Chapter 6

Martingale et al.

SECTION 1 gives some examples of martingales, submartingales, and supermartingales.
SECTION 2 introduces stopping times and the sigma-fields corresponding to "information available at a random time." A most important Stopping Time Lemma is proved, extending the martingale properties to processes evaluted at stopping times.
SECTION 3 shows that positive supermartingales converge almost surely.
SECTION 4 presents a condition under which a submartingale can be written as a difference between a positive martingale and a positive supermartingale (the Krickeberg decomposition). A limit theorem for submartingales then follows.
*SECTION *5 proves the Krickeberg decomposition.*
*SECTION *6 defines uniform integrability and shows how uniformly integrable martingales are particularly well behaved.*
*SECTION *7 show that martingale theory works just as well when time is reversed.*
*SECTION *8 uses reverse martingale theory to study exchangeable probability measures on infinite product spaces. The de Finetti representation and the Hewitt-Savage zero-one law are proved.*

1. What are they?

The theory of martingales (and submartingales and supermartingales and other related concepts) has had a profound effect on modern probability theory. Whole branches of probability, such as stochastic calculus, rest on martingale foundations. The theory is elegant and powerful: amazing consequences flow from an innocuous assumption regarding conditional expectations. Every serious user of probability needs to know at least the rudiments of martingale theory.

A little notation goes a long way in martingale theory. A fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$ sits in the background. The key new ingredients are:

(i) a subset T of the extended real line $\overline{\mathbb{R}}$;

(ii) a ***filtration*** $\{\mathcal{F}_t : t \in T\}$, that is, a collection of sub-sigma-fields of $\mathcal{F}$ for which $\mathcal{F}_s \subseteq \mathcal{F}_t$ if $s < t$;

(iii) a family of integrable random variables $\{X_t : t \in T\}$ ***adapted*** to the filtration, that is, X_t is $\mathcal{F}_t$-measurable for each t in T.

The set T has the interpretation of time, the sigma-field $\mathcal{F}_t$ has the interpretation of *information available at time* t, and X_t denotes some random quantity whose value $X_t(\omega)$ is revealed at time t.

<1> **Definition.** *A family of integrable random variables* $\{X_t : t \in T\}$ *adapted to a filtration* $\{\mathcal{F}_t : t \in T\}$ *is said to be a* ***martingale*** *(for that filtration) if*

$$\text{(MG)} \qquad X_s \underset{a.s.}{=} \mathbb{P}(X_t \mid \mathcal{F}_s) \qquad \text{for all } s < t.$$

Equivalently, the random variables should satisfy

$$\text{(MG)}' \qquad \mathbb{P}X_s F = \mathbb{P}X_t F \qquad \text{for all } F \in \mathcal{F}_s, \text{ all } s < t.$$

REMARK. Often the filtration is fixed throughout an argument, or the particular choice of filtration is not important for some assertion about the random variables. In such cases it is easier to talk about a martingale $\{X_t : t \in T\}$ without explicit mention of that filtration. If in doubt, we could always work with the ***filtration!natural***, $\mathcal{F}_t := \sigma\{X_s : s \le t\}$, which takes care of adaptedness, by definition.

Analogously, if there is a need to identify the filtration explicitly, it is convenient to speak of a martingale $\{(X_t, \mathcal{F}_t) : t \in T\}$, and so on.

Property (MG) has the interpretation that X_s is the best predictor for X_t based on the information available at time s. The equivalent formulation (MG)$'$ is a minor repackaging of the definition of the conditional expectation $\mathbb{P}(X_t \mid \mathcal{F}_s)$. The $\mathcal{F}_s$-measurability of X_s comes as part of the adaptation assumption. Approximation by simple functions, and a passage to the limit, gives another equivalence,

$$\text{(MG)}'' \qquad \mathbb{P}X_s Z = \mathbb{P}X_t Z \qquad \text{for all } Z \in \mathcal{M}_{\text{bdd}}(\mathcal{F}_s), \text{ all } s < t,$$

where $\mathcal{M}_{\text{bdd}}(\mathcal{F}_s)$ denotes the set of all bounded, $\mathcal{F}_s$-measurable random variables. The formulations (MG)$'$ and (MG)$''$ have the advantage of removing the slippery concept of conditioning on sigma-fields from the definition of a martingale. One could develop much of the basic theory without explicit mention of conditioning, which would have some pedagogic advantages, even though it would obscure one of the important ideas behind the martingale concept.

Several of the desirable properties of martingales are shared by families of random variables for which the defining equalities (MG) and (MG)$'$ are relaxed to inequalities. I find that one of the hardest things to remember about these martingale relatives is which name goes with which direction of the inequality.

<2> **Definition.** *A family of integrable random variables* $\{X_t : t \in T\}$ *adapted to a filtration* $\{\mathcal{F}_t : t \in T\}$ *is said to be a* ***submartingale*** *(for that filtration) if it satisfies any (and hence all) of the following equivalent conditions:*

$$\textit{(subMG)} \qquad X_s \le \mathbb{P}(X_t \mid \mathcal{F}_s) \qquad \text{for all } s < t, \qquad \text{almost surely}$$

$$\text{(subMG)}' \qquad \mathbb{P}X_s F \le \mathbb{P}X_t F \qquad \text{for all } F \in \mathcal{F}_s, \text{ all } s < t.$$

$$\text{(subMG)}'' \qquad \mathbb{P}X_s Z \le \mathbb{P}X_t Z \qquad \text{for all } Z \in \mathcal{M}^+_{\text{bdd}}(\mathcal{F}_s), \text{ all } s < t,$$

The family is said to be a ***supermartingale*** *(for that filtration) if* $\{-X_t : t \in T\}$ *is a submartingale. That is, the analogous requirements (superMG), (superMG)$'$, and (superMG)$''$ reverse the direction of the inequalities.*

REMARK. It is largely a matter of taste, or convenience of notation for particular applications, whether one works primarily with submartingales or supermartingales.

For most of this Chapter, the index set T will be ***discrete***, either finite or equal to $\mathbb{N}$, the set of positive integers, or equal to one of

$$\mathbb{N}_0 := \{0\} \cup \mathbb{N} \qquad \text{or} \qquad \overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\} \qquad \text{or} \qquad \overline{\mathbb{N}}_0 := \{0\} \cup \mathbb{N} \cup \{\infty\}.$$

For some purposes it will be useful to have a distinctively labelled first or last element in the index set. For example, if a limit $X_\infty := \lim_{n\in\mathbb{N}} X_n$ can be shown to exist, it is natural to ask whether $\{X_n : n \in \overline{\mathbb{N}}\}$ also has sub- or supermartingale properties. Of course such a question only makes sense if a corresponding sigma-field $\mathcal{F}_\infty$ exists. If it is not otherwise defined, I will take $\mathcal{F}_\infty$ to be the sigma-field $\sigma\left(\cup_{i<\infty}\mathcal{F}_i\right)$.

Continuous time theory, where T is a subinterval of $\overline{\mathbb{R}}$, tends to be more complicated than discrete time. The difficulties arise, in part, from problems related to management of uncountable families of negligible sets associated with uncountable collections of almost sure equality or inequality assertions. A nontrivial part of the continuous time theory deals with sample path properties, that is, with the behavior of a process $X_t(\omega)$ as a function of t for fixed ω or with properties of X as a function of two variables. Such properties are typically derived from probabilistic assertions about finite or countable subfamilies of the $\{X_t\}$ random variables. An understanding of the discrete-time theory is an essential prerequisite for more ambitious undertakings in continuous time—see Appendix E.

For discrete time, the (MG)$'$ property becomes

$$\mathbb{P}X_n F = \mathbb{P}X_m F \qquad \text{for all } F \in \mathcal{F}_n, \text{ all } n < m.$$

It suffices to check the equality for $m = n + 1$, with $n \in \mathbb{N}_0$, for then repeated appeals to the special case extend the equality to $m = n + 2$, then $m = n + 3$, and so on. A similar simplification applies to submartingales and supermartingales.

<3> **Example.** Martingales generalize the theory for sums of independent random variables. Let $\xi_1, \xi_2, \ldots$ be independent, integrable random variables with $\mathbb{P}\xi_n = 0$ for $n \geq 1$. Define $X_0 := 0$ and $X_n := \xi_1 + \ldots + \xi_n$. The sequence $\{X_n : n \in \mathbb{N}_0\}$ is a martingale with respect to the natural filtration, because for $F \in \mathcal{F}_{n-1}$,

$$\mathbb{P}(X_n - X_{n-1})F = (\mathbb{P}\xi_n)(\mathbb{P}F) = 0 \qquad \text{by independence.}$$

You could write F as a measurable function of $X_1, \ldots, X_{n-1}$, or of $\xi_1, \ldots, \xi_{n-1}$, if you prefer to work with random variables. □

<4> **Example.** Let $\{X_n : n \in \mathbb{N}_0\}$ be a martingale and let Ψ be a convex function for which each $\Psi(X_n)$ is integrable. Then $\{\Psi(X_n) : n \in \mathbb{N}_0\}$ is a submartingale: the required almost sure inequality, $\mathbb{P}\left(\Psi(X_n) \mid \mathcal{F}_{n-1}\right) \geq \Psi(X_{n-1})$, is a direct application of the conditional expectation form of Jensen's inequality.

The companion result for submartingales is: if the convex Ψ function is increasing, if $\{X_n\}$ is a submartingale, and if each $\Psi(X_n)$ is integrable, then $\{\Psi(X_n) : n \in \mathbb{N}_0\}$ is a submartingale, because

$$\mathbb{P}\left(\Psi(X_n) \mid \mathcal{F}_{n-1}\right) \underset{a.s.}{\geq} \Psi(\mathbb{P}(X_n \mid \mathcal{F}_{n-1}) \underset{a.s.}{\geq} \Psi(X_{n-1}).$$

Two good examples to remember: if $\{X_n\}$ is a martingale and each X_n is square integrable then $\{X_n^2\}$ is a submartingale; and if $\{X_n\}$ is a submartingale then $\{X_n^+\}$ is also a submartingale. □

<5> **Example.** Let $\{X_n : n \in \mathbb{N}_0\}$ be a martingale written as a sum of increments, $X_n := X_0 + \xi_1 + \ldots + \xi_n$. Not surprisingly, the $\{\xi_i\}$ are called ***martingale differences***. Each ξ_n is integrable and $\mathbb{P}(\xi_n \mid \mathcal{F}_{n-1}) \underset{a.s.}{=} 0$ for $n \in \mathbb{N}_0^+$.

A new martingale can be built by weighting the increments using ***predictable*** functions $\{H_n : n \in \mathbb{N}\}$, meaning that each H_n should be an $\mathcal{F}_{n-1}$-measurable random variable, a more stringent requirement than adaptedness. The value of the weight becomes known before time n; it is known before it gets applied to the next increment.

If we assume that each $H_n\xi_n$ is integrable then the sequence

$$Y_n := X_0 + H_1\xi_1 + \ldots + H_n\xi_n$$

is both integrable and adapted. It is a martingale, because

$$\mathbb{P}H_i\xi_i F = \mathbb{P}(X_i - X_{i-1})(H_i F),$$

which equals zero by a simple generalization of (MG)″. (Use Dominated Convergence to accommodate integrable Z.) If $\{X_n : n \in \mathbb{N}_0\}$ is just a submartingale, a similar argument shows that the new sequence is also a submartingale, provided the predictable weights are also nonnegative. □

<6> **Example.** Suppose X is an integrable random variable and $\{\mathcal{F}_t : t \in T\}$ is a filtration. Define $X_t := \mathbb{P}(X \mid \mathcal{F}_t)$. Then the family $\{X_t : t \in T\}$ is a martingale with respect to the filtration, because for $s < t$,

$$\begin{aligned} \mathbb{P}(X_t F) &= \mathbb{P}(XF) && \text{if } F \in \mathcal{F}_t \\ &= \mathbb{P}(X_s F) && \text{if } F \in \mathcal{F}_s \end{aligned}$$

(We have just reproved the formula for conditioning on nested sigma-fields.) □

<7> **Example.** Every sequence $\{X_n : n \in \mathbb{N}_0\}$ of integrable random variables adapted to a filtration $\{\mathcal{F}_n : n \in \mathbb{N}_0\}$ can be broken into a sum of a martingale plus a sequence of accumulated conditional expectations. To establish this fact, consider the increments $\xi_n := X_n - X_{n-1}$. Each ξ_n is integrable, but it need not have zero conditional expectation given $\mathcal{F}_{n-1}$, the property that characterizes martingale differences. Extraction of the martingale component is merely a matter of recentering the increments to zero conditional expectations. Define $\eta_n := \mathbb{P}(\xi_n \mid \mathcal{F}_{n-1})$ and

$$\begin{aligned} M_n &:= X_0 + (\xi_1 - \eta_1) + \ldots + (\xi_n - \eta_n) \\ A_n &:= \eta_1 + \ldots + \eta_n. \end{aligned}$$

Then $X_n = M_n + A_n$, with $\{M_n\}$ a martingale and $\{A_n\}$ a predictable sequence.

Often $\{A_n\}$ will have some nice behavior, perhaps due to the smoothing involved in the taking of a conditional expectation, or perhaps due to some other special property of the $\{X_n\}$. For example, if $\{X_n\}$ were a submartingale the η_i would all be nonnegative (almost surely) and $\{A_n\}$ would be an increasing sequence of random variables. Such properties are useful for establishing limit theory and inequalities—see Example <18> for an illustration of the general method. □

REMARK. The representation of a submartingale as a martingale plus an increasing, predictable process is sometimes called the ***Doob decomposition***. The corresponding representation for continuous time, which is exceedingly difficult to establish, is called the ***Doob-Meyer decomposition***.

2. Stopping times

The martingale property requires equalities $\mathbb{P}X_s F = \mathbb{P}X_t F$, for $s < t$ and $F \in \mathcal{F}_s$. Much of the power of the theory comes from the fact that analogous inequalities hold when s and t are replaced by certain types of random times. To make sense of the broader assertion, we need to define objects such as $\mathcal{F}_\tau$ and X_τ for random times τ.

<8> **Definition.** *A random variable τ taking values in $\overline{T} := T \cup \{\infty\}$ is called a **stopping time** for a filtration $\{\mathcal{F}_t : t \in T\}$ if $\{\tau \le t\} \in \mathcal{F}_t$ for each t in T.*

In discrete time, with $T = \mathbb{N}_0$, the defining property is equivalent to

$$\{\tau = n\} \in \mathcal{F}_n \qquad \text{for each } n \text{ in } \overline{\mathbb{N}}_0,$$

because $\{\tau \le n\} = \bigcup_{i \le n}\{\tau = i\}$ and $\{\tau = n\} = \{\tau \le n\}\{\tau \le n-1\}^c$.

<9> **Example.** Let $\{X_n : n \in \mathbb{N}_0\}$ be adapted to a filtration $\{\mathcal{F}_n : n \in \mathbb{N}_0\}$, and let B be a Borel subset of $\mathbb{R}$. Define $\tau(\omega) := \inf\{n : X_n(\omega) \in B\}$, with the interpretation that the infimum of the empty set equals $+\infty$. That is, $\tau(\omega) = +\infty$ if $X_n(\omega) \notin B$ for all n. The extended-real-valued random variable τ is a stopping time because

$$\{\tau \le n\} = \bigcup_{i \le n}\{X_i \in B\} \in \mathcal{F}_n \qquad \text{for } n \in \mathbb{N}_0.$$

It is called the ***first hitting time*** of the set B. Do you see why it is convenient to allow stopping times to take the value $+\infty$? □

If $\mathcal{F}_i$ corresponds to the information available up to time i, how should we define a sigma-field $\mathcal{F}_\tau$ to correspond to information available up to a random time τ? Intuitively, on the part of Ω where $\tau = i$ the sets in the sigma-field $\mathcal{F}_\tau$ should be the same as the sets in the sigma-field $\mathcal{F}_i$. That is, we could hope that

$$\{F\{\tau = i\} : F \in \mathcal{F}_\tau\} = \{F\{\tau = i\} : F \in \mathcal{F}_i\} \qquad \text{for each } i.$$

These equalities would be suitable as a definition of $\mathcal{F}_\tau$ in discrete time; we could define $\mathcal{F}_\tau$ to consist of all those F in $\mathcal{F}$ for which

<10> $$F\{\tau = i\} \in \mathcal{F}_i \qquad \text{for all } i \in \overline{\mathbb{N}}_0.$$

For continuous time such a definition could become vacuous if all the sets $\{\tau = t\}$ were negligible, as sometimes happens. Instead, it is better to work with a definition that makes sense in both discrete and continuous time, and which is equivalent to <10> in discrete time.

<11> **Definition.** *Let τ be a stopping time for a filtration $\{\mathcal{F}_t : t \in T\}$, taking values in $\overline{T} := T \cup \{\infty\}$. If the sigma-field $\mathcal{F}_\infty$ is not already defined, take it to be $\sigma(\cup_{t \in T}\mathcal{F}_t)$. The pre-$\tau$ sigma-field $\mathcal{F}_\tau$ is defined to consist of all F for which $F\{\tau \le t\} \in \mathcal{F}_t$ for all $t \in \overline{T}$.*

The class $\mathcal{F}_\tau$ would not be a sigma-field if τ were not a stopping time: the property $\Omega \in \mathcal{F}_\tau$ requires $\{\tau \le t\} \in \mathcal{F}_t$ for all t.

> REMARK. Notice that $\mathcal{F}_\tau \subseteq \mathcal{F}_\infty$ (because $F\{\tau \le \infty\} \in \mathcal{F}_\infty$ if $F \in \mathcal{F}_\tau$), with equality when $\tau \equiv \infty$. More generally, if τ takes a constant value, t, then $\mathcal{F}_\tau = \mathcal{F}_t$. It would be very awkward if we had to distinguish between random variables taking constant values and the constants themselves.

<12> **Example.** The stopping time τ is measurable with respect to $\mathcal{F}_\tau$, because, for each $\alpha \in \mathbb{R}^+$ and $t \in T$,

$$\{\tau \le \alpha\}\{\tau \le t\} = \{\tau \le \alpha \wedge t\} \in \mathcal{F}_{\alpha\wedge t} \subseteq \mathcal{F}_t.$$

That is, $\{\tau \le \alpha\} \in \mathcal{F}_\tau$ for all $\alpha \in \mathbb{R}^+$, from which the $\mathcal{F}_\tau$-measurability follows by the usual generating class argument. It would be counterintuitive if the information corresponding to the sigma-field $\mathcal{F}_\tau$ did not include the value taken by τ itself. □

<13> **Example.** Suppose σ and τ are both stopping times, for which $\sigma \le \tau$ always. Then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$ because

$$F\{\tau \le t\} = \big(F\{\sigma \le t\}\big)\{\tau \le t\} \qquad \text{for all } t \in \overline{T},$$

and both sets on the right-hand side are $\mathcal{F}_t$-measurable if $F \in \mathcal{F}_\sigma$. □

<14> **Exercise.** Show that a random variable Z is $\mathcal{F}_\tau$-measurable if and only if $Z\{\tau \le t\}$ is $\mathcal{F}_t$-measurable for all t in $\overline{T}$.

SOLUTION: For necessity, write Z as a pointwise limit of $\mathcal{F}_\tau$-measurable simple functions Z_n, then note that each $Z_n\{\tau \le t\}$ is a linear combination of indicator functions of $\mathcal{F}_t$-measurable sets.

For sufficiency, it is enough to show that $\{Z > \alpha\} \in \mathcal{F}_\tau$ and $\{Z < -\alpha\} \in \mathcal{F}_\tau$, for each $\alpha \in \mathbb{R}^+$. For the first requirement, note that $\{Z > \alpha\}\{\tau \le t\} = \{Z\{\tau \le t\} > \alpha\}$, which belongs to $\mathcal{F}_t$ for each t, because $Z\{\tau \le t\}$ is assumed to be $\mathcal{F}_t$-measurable. Thus $\{Z > \alpha\} \in \mathcal{F}_\tau$. Argue similarly for the other requirement. □

The definition of X_τ is almost straightforward. Given random variables $\{X_t : t \in T\}$ and a stopping time τ, we should define X_τ as the function taking the value $X_t(\omega)$ when $\tau(\omega) = t$. If τ takes only values in T there is no problem. However, a slight embarrassment would occur when $\tau(\omega) = \infty$ if ∞ were not a point of T, for then $X_\infty(\omega)$ need not be defined. In the happy situation when there is a natural candidate for X_∞, the embarrassment disappears with little fuss; otherwise it is wiser to avoid the difficulty altogether by working only with the random variable $X_\tau\{\tau < \infty\}$, which takes the value zero when τ is infinite.

Measurability of $X_\tau\{\tau < \infty\}$, even with respect to the sigma-field $\mathcal{F}$, requires further assumptions about the $\{X_t\}$ for continuous time. For discrete time the task is much easier. For example, if $\{X_n : n \in \mathbb{N}_0\}$ is adapated to a filtration $\{\mathcal{F}_n : n \in \mathbb{N}_0\}$, and τ is a stopping time for that filtration, then

$$X_\tau\{\tau < \infty\}\{\tau \le t\} = \sum_{i\in\mathbb{N}_0} X_i\{i = \tau \le t\}.$$

For $i > t$ the ith summand is zero; for $i \le t$ it equals $X_i\{\tau = i\}$, which is $\mathcal{F}_i$-measurable. The $\mathcal{F}_\tau$-measurability of $X_\tau\{\tau < \infty\}$ then follows by Exercise <14>

The next Exercise illustrates the use of stopping times and the σ-fields they define. The discussion does not directly involve martingales, but they are lurking in the background.

<15> **Exercise.** A deck of 52 cards (26 reds, 26 blacks) is dealt out one card at a time, face up. Once, and only once, you will be allowed to predict that the next card will be red. What strategy will maximize your probability of predicting correctly?

SOLUTION: Write R_i for the event $\{i\text{th card is red}\}$. Assume all permutations of the deck are equally likely initially. Write $\mathcal{F}_n$ for the σ-field generated by $R_1, \dots, R_n$. A strategy corresponds to a stopping time τ that takes values in $\{0, 1, \dots, 51\}$: you should try to maximize $\mathbb{P}R_{\tau+1}$.

Surprisingly, $\mathbb{P}R_{\tau+1} = 1/2$ for all such stopping rules. The intuitive explanation is that you should always be indifferent, given that you have observed cards $1, 2, \dots, \tau$, between choosing card $\tau + 1$ or choosing card 52. That is, it should be true that $\mathbb{P}(R_{\tau+1} \mid \mathcal{F}_\tau) = \mathbb{P}(R_{52} \mid \mathcal{F}_\tau)$ almost surely; or, equivalently, that $\mathbb{P}R_{\tau+1}F = \mathbb{P}R_{52}F$ for all $F \in \mathcal{F}_\tau$; or, equivalently, that

$$\mathbb{P}R_{k+1}F\{\tau = k\} = \mathbb{P}R_{52}F\{\tau = k\} \qquad \text{for all } F \in \mathcal{F}_\tau \text{ and } k = 0, 1, \dots, 51.$$

We could then deduce that $\mathbb{P}R_{\tau+1} = \mathbb{P}R_{52} = 1/2$. Of course, we only need the case $F = \Omega$, but I'll carry along the general F as an illustration of technique while proving the assertion in the last display. By definition of $\mathcal{F}_\tau$,

$$F\{\tau = k\} = F\{\tau \le k-1\}^c\{\tau \le k\} \in \mathcal{F}_k.$$

That is, $F\{\tau = k\}$ must be of the form $\{(R_1, \dots, R_k) \in B\}$ for some Borel subset B of $\mathbb{R}^k$. Symmetry of the joint distribution of $R_1, \dots, R_{52}$ implies that the random vector $(R_1, \dots, R_k, R_{k+1})$ has the same distribution as the random vector $(R_1, \dots, R_k, R_{52})$, whence

$$\mathbb{P}R_{k+1}\{(R_1, \dots, R_k) \in B\} = \mathbb{P}R_{52}\{(R_1, \dots, R_k) \in B\}.$$

□ See Section 8 for more about symmetry and martingale properties.

The hidden martingale in the previous Exercise is X_n, the proportion of red cards remaining in the deck after n cards have been dealt. You could check the martingale property by first verifying that $\mathbb{P}(R_{n+1} \mid \mathcal{F}_n) = X_n$ (an equality that is obvious if one thinks in terms of conditional distributions), then calculating

$$(52-n-1)\mathbb{P}(X_{n+1} \mid \mathcal{F}_n) = \mathbb{P}\big((52-n)X_n - R_{n+1} \mid \mathcal{F}_n\big) = (52-n)X_n - \mathbb{P}(R_{n+1} \mid \mathcal{F}_n).$$

The problem then asks for the stopping time to maximaize

$$\begin{aligned}\mathbb{P}R_{\tau+1} &= \textstyle\sum_{i=0}^{51} \mathbb{P}\left(R_{i+1}\{\tau = i\}\right)\\ &= \textstyle\sum_{i=0}^{51} \mathbb{P}\left(X_i\{\tau = i\}\right) \qquad \text{because } \{\tau = i\} \in \mathcal{F}_i\\ &= \mathbb{P}X_\tau.\end{aligned}$$

The martingale property tells us that $\mathbb{P}X_0 = \mathbb{P}X_i$ for $i = 1, \dots, 51$. If we could extend the equality to random i, by showing that $\mathbb{P}X_\tau = \mathbb{P}X_0$, then the surprising conclusion from the Exercise would follow.

Clearly it would be useful if we could always assert that $\mathbb{P}X_\sigma = \mathbb{P}X_\tau$ for every martingale, and every pair of stopping times. Unfortunately (Or should I say fortunately?) the result is not true without some extra assumptions. The simplest and most useful case concerns finite time sets. If σ takes values in a finite set T, and if each X_t is integrable, then $|X_\sigma| \le \sum_{t\in T}|X_t|$, which eliminates any integrability difficulties. For an infinite index set, the integrability of X_σ is not automatic.

<16> **Stopping Time Lemma.** *Suppose σ and τ are stopping times for a filtration $\{\mathcal{F}_t : t \in T\}$, with T finite. Suppose both stopping times take only values in T. Let F be a set in $\mathcal{F}_\sigma$ for which $\sigma(\omega) \le \tau(\omega)$ when $\omega \in F$. If $\{X_t : t \in T\}$ is a submartingale, then $\mathbb{P}X_\sigma F \le \mathbb{P}X_\tau F$. For supermartingales, the inequality is reversed. For martingales, the inequality becomes an equality.*

Proof. Consider only the submartingale case. For simplicity of notation, suppose $T = \{0, 1, \ldots, N\}$. Write each X_n as a sum of increments, $X_n = X_0 + \xi_1 + \ldots + \xi_n$. The inequality $\sigma \le \tau$, on F, lets us write

$$
\begin{aligned}
X_\tau F - X_\sigma F &= \left(X_0 F + \sum_{1\le i\le N}\{i \le \tau\}F\xi_i\right) - \left(X_0 F + \sum_{1\le i\le N}\{i \le \sigma\}F\xi_i\right)\\
&= \sum_{1\le i\le N}\{\sigma < i \le \tau\}F\xi_i.
\end{aligned}
$$

Note that $\{\sigma < i \le \tau\}F = \big(\{\sigma \le i-1\}F\big)\{\tau \le i-1\}^c \in \mathcal{F}_{i-1}$. The expected value of each summand is nonnegative, by (subMG)$'$. □

> Remark. If $\sigma \le \tau$ everywhere, the inequality for all F in $\mathcal{F}_\sigma$ implies that $X_\sigma \le \mathbb{P}\left(X_\tau \mid \mathcal{F}_\sigma\right)$ almost surely. That is, the submartingale (or martingale, or supermartingale) property is preserved at bounded stopping times.

The Stopping Time Lemma, and its extensions to various cases with infinite index sets, is basic to many of the most elegant martingale properties. Results for a general stopping time τ, taking values in $\overline{\mathbb{N}}$ or $\overline{\mathbb{N}}_0$, can often be deduced from results for $\tau \wedge N$, followed by a passage to the limit as N tends to infinity. (The random variable $\tau \wedge N$ is a stopping time, because $\{\tau \wedge N \le n\}$ equals the whole of Ω when $N \le n$, and equals $\{\tau \le n\}$ when $N > n$.) As Problem [1] shows, the finiteness assumption on the index set T is not just a notational convenience; the Lemma <16> can fail for infinite T.

It is amazing how many of the classical inequalities of probability theory can be derived by invoking the Lemma for a suitable martingale (or submartingale or supermartingale).

<17> **Exercise.** Let $\xi_1, \ldots, \xi_N$ be independent random variables (or even just martingale increments) for which $\mathbb{P}\xi_i = 0$ and $\mathbb{P}\xi_i^2 < \infty$ for each i. Define $S_i := \xi_1 + \ldots + \xi_i$. Prove the maximal inequality ***Kolmogorov inequality***: for each $\epsilon > 0$,

$$
\mathbb{P}\left\{\max_{1\le i\le N}|S_i| \ge \epsilon\right\} \le \mathbb{P}S_N^2/\epsilon^2.
$$

Solution: The random variables $X_i := S_i^2$ form a submartingale, for the natural filtration. Define stopping times $\tau \equiv N$ and $\sigma :=$ first i such that $|S_i| \ge \epsilon$, with the convention that $\sigma = N$ if $|S_i| < \epsilon$ for every i. Why is σ a stopping time? Check the pointwise bound,

$$
\epsilon^2\{\max_i |S_i| \ge \epsilon\} = \epsilon^2\{X_\sigma \ge \epsilon^2\} \le X_\sigma.
$$

What happens in the case when σ equals N because $|S_i| < \epsilon$ for every i? Take expectations, then invoke the Stopping Time Lemma (with $F = \Omega$) for the submartingale $\{X_i\}$, to deduce

$$\epsilon^2 \mathbb{P}\{\max_i |S_i| \geq \epsilon\} \leq \mathbb{P}X_\sigma \leq \mathbb{P}X_\tau = \mathbb{P}S_N^2,$$

□ as asserted.

Notice how the Kolmogorov inequality improves upon the elementary bound $\mathbb{P}\{|S_N| \geq \epsilon\} \leq \mathbb{P}S_N^2/\epsilon^2$. Actually it is the same inequality, applied to S_σ instead of S_N, supplemented by a useful bound for $\mathbb{P}S_\sigma^2$ made possible by the submartingale property. Kolmogorov (1928) established his inequality as the first step towards a proof of various convergence results for sums of independent random variables. More versatile maximal inequalities follow from more involved appeals to the Stopping Time Lemma. For example, a strong law of large numbers can be proved quite efficiently (Bauer 1981, Section 6.3) by an appeal to the next inequality.

<18> **Exercise.** Let $0 = S_0, \ldots, S_N$ be a martingale with $v_i := \mathbb{P}(S_i - S_{i-1})^2 < \infty$ for each i. Let $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_N$ be nonnegative constants. Prove the ***Hájek-Rényi inequality***:

<19>
$$\mathbb{P}\left\{\max_{1\leq i\leq N} \gamma_i |S_i| \geq 1\right\} \leq \sum_{1\leq i\leq N} \gamma_i^2 v_i.$$

SOLUTION: Define $\mathcal{F}_i := \sigma(S_1, \ldots, S_i)$. Write $\mathbb{P}_i$ for $\mathbb{P}(\cdot \mid \mathcal{F}_i)$. Define $\eta_i := \gamma_i^2 S_i^2 - \gamma_{i-1}^2 S_{i-1}^2$, and $\Delta_i := \mathbb{P}_{i-1}\eta_i$. By the Doob decomposition from Example <7>, the sequence $M_k := \sum_{i=1}^k (\eta_i - \Delta_i)$ is a martingale with respect to the filtration $\{\mathcal{F}_i\}$; and $\gamma_k^2 S_k^2 = (\Delta_1 + \ldots + \Delta_k) + M_k$. Define stopping times $\sigma \equiv 0$ and

$$\tau = \begin{cases} \text{first } i \text{ such that } \gamma_i |S_i| \geq 1, \\ N \text{ if } \gamma_i |S_i| < 1 \text{ for all } i. \end{cases}$$

The main idea is to bound each $\Delta_1 + \ldots + \Delta_k$ by a single random variable Δ, whose expectation will become the right-hand side of the asserted inequality.

Construct Δ from the martingale differences $\xi_i := S_i - S_{i-1}$ for $i = 1, \ldots, N$. For each i, use the fact that S_{i-1} is $\mathcal{F}_{i-1}$ measurable to bound the contribution of Δ_i:

$$\begin{aligned} \Delta_i &= \mathbb{P}_{i-1}\left(\gamma_i^2 S_i^2 - \gamma_{i-1}^2 S_{i-1}^2\right) \\ &= \gamma_i^2 \mathbb{P}_{i-1}\left(\xi_i^2 + 2\xi_i S_{i-1} + S_{i-1}^2\right) - \gamma_{i-1}^2 S_{i-1}^2 \\ &= \gamma_i^2 \mathbb{P}_{i-1}\xi_i^2 + 2\gamma_i^2 S_{i-1}\mathbb{P}_{i-1}\xi_i + (\gamma_i^2 - \gamma_{i-1}^2) S_{i-1}^2. \end{aligned}$$

The middle term on the last line vanishes, by the martingale difference property, and the last term is negative, because $\gamma_i^2 \leq \gamma_{i-1}^2$. The sum of the three terms is less than the nonnegative quantity $\gamma_i^2 \mathbb{P}(\xi_i^2 \mid \mathcal{F}_{i-1})$, and

$$\Delta := \sum_{i\leq N} \gamma_i^2 \mathbb{P}_{i-1}\xi_i^2 \geq \sum_{i\leq k} \Delta_i,$$

for each k, as required.

The asserted inequality now follows via the Stopping Time Lemma:

$$\begin{aligned}\mathbb{P}\{\max_i \gamma_i |S_i| \geq 1\} &= \mathbb{P}\{\gamma_\tau |S_\tau| \geq 1\}\\ &\leq \mathbb{P}\gamma_\tau^2 S_\tau^2\\ &= \mathbb{P}M_\tau + \mathbb{P}(\Delta_1 + \ldots + \Delta_\tau)\\ &\leq \mathbb{P}\Delta,\end{aligned}$$

□ because $\Delta_1 + \ldots + \Delta_\tau \leq \Delta$ and $\mathbb{P}M_\tau = \mathbb{P}M_\sigma = 0$.

The method of proof in Example <18> is worth remembering; it can be used to derive several other bounds.

3. Convergence of positive supermartingales

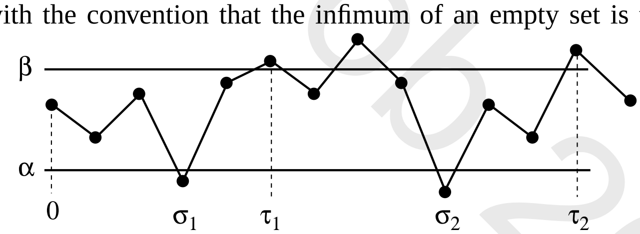
In several respects the theory for positive (meaning nonnegative) supermartingales $\{X_n : n \in \mathbb{N}_0\}$ is particularly elegant. For example (Problem [5]), the Stopping Time Lemma extends naturally to pairs of unbounded stopping times for positive supermartingales. Even more pleasantly surprising, positive supermartingales converge almost surely, to an integrable limit—as will be shown in this Section.

The key result for the proof of convergence is an elegant lemma (Dubins's Inequality) that shows why a positive supermartingale $\{X_n\}$ cannot oscillate between two levels infinitely often.

For fixed constants α and β with $0 \leq \alpha < \beta < \infty$ define increasing sequences of random times at which the process might drop below α or rise above β:

$$\begin{aligned}\sigma_1 &:= \inf\{i \geq 0 : X_i \leq \alpha\}, & \tau_1 &:= \inf\{i \geq \sigma_1 : X_i \geq \beta\},\\ \sigma_2 &:= \inf\{i \geq \tau_1 : X_i \leq \alpha\}, & \tau_2 &:= \inf\{i \geq \sigma_2 : X_i \geq \beta\},\end{aligned}$$

and so on, with the convention that the infimum of an empty set is taken as $+\infty$.



Because the $\{X_i\}$ are adapted to $\{\mathcal{F}_i\}$, each σ_i and τ_i is a stopping time for the filtration. For example,

$$\{\tau_1 \leq k\} = \{X_i \leq \alpha,\ X_j \geq \beta \text{ for some } i \leq j \leq k\ \},$$

which could be written out explicitly as a finite union of events involving only $X_0, \ldots, X_k$.

When τ_k is finite, the segment $\{X_i : \sigma_k \leq i \leq \tau_k\}$ is called the kth ***upcrossing*** of the interval $[\alpha, \beta]$ by the process $\{X_n : n \in \mathbb{N}_0\}$. The event $\{\tau_k \leq N\}$ may be described, slightly informally, by saying that the process completes at least k upcrossings of $[\alpha, \beta]$ up to time N.

<20> **Dubins's inequality.** *For a positive supermartingale* $\{(X_n, \mathcal{F}_n) : n \in \mathbb{N}_0\}$ *and constants* $0 \le \alpha < \beta < \infty$*, and stopping times as defined above,*

$$\mathbb{P}\{\tau_k < \infty\} \le (\alpha/\beta)^k \qquad \text{for } k \in \mathbb{N}.$$

Proof. Choose, and temporarily hold fixed, a finite positive integer N. Define τ_0 to be identically zero. For $k \ge 1$, using the fact that $X_{\tau_k} \ge \beta$ when $\tau_k < \infty$ and $X_{\sigma_k} \le \alpha$ when $\sigma_k < \infty$, we have

$$\begin{aligned}\mathbb{P}\left(\beta\{\tau_k \le N\} + X_N\{\tau_k > N\}\right) &\le \mathbb{P}X_{\tau_k \wedge N} \\ &\le \mathbb{P}X_{\sigma_k \wedge N} \qquad \text{Stopping Time Lemma} \\ &\le \mathbb{P}\left(\alpha\{\sigma_k \le N\} + X_N\{\sigma_k > N\}\right),\end{aligned}$$

which rearranges to give

$$\begin{aligned}\beta\mathbb{P}\{\tau_k \le N\} &\le \alpha\mathbb{P}\{\sigma_k \le N\} + \mathbb{P}X_N\left(\{\sigma_k > N\} - \{\tau_k > N\}\right) \\ &\le \alpha\mathbb{P}\{\tau_{k-1} \le N\} \qquad \text{because } \tau_{k-1} \le \sigma_k \le \tau_k \text{ and } X_N \ge 0.\end{aligned}$$

That is,

$$\mathbb{P}\{\tau_k \le N\} \le \frac{\alpha}{\beta}\mathbb{P}\{\tau_{k-1} \le N\} \qquad \text{for } k \ge 1.$$

Repeated appeals to this inequality, followed by a passage to the limit as $N \to \infty$, leads to Dubins's Inequality. □

REMARK. When $0 = \alpha < \beta$ we have $\mathbb{P}\{\tau_1 < \infty\} = 0$. By considering a sequence of β values decreasing to zero, we deduce that on the set $\{\sigma_1 < \infty\}$ we must have $X_n = 0$ for all $n \ge \sigma_1$. That is, if a positive supermartingale hits zero then it must stay there forever.

Notice that the main part of the argument, before N was sent off to infinity, involved only the variables $X_0, \dots, X_N$. The result may fruitfully be reexpressed as an assertion about positive supermartingales with a finite index set.

<21> **Corollary.** *Let* $\{(X_n, \mathcal{F}_n) : n = 0, 1, \dots, N\}$ *be a positive supermartingale with a finite index set. For each pair of constants* $0 < \alpha < \beta < \infty$*, the probability that the process completes at least* k *upcrossings is less than* $(\alpha/\beta)^k$.

<22> **Theorem.** *Every positive supermartingale converges almost surely to a nonnegative, integrable limit.*

Proof. To prove almost sure convergence (with possibly an infinite limit) of the sequence $\{X_n\}$, it is enough to show that the event

$$D = \{\omega : \limsup X_n(\omega) > \liminf X_n(\omega)\}$$

is negligible. Decompose D into a countable union of events

$$D_{\alpha,\beta} = \{\limsup X_n > \beta > \alpha > \liminf X_n\},$$

with α, β ranging over all pairs of rational numbers. On $D_{\alpha,\beta}$ we must have $\tau_k < \infty$ for every k. Thus $\mathbb{P}D_{\alpha,\beta} \le (\alpha/\beta)^k$ for every k, which forces $\mathbb{P}D_{\alpha,\beta} = 0$, and $\mathbb{P}D = 0$.

The sequence X_n converges to $X_\infty := \liminf X_n$ on the set D^c. Fatou's lemma, and the fact that $\mathbb{P}X_n$ is nonincreasing, ensure that X_∞ is integrable. □

<23> **Exercise.** Suppose $\{\xi_i\}$ are independent, identically distributed random variables ξ_i with $\mathbb{P}\{\xi_i = +1\} = p$ and $\mathbb{P}\{\xi_i = -1\} = 1 - p$. Define the partial

sums $S_0 = 0$ and $S_i = \xi_1 + \ldots + \xi_i$ for $i \geq 1$. For $1/2 \leq p < 1$, show that $\mathbb{P}\{S_i = -1 \text{ for at least one } i\} = (1-p)/p$.

SOLUTION: Consider a fixed p with $1/2 < p < 1$. Define $\theta = (1-p)/p$. Define $\tau = \inf\{i \in \mathbb{N} : S_i = -1\}$. We are trying to show that $\mathbb{P}\{\tau < \infty\} = \theta$. Observe that $X_n = \theta^{S_n}$ is a positive martingale with respect to the filtration $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$: by independence and the equality $\mathbb{P}\theta^{\xi_i} = 1$,

$$\mathbb{P}X_n F = \mathbb{P}\theta^{\xi_n}\theta^{S_{n-1}}F = \mathbb{P}\theta^{\xi_n}\mathbb{P}X_{n-1}F = \mathbb{P}X_{n-1}F \qquad \text{for } F \text{ in } \mathcal{F}_{n-1}.$$

The sequence $\{X_{\tau\wedge n}\}$ is a positive martingale (Problem [3]). It follows that there exists an integrable X_∞ such that $X_{\tau\wedge n} \to X_\infty$ almost surely. The sequence $\{S_n\}$ cannot converge to a finite limit because $|S_n - S_{n-1}| = 1$ for all n. On the set where $\tau = \infty$, convergence of θ^{S_n} to a finite limit is possible only if $S_n \to \infty$ and $\theta^{S_n} \to 0$. Thus,

$$X_{\tau\wedge n} \to \theta^{-1}\{\tau < \infty\} + 0\{\tau = \infty\} \qquad \text{almost surely.}$$

The bounds $0 \leq X_{\tau\wedge n} \leq \theta^{-1}$ allow us to invoke Dominated Convergence to deduce that $1 = \mathbb{P}X_{\tau\wedge n} \to \theta^{-1}\mathbb{P}\{\tau < \infty\}$.

□ Monotonicity of $\mathbb{P}\{\tau < \infty\}$ as a function of p extends the solution to $p = 1/2$.

The almost sure limit X_∞ of a positive supermartingale $\{X_n\}$ satisfies the inequality $\liminf \mathbb{P}X_n \geq \mathbb{P}X_\infty$, by Fatou. The sequence $\{\mathbb{P}X_n\}$ is decreasing. Under what circumstances do we have it converging to $\mathbb{P}X_\infty$? Equality certainly holds if $\{X_n\}$ converges to X_∞ in L^1 norm. In fact, convergence of expectations is equivalent to L^1 convergence, because

$$\begin{aligned}\mathbb{P}|X_n - X_\infty| &= \mathbb{P}(X_\infty - X_n)^+ + \mathbb{P}(X_\infty - X_n)^- \\ &= 2\mathbb{P}(X_\infty - X_n)^+ - (\mathbb{P}X_\infty - \mathbb{P}X_n)\,.\end{aligned}$$

On the right-hand side the first contribution tends to zero, by Dominated Convergence, because $X_\infty \geq (X_\infty - X_n)^+ \to 0$ almost surely. (I just reproved Scheffé's lemma.)

<24> **Corollary.** *A positive supermartingale $\{X_n\}$ converges in L^1 to its limit X_∞ if and only if $\mathbb{P}X_n \to \mathbb{P}X_\infty$.*

<25> **Example.** Female bunyips reproduce once every ten years, according to a fixed offspring distribution P on $\mathbb{N}$. Different bunyips reproduce independently of each other. What is the behavior of the number Z_n of nth generation offspring from Lucy bunyip, the first of the line, as n gets large? (The process $\{Z_n : n \in \mathbb{N}_0\}$ is usually called a ***branching process***.)

Write μ for the expected number of offspring for a single bunyip. If reproduction went strictly according to averages, the nth generation size would equal μ^n. Intuitively, if $\mu > 1$ there could be an explosion of the bunyip population; if $\mu < 1$ bunyips would be driven to extinction; if $\mu = 1$, something else might happen. A martingale argument will lead rigorously to a similar conclusion.

Given $Z_{n-1} = k$, the size of the nth generation is a sum of k independent random variables, each with distribution P. Perhaps we could write $Z_n = \sum_{i=1}^{Z_{n-1}} \xi_{ni}$, with the $\{\xi_{ni} : i = 1, \ldots, Z_{n-1}\}$ (conditionally) independently distributed like P. I have a few difficulties with that representation. For example, where is ξ_{n3} defined?

Just on $\{Z_{n-1} \ge 3\}$? On all of Ω? Moreover, the notation invites the blunder of ignoring the randomness of the range of summation, leading to an absurd assertion that $\mathbb{P}\sum_{i=1}^{Z_{n-1}} \xi_{ni}$ equals $\sum_{i=1}^{Z_{n-1}} \mathbb{P}\xi_{ni} = Z_{n-1}\mu$. The corresponding assertion for an expectation conditional on Z_{n-1} is correct, but some of the doubts still linger.

It is much better to start with an entire family $\{\xi_{ni} : n \in \mathbb{N}, i \in \mathbb{N}\}$ of independent random variables, each with distribution P, then define $Z_0 = 1$ and $Z_n := \sum_{i\in\mathbb{N}} \xi_{ni}\{i \le Z_{n-1}\}$ for $n \ge 1$. The random variable Z_n is measurable with respect to the sigma-field $\mathcal{F}_n = \sigma\{\xi_{ki} : k \le n, i \in \mathbb{N}\}$, and, almost surely,

$$\begin{aligned}
\mathbb{P}(Z_n \mid \mathcal{F}_{n-1}) &= \sum_{i\in\mathbb{N}} \mathbb{P}(\xi_{ni}\{i \le Z_{n-1}\} \mid \mathcal{F}_{n-1}) \\
&= \sum_{i\in\mathbb{N}} \{i \le Z_{n-1}\}\mathbb{P}(\xi_{ni} \mid \mathcal{F}_{n-1}) \quad \text{because } Z_{n-1} \text{ is } \mathcal{F}_{n-1}\text{-measurable} \\
&= \sum_{i\in\mathbb{N}} \{i \le Z_{n-1}\}\mathbb{P}(\xi_{ni}) \quad \text{because } \xi_{ni} \text{ is independent of } \mathcal{F}_{n-1} \\
&= Z_{n-1}\mu .
\end{aligned}$$

If $\mu \le 1$, the $\{Z_n\}$ sequence is a positive supermartingale with respect to the $\{\mathcal{F}_n\}$ filtration. By Theorem <22>, there exists an integrable random variable Z_∞ with $Z_n \to Z_\infty$ almost surely.

A sequence of integers $Z_n(\omega)$ can converge to a finite limit k only if $Z_n(\omega) = k$ for all n large enough. If $k > 0$, the convergence would imply that, with nonzero probability, only finitely many of the independent events $\{\sum_{i\le k} \xi_{ni} \ne k\}$ can occur. By the converse to the Borel-Cantelli lemma, it would follow that $\sum_{i\le k} \xi_{ni} = k$ almost surely, which can happen only if $P\{1\} = 1$. In that case, $Z_n \equiv 1$ for all n. If $P\{1\} < 1$, then Z_n must converge to zero, with probability one, if $\mu \le 1$. The bunyips die out if the average number of offspring is less than or equal to 1.

If $\mu > 1$ the situation is more complex. If $P\{0\} = 0$ the population cannot decrease and Z_n must then diverge to infinity with probability one. If $P\{0\} > 0$ the convex function $g(t) := P^x t^x$ must have a unique value θ with $0 < \theta < 1$ for which $g(\theta) = \theta$: the strictly convex function $h(t) := g(t) - t$ has $h(0) = P\{0\} > 0$ and $h(1) = 0$, and its left-hand derivative $P^x(xt^{x-1} - 1)$ converges to $\mu - 1 > 0$ as t increases to 1. The sequence $\{\theta^{Z_n}\}$ is a positive martingale:

$$\begin{aligned}
\mathbb{P}(\theta^{Z_n} \mid \mathcal{F}_{n-1}) &= \{Z_{n-1} = 0\} + \sum_{k\ge1} \mathbb{P}\left(\theta^{\xi_{n,1}+\dots+\xi_{n,k}}\{Z_{n-1} = k\} \mid \mathcal{F}_{n-1}\right) \\
&= \{Z_{n-1} = 0\} + \sum_{k\ge1} \{Z_{n-1} = k\}\mathbb{P}(\theta^{\xi_{n1}+\dots+\xi_{n,k-1}} \mid \mathcal{F}_{n-1}) \\
&= \sum_{k\in\mathbb{N}_0} \{Z_{n-1} = k\} g(\theta)^k \\
&= g(\theta)^{Z_{n-1}} = \theta^{Z_{n-1}} \qquad \text{because } g(\theta) = \theta .
\end{aligned}$$

The positive martingale $\{\theta^{Z_n}\}$ has an almost sure limit, W. The sequence $\{Z_n\}$ must converge almost surely, with an infinite limit when $W = 0$. As with the situation when $\mu \le 1$, the only other possible limit for Z_n is 0, corresponding to $W = 1$. Because $0 \le \theta^{Z_n} \le 1$ for every n, Dominated Convergence and the martingale property give $\mathbb{P}\{W = 1\} = \lim_{n\to\infty} \mathbb{P}\theta^{Z_n} = \mathbb{P}\theta^{Z_0} = \theta$.

In summary: On the set $D := \{W = 1\}$, which has probability θ, the bunyip population eventually dies out; on D^c, the population explodes, that is, $Z_n \to \infty$.

It is possible to say a lot more about the almost sure behavior of the process $\{Z_n\}$ when $\mu > 1$. For example, the sequence $X_n := Z_n/\mu^n$ is a positive martingale, which must converge to an integrable random variable X. On the set $\{X > 0\}$, the process $\{Z_n\}$ grows geometrically fast, like μ^n. On the set D we must have $X = 0$, but it is not obvious whether or not we might have $X = 0$ for some realizations where the process does not die out.

There is a simple argument to show that, in fact, either $X = 0$ almost surely or $X > 0$ almost surely on D. With a little extra bookkeeping we could keep track of the first generation ancestor of each bunyip in the later generations. If we write $Z_n^{(j)}$ for the members of the nth generation descended from the jth (possibly hypothetical) member of the first generation, then $Z_n = \sum_{j\in\mathbb{N}} Z_n^{(j)}\{j \le Z_1\}$. The $Z_n^{(j)}$, for $j = 1, 2, \ldots$, are independent random variables, each with the same distribution as Z_{n-1}, and each independent of Z_1. In particular, for each j, we have $Z_n^{(j)}/\mu^{n-1} \to X^{(j)}$ almost surely, where the $X^{(j)}$, for $j = 1, 2, \ldots$, are independent random variables, each distributed like X, and

$$\mu X = \sum_{j\in\mathbb{N}} X^{(j)}\{j \le Z_1\} \qquad \text{almost surely.}$$

Write ϕ for $\mathbb{P}\{X = 0\}$. Then, by independence,

$$\mathbb{P}\{X = 0 \mid Z_1 = k\} = \prod_{j=1}^{k} \mathbb{P}\{X^{(j)} = 0\} = \phi^k,$$

whence $\phi = \sum_{k\in\mathbb{N}_0} \phi^k \mathbb{P}\{Z_1 = k\} = g(\phi)$. We must have either $\phi = 1$, meaning that $X = 0$ almost surely, or else $\phi = \theta$, in which case $X > 0$ almost surely on D^c. The latter must be the case if X is nondegenerate, that is, if $\mathbb{P}\{X > 0\} > 0$, which happens if and only if $P^x\,(x\log(1+x)) < \infty$—see Problem [14]. □

4. Convergence of submartingales

Theorem <22> can be extended to a large class of submartingales by means of the following decomposition Theorem, whose proof appears in the next Section.

<26> **Krickeberg decomposition.** *Let $\{S_n : n \in \mathbb{N}_0\}$ be a submartingale for which $\sup_n \mathbb{P}S_n^+ < \infty$. Then there exists a positive martingale $\{M_n\}$ and a positive supermartingale $\{X_n\}$ such that $S_n = M_n - X_n$ almost surely, for each n.*

<27> **Corollary.** *A submartingale with $\sup_n \mathbb{P}S_n^+ < \infty$ converges almost surely to an integrable limit.*

For a direct proof of this convergence result, via an upcrossing inequality for supermartingales that are not necessarily nonnegative, see Problem [11].

REMARK. Finiteness of $\sup_n \mathbb{P}S_n^+$ is equivalent to the finiteness of $\sup_n \mathbb{P}|S_n|$, because $|S_n| = 2S_n^+ - (S_n^+ - S_n^-)$ and by the submartingale property, $\mathbb{P}(S_n^+ - S_n^-) = \mathbb{P}S_n$ increases with n.

<28> **Example.** (Section 68 of Lévy 1937.) Let $\{M_n : n \in \mathbb{N}_0\}$ be a martingale such that $|M_n - M_{n-1}| \le 1$, for all n, and $M_0 = 0$. In order that $M_n(\omega)$ converges to a

finite limit, it is necessary that $\sup_n M_n(\omega)$ be finite. In fact, it is also a sufficicient condition. More precisely

$$\{\omega : \lim_n M_n(\omega) \text{ exists as a finite limit}\} = \{\omega : \sup M_n(\omega) < \infty\} \qquad \text{almost surely.}$$

To establish that the right-hand side is (almost surely) a subset of the left-hand side, for a fixed positive C define τ as the first n for which that $M_n > C$, with $\tau = \infty$ when $\sup_n M_n \le C$. The martingale $X_n := M_{\tau\wedge n}$ is bounded above by the constant $C + 1$, because the increment (if any) that pushes M_n above C cannot be larger than 1. In particular, $\sup_n \mathbb{P}X_n^+ < \infty$, which ensures that $\{X_n\}$ converges almost surely to a finite limit. On the set $\{\sup_n M_n \le C\}$ we have $M_n = X_n$ for all n, and hence M_n also converges almost surely to a finite limit on that set. Take a union over a sequence of C values increasing to ∞ to complete the argument.

REMARK. Convergence of $M_n(\omega)$ to a finite limit also implies that $\sup_n |M_n(\omega)| < \infty$. The result therefore contains the surprising asssertion that, almost surely, finiteness of $\sup_n M_n(\omega)$ implies finiteness of $\sup_n |M_n(\omega)|$.

As a special case, consider a sequence $\{A_n\}$ of events adapted to a filtration $\{\mathcal{F}_n\}$. The martingale $M_n := \sum_{i=1}^n (A_i - \mathbb{P}(A_i \mid \mathcal{F}_{i-1}))$ has increments bounded in absolute value by 1. For almost all ω, finiteness of $\sum_{n\in\mathbb{N}}\{\omega \in A_n\}$ implies $\sup_n M_n(\omega) < \infty$, and hence convergence of the sum of conditional probabilities. Argue similarly for the martingale $\{-M_n\}$ to conclude that

$$\{\omega : \textstyle\sum_{n=1}^\infty A_n < \infty\} = \{\omega : \sum_{n=1}^\infty \mathbb{P}(A_n \mid \mathcal{F}_{n-1}) < \infty\} \qquad \text{almost surely,}$$

a remarkable generalization of the Borel-Cantelli lemma for sequences of independent events. □

*5. Proof of the Krickeberg decomposition

It is easiest to understand the proof by reinterpreting the result as assertion about measures. To each integrable random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ there corresponds a signed measure μ defined on $\mathcal{F}$ by $\mu F := \mathbb{P}(XF)$ for $F \in \mathcal{F}$. The measure can also be written as a difference of two nonnegative measures μ^+ and μ^-, defined by $\mu^+ F := \mathbb{P}(X^+ F)$ and $\mu^- F := \mathbb{P}(X^- F)$, for $f \in \mathcal{F}$.

By equivalence (MG)′, a sequence of integrable random variables $\{X_n : n \in \mathbb{N}_0\}$ adapted to a filtration $\{\mathcal{F}_n : n \in \mathbb{N}_0\}$ is a martingale if and only if the corresponding sequence of measures $\{\mu_n\}$ on $\mathcal{F}$ has the property

<29> $$\mu_{n+1}\big|_{\mathcal{F}_n} = \mu_n\big|_{\mathcal{F}_n} \qquad \text{for each } n,$$

where, in general, $\nu\big|_{\mathcal{G}}$ denotes the restriction of a measure ν to a sub-sigma-field $\mathcal{G}$. Similarly, the defining inequality (subMG)′ for a submartingale, $\mu_{n+1}F := \mathbb{P}(X_{n+1}F) \ge \mathbb{P}X_n F =: \mu_n F$ for all $F \in \mathcal{F}_n$, is equivalent to

<30> $$\mu_{n+1}\big|_{\mathcal{F}_n} \ge \mu_n\big|_{\mathcal{F}_n} \qquad \text{for each } n.$$

Now consider the submartingale $\{S_n : n \in \mathbb{N}_0\}$ from the statement of the Krickeberg decomposition. Define an increasing functional $\lambda : \mathcal{M}^+(\mathcal{F}) \to [0, \infty]$ by

$$\lambda f := \limsup_n \mathbb{P}(S_n^+ f) \qquad \text{for } f \in \mathcal{M}^+(\mathcal{F}).$$

Notice that $\lambda 1 = \limsup_n \mathbb{P}S_n^+$, which is finite, by assumption. The functional also has a property analogous to absolute continuity: if $\mathbb{P}f = 0$ then $\lambda f = 0$.

Write λ_k for the restriction of λ to $\mathcal{M}^+(\mathcal{F}_k)$. For f in $\mathcal{M}^+(\mathcal{F}_k)$, the submartingale property for $\{S_n^+\}$ ensures that $\mathbb{P}S_n^+ f$ increases with n for $n \geq k$. Thus

<31>
$$\lambda_k f := \lambda f = \lim_n \mathbb{P}(S_n^+ f) = \sup_{n \geq k} \mathbb{P}(S_n^+ f) \qquad \text{if } f \in \mathcal{M}^+(\mathcal{F}_k).$$

The increasing functional λ_k is linear (because linearity is preserved by limits), and it inherits the Monotone Convergence property from $\mathbb{P}$: for functions in $\mathcal{M}^+(\mathcal{F}_k)$ with $0 \leq f_i \uparrow f$,

$$\sup_i \lambda_k f_i = \sup_i \sup_{n \geq k} \mathbb{P}\left(S_n^+ f_i\right) = \sup_{n \geq k} \sup_i \mathbb{P}\left(S_n^+ f_i\right) = \sup_{n \geq k} \mathbb{P}\left(S_n^+ f\right) = \lambda_k f.$$

It defines a finite measure on $\mathcal{F}_k$ that is absolutely continuous with respect to $\mathbb{P}\big|_{\mathcal{F}_k}$. Write M_k for the corresponding density in $\mathcal{M}^+(\Omega, \mathcal{F}_k)$.

The analog of <29> identifies $\{M_k\}$ as a nonnegative martingale, because $\lambda_{k+1}\big|_{\mathcal{F}_k} = \lambda\big|_{\mathcal{F}_k} = \lambda_k$. Moreover, $M_k \geq S_k^+$ almost surely because

$$\mathbb{P}M_k\{M_k < S_k^+\} = \lambda_k\{M_k < S_k^+\} \geq \mathbb{P}S_k^+\{M_k < S_k^+\},$$

the last inequality following from <31> with $f := \{M_k < S_k^+\}$. The random variables $X_k := M_k - S_k$ are almost surely nonnegative. Also, for $F \in \mathcal{F}_k$,

$$\mathbb{P}X_k F = \mathbb{P}M_k F - \mathbb{P}S_k F \geq \mathbb{P}M_{k+1} F - \mathbb{P}S_{k+1} F = \mathbb{P}X_{k+1} F,$$

because $\{M_k\}$ is a martingale and $\{S_k\}$ is a submartingale. It follows that $\{X_k\}$ is a supermartingale, as required for the Krickeberg decomposition.

*6. Uniform integrability

Corollary <27> gave a sufficient condition for a submartingale $\{X_n\}$ to converge almost surely to an integrable limit X_∞. If $\{X_n\}$ happens to be a martingale, we know that $X_n = \mathbb{P}(X_{n+m} \mid \mathcal{F}_n)$ for arbitrarily large m. It is tempting to leap to the conclusion that

<32>
$$X_n \overset{?}{=} \mathbb{P}(X_\infty \mid \mathcal{F}_n),$$

as suggested by a purely formal passage to the limit as m tends to infinity. One should perhaps look before one leaps.

<33> **Example.** Reconsider the limit behavior of the partial sums $\{S_n\}$ from Example <23> but with $p = 1/3$ and $\theta = 2$. The sequence $X_n = 2^{S_n}$ is a positive martingale. By the strong law of large numbers, $S_n/n \to -1/3$ almost surely, which gives $S_n \to -\infty$ almost surely and $X_\infty = 0$ as the limit of the martingale. Clearly X_n is not equal to $\mathbb{P}(X_\infty \mid \mathcal{F}_n)$. □

REMARK. The branching process of Example <25> with $\mu = 1$ provides another case of a nontrivial martingale converging almost surely to zero.

As you will learn in this Section, the condition for the validity of <32> (without the cautionary question mark) is ***uniform integrability***. Remember that a family of random variables $\{Z_t : t \in T\}$ is said to be uniformly integrable if $\sup_{t\in T} \mathbb{P}|Z_t|\{|Z_t| > M\} \to 0$ as $M \to \infty$. Remember also the following characterization of $\mathcal{L}^1$ convergence, which was proved in Section 2.8.

<34> **Theorem.** *Let $\{Z_n : n \in \mathbb{N}\}$ be a sequence of integrable random variables. The following two conditions are equivalent.*

(i) The sequence is uniformly integrable and it converges in probability to a random variable Z_∞, which is necessarily integrable.

(ii) The sequence converges in $\mathcal{L}^1$ norm, $\mathbb{P}|Z_n - Z_\infty| \to 0$, to an integrable random variable Z_∞.

The necessity of uniform integrability for <32> follows immediately from a general property of conditional expectations.

<35> **Lemma.** *For a fixed integrable random variable Z, the family of all conditional expectations* $\{\mathbb{P}(Z \mid \mathcal{G}) : \mathcal{G}$ a sub-sigma-field of $\mathcal{F}\}$ *is uniformly integrable.*

Proof. Write $Z_\mathcal{G}$ for $\mathbb{P}(Z \mid \mathcal{G})$. With no loss of generality, we may suppose $Z \geq 0$, because $|Z_\mathcal{G}| \leq \mathbb{P}\left(|Z| \mid \mathcal{G}\right)$. Invoke the defining property of the conditional expectation, and the fact that $\{Z_\mathcal{G} > M^2\} \in \mathcal{G}$, to rewrite $\mathbb{P}Z_\mathcal{G}\{Z_\mathcal{G} > M^2\}$ as

$$\mathbb{P}Z\{Z_\mathcal{G} > M^2\} \leq M\mathbb{P}\{Z_\mathcal{G} > M^2\} + \mathbb{P}Z\{Z > M\}.$$

The first term on the right-hand side is less than $M\mathbb{P}Z_\mathcal{G}/M^2 = \mathbb{P}Z/M$, which tends to zero as $M \to \infty$. The other term also tends to zero, because Z is integrable. □

More generally, if X is an integrable random variable and $\{\mathcal{F}_n : n \in \mathbb{N}_0\}$ is a filtration then $X_n := \mathbb{P}(X \mid \mathcal{F}_n)$ defines a uniformly integrable martingale. In fact, every uniformly integrable martingale must be of this form.

<36> **Theorem.** *Every uniformly integrable martingale $\{X_n : n \in \mathbb{N}_0\}$ converges almost surely and in $\mathcal{L}^1$ to an integrable random variable X_∞, for which $X_n = \mathbb{P}(X_\infty \mid \mathcal{F}_n)$. Moreover, if $X_n := \mathbb{P}(X \mid \mathcal{F}_n)$ for some integrable X then $X_\infty = \mathbb{P}(X \mid \mathcal{F}_\infty)$, where $\mathcal{F}_\infty := \sigma\left(\cup_{n\in\mathbb{N}}\mathcal{F}_n\right)$.*

Proof. Uniform integrability implies finiteness of $\sup_n \mathbb{P}|X_n|$, which lets us deduce via Corollary <27> the almost sure convergence to the integrable limit X_∞. Almost sure convergence implies convergence in probability, which uniform integrability and Theorem <34> strengthen to $\mathcal{L}^1$ convergence. To show that $X_n = \mathbb{P}(X_\infty \mid \mathcal{F}_n)$, fix an F in $\mathcal{F}_n$. Then, for all positive m,

$$|\mathbb{P}X_\infty F - \mathbb{P}X_n F| \leq \mathbb{P}|X_\infty - X_{n+m}| + |\mathbb{P}X_{n+m}F - \mathbb{P}X_n F|.$$

The $\mathcal{L}^1$ convergence makes the first term on the right-hand side converge to zero as m tends to infinity. The second term is zero for all positive m, by the martingale property. Thus $\mathbb{P}X_\infty F = \mathbb{P}X_n F$ for every F in $\mathcal{F}_n$.

If $\mathbb{P}(X \mid \mathcal{F}_n) = X_n = \mathbb{P}(X_\infty \mid \mathcal{F}_n)$ then $\mathbb{P}XF = \mathbb{P}X_\infty F$ for each F in $\mathcal{F}_n$. A generating class argument then gives the equality for all F in $\mathcal{F}_\infty$, which characterizes the $\mathcal{F}_\infty$-measurable random variable X_∞ as the conditional expectation $\mathbb{P}(X \mid \mathcal{F}_\infty)$. □

REMARK. More concisely: the uniformly integrable martingales $\{X_n : n \in \mathbb{N}\}$ are precisely those that can be extended to martingales $\{X_n : n \in \overline{\mathbb{N}}\}$. Such a martingale is sometimes said to be ***closed on the right***.

<37> **Example.** Classical statistical models often consist of a parametric family $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ of probability measures that define joint distributions of infinite sequences $\omega := (\omega_1, \omega_2, \ldots)$ of possible observations. More formally, each $\mathbb{P}_\theta$ could be thought of as a probability measure on $\mathbb{R}^{\mathbb{N}}$, with random variables X_i as the coordinate maps.

For simplicity, suppose Θ is a Borel subset of a Euclidean space. The parameter θ is said to be ***consistently estimated*** by a sequence of measurable functions $\widehat{\theta}_n = \widehat{\theta}_n(\omega_0, \ldots, \omega_n)$ if

<38> $$\mathbb{P}_\theta\{\,|\widehat{\theta}_n - \theta| > \epsilon\} \to 0 \qquad \text{for each } \epsilon > 0 \text{ and each } \theta \text{ in } \Theta.$$

A Bayesian would define a joint distribution $\mathbb{Q} := \pi \otimes \mathcal{P}$ for θ and ω by equipping Θ with a prior probability distribution π. The conditional distributions $\mathbb{Q}_{n,t}$ given the random vectors $T_n := (X_0, \ldots, X_n)$ are called posterior distributions. We could also regard $\pi_{n\omega}(\cdot) := \mathbb{Q}_{n,T_n(\omega)}(\cdot)$ as random probability measures on the product space. An expectation with respect to $\pi_{n\omega}$ is a version of a conditional expectation given the the sigma-field $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$.

A mysterious sounding result of Doob (1949) asserts that mere existence of some consistent estimator for θ ensures that the $\pi_{n\omega}$ distributions will concentrate around the right value, in the delicate sense that for π-almost all θ, the $\pi_{n\omega}$ measure of each neighborhood of θ tends to one for $\mathbb{P}_\theta$-almost all ω.

The mystery dissipates when one understands the role of the consistent estimator. When averaged out over the prior, property <38> implies (via Dominated Convergence) that $\mathbb{Q}\{(\theta, \omega) : |\widehat{\theta}_n(\omega) - \theta| > \epsilon\} \to 0$. A $\mathbb{Q}$-almost surely convergent subsequence identifies θ as an $\mathcal{F}_\infty$-measurable random variable, $\tau(\omega)$, on the product space, up to a $\mathbb{Q}$ equivalence. That is, $\theta = \tau(\omega)$ a.e. $[\mathbb{Q}]$.

Let $\mathcal{U}$ be a countable collection of open sets generating the topology of Θ. That is, each open set should equal the union of the $\mathcal{U}$-sets that it contains. For each U in $\mathcal{U}$, the sequence of posterior probabilities $\pi_{n\omega}\{\theta \in U\} = \mathbb{Q}\{\theta \in U \mid \mathcal{F}_n\}$ defines a uniformly integrable martingale, which converges $\mathbb{Q}$-almost surely to

$$\mathbb{Q}\{\theta \in U \mid \mathcal{F}_\infty\} = \{\theta \in U\} \qquad \text{because } \{\theta \in U\} \underset{a.s.}{=} \{\tau(\omega) \in U\} \in \mathcal{F}_\infty.$$

Cast out a sequence of $\mathbb{Q}$-negligible sets, leaving a set E with $\mathbb{Q}E = 1$ and $\pi_{n\omega}\{\theta \in U\} \to \{\theta \in U\}$ for all U in $\mathcal{U}$, all $(\theta, \omega) \in E$, which implies Doob's result. □

*7. Reversed martingales

Martingale theory gets easier when the index set T has a largest element, as in the case $T = -\mathbb{N}_0 := \{-n : n \in \mathbb{N}_0\}$. Equivalently, one can reverse the "direction of time," by considering families of integrable random variables $\{X_t : t \in T\}$ adapted to ***decreasing filtrations***, families of sub-sigma-fields $\{\mathcal{G}_t : t \in T\}$ for which $\mathcal{G}_s \supseteq \mathcal{G}_t$ when $s < t$. For such a family, it is natural to define $\mathcal{G}_\infty := \cap_{t\in T}\mathcal{G}_t$ if it is not already defined.

<39> **Definition.** *Let $\{X_n : n \in \mathbb{N}_0\}$ be a sequence of integrable random variables, adapted to a decreasing filtration $\{\mathcal{G}_n : n \in \mathbb{N}_0\}$. Call $\{(X_n, \mathcal{G}_n) : n \in \mathbb{N}_0\}$ a **reversed supermartingale** if $\mathbb{P}(X_n \mid \mathcal{G}_{n+1}) \le X_{n+1}$ almost surely, for each n. Define reversed submartingales and reversed martingales analogously.*

That is, $\{(X_n, \mathcal{G}_n) : n \in \mathbb{N}_0\}$ is a reversed supermartingale if and only if $\{(X_{-n}, \mathcal{G}_{-n}) : n \in -\mathbb{N}_0\}$ is a supermartingale. In particular, for each fixed N, the finite sequence $X_N, X_{N-1}, \ldots, X_0$ is a supermartingale with respect to the filtration $\mathcal{G}_N \subseteq \mathcal{G}_{N-1} \subseteq \ldots \subseteq \mathcal{G}_0$.

<40> **Example.** If $\{\mathcal{G}_n : n \in \mathbb{N}_0\}$ is a decreasing filtration and X is an integrable random variable, the sequence $X_n := \mathbb{P}(X \mid \mathcal{G}_n)$ defines a uniformly integrable, (Lemma <35>) reversed martingale. □

The theory for reversed positive supermartingales is analogous to the theory from Section 3, except for the slight complication that the sequence $\{\mathbb{P}X_n : n \in \mathbb{N}_0\}$ might be increasing, and therefore it is not automatically bounded.

<41> **Theorem.** *For every reversed, positive supermartingale $\{(X_n, \mathcal{G}_n) : n \in \mathbb{N}_0\}$:*

(i) there exists an X_∞ in $\mathcal{M}^+(\Omega, \mathcal{G}_\infty)$ for which $X_n \to X_\infty$ almost surely;

(ii) $\mathbb{P}(X_n \mid \mathcal{G}_\infty) \uparrow X_\infty$ almost surely;

(iii) $\mathbb{P}|X_n - X_\infty| \to 0$ if and only if $\sup_n \mathbb{P}X_n < \infty$.

Proof. The Corollary <21> to Dubins's Inequality bounds by $(\alpha/\beta)^k$ the probability that $X_N, X_{N-1}, \ldots, X_0$ completes at least k upcrossings of the interval $[\alpha, \beta]$, no matter how large we take N. As in the proof of Theorem <22>, it then follows that $\mathbb{P}\{\limsup X_n > \beta > \alpha > \liminf X_n\} = 0$, for each pair $0 \le \alpha < \beta < \infty$, and hence X_n converges almost surely to a nonnegative limit X_∞, which is necessarily $\mathcal{G}_\infty$-measurable.

I will omit most of the "almost sure" qualifiers for the remainder of the proof. Temporarily abbreviate $\mathbb{P}(\cdot \mid \mathcal{G}_n)$ to $\mathbb{P}_n(\cdot)$, for $n \in \overline{\mathbb{N}}_0$, and write Z_n for $\mathbb{P}_\infty X_n$. From the reversed supermartingale property, $\mathbb{P}_{n+1}X_n \le X_{n+1}$, and the rule for iterated conditional expectations we get

$$Z_n = \mathbb{P}_\infty X_n = \mathbb{P}_\infty(\mathbb{P}_{n+1}X_n) \le \mathbb{P}_\infty X_{n+1} = Z_{n+1}.$$

Thus $Z_n \uparrow Z_\infty := \limsup_n Z_n$, which is $\mathcal{G}_\infty$-measurable.

For (ii) we need to show $Z_\infty = X_\infty$, almost surely. Equivalently, as both variables are $\mathcal{G}_\infty$-measurable, we need to show $\mathbb{P}(Z_\infty G) = \mathbb{P}(X_\infty G)$ for each G in $\mathcal{G}_\infty$. For such a G,

$$\begin{aligned} \mathbb{P}(Z_\infty G) &= \sup_{n\in\mathbb{N}_0} \mathbb{P}(Z_n G) && \text{Monotone Convergence} \\ &= \sup_n \mathbb{P}(X_n G) && \text{definition of } Z_n := \mathbb{P}_\infty X_n \\ &= \sup_n \sup_{m\in\mathbb{N}} \mathbb{P}((X_n \wedge m)G) && \text{Monotone Convergence, for fixed } n \\ &= \sup_m \sup_n \mathbb{P}((X_n \wedge m)G). \end{aligned}$$

The sequence $\{X_n \wedge m : n \in \mathbb{N}_0\}$ is a uniformly bounded, reversed positive supermartingale, for each fixed m. Thus $\mathbb{P}((X_n \wedge m)G)$ increases with n, and, by Dominated Convergence, its limit equals $\mathbb{P}((X_\infty \wedge m)G)$. Thus

$$\mathbb{P}(Z_\infty G) = \sup_m \mathbb{P}((X_\infty \wedge m)G) = \mathbb{P}(X_\infty G),$$

the final equality by Monotone Convergence. Assertion (ii) follows.

Monotone Convergence and (ii) imply that $\mathbb{P}X_\infty = \sup_n \mathbb{P}X_n$. Finiteness of the supremum is equivalent to integrability of X_∞. The sufficiency in assertion (iii) follows by the usual Dominated Convergence trick (also known as Scheffé's lemma):

$$\mathbb{P}|X_\infty - X_n| = 2\mathbb{P}(X_\infty - X_n)^+ - (\mathbb{P}X_\infty - \mathbb{P}X_n) \to 0.$$

For the necessity, just note that an $\mathcal{L}^1$ limit of a sequence of integrable random variables is integrable. □

The analog of the Krickeberg decomposition extends the result to reversed submartingales $\{(X_n, \mathcal{G}_n) : n \in \mathbb{N}_0\}$. The sequence $M_n := \mathbb{P}(X_0^+ \mid \mathcal{G}_n)$ is a reversed positive martingale for which $M_n \geq \mathbb{P}(X_0 \mid \mathcal{G}_n) \geq X_n$ almost surely. Thus $X_n = M_n - (M_n - X_n)$ decomposes X_n into a difference of a reversed positive martingale and a reversed positive supermartingale.

<42> **Corollary.** *Every reversed submartingale $\{(X_n, \mathcal{G}_n) : n \in \mathbb{N}_0\}$ converges almost surely. The limit is integrable, and the sequence also converges in the $\mathcal{L}^1$ sense, if $\inf_n \mathbb{P}X_n > -\infty$.*

Proof. Apply Theorem <41> to both reversed positive supermartingales $\{M_n - X_n\}$ and $\{M_n\}$, noting that $\sup_n \mathbb{P}M_n = \mathbb{P}X_0^+$ and $\sup_n \mathbb{P}(M_n - X_n) = \mathbb{P}X_0^+ - \inf_n \mathbb{P}X_n$. □

<43> **Corollary.** *Every reversed martingale $\{(X_n, \mathcal{G}_n) : n \in \mathbb{N}_0\}$ converges almost surely, and in $\mathcal{L}^1$, to the limit $X_\infty := \mathbb{P}(X_0 \mid \mathcal{G}_\infty)$, where $\mathcal{G}_\infty := \cap_{n\in\mathbb{N}_0}\mathcal{G}_n$.*

Proof. The identification of the limit as the conditional expectation follows from the facts that $\mathbb{P}(X_0 G) = \mathbb{P}(X_n G)$, for each n, and $|\mathbb{P}(X_n G) - \mathbb{P}(X_\infty G)| \leq \mathbb{P}|X_n - X_\infty| \to 0$, for each G in $\mathcal{G}_\infty$. □

Reversed martingales arise naturally from symmetry arguments.

<44> **Example.** Let $\{\xi_i : i \in \mathbb{N}\}$ be a sequence of independent random elements taking values in a set $\mathcal{X}$ equipped with a sigma-field $\mathcal{A}$. Suppose each ξ_i induces the same distribution P on $\mathcal{A}$, that is, $\mathbb{P}f(\xi_i) = Pf$ for each f in $\mathcal{M}^+(\mathcal{X}, \mathcal{A})$. For each n define the ***empirical measure*** $P_{n,\omega}$ (or just P_n, if there is no need to emphasize the dependence on ω) on $\mathcal{X}$ as the probability measure that places mass n^{-1} at each of the points $\xi_1(\omega), \ldots, \xi_n(\omega)$. That is, $P_{n,\omega}f := n^{-1}\sum_{1\leq i\leq n} f(\xi_i(\omega))$.

Intuitively speaking, knowledge of P_n tells us everything about the values $\xi_1(\omega), \ldots, \xi_n(\omega)$ except for the order in which they were produced. Conditionally on P_n, we know that ξ_1 should be one of the points supporting P_n, but we don't know which one. The conditional distribution of ξ_1 given P_n should put probability n^{-1} at each support point, and it seems we should then have

<45>
$$\mathbb{P}(f(\xi_1) \mid P_n) = P_n f.$$

REMARK. Here I am arguing, heuristically, assuming P_n concentrates on n distinct points. A similar heuristic could be developed when there are ties, but there is no point in trying to be too precise at the moment. The problem of ties would disappear from the formal argument.

Similarly, if we knew all P_i for $i \geq n$ then we should be able to locate $\xi_i(\omega)$ exactly for $i \geq n+1$, but the values of $\xi_1(\omega), \ldots, \xi_n(\omega)$ would still be known only up to random relabelling of the support points of P_n. The new information would tell us no more about ξ_1 than we already knew from P_n. In other words, we should have

<46> $$\mathbb{P}\left(f(\xi_1) \mid \mathcal{G}_n\right) = \ldots = \mathbb{P}\left(f(\xi_n) \mid \mathcal{G}_n\right) = P_n f \qquad \text{where } \mathcal{G}_n \overset{?}{:=} \sigma(P_n, P_{n+1}, \ldots),$$

which would then give

$$\mathbb{P}\left(P_{n-1} f \mid \mathcal{G}_n\right) = \frac{1}{n-1}\sum_{i=1}^{n-1} \mathbb{P}\left(f(\xi_i) \mid \mathcal{G}_n\right) = \frac{1}{n-1}\sum_{i=1}^{n-1} P_n f = P_n f.$$

That is, $\{(P_n f, \mathcal{G}_n) : n \in \mathbb{N}\}$ would be a reversed martingale, for each fixed f.

It is possible to define $\mathcal{G}_n$ rigorously, then formalize the preceeding heuristic argument to establish the reversed martingale property. I will omit the details, because it is simpler to replace $\mathcal{G}_n$ by the closely related **n-*symmetric sigma-field*** $\mathcal{S}_n$, to be defined in the next Section, then invoke the more general symmetry arguments (Example <50>) from that Section to show that $\{(P_n f, \mathcal{S}_n) : n \in \mathbb{N}\}$ is a reversed martingale for each P-integrable f.

Corollary <43> ensures that $P_n f \to \mathbb{P}(f(\xi_1) \mid \mathcal{S}_\infty)$ almost surely. As you will see in the next Section (Theorem <51>, to be precise), the sigma-field $\mathcal{S}_\infty$ is trivial—it contains only events with probability zero or one—and $\mathbb{P}(X \mid \mathcal{S}_\infty) = \mathbb{P}X$, almost surely, for each integrable random variable X. In particular, for each P-integrable function f we have

$$\frac{f(\xi_1) + \ldots + f(\xi_n)}{n} = P_n f \to Pf \qquad \text{almost surely.}$$

The special case $\mathcal{X} := \mathbb{R}$ and $\mathbb{P}|\xi_1| < \infty$ and $f(x) \equiv x$ recovers the Strong Law of Large Numbers (SLLN) for independent, identically distributed summands.

In statistical problems it is sometimes necessary to prove a uniform analog of the SLLN (a USLLN):

$$\Delta_n := \sup_\theta |P_n f_\theta - P f_\theta| \to 0 \qquad \text{almost surely,}$$

where $\{f_\theta : \theta \in \Theta\}$ is a class of P-integrable functions on $\mathcal{X}$. Corollary <41> can greatly simplify the task of establishing such a USLLN.

To avoid measurability difficulties, let me consider only the case where Θ is countable. Write $X_{n,\theta}$ for $P_n f_\theta - P f_\theta$. Also, assume that the ***envelope*** $F := \sup_\theta |f_\theta|$ is P-integrable, so that $\mathbb{P}\Delta_n \leq \mathbb{P}\left(P_n F + PF\right) = 2PF < \infty$.

For each fixed θ, we know that $\{(X_{n,\theta}, \mathcal{S}_n) : n \in \mathbb{N}\}$ is a reversed martingale, and hence

$$\mathbb{P}\left(\Delta_n \mid \mathcal{S}_{n+1}\right) = \mathbb{P}\left(\sup_\theta |X_{n,\theta}| \mid \mathcal{S}_{n+1}\right) \geq \sup_\theta \left|\mathbb{P}\left(X_{n,\theta} \mid \mathcal{S}_{n+1}\right)\right| = \Delta_{n+1}.$$

That is, $\{(\Delta_n, \mathcal{S}_n) : n \in \mathbb{N}\}$ is a reversed submartingale. From Corollary <41>, Δ_n converges almost surely to a $\mathcal{S}_\infty$-measurable random variable Δ_∞, which by

the triviality of $\mathcal{S}_\infty$ (Theorem <51> again) is (almost surely) constant. To prove the USLLN, we have only to show that the constant is zero. For example, it would suffice to show $\mathbb{P}\Delta_n \to 0$, a great simplification of the original task. See
□ Pollard (1984, Section II.5) for details.

*8. Symmetry and exchangeability

The results in this Section involve probability measures on infinite product spaces. You might want to consult Section 4.8 *for notation and the construction of product measures.*

The symmetry arguments from Example <44> did not really require an assumption of independence. The reverse martingale methods can be applied to more general situations where probability distributions have symmetry properties.

Rather than work with random elements of a space $(\mathcal{X}, \mathcal{A})$, it is simpler to deal with their joint distribution as a probability measure on the product sigma-field $\mathcal{A}^{\mathbb{N}}$ of the product space $\mathcal{X}^{\mathbb{N}}$, the space of all sequences, $\mathbf{x} := (x_1, x_2, \ldots)$, on $\mathcal{X}$. We can think of the coordinate maps $\xi_i(\mathbf{x}) := x_i$ as a sequence of random elements of $(\mathcal{X}, \mathcal{A})$, when it helps.

<47> **Definition.** *Call a one-to-one map π from $\mathbb{N}$ onto itself an **n-permutation** if $\pi(i) = i$ for $i > n$. Write $\mathcal{R}(n)$ for the set of all $n!$ distinct n-permutations. Call $\cup_{n\in\mathbb{N}}\mathcal{R}(n)$ the set of all **finite permutations** of $\mathbb{N}$. Write S_π for the map, from $\mathcal{X}^{\mathbb{N}}$ back onto itself, defined by*

$$
\begin{aligned}
S_\pi(x_1, x_2, \ldots, x_n, \ldots) &:= (x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}, \ldots) \\
&:= (x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}, x_{n+1}, \ldots) \qquad \text{if } \pi \in \mathcal{R}(n).
\end{aligned}
$$

*Say that a function h on $\mathcal{X}^{\mathbb{N}}$ is **n-symmetric** if it is unaffected all n-permutations, that is, if $h_\pi(\mathbf{x}) := h(S_\pi\mathbf{x}) = h(\mathbf{x})$ for every n-permutation π.*

<48> **Example.** Let f be a real valued function on $\mathcal{X}$. Then the function $\sum_{i=1}^m f(x_i)$ is n-symmetric for every $m \ge n$, and the function $\limsup_{m\to\infty} \sum_{i=1}^m f(x_i)/m$ is n-symmetric for every n.

Let g be a real valued function on $\mathcal{X} \otimes \mathcal{X}$. Then $g(x_1, x_2) + g(x_2, x_1)$ is 2-symmetric. More generally, $\sum_{1\le i\ne j\le m} g(x_i, x_j)$ is n-symmetric for every $m \ge n$.

For every real valued function f on $\mathcal{X}^{\mathbb{N}}$, the function

$$
F(\mathbf{x}) := \frac{1}{n!}\sum_{\pi\in\mathcal{R}(n)} f_\pi(\mathbf{x}) = \frac{1}{n!}\sum_{\pi\in\mathcal{R}(n)} f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}, x_{n+1}, \ldots)
$$

□ is n-symmetric.

<49> **Definition.** *A probability measure $\mathbb{P}$ on $\mathcal{A}^{\mathbb{N}}$ is said to be **exchangeable** if it is invariant under S_π for every finite permutation π, that is, if $\mathbb{P}h = \mathbb{P}h_\pi$ for every h in $\mathcal{M}^+(\mathcal{X}^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$ and every finite permutation π. Equivalently, under $\mathbb{P}$ the random vector $\left(\xi_{\pi(1)}, \xi_{\pi(2)}, \ldots, \xi_{\pi(n)}\right)$ has the same distribution as $(\xi_1, \xi_2, \ldots, \xi_n)$, for every n-permutation π, and every n.*

The collection of all sets in $\mathcal{A}^{\mathbb{N}}$ whose indicator functions are n-symmetric forms a sub-sigma-field $\mathcal{S}_n$ of $\mathcal{A}^{\mathbb{N}}$, the ***n-symmetric sigma-field***. The $\mathcal{S}_n$-measurable functions are those $\mathcal{A}^{\mathbb{N}}$-measurable functions that are n-symmetric. The sigma-fields $\{\mathcal{S}_n : n \in \mathbb{N}\}$ form a decreasing filtration on $\mathcal{X}^{\mathbb{N}}$, with $\mathcal{S}_1 = \mathcal{A}^{\mathbb{N}}$.

<50> **Example.** Suppose $\mathbb{P}$ is exchangeable. Let f be a fixed $\mathbb{P}$-integrable function on $\mathcal{X}^{\mathbb{N}}$. Then a symmetry argument will show that

$$\mathbb{P}(f \mid \mathcal{S}_n) = \frac{1}{n!} \sum_{\pi \in \mathcal{R}(n)} f_\pi(\mathbf{x}).$$

The function—call it $F(\mathbf{x})$—on the right-hand side is n-symmetric, and hence $\mathcal{S}_n$-measurable. Also, for each bounded, $\mathcal{S}_n$-measurable function H,

$$\begin{aligned} \mathbb{P}\left(f(\mathbf{x})H(\mathbf{x})\right) &= \mathbb{P}\left(f_\pi H_\pi\right) \qquad \text{for all } \pi, \text{ by exchangeability} \\ &= \mathbb{P}\left(f_\pi H\right) \qquad \text{for all } \pi \text{ in } \mathcal{R}(n) \\ &= \frac{1}{n!} \sum_{\pi \in \mathcal{R}(n)} \mathbb{P}\left(f_\pi(\mathbf{x})H(\mathbf{x})\right) = \mathbb{P}\left(F(\mathbf{x})H(\mathbf{x})\right). \end{aligned}$$

As a special case, if f depends only on the first coordinate then we have

$$\mathbb{P}\left(f(x_1) \mid \mathcal{S}_n)\right) = \frac{1}{n!} \sum_{\pi \in \mathcal{R}(n)} f\left(x_{\pi(1)}\right) = P_n f,$$

□ where P_n denotes the empirical measure, as in Example <44>.

When the coordinate maps are independent under an exchangeable $\mathbb{P}$, the symmetric sigma-field $\mathcal{S}_\infty$ becomes trivial, and conditional expectations (such as $\mathbb{P}\left(f(x_1) \mid \mathcal{S}_\infty\right)$) reduce to constants.

<51> **Theorem.** *(Hewitt-Savage zero-one law) If $\mathbb{P} = P^{\mathbb{N}}$, the symmmetric sigma-field $\mathcal{S}_\infty$ is trivial: for each F in $\mathcal{S}_\infty$, either $\mathbb{P}F = 0$ or $\mathbb{P}F = 1$.*

Proof. Write $h(\mathbf{x})$ for the indicator function of F, a set in $\mathcal{S}_\infty$. By definition, $h_\pi = h$ for every finite permutation. Equip $\mathcal{X}^{\mathbb{N}}$ with the filtration $\mathcal{F}_n = \sigma\{x_i : i \le n\}$. Notice that $\mathcal{F}_\infty := \sigma\left(\cup_{n \in \mathbb{N}} \mathcal{F}_n\right) = \mathcal{A}^{\mathbb{N}} = \mathcal{S}_1$.

The martingale $Y_n := \mathbb{P}(F \mid \mathcal{F}_n)$ converges almost surely to $\mathbb{P}(F \mid \mathcal{F}_\infty) = F$, and also, by Dominated Convergence, $\mathbb{P}|h - Y_n|^2 \to 0$.

The $\mathcal{F}_n$-measurable random variable Y_n may be written as $h_n(x_1, \dots, x_n)$, for some $\mathcal{A}^n$-measurable h_n on $\mathcal{X}^n$. The random variable $Z_n := h_n(x_{n+1}, \dots, x_{2n})$ is independent of Y_n, and it too converges in $\mathcal{L}^2$ to h: if π denotes the $2n$-permutation that interchanges i and $i+n$, for $1 \le i \le n$, then, by exchangeability,

$$\mathbb{P}|h(\mathbf{x}) - Z_n|^2 = \mathbb{P}|h_\pi(\mathbf{x}) - h_n\left(x_{\pi(n+1)}, \dots, x_{\pi(2n)}\right)|^2 = \mathbb{P}|h(\mathbf{x}) - Y_n|^2 \to 0.$$

The random variables Z_n and Y_n are independent, and they both converge in $\mathcal{L}^2(\mathbb{P})$-norm to F. Thus

$$0 = \lim_{n\to\infty} \mathbb{P}|Y_n - Z_n|^2 = \lim_{n\to\infty} \left(\mathbb{P}Y_n^2 - 2\left(\mathbb{P}Y_n\right)\left(\mathbb{P}Z_n\right) + \mathbb{P}Z_n^2\right) = \mathbb{P}F - 2(\mathbb{P}F)^2 + \mathbb{P}F.$$

□ It follows that either $\mathbb{P}F = 0$ or $\mathbb{P}F = 1$.

In a sense made precise by Problem [17], the product measures $P^{\mathbb{N}}$ are the extreme examples of exchangeable probability measures—they are the extreme points in the convex set of all exchangeable probability measures on $\mathcal{A}^{\mathbb{N}}$. A celebrated result of de Finetti (1937) asserts that all the exchangeable probabilities can be built up from mixtures of product measures, in various senses. The simplest general version of the de Finetti result is expressed as an assertion of conditional independence.

<52> **Theorem.** *Under an exchangeable probability distribution $\mathbb{P}$ on $(\mathcal{X}^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$, the coordinate maps are conditionally independent given the symmetric sigma-field $\mathcal{S}_\infty$. That is, for all sets A_i in $\mathcal{A}$,*

$$\mathbb{P}\left(x_1 \in A_1, \ldots, x_m \in A_m \mid \mathcal{S}_\infty\right) = \mathbb{P}\left(x_1 \in A_1 \mid \mathcal{S}_\infty\right) \times \ldots \times \mathbb{P}\left(x_m \in A_m \mid \mathcal{S}_\infty\right)$$

almost surely, for every m.

Proof. Consider only the typical case where $m = 3$. The proof of the general case is similar. Write f_i for the indicator function of A_i. Abbreviate $\mathbb{P}(\cdot \mid \mathcal{S}_n)$ to $\mathbb{P}_n$, for $n \in \overline{\mathbb{N}}$. From Example <50>, for $n \geq 3$,

$$n^3 \left(\mathbb{P}_n f_1(x_1)\right) \left(\mathbb{P}_n f_2(x_2)\right) \left(\mathbb{P}_n f_3(x_3)\right) = \sum \{1 \leq i, j, k \leq n\} f_1(x_i) f_2(x_j) f_3(x_k).$$

On the right-hand side, there are $n(n-1)(n-2)$ triples of distinct subscripts (i, j, k), leaving $O(n^2)$ of them with at least one duplicated subscript. The latter contribute a sum bounded in absolute value by a multiple of n^2; the former appear in the sum that Example <50> identifies as $\mathbb{P}_n \left(f_1(x_1) f_2(x_2) f_3(x_3)\right)$. Thus

$$\left(\mathbb{P}_n f_1(x_1)\right) \left(\mathbb{P}_n f_2(x_2)\right) \left(\mathbb{P}_n f_3(x_3)\right) = \frac{n(n-1)(n-2)}{n^3} \mathbb{P}_n \left(f_1(x_1) f_2(x_2) f_3(x_3)\right) + O(n^{-1}).$$

By the convergence of reverse martingales, in the limit we get

$$\left(\mathbb{P}_\infty f_1(x_1)\right) \left(\mathbb{P}_\infty f_2(x_2)\right) \left(\mathbb{P}_\infty f_3(x_3)\right) = \mathbb{P}_\infty \left(f_1(x_1) f_2(x_2) f_3(x_3)\right),$$

□ the desired factorization.

When conditional distributions exist, it is easy to extract from Theorem <52> the representation of $\mathbb{P}$ as a mixture of product measures.

<53> **Theorem.** *Let $\mathcal{A}$ be the Borel sigma-field of a separable metric space $\mathcal{X}$. Let $\mathbb{P}$ be an exchangeable probability measure on $\mathcal{A}^{\mathbb{N}}$, under which the distribution P of x_1 is tight. Then there exists an $\mathcal{S}_\infty$-measurable map T into $[0,1]^{\mathbb{N}}$, with distribution $\mathbb{Q}$, for which conditional distributions $\{\mathbb{P}_t : t \in \mathcal{T}\}$ exist, and $\mathbb{P}_t = P_t^{\mathbb{N}}$, a product measure, for $\mathbb{Q}$ almost all t.*

Proof. Let $\mathcal{E} := \{E_i : i \in \mathbb{N}\}$ be a countable generating class for the sigma-field $\mathcal{A}$, stable under finite intersections and containing $\mathcal{X}$. For each i let $T_i(\mathbf{x})$ be a version of $\mathbb{P}\left(x_1 \in E_i \mid \mathcal{S}_\infty\right)$. By symmetry, $T_i(\mathbf{x})$ is also a version of $\mathbb{P}\left(x_j \in E_i \mid \mathcal{S}_\infty\right)$, for every j. Define T as the map from $\mathcal{X}^{\mathbb{N}}$ into $\mathcal{T} := [0,1]^{\mathbb{N}}$ for which $T(\mathbf{x})$ has ith coordinate $T_i(\mathbf{x})$.

The joint distribution of x_1 and T is a probability measure Γ on the product sigma-field of $\mathcal{X} \times \mathcal{T}$, with marginals P and $\mathbb{Q}$. As shown in Section 1 of Appendix F, the assumptions on P ensure existence of a probability kernel $\mathcal{P} := \{P_t : t \in \mathcal{T}\}$ for which

$$\mathbb{P} g(x_1, T) = \Gamma^{x,t} g(x,t) = \mathbb{Q}^t P_t^x g(x,t) \qquad \text{for all } g \text{ in } \mathcal{M}^+(\mathcal{X} \times \mathcal{T}).$$

In particular, by definition of T_i and the $\mathcal{S}_\infty$-measurability of T,

$$\mathbb{Q}^t \left(t_i h(t)\right) = \mathbb{P}\left(T_i h(T)\right) = \mathbb{P}\left(\{x_1 \in E_i\} h(T)\right) = \mathbb{Q}^t \left(h(t) P_t E_i\right)$$

for all h in $\mathcal{M}^+(\mathcal{T})$, which implies that $P_t E_i = t_i$ a.e. $[\mathbb{Q}]$, for each i.

For every finite subcollection $\{E_{i_1}, \dots, E_{i_n}\}$ of $\mathcal{E}$, Theorem <52> asserts

$$\mathbb{P}\{x_1 \in E_{i_1}, \dots, x_n \in E_{i_n} \mid \mathcal{S}_\infty\} = \prod\nolimits_{j=1}^n \mathbb{P}\{x_j \in E_{i_j} \mid \mathcal{S}_\infty\} = \prod\nolimits_{j=1}^n T_{i_j}(\mathbf{x}) \qquad \text{a.e. } [\mathbb{P}],$$

which integrates to

$$\begin{aligned}\mathbb{P}\{x_1 \in E_{i_1}, \dots, x_n \in E_{i_n}\} &= \mathbb{P}\left(\prod\nolimits_{j=1}^n T_{i_j}(\mathbf{x})\right)\\ &= \mathbb{Q}^t\left(\prod\nolimits_{j=1}^n t_{i_j}\right)\\ &= \mathbb{Q}^t\left(\prod\nolimits_{j=1}^n P_t E_{i_j}\right) = \mathbb{Q}^t \mathbb{P}_t\{x_1 \in E_{i_1}, \dots, x_n \in E_{i_n}\}.\end{aligned}$$

□ A routine generating class argument completes the proof.

9. Problems

[1] Follow these steps to construct an example of a martingale $\{Z_i\}$ and a stopping time τ for which $\mathbb{P}Z_0 \neq \mathbb{P}Z_\tau\{\tau < \infty\}$.

(i) Let $\xi_1, \xi_2, \dots$ be independent, identically distributed random variables with $\mathbb{P}\{\xi_i = +1\} = 1/3$ and $\mathbb{P}\{\xi_i = -1\} = 2/3$. Define $X_0 = 0$ and $X_i := \xi_1 + \dots + \xi_i$ and $Z_i := 2^{X_i}$, for $i \geq 1$. Show that $\{Z_i\}$ is a martingale with respect to an appropriate filtration.

(ii) Define $\tau := \inf\{i : X_i = -1\}$. Show that τ is a stopping time, finite almost everywhere. Hint: Use SLLN.

(iii) Show that $\mathbb{P}Z_0 > \mathbb{P}Z_\tau$. (Should you worry about what happens on the set $\{\tau = \infty\}$?)

[2] Let τ be a stopping time for the natural filtration generated by a sequence of random variables $\{Z_n : n \in \mathbb{N}\}$. Show that $\mathcal{F}_\tau = \sigma\{Z_{\tau\wedge n} : n \in \mathbb{N}\}$.

[3] Let $\{(Z_n, \mathcal{F}_n) : n \in \mathbb{N}_0\}$ be a (sub)martingale and τ be a stopping time. Show that $\{(Z_{\tau\wedge n}, \mathcal{F}_n) : n \in \mathbb{N}_0\}$ is also a (sub)martingale. Hint: For F in $\mathcal{F}_{n-1}$, consider separately the contributions to $\mathbb{P}Z_{n\wedge\tau}F$ and $\mathbb{P}Z_{(n-1)\wedge\tau}F$ from the regions $\{\tau \leq n-1\}$ and $\{\tau \geq n\}$.

[4] Let τ be a stopping time for a filtration $\{\mathcal{F}_i : i \in \overline{\mathbb{N}}_0\}$. For an integrable random variable X, define $X_i := \mathbb{P}(X \mid \mathcal{F}_i)$. Show that

$$\mathbb{P}(X \mid \mathcal{F}_\tau) = \sum_{i \in \overline{\mathbb{N}}_0} \{\tau = i\} X_i = X_\tau \qquad \text{almost surely.}$$

Hint: Start with $X \geq 0$, so that there are no convergence problems.

[5] Let $\{(X_n, \mathcal{F}_n) : n \in \mathbb{N}_0\}$ be a positive supermartingale, and let σ and τ be stopping times (not necessarily bounded) for which $\sigma \leq \tau$ on a set F in $\mathcal{F}_\sigma$. Show that $\mathbb{P}X_\sigma\{\sigma < \infty\}F \geq \mathbb{P}X_\tau\{\tau < \infty\}F$. Hint: For each positive integer N, show that $F_N := F\{\sigma \leq N\} \in \mathcal{F}_{\sigma\wedge N}$. Use the Stopping Time Lemma to prove that $\mathbb{P}X_{\sigma\wedge N}F_N \geq \mathbb{P}X_{\tau\wedge N}F_N \geq \mathbb{P}X_\tau\{\tau \leq N\}F$, then invoke Monotone Convergence.

[6] For each positive supermartingale $\{(X_n, \mathcal{F}_n) : n \in \mathbb{N}_0\}$, and stopping times $\sigma \le \tau$, show that $\mathbb{P}\left(X_\tau\{\tau < \infty\} \mid \mathcal{F}_\sigma\right) \le X_\sigma\{\sigma < \infty\}$ almost surely.

[7] (Kolmogorov 1928) Let $\xi_1, \dots, \xi_n$ be independent random variables with $\mathbb{P}\xi_i = 0$ and $|\xi_i| \le 1$ for each i. Define $X_i := \xi_1 + \ldots + \xi_i$ and $V_i := \mathbb{P}X_i^2$. For each $\epsilon > 0$ show that $\mathbb{P}\left\{\max_{i \le n} |X_i| \le \epsilon\right\} \le (1+\epsilon)^2 / V_n$. Note the direction of the inequalities. Hint: Define a stopping time τ for which $V_n\left\{\max_{i \le n} |X_i| \le \epsilon\right\} \le V_\tau\{\tau = n\}$. Show that $\mathbb{P}V_\tau = \mathbb{P}X_\tau^2 \le (1+\epsilon)^2$.

[8] (Birnbaum & Marshall 1961) Let $0 = X_0, X_1, \dots$ be nonnegative integrable random variables that are adapted to a filtration $\{\mathcal{F}_i\}$. Suppose there exist constants θ_i, with $0 \le \theta_i \le 1$, for which

$$(*) \qquad \mathbb{P}(X_i \mid \mathcal{F}_{i-1}) \ge \theta_i X_{i-1} \qquad \text{for } i \ge 1.$$

Let $C_1 \ge C_2 \ge \ldots \ge C_{N+1} = 0$ be constants. Prove the inequality

$$(**) \qquad \mathbb{P}\{\max_{i \le N} C_i X_i \ge 1\} \le \sum_{i=1}^{N} (C_i - \theta_{i+1} C_{i+1}) \mathbb{P}X_i,$$

by following these steps.

(i) Interpret $(*)$ to mean that there exist nonnegative, $\mathcal{F}_{i-1}$-measurable random variables Y_{i-1} for which $\mathbb{P}(X_i \mid \mathcal{F}_{i-1}) = Y_{i-1} + \theta_i X_{i-1}$ almost surely. Put $Z_i := X_i - Y_{i-1} - \theta_i X_{i-1}$. Show that $C_i X_i \le C_{i-1} X_{i-1} + C_i Z_i + C_i Y_{i-1}$ almost surely.

(ii) Deduce that $C_i X_i \le M_i + \Delta$, where M_i is a martingale with $M_0 = 0$ and $\Delta := \sum_{i=1}^{N} C_i Y_{i-1}$.

(iii) Show that the left-hand side of inequality $(**)$ is less than $\mathbb{P}C_\tau X_\tau$ for an appropriate stopping time τ, then rearrange the sum for $\mathbb{P}\Delta$ to get the asserted upper bound.

[9] (Doob 1953, page 317) Suppose $S_1, \dots, S_n$ is a nonnegative submartingale, with $\mathbb{P}S_i^p < \infty$ for some fixed $p > 1$. Let $q > 1$ be defined by $p^{-1} + q^{-1} = 1$. Show that $\mathbb{P}\left(\max_{i \le n} S_i^p\right) \le q^p \mathbb{P}S_n^p$, by following these steps.

(i) Write M_n for $\max_{i \le n} S_i$. For fixed $x > 0$, and an appropriate stopping time τ, apply the Stopping Time Lemma to show that

$$x\mathbb{P}\{M_n \ge x\} \le \mathbb{P}S_\tau\{S_\tau \ge x\} \le \mathbb{P}S_n\{M_n \ge x\}.$$

(ii) Show that $\mathbb{P}X^p = \int_0^\infty p x^{p-1} \mathbb{P}\{X \ge x\}\, dx$ for each nonnegative random variable X.

(iii) Show that $\mathbb{P}M_n^p \le q \mathbb{P}S_n M_n^{p-1}$.

(iv) Bound the last product using Hölder's inequality, then rearrange to get the stated inequality. (Any problems with infinite values?)

[10] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space such that $\mathcal{F}$ is *countably generated*: that is, $\mathcal{F} = \sigma\{B_1, B_2, \ldots\}$ for some sequence of sets $\{B_i\}$. Let μ be a finite measure on $\mathcal{F}$, dominated by $\mathbb{P}$. Let $\mathcal{F}_n := \sigma\{B_1, \dots, B_n\}$.

(i) Show that there is a partition π_n of Ω into at most 2^n disjoint sets from $\mathcal{F}_n$ such that each F in $\mathcal{F}_n$ is a union of sets from π_n.

(ii) Define $\mathcal{F}_n$-measurable random variables X_n by: for $\omega \in A \in \pi_n$,

$$X_n(\omega) = \begin{cases} \mu A/\mathbb{P}A & \text{if } \mathbb{P}A > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\mathbb{P}X_n F = \mu F$ for all F in $\mathcal{F}_n$.

(iii) Show that $(X_n, \mathcal{F}_n)$ is a positive martingale.

(iv) Show that $\{X_n\}$ is uniformly integrable. Hint: What do you know about $\mu\{X_n \geq M\}$?

(v) Let X_∞ denote the almost sure limit of the $\{X_n\}$. Show that $\mathbb{P}X_\infty F = \mu F$ for all F in $\mathcal{F}$. That is, show that X_∞ is a density for μ with respect to $\mathbb{P}$.

[11] Let $\{(X_n, \mathcal{F}_n) : n \in \mathbb{N}_0\}$ be a submartingale. For fixed constants $\alpha < \beta$ (not necessarily nonnegative), define stopping times $\sigma_1 \leq \tau_1 \leq \sigma_2 \leq \ldots$, as in Section 3. Establish the upcrossing inequality,

$$\mathbb{P}\{\tau_k \leq N\} \leq \frac{\mathbb{P}\,(X_N - \alpha)^+}{k(\beta - \alpha)}$$

for each positive integer N, by following these steps.

(i) Show that $Z_n = (X_n - \alpha)^+$ is a positive submartingale, with $Z_{\sigma_i} = 0$ if $\sigma_i < \infty$ and $Z_{\tau_i} \geq \beta - \alpha$ if $\tau_i < \infty$.

(ii) For each i, show that $Z_{\tau_i \wedge N} - Z_{\sigma_i \wedge N} \geq (\beta - \alpha)\{\tau_i \leq N\}$. Hint: Consider separately the three cases $\sigma_i > N$, $\sigma_i \leq N < \tau_i$, and $\tau_i \leq N$.

(iii) Show that $-\mathbb{P}Z_{\sigma_1 \wedge N} + \mathbb{P}Z_{\tau_k \wedge N} \geq k(\beta - \alpha)\mathbb{P}\{\tau_k \leq N\}$. Hint: Take expectations then sum over i in the inequality from part (ii). Use the Stopping Time Lemma for submartingales to prove $\mathbb{P}Z_{\tau_i \wedge N} - \mathbb{P}Z_{\sigma_{i+1} \wedge N} \leq 0$.

(iv) Show that $\mathbb{P}Z_{\tau_k \wedge N} \leq \mathbb{P}Z_N = \mathbb{P}\,(X_N - \alpha)^+$.

[12] Reprove Corollary <27> (a submartingale $\{X_n : n \in \mathbb{N}_0\}$ converges almost surely to an integrable limit if $\sup_n \mathbb{P}X_n^+ < \infty$) by following these steps.

(i) For fixed $\alpha < \beta$, use the upcrossing inequality from Problem [11] to prove that

$$\mathbb{P}\{\liminf_n X_n < \alpha < \beta < \limsup_n X_n\} = 0$$

(ii) Deduce that $\{X_n\}$ converges almost surely to a limit random variable X that might take the values $\pm\infty$.

(iii) Prove that $\mathbb{P}|X_n| \leq 2\mathbb{P}X_n^+ - \mathbb{P}X_1$ for every n. Deduce via Fatou's lemma that $\mathbb{P}|X| < \infty$.

[13] Suppose the offspring distribution in Example <25> has finite mean $\mu > 1$ and variance σ^2.

(i) Show that $\text{var}(Z_n) = \sigma^2\mu^{n-1} + \mu^2\text{var}(Z_{n-1})$.

(ii) Write X_n for the martingale Z_n/μ^n. Show that $\sup_n \text{var}(X_n) < \infty$.

(iii) Deduce that X_n converges both almost surely and in $\mathcal{L}^1$ to the limit X, and hence $\mathbb{P}X = 1$. In particular, the limit X cannot be degenerate at 0.

[14] Suppose the offspring distribution P from Example <25> has has finite mean $\mu > 1$. Write X_n for the martingale Z_n/μ^n, which converges almost surely to an integrable limit random variable X. Show that the limit X is nondegenerate if and only if the condition

$$(\text{XLOGX}) \qquad\qquad P^x\left(x\log(1+x)\right) < \infty,$$

holds. Follow these steps. Write μ_n for $P^x\left(x\{x \le \mu^n\}\right)$ and $\mathbb{P}_n(\cdot)$ for expectations conditional on $\mathcal{F}_n$.

(i) Show that $\sum_n(\mu - \mu_n) = P^x\left(x\sum_n\{x > \mu^n\}\right)$, which converges to a finite limit if and only if (XLOGX) holds.

(ii) Define $\widetilde{X}_n := \mu^{-n}\sum_i \xi_{n,i}\{\xi_{n,i} \le \mu^n\}\{i \le Z_{n-1}\}$. Show that $\mathbb{P}_{n-1}\widetilde{X}_n = \mu_n X_{n-1}/\mu$ almost surely. Show also that

$$X - X_N = \sum\nolimits_{n=N+1}^{\infty}(X_n - X_{n-1}) \ge \sum\nolimits_{n=N+1}^{\infty}\left(\widetilde{X}_n - X_{n-1}\right) \qquad \text{almost surely.}$$

(iii) Show that, for some constant C_1,

$$\sum\nolimits_n \mathbb{P}\{\widetilde{X}_n \ne X_n\} \le \sum\nolimits_n \mu^{n-1}P\{x > \mu^n\} \le C_1\mu < \infty.$$

Deduce that $\sum_n\left(\widetilde{X}_n - X_n\right)$ converges almost surely to a finite limit.

(iv) Write var_{n-1} for the conditional variance corresponding to $\mathbb{P}_{n-1}$. Show that

$$\text{var}_{n-1}(\widetilde{X}_n) = \mu^{-2n}\sum\nolimits_i\{i \le Z_{n-1}\}\text{var}_{n-1}\left(\xi_{n,i}\{\xi_{n,i} \le \mu^n\}\right).$$

Deduce, via (ii), that

$$\sum\nolimits_n \mathbb{P}\left(\widetilde{X}_n - \mu_n X_{n-1}/\mu\right)^2 \le \sum\nolimits_n \mu^{-n-1}Px^2\{x \le \mu^n\} \le C_2\mu < \infty,$$

for some constant C_2. Conclude that $\sum_n\left(\widetilde{X}_n - \mu_n X_{n-1}/\mu\right)$ is a martingale, which converges both almost surely and in $\mathcal{L}^1$.

(v) Deduce from (iii), (iv), and the fact that $\sum_n(X_n - X_{n-1})$ converges almost surely, that $\sum_n X_{n-1}(1 - \mu_n/\mu)$ converges almost surely to a finite limit.

(vi) Suppose $\mathbb{P}\{X > 0\} > 0$. Show that there exists an ω for which both $\sum_n X_{n-1}(\omega)(1 - \mu_n/\mu) < \infty$ and $\lim X_{n-1}(\omega) > 0$. Deduce via (i) that (XLOGX) holds.

(vii) Suppose (XLOGX) holds. From (i) deduce that $\mathbb{P}\left(\sum_n X_{n-1}(1 - \mu_n/\mu)\right) < \infty$. Deduce via (iv) that $\sum_n(\widetilde{X}_n - X_{n-1})$ converges in $\mathcal{L}^1$. Deduce via (ii) that $\mathbb{P}X \ge \mathbb{P}\left(X_N + \sum_{n=N+1}^{\infty}(\widetilde{X}_n - X_{n-1})\right) = 1 - o(1)$ as $N \to \infty$, from which it follows that X is nondegenerate. (In fact, $\mathbb{P}|X_n - X| \to 0$. Why?)

[15] Let $\{\xi_i : i \in \mathbb{N}\}$ be a martingale difference array for which $\sum_{i\in\mathbb{N}}\mathbb{P}\left(\xi_i^2/i^2\right) < \infty$.

(i) Define $X_n := \sum_{i=1}^n \xi_i/i$. Show that $\sup_n \mathbb{P}X_n^2 < \infty$. Deduce that $X_n(\omega)$ converges to a finite limit for almost all ω.

(ii) Invoke Kronecker's lemma to deduce that $n^{-1}\sum_{i=1}^n \xi_i \to 0$ almost surely.

[16] Suppose $\{X_n : n \in \mathbb{N}\}$ is an exchangeable sequence of square-integrable random variables. Show that $\text{cov}(X_1, X_2) \ge 0$. Hint: Each X_i must have the same variance, V; each pair X_i, X_j, for $i \ne j$, must have the same covariance, C. Consider $\text{var}\left(\sum_{i\le n} X_i\right)$ for arbitrarily large n.

[17] (Hewitt & Savage 1955, Section 5) Let $\mathbb{P}$ be exchangeable, in the sense of Definition <49>.

(i) Let f be a bounded, $\mathcal{A}^n$-measurable function on $\mathcal{X}^n$. Define $X := f(x_1, \ldots, x_n)$ and $Y := f(x_{n+1}, \ldots, x_{2n})$. Use Problem [16] to show that $\mathbb{P}(XY) \geq (\mathbb{P}X)(\mathbb{P}Y)$, with equality if $\mathbb{P}$ is a product measure.

(ii) Suppose $\mathbb{P} = \alpha_1 \mathbb{Q}_1 + \alpha_2 \mathbb{Q}_2$, with $\alpha_i > 0$ and $\alpha_1 + \alpha_2 = 1$, where $\mathbb{Q}_1$ and $\mathbb{Q}_2$ are distinct exchangeable probability measures. Let f be a bounded, measurable function on some $\mathcal{X}^n$ for which $\mu_1 := \mathbb{Q}_1 f(x_1, \ldots, x_n) \neq \mathbb{Q}_2 f(x_1, \ldots, x_n) =: \mu_2$. Define X and Y as in part (i). Show that $\mathbb{P}(XY) > (\mathbb{P}X)(\mathbb{P}Y)$. Hint: Use strict convexity of the square function to show that $\alpha_1 \mu_1^2 + \alpha_2 \mu_2^2 > (\alpha_1 \mu_1 + \alpha_2 \mu_2)^2$. Deduce that $\mathbb{P}$ is not a product measure.

(iii) Suppose $\mathbb{P}$ is not a product measure. Explain why there exists an $E \in \mathcal{A}^n$ and a bounded measurable function g for which

$$\mathbb{P}\big(\{\mathbf{z} \in E\} g(x_{n+1}, x_{n+2}, \ldots)\big) \neq (\mathbb{P}\{\mathbf{z} \in E\})\,(\mathbb{P} g(x_{n+1}, x_{n+2}, \ldots)),$$

where $\mathbf{z} := (x_1, \ldots, x_n)$. Define $\alpha = \mathbb{P}\{\mathbf{z} \in E\}$. Show that $0 < \alpha < 1$. For each $h \in \mathcal{M}^+(\mathcal{X}^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$, define

$$\begin{aligned} \mathbb{Q}_1 h &:= \mathbb{P}\left(\{\mathbf{z} \in E\} h(x_{n+1}, x_{n+2}, \ldots)\right)/\alpha, \\ \mathbb{Q}_2 h &:= \mathbb{P}\big(\{\mathbf{z} \in E^c\} h(x_{n+1}, x_{n+2}, \ldots)\big)/(1-\alpha). \end{aligned}$$

Show that $\mathbb{Q}_1$ and $\mathbb{Q}_2$ correspond to distinct exchangeable probability measures for which $\mathbb{P} = \alpha \mathbb{Q}_1 + (1-\alpha)\mathbb{Q}_2$. That is, $\mathbb{P}$ is not an extreme point of the set of all exchangeable probability measures on $\mathcal{A}^{\mathbb{N}}$.

10. Notes

De Moivre used what would now be seen as a martingale method in his solution of the gambler's ruin problem. (Apparently first published in 1711, according to Thatcher (1957). See pages 51–53 of the 1967 reprint of the third edition of de Moivre (1718).)

The name *martingale* is due to Ville (1939). Lévy (1937, chapter VIII), expanding on earlier papers (Lévy 1934, 1935*a*, 1935*b*), had treated martingale differences, identifying them as sequences satisfying his condition ($\mathcal{C}$). He extended several results for sums of independent variables to martingales, including Kolmogorov's maximal inequality and strong law of large numbers (the version proved in Section 4.6), and even a central limit theorem, extending Lindeberg's method (to be discussed, for independent summands, in Section 7.2). He worked with martingales stopped at random times, in order to have sums of conditional variances close to specified constant values.

Doob (1940) established convergence theorems (without using stopping times) for martingales and reversed martingales, calling them sequences with "property $\mathcal{E}$." He acknowledged (footnote to page 458) that the basic maximal inequalities were "implicit in the work of Ville" and that the method of proof he used "was used by Lévy (1937), in a related discussion." It was Doob, especially with his stochastic processes book (Doob 1953—see, in particular the historical notes to Chapter VII, starting page 629), who was the major driving force behind the recognition of martingales as one of the most important tools of probability theory. See Lévy's comments in Note II of the 1954 edition of Lévy (1937) and in Lévy (1970, page 118) for the relationship between his work and Doob's.

I first understood some martingale theory by reading the superb text of Ash (1972, Chapter 7), and from conversations with Jim Pitman. The material in Section 3 on positive supermartingales was inspired by an old set of notes for lectures given by Pitman at Cambridge. I believe the lectures were based in part on the original French edition of the book Neveu (1975). I have also borrowed heavily from that book, particularly so for Theorems <26> and <41>. The book of Hall & Heyde (1980), although aimed at central limit theory and its application, contains much about martingales in discrete time. Dellacherie & Meyer (1982, Chapter V) covered discrete-time martingales as a preliminary to the detailed study of martingales in continuous time.

Exercise <15> comes from Aldous (1983, p. 47).

Inequality <20> is due to Dubins (1966). The upcrossing inequality of Problem [11] comes from the same paper, slightly weakening an analogous inequality of Doob (1953, page 316). Krickeberg (1963, Section IV.3) established the decomposition (Theorem <26>) of submartingales as differences of positive supermartingales.

I adapted the branching process result of Problem [14], which is due to Kesten & Stigum (1966), from Asmussen & Hering (1983, Chapter II).

The reversed submartingale part of Example <44> comes from Pollard (1981). The zero-one law of Theorem <51> for symmetric events is due to Hewitt & Savage (1955). The study of exchangeability has progressed well beyond the original representation theorem. Consult Aldous (1983) if you want to know more.

References

Aldous, D. (1983), 'Exchangeability and related topics', *Springer Lecture Notes in Mathematics* **1117**, 1–198.

Ash, R. B. (1972), *Real Analysis and Probability*, Academic Press, New York.

Asmussen, S. & Hering, H. (1983), *Branching processes*, Birkhäuser.

Bauer, H. (1981), *Probability Theory and Elements of Measure Theory*, second english edn, Academic Press.

Birnbaum, Z. W. & Marshall, A. W. (1961), 'Some multivariate Chebyshev inequalities with extensions to continuous parameter processes', *Annals of Mathematical Statistics* pp. 687–703.

de Finetti, B. (1937), 'La prévision: ses lois logiques, ses sources subjectives', *Annales de l'Institut Henri Poincaré* **7**, 1–68. English translation by H. Kyburg in Kyberg & Smokler 1980.

de Moivre, A. (1718), *The Doctrine of Chances*, first edn. Second edition 1738. Third edition, from 1756, reprinted in 1967 by Chelsea, New York.

Dellacherie, C. & Meyer, P. A. (1982), *Probabilities and Potential B: Theory of Martingales*, North-Holland, Amsterdam.

Doob, J. L. (1940), 'Regularity properties of certain families of chance variables', *Transactions of the American Mathematical Society* **47**, 455–486.

Doob, J. L. (1949), 'Application of the theory of martingales', *Colloques Internationaux du Centre National de la Recherche Scientifique* pp. 23–27.

Doob, J. L. (1953), *Stochastic Processes*, Wiley, New York.

Dubins, L. E. (1966), 'A note on upcrossings of semimartingales', *Annals of Mathematical Statistics* **37**, 728.

Hall, P. & Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, Academic Press, New York, NY.

Hewitt, E. & Savage, L. J. (1955), 'Symmetric measures on cartesian products', *Transactions of the American Mathematical Society* **80**, 470–501.

Kesten, H. & Stigum, B. P. (1966), 'Additional limit theorems for indecomposable multidimensional Galton-Watson process', *Annals of Mathematical Statistics* **37**, 1463–1481.

Kolmogorov, A. (1928), 'Über die Summen durch den Zufall bestimmter unabhängiger Größen', *Mathematische Annalen* **99**, 309–319. Corrections: same journal, volume 102, 1929, pages 484–488.

Krickeberg, K. (1963), *Wahrscheinlichkeitstheorie*, Teubner. English translation, 1965, Addison-Wesley.

Kyberg, H. E. & Smokler, H. E. (1980), *Studies in Subjective Probability*, second edn, Krieger, Huntington, New York. Reprint of the 1964 Wiley edition.

Lévy, P. (1934), 'L'addition de variables aléatoires enchaînées et la loi de Gauss', *Bull. Soc. Math. France* **62**, 42–43.

Lévy, P. (1935*a*), 'Propriétés asymptotiques des sommes de variables aléatoires enchaînées', *Comptes Rendus de l'Academie des Sciences, Paris* **199**, 627–629.

Lévy, P. (1935*b*), 'Propriétés asymptotiques des sommes de variables aléatoires enchaînées', *Bull. Soc. math* **59**, 1–32.

Lévy, P. (1937), *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris. Second edition, 1954.

Lévy, P. (1970), *Quelques Aspects de la Pensée d'un Mathématicien*, Blanchard, Paris.

Neveu, J. (1975), *Discrete-Parameter Martingales*, North-Holland, Amsterdam.

Pollard, D. (1981), 'Limit theorems for empirical processes', *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **57**, 181–195.

Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer, New York.

Thatcher, A. R. (1957), 'A note on the early solutions of the problem of the duration of play', *Biometrika* **44**, 515–518.

Ville, J. (1939), *Etude Critique de la Notion de Collectif*, Gauthier-Villars, Paris.