$$\sum_i X_{ni}\, I_{[1 \leq i \leq \bar{k}_n(\cdot)]} \xrightarrow{D} W(\cdot) \quad \text{in} \quad D[0,F(1)] \quad \text{as} ; n \longrightarrow \infty \tag{4.7}$$

But continuity of F means that $f \mapsto f \circ F$ is a continuous functional from $D[0,F(1)]$ to $D[0,1]$ with respect to the topologies of uniform convergence. Hence by Corollary 2.7 and (4.7), as $n \longrightarrow \infty$

$$\sum_i X_{ni}\, I_{[1 \leq i \leq \bar{k}_n(F(\cdot))]} \xrightarrow{D} W \circ F \quad \text{in} \quad D[0,1]$$

In order to conclude $\sum_{i \leq k_n(\cdot)} X_{ni} \xrightarrow{D} W \circ F$ in $D[0,1]$, we will show as $n \longrightarrow \infty$,

$$\sup_{0 \leq t \leq 1} \Big| \sum_i X_{ni}\, I_{[k_n(F^{-1}(F(t))) < i \leq k_n(t)]} \Big| \xrightarrow{P} 0 \tag{4.8}$$

Making use of the same reductions as in the proof of Theorem 4.1, we assume without loss of generality that $E_{i-1} X_{ni} = 0$ and $|X_{ni}| \leq 1$ almost surely. Also, by replacing X_{ni} if necessary with $X_{ni} I_{[\sum_{j \leq i} E_{i-1}(X_{ni}^2) \leq F(1)+1]}$, we can assume $\sum_i E_{i-1} X_{ni}^2 \leq F(1)+2$ almost surely. Next, since $F^{-1}(F(t)) \geq t$ and F is constant on the interval $[t, F^{-1}(F(t))]$, for each $\delta > 0$ there exists a finite system $\{(a_i, b_i]\}_{i=1}^p$ of disjoint intervals in $[0,1]$ such that $\sum_i (F(b_i) - F(a_i)) \leq \delta$, and $\{t \in [0,1] : t < F^{-1}(F(t))\} \subset \bigcup_{i=1}^p (a_i, b_i]$. To see this, note that the measure μ on $[0,\infty)$ with $F(t) = \mu([0,t])$ must assign measure 0 to $[t, F^{-1}(F(t))]$. Now define a martingale with respect to $\mathcal{G}_n(t) \equiv \mathcal{F}_{n,k_n(t)}$ by

$$S_{nk} \equiv \sum_{j=1}^m \sum_i X_{ni}\, I_{[k_n(a_j) < i \leq k \wedge k_n(b_j)]}$$

again using $\mathcal{F}_{n,i-1}$-measurability of each event $[k_n(s) < i]$, and note that the left-hand side of (4.8) is $\leq 2 \max_{1 \leq k \leq k_n(1)} |S_{nk}|$. But by the SubmartingaleMaximal Inequality,

$$E \max_k |S_{nk}|^2 = E \max_k |S_{n,k \wedge k_n(1)}|^2 \leq 4\, E S_{n,k_n(1)}^2 = 4\, E \sum_k (S_{nk} - S_{n,k-1})^2\, I_{[k \leq k_n(1)]} \qquad [\text{by}(3.1)] = 4 \sum_{j=1}^m E \sum$$

which converges in probability as $n \longrightarrow \infty$ to $4 \sum_{j=1}^{m} (F(b_j) - F(a_j))$, which is $\leq 4\delta$. Therefore (4.8) holds, and we have proved

Theorem 4.3 (Modified McLeish Theorem) *Let* $\{X_{ni}\}$ *be an array of square-integrable random variables on* $(\Omega, \mathcal{F}, P)$ *and* $\{\mathcal{F}_{ni}\}$ *an array of sub-*σ*-algebras of* $\mathcal{F}$ *such that* $\mathcal{F}_{ni} \subset \mathcal{F}_{n,i+1}$ *and* X_{ni} *is* $\mathcal{F}_{ni}$*-measurable for all* n *and* i. *Let* $k_n(\cdot)$ *for each* n *be a non-decreasing right-continuous integer-valued process of* $\{\mathcal{F}_{ni}\}$ *stopping times with* $k_n(0) = 0$; *and assume as* $n \longrightarrow \infty$,

$$(i) \quad \text{for all } \delta > 0, \qquad \sum_{i \leq k_n(1)} E_{i-1}\{X_{ni}^2 I_{[|X_{ni}|>\delta]} \stackrel{P}{\longrightarrow} 0$$

$$(ii) \quad \text{for } t \in [0,1], \qquad \sum_{i \leq k_n(t)} E_{i-1} X_{ni}^2 \stackrel{P}{\longrightarrow} F(t)$$

$$(iii) \quad \sum_{i \leq k_n(1)} |E_{i-1}(X_{ni})| \stackrel{P}{\longrightarrow} 0$$

where $F(\cdot)$ *is a nonrandom continuous function with* $F(0) = 0$. *Then as* $n \longrightarrow \infty$,

$$\sum_i X_{ni}\, I_{[i \leq k_n(\cdot)]} \stackrel{D}{\longrightarrow} W \circ F \quad \text{in} \quad D[0,1]$$

where $W(\cdot)$ *is Wiener process on* $[0, F(1)]$.

An immediate corollary of Theorem 4.1 places it squarely in the context of the heuristic principle (3.2). For a reference to this Corollary, where it is attributed to D. Freedman and used to prove extensions of Theorem 4.3 to cases where the variance processes in condition (ii) converge to possibly random limits in probability, see Durrett and Resnick (1978).

Corollary 4.4 (Freedman) *Let* $\{X_{ni}, \mathcal{F}_{ni}\}$ *be as in Theorem 4.3, except that condition (ii) need not hold, and let* k_n *for each* n *be a stopping time for which* $\sum_{i \leq k_n} E_{i-1}(X_{ni}^2) \longrightarrow \infty$ *and for each* $\delta > 0$,

$$\sum_{i \leq k_n} E_{i-1}\{X_{ni}^2\, I_{[|X_{ni}|>\delta]} \stackrel{P}{\longrightarrow} 0 \quad as \quad n \longrightarrow \infty$$

Then for **either** *of the two time-scales*

$$\overline{k}_n(t) \equiv \inf\{j : \sum_{i=1}^{j+1} E_{i-1}(X_{ni}^2) > t\}$$

or

$$\overline{k}_n^*(t) \equiv \inf\{j : \sum_{i=1}^{j} X_{ni}^2 \geq t\}$$

in place of $k_n(\cdot)$ *on* $[0,1]$,

$$\sum_i X_{ni}\, I_{[i\le k_n(\cdot)]} \xrightarrow{D} W(\cdot) \quad in\;\; D[0,1] \quad as\;\; n\longrightarrow\infty$$

Proof. The $k_n(\cdot)$ so defined are as required in Theorem 4.3, and (i)-(iii) are obviously satisfied under the definition $k_n(\cdot)\equiv \overline{k}_n(\cdot)$. The proof is completed by

Exercise 9 *Using the definition* $k_n(\cdot)\equiv \overline{k}^*_n(\cdot)$ *in Corollary 4.4, along with any of the proof ideas and reductions of this Section, show that (i)-(iii) of the Modified McLeish Theorem 4.3 are satisfied.* □

Remark 4.2 *If we really are in a situation where we can say only that* $\sum_{i\le k_n} E_{i-1}X^2_{ni}\longrightarrow\infty$, *without being able to say how large* $k_n(t)$ *is asymptotically, we will be wasting information in collecting a potential experimental data sample* $\{X_{ni}\}$ *of size* k_n *while basing stopping-decisions and inference on* $k_n(1)$ *observations. The waste occurs because* $k_n(1)$ *may well be of a smaller order of magnitude than* k_n. □

4.3 Continuous time: the Rebolledo Theorem

As in Chapter 3, the passage from discrete- to continuous-time theorems can be accomplished by limiting operations in probability and in the mean once we have restricted consideration to martingales whose squares have calculable compensators. The idea of proving Rebolledo's theorem in this way is due to Helland (1982).

Theorem 4.5 (Rebolledo, 1977, 1980) *Suppose that for each* $n\ge 1$, $M_n(\cdot)$ *is a locally square-integrable* $\{\mathcal{F}_n(t)\}_t$ *martingale on* $[0,T)$, *almost surelyin* $D[0,t]$ *for each* $t<T$, *and which satisfies any of the conditions [of Theorems 3.3 and 3.5, or of Theorem 5.4 below] for "calculability" of the variance-process* $<M_n>(\cdot)$. *Suppose that for each* $\delta>0$ *and fixed* $t_0<T$,

$$\sum_{s:\, s\le t_0} |\Delta M_n(s)|^2\, I_{[|\Delta M_n(s)|\ge\delta]} \xrightarrow{P} 0 \qquad as \quad n\longrightarrow\infty \tag{4.9}$$

and that for a nonrandom continuous nondecreasing function $F(\cdot)$ *with* $F(0)=0$,

$$<M_n>(s) \xrightarrow{P} F(s) \qquad as \quad n\longrightarrow\infty\;, \quad for\ each\ \ s\in[0,t_0] \tag{4.10}$$

Then $M_n \xrightarrow{D} W\circ F$ *in* $D[0,t_0]$ *as* $n\longrightarrow\infty$. *If (4.9) holds, but the convergence in (4.10) is assumed to hold only for a single fixed* $s\in(0,t_0]$, *then* $M_n(s)\xrightarrow{D} W(F(s))$ *in* $\mathbb{R}$.

Remark 4.3 *Assumption (4.9), which is equivalent to "uniform asymptotic negligibility in probability of jumps" ,i.e., to* $\sup_{0 \leq s \leq t_0} |\Delta M_n(s)| \xrightarrow{P} 0$, *was called by Rebolledo an* **Asymptotic Rarefaction of Jumps** *(ARJ) Condition.* □