

17.6: Backwards Martingales

Basic Theory

A *backwards martingale* is a stochastic process that satisfies the martingale property reversed in time, in a certain sense. In some ways, backward martingales are simpler than their forward counterparts, and in particular, satisfy a convergence theorem similar to the convergence theorem for ordinary martingales. The importance of backward martingales stems from their numerous applications. In particular, some of the fundamental theorems of classical probability can be formulated in terms of backward martingales.

Definitions

As usual, we start with a stochastic process $\mathbf{Y} = \{Y_t : t \in T\}$ on an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, having state space \mathbb{R} , and where the index set T (representing time) is either \mathbb{N} (discrete time) or $[0, \infty)$ (continuous time). So to review what all this means, Ω is the sample space, \mathcal{F} the σ -algebra of events, \mathbb{P} the probability measure on (Ω, \mathcal{F}) , and Y_t is a random variable with values in \mathbb{R} for each $t \in T$. But at this point our formulation diverges. Suppose that \mathcal{G}_t is a sub σ -algebra of \mathcal{F} for each $t \in T$, and that $\mathfrak{G} = \{\mathcal{G}_t : t \in T\}$ is *decreasing* so that if $s, t \in T$ with $s \leq t$ then $\mathcal{G}_t \subseteq \mathcal{G}_s$. Let $\mathcal{G}_\infty = \bigcap_{t \in T} \mathcal{G}_t$. We assume that Y_t is measurable with respect to \mathcal{G}_t and that $\mathbb{E}(|Y_t|) < \infty$ for each $t \in T$.

The process $\mathbf{Y} = \{Y_t : t \in T\}$ is a *backwards martingale* (or *reversed martingale*) with respect to $\mathfrak{G} = \{\mathcal{G}_t : t \in T\}$ if $\mathbb{E}(Y_s | \mathcal{G}_t) = Y_t$ for all $s, t \in T$ with $s \leq t$.

A backwards martingale can be formulated as an ordinary martingale by using negative times as the indices. Let $T^- = \{-t : t \in T\}$, so that if $T = \mathbb{N}$ (the discrete case) then T^- is the set of non-positive integers, and if $T = [0, \infty)$ (the continuous case) then $T^- = (-\infty, 0]$. Recall also that the standard martingale definitions make sense for any totally ordered index set.

Suppose again that $\mathbf{Y} = \{Y_t : t \in T\}$ is a backwards martingale with respect to $\mathfrak{G} = \{\mathcal{G}_t : t \in T\}$. Let $X_t = Y_{-t}$ and $\mathcal{F}_t = \mathcal{G}_{-t}$ for $t \in T^-$. Then $\mathbf{X} = \{X_t : t \in T^-\}$ is a martingale with respect to $\mathfrak{F} = \{\mathcal{F}_t : t \in T^-\}$.

Proof

Since \mathfrak{G} is a decreasing family of sub σ -algebras of \mathcal{F} , the collection \mathfrak{F} is an increasing family of sub σ -algebras of \mathcal{F} , and hence is a filtration. Next, $X_t = Y_{-t}$ is measurable with respect to $\mathcal{G}_{-t} = \mathcal{F}_t$ for $t \in T^-$, so \mathbf{X} is adapted to \mathfrak{F} . Finally, if $s, t \in T^-$ with $s \leq t$ then $-t \leq -s$ so

$$\mathbb{E}(X_t | \mathcal{F}_s) = \mathbb{E}(Y_{-t} | \mathcal{G}_{-s}) = Y_{-s} = X_s \quad (17.6.1)$$

Most authors define backwards martingales with negative indices, as above, in the first place. There are good reasons for doing so, since some of the fundamental theorems of martingales apply immediately to backwards martingales. However, for the *applications* of backwards martingales, this notation is artificial and clunky, so for the most part, we will use our original definition. The next result is another way to view a backwards martingale as an ordinary martingale. This one preserves nonnegative time, but introduces a finite time horizon. For $t \in T$, let $T_t = \{s \in T : s \leq t\}$, a notation we have used often before.

Suppose again that $\mathbf{Y} = \{Y_t : t \in T\}$ is a backwards martingale with respect to $\mathfrak{G} = \{\mathcal{G}_t : t \in T\}$. Fix $t \in T$ and define $X_s^t = Y_{t-s}$ and $\mathcal{F}_s^t = \mathcal{G}_{t-s}$ for $s \in T_t$. Then $\mathbf{X}^t = \{X_s^t : s \in T_t\}$ is a martingale relative to $\mathfrak{F}^t = \{\mathcal{F}_s^t : s \in T_t\}$.

Proof

The proof is essentially the same as for the previous result. Since \mathfrak{G} is a decreasing family of sub σ -algebras of \mathcal{F} , the collection \mathfrak{F}^t is an increasing family of sub σ -algebras of \mathcal{F} , and hence is a filtration. Next, $X_s^t = Y_{t-s}$ is measurable with respect to $\mathcal{G}_{t-s} = \mathcal{F}_s^t$ for $s \in T_t$, so \mathbf{X}^t is adapted to \mathfrak{F}^t . Finally, if $r, s \in T_t$ with $r \leq s$ then $t-s \leq t-r$ so

$$\mathbb{E}(X_s^t | \mathcal{F}_r^t) = \mathbb{E}(Y_{t-s} | \mathcal{G}_{t-r}) = Y_{t-r} = X_r^t \quad (17.6.2)$$

Properties

Backwards martingales satisfy a simple and important property.

Suppose that $\mathbf{Y} = \{Y_t : t \in T\}$ is a backwards martingale with respect to $\mathfrak{G} = \{\mathcal{G}_t : t \in T\}$. Then $Y_t = \mathbb{E}(Y_0 | \mathcal{G}_t)$ for $t \in T$ and hence \mathbf{Y} is uniformly integrable.

Proof

The fact that $Y_t = \mathbb{E}(Y_0 | \mathcal{G}_t)$ for $t \in T$ follows directly from the definition of a backwards martingale. Since we have assumed that $\mathbb{E}(|Y_0|) < \infty$, it follows from a basic property that \mathbf{Y} is uniformly integrable.

Here is the *Doob backwards martingale*, analogous to the ordinary Doob martingale, and of course named for Joseph Doob. In a sense, this is the converse to the previous result.

Suppose that Y is a random variable on our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}(|Y|) < \infty$, and that $\mathfrak{G} = \{\mathcal{G}_t : t \in T\}$ is a decreasing family of sub σ -algebras of \mathcal{F} , as above. Let $Y_t = \mathbb{E}(Y | \mathcal{G}_t)$ for $t \in T$. Then $\mathbf{Y} = \{Y_t : t \in T\}$ is a backwards martingale with respect to \mathfrak{G} .

Proof

By definition, $Y_t = \mathbb{E}(Y | \mathcal{G}_t)$ is measurable with respect to \mathcal{G}_t . Also,

$$\mathbb{E}(|Y_t|) = \mathbb{E}[\mathbb{E}(|Y| | \mathcal{G}_t)] \leq \mathbb{E}[\mathbb{E}(|Y| | \mathcal{G}_t)] = \mathbb{E}(|Y|) < \infty, \quad t \in T \quad (17.6.3)$$

Next, suppose that $s, t \in T$ with $s \leq t$. Then $\mathcal{G}_t \subseteq \mathcal{G}_s$ so by the tower property of conditional expected value,

$$\mathbb{E}(Y_s | \mathcal{G}_t) = \mathbb{E}[\mathbb{E}(Y | \mathcal{G}_s) | \mathcal{G}_t] = \mathbb{E}(Y | \mathcal{G}_t) = Y_t \quad (17.6.4)$$

The convergence theorems are the most important results for the applications of backwards martingales. Recall once again that for $k \in [1, \infty)$, the k -norm of a real-valued random variable X is

$$\|X\|_k = \left[\mathbb{E}(|X|^k) \right]^{1/k} \quad (17.6.5)$$

and the normed vector space \mathcal{L}_k consists of all X with $\|X\|_k < \infty$. Convergence in the space \mathcal{L}_1 is also referred to as *convergence in mean*, and convergence in the space \mathcal{L}_2 is referred to as *convergence in mean square*. Here is the primary *backwards martingale convergence theorem*:

Suppose again that $\mathbf{Y} = \{Y_t : t \in T\}$ is a backwards martingale with respect to $\mathfrak{G} = \{\mathcal{G}_t : t \in T\}$. Then there exists a random variable Y_∞ such that

1. $Y_t \rightarrow Y_\infty$ as $t \rightarrow \infty$ with probability 1.
2. $Y_t \rightarrow Y_\infty$ as $t \rightarrow \infty$ in mean.
3. $Y_\infty = \mathbb{E}(Y_0 | \mathcal{G}_\infty)$.

Proof

The proof is essentially the same as the ordinary martingale convergence theorem if we use the martingale constructed from \mathbf{Y} above. So, fix $t \in T$ and let $T_t = \{s \in T : s \leq t\}$. Let $X_s^t = Y_{t-s}$ and $\mathcal{F}_s^t = \mathcal{G}_{t-s}$ for $s \in T_t$, so that $\mathbf{X}^t = \{X_s^t : s \in T_t\}$ is a martingale relative to $\mathfrak{F}^t = \{\mathcal{F}_s^t : s \in T_t\}$. Now, for $a, b \in \mathbb{R}$ with $a < b$, let $U_t(a, b)$ denote the number of up-crossings of $[a, b]$ by \mathbf{X}^t on T_t . Note that $U_t(a, b)$ is also the number of down-crossings of $[a, b]$ by \mathbf{Y} on T_t . By the up-crossing inequality applied to the martingale \mathbf{X}^t ,

$$\mathbb{E}[U_t(a, b)] \leq \frac{1}{b-a} [\mathbb{E}(|X_t|) + |a|] = \frac{1}{b-a} [\mathbb{E}(|Y_0|) + |a|] \quad (17.6.6)$$

Now let $U_\infty(a, b)$ denote the number of down-crossings of $[a, b]$ by \mathbf{Y} on all of T . Since $U_t \uparrow U_\infty$ as $t \rightarrow \infty$ it follows from the monotone convergence theorem that

$$\mathbb{E}[U_\infty(a, b)] \leq \frac{1}{b-a} [\mathbb{E}(|Y_0|) + |a|] \quad (17.6.7)$$

Hence with probability 1, $U_\infty(a, b) < \infty$ for every $a, b \in \mathbb{Q}$ with $a < b$. By the characterization of convergence in terms of down-crossings (completely analogous to the one for up-crossings), there exists a random variable Y_∞ with values in $\mathbb{R}^* = \mathbb{R} \cup \{-\infty, \infty\}$ such that $Y_t \rightarrow Y_\infty$ as $t \rightarrow \infty$. By Fatou's lemma,

$$\mathbb{E}(|Y_\infty|) \leq \liminf_{t \rightarrow \infty} \mathbb{E}(|Y_t|) \leq \mathbb{E}(|Y_0|) < \infty \quad (17.6.8)$$

In particular, $\mathbb{P}(Y_\infty \in \mathbb{R}) = 1$. Since \mathbf{Y} is uniformly integrable, and $Y_\infty \in \mathcal{L}_1$, it follows that $Y_t \rightarrow Y_\infty$ as $t \rightarrow \infty$ in \mathcal{L}_1 also.

It remains to show that $Y_\infty = \mathbb{E}(Y_0 | \mathcal{G}_\infty)$. Let $A \in \mathcal{G}_\infty$. Then $A \in \mathcal{G}_t$ for every $t \in T$. Since $Y_t = \mathbb{E}(Y_0 | \mathcal{G}_t)$ it follows by definition that $\mathbb{E}(Y_t; A) = \mathbb{E}(Y_0; A)$ for every $t \in T$. Letting $t \rightarrow \infty$ and using the dominated convergence theorem, gives $\mathbb{E}(Y_\infty; A) = \mathbb{E}(Y_0; A)$. Hence $Y_\infty = \mathbb{E}(Y_0 | \mathcal{G}_\infty)$.

As a simple extension of the last result, if $Y_0 \in \mathcal{L}_k$ for some $k \in [1, \infty)$ then the convergence is in \mathcal{L}_k also.

Suppose again that $\mathbf{Y} = \{Y_t : t \in T\}$ is a backwards martingale relative to $\mathfrak{G} = \{\mathcal{G}_t : t \in T\}$. If $Y_0 \in \mathcal{L}_k$ for some $k \in [1, \infty)$ then $Y_t \rightarrow Y_\infty$ as $t \rightarrow \infty$ in \mathcal{L}_k .

Proof

The previous result applies, of course, so we know that there exists a random variable $Y_\infty \in \mathcal{L}_1$ such that $Y_t \rightarrow Y_\infty$ as $t \rightarrow \infty$ with probability 1 and in \mathcal{L}_1 . The function $x \mapsto |x|^k$ is convex on \mathbb{R} so by Jensen's inequality for conditional expected value,

$$\mathbb{E}(|Y_t|^k) = \mathbb{E}[\mathbb{E}(|Y_t|^k | \mathcal{G}_t)] \leq \mathbb{E}[\mathbb{E}(|Y_0|^k | \mathcal{G}_t)] = \mathbb{E}(|Y_0|^k) < \infty \quad (17.6.9)$$

so $Y_t \in \mathcal{L}_k$ for every $t \in T$. By Fatou's lemma,

$$\mathbb{E}(|Y_\infty|^k) \leq \liminf_{t \rightarrow \infty} \mathbb{E}(|Y_t|^k) \leq \mathbb{E}(|Y_0|^k) < \infty \quad (17.6.10)$$

so $Y_\infty \in \mathcal{L}_k$ also. Next, since $Y_t = \mathbb{E}(Y_0 | \mathcal{G}_t)$ and Y_∞ is measurable with respect to \mathcal{G}_t , we can use Jensen's inequality again to get

$$|Y_t - Y_\infty|^k = |\mathbb{E}(Y_0 - Y_\infty | \mathcal{G}_t)|^k \leq \mathbb{E}(|Y_0 - Y_\infty|^k | \mathcal{G}_t) \quad (17.6.11)$$

It follows that the family of random variables $\{|Y_t - Y_\infty|^k : t \in T\}$ is uniformly integrable, and hence $\mathbb{E}(|Y_t - Y_\infty|^k) \rightarrow 0$ as $t \rightarrow \infty$.

Applications

The Strong Law of Large Numbers

The strong law of large numbers is one of the fundamental theorems of classical probability. Our previous proof required that the underlying distribution have finite variance. Here we present an elegant proof using backwards martingales that does not require this extra assumption. So, suppose that $\mathbf{X} = \{X_n : n \in \mathbb{N}_+\}$ is a sequence of independent, identically distributed random variables with common mean $\mu \in \mathbb{R}$. In statistical terms, \mathbf{X} corresponds to *sampling* from the underlying distribution. Next let

$$Y_n = \sum_{i=1}^n X_i, \quad n \in \mathbb{N} \quad (17.6.12)$$

so that $\mathbf{Y} = \{Y_n : n \in \mathbb{N}\}$ is the partial sum process associated with \mathbf{X} . Recall that the sequence \mathbf{Y} is also a discrete-time random walk. Finally, let $M_n = Y_n/n$ for $n \in \mathbb{N}_+$ so that $\mathbf{M} = \{M_n : n \in \mathbb{N}_+\}$ is the sequence of *sample means*.

The law of large numbers

1. $M_n \rightarrow \mu$ as $n \rightarrow \infty$ with probability 1.
2. $M_n \rightarrow \mu$ as $n \rightarrow \infty$ in mean.

Proof

As usual, let $(\Omega, \mathcal{F}, \mathbb{P})$ denote the underlying probability space. Also, equalities involving random variables (and particularly conditional expected values) are assumed to hold with probability 1. Now, for $n \in \mathbb{N}$, let

$$\mathcal{G}_n = \sigma\{Y_n, Y_{n+1}, Y_{n+2}, \dots\} = \sigma\{Y_n, X_{n+1}, X_{n+2}, \dots\} \quad (17.6.13)$$

so that $\mathcal{G} = \{\mathcal{G}_n : n \in \mathbb{N}\}$ is a decreasing family of sub σ -algebras of \mathcal{F} . The core of the proof is to show that \mathbf{M} is a backwards martingale relative to \mathcal{G} . Let $n \in \mathbb{N}_+$. Clearly M_n is measurable with respect to \mathcal{G}_n . By independence, $\mathbb{E}(X_i | \mathcal{G}_n) = \mathbb{E}(X_i | Y_n)$ for $i \in \{1, 2, \dots, n\}$. By symmetry (the sequence \mathbf{X} is exchangeable), $\mathbb{E}(X_i | Y_n) = \mathbb{E}(X_j | Y_n)$ for $i, j \in \{1, 2, \dots, n\}$. Hence for $i \in \{1, 2, \dots, n\}$

$$Y_n = \mathbb{E}(Y_n | \mathcal{G}_n) = \sum_{j=1}^n \mathbb{E}(X_j | \mathcal{G}_n) = \sum_{j=1}^n \mathbb{E}(X_i | \mathcal{G}_n) = n\mathbb{E}(X_i | \mathcal{G}_n) \quad (17.6.14)$$

so that $\mathbb{E}(X_i | \mathcal{G}_n) = Y_n/n = M_n$ for each $i \in \{1, 2, \dots, n\}$. Next,

$$\mathbb{E}(Y_n | \mathcal{G}_{n+1}) = \mathbb{E}(Y_{n+1} - X_{n+1} | \mathcal{G}_{n+1}) = Y_{n+1} - \mathbb{E}(X_{n+1} | \mathcal{G}_{n+1}) = Y_{n+1} - \frac{1}{n+1} Y_{n+1} = \frac{n}{n+1} Y_{n+1} \quad (17.6.15)$$

Dividing by n gives $\mathbb{E}(M_n | \mathcal{G}_{n+1}) = M_{n+1}$ and hence \mathbf{M} is a backwards martingale with respect to \mathcal{G} . From the backwards martingale convergence theorem, there exists M_∞ such that $M_n \rightarrow M_\infty$ as $n \rightarrow \infty$ with probability 1 and in mean. Next, for $n, k \in \mathbb{N}_+$ simple algebra gives

$$M_{n+k} = \frac{1}{n+k} \sum_{i=1}^k X_i + \frac{n}{n+k} \frac{1}{n} \sum_{i=k+1}^{k+n} X_i \quad (17.6.16)$$

Letting $n \rightarrow \infty$ then shows that

$$M_\infty = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=k+1}^{k+n} X_i \quad (17.6.17)$$

for every $k \in \mathbb{N}_+$. Hence M_∞ is a tail random variable for the IID sequence \mathbf{X} . From the Kolmogorov 0-1 law, M_∞ must be a constant. Finally, convergence in mean implies that the means converge, and since $\mathbb{E}(M_n) = \mu$ for each n , it follows that $M_\infty = \mu$.

Exchangeable Variables

We start with a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and another measurable space (S, \mathcal{S}) . Suppose that $\mathbf{X} = (X_1, X_2, \dots)$ is a sequence of random variables each taking values in S . Recall that \mathbf{X} is *exchangeable* if for every $n \in \mathbb{N}$, every permutation of (X_1, X_2, \dots, X_n) has the same distribution on (S^n, \mathcal{S}^n) (where \mathcal{S}^n is the n -fold product σ -algebra). Clearly if \mathbf{X} is a sequence of independent, identically distributed variables, then \mathbf{X} is exchangeable. Conversely, if \mathbf{X} is exchangeable then the variables are identically distributed (by definition), but are not necessarily independent. The most famous example of a sequence that is exchangeable but not independent is Pólya's urn process, named for George Pólya. On the other hand, *conditionally* independent and identically distributed sequences *are* exchangeable. Thus suppose that (T, \mathcal{T}) is another measurable space and that Θ is a random variable taking values in T .

If \mathbf{X} is conditionally independent and identically distributed given Θ , then \mathbf{X} is exchangeable.

Proof

Implicit in the statement is that the variables in the sequence have a regular conditional distribution μ_Θ given Θ . Then for every $n \in \mathbb{N}_+$, the conditional distribution of every permutation of (X_1, X_2, \dots, X_n) , given Θ , is μ_Θ^n on (S^n, \mathcal{S}^n) , where μ_Θ^n is the n -fold product measure. Unconditionally, the distribution of any permutation is $B \mapsto \mathbb{E}[\mu_\Theta^n(B)]$ for $B \in \mathcal{S}^n$.

Often the setting of this theorem arises when we start with a sequence of independent, identically distributed random variables that are governed by a parametric distribution, and then randomize one of the parameters. In a sense, we can always think of the setting in this way: Imagine that $\theta \in T$ is a parameter for a distribution on S . A special case is the beta-Bernoulli process, in which the success parameter p in sequence of Bernoulli trials is randomized with the beta distribution. On the other hand, Pólya's urn process is an example of an exchangeable sequence that does not at first seem to have anything to do with randomizing parameters. But in fact, we know that Pólya's urn process is a special case of the beta-Bernoulli process. This connection gives a hint of *de Finetti's theorem*, named for Bruno de Finetti, which we consider next. This theorem states any exchangeable sequence of indicator random variables corresponds to randomizing the success parameter in a sequence of Bernoulli trials.

de Finetti's Theorem. Suppose that $\mathbf{X} = (X_1, X_2, \dots)$ is an exchangeable sequence of random variables, each taking values in $\{0, 1\}$. Then there exists a random variable P with values in $[0, 1]$, such that given $P = p \in [0, 1]$, \mathbf{X} is a sequence of Bernoulli trials with success parameter p .

Proof

As usual, we need some notation. First recall the falling power notation $r^{(j)} = r(r-1) \cdots (r-j+1)$ for $r \in \mathbb{R}$ and $j \in \mathbb{N}$. Next for $n \in \mathbb{N}_+$ and $k \in \{0, 1, \dots, n\}$, let

$$B_k^n = \left\{ (x_1, x_2, \dots, x_n) \in \{0, 1\}^n : \sum_{i=1}^n x_i = k \right\} \quad (17.6.18)$$

That is, B_k^n is the set of bit strings of length n with 1 occurring exactly k times. Of course, $\#(B_k^n) = \binom{n}{k} = n^{(k)}/k!$.

Suppose now that $\mathbf{X} = (X_1, X_2, \dots)$ is an exchangeable sequence of variables with values in $\{0, 1\}$. For $n \in \mathbb{N}_+$ let $Y_n = \sum_{i=1}^n X_i$ and $M_n = Y_n/n$. So $\mathbf{Y} = \{Y_n : n \in \mathbb{N}_+\}$ is the partial sum process associated with \mathbf{X} and $\mathbf{M} = \{M_n : n \in \mathbb{N}_+\}$ the sequence of sample means. Let $\mathcal{G}_n = \sigma\{Y_n, Y_{n+1}, \dots\}$ and $\mathcal{G}_\infty = \bigcap_{n=0}^\infty \mathcal{G}_n$. The family of σ -algebras $\mathcal{G} = \{\mathcal{G}_n : n \in \mathbb{N}_+\}$ is decreasing. The key to the proof is to find two backwards martingales and use the [backwards martingale convergence theorem](#).

Let $m \in \mathbb{N}_+$ and $k \in \{0, 1, \dots, m\}$. The crucial insight is that by exchangeability, given $Y_m = k$, the random vector (X_1, X_2, \dots, X_m) is uniformly distributed on B_k^m . So if $n \in \mathbb{N}_+$ and $n \leq m$, the random vector (X_1, X_2, \dots, X_n) , again given $Y_m = k$, fits the hypergeometric model: a sample of size n chosen at random and without replacement from a population of m objects of which k are type 1 and $m-k$ are type 0. Thus, if $j \in \{0, 1, \dots, n\}$ and $(x_1, x_2, \dots, x_n) \in B_j^n$ then

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid Y_m = k) = \frac{k^{(j)}(m-k)^{(n-j)}}{m^{(n)}} \quad (17.6.19)$$

Equivalently,

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid Y_m) = \frac{Y_m^{(j)}(m - Y_m)^{(n-j)}}{m^{(n)}} \quad (17.6.20)$$

Given Y_m , the variables $(Y_{m+1}, Y_{m+2}, \dots)$ give no additional information about the distribution of (X_1, X_2, \dots, X_n) and hence

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid \mathcal{G}_m) = \mathbb{E}[\mathbf{1}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \mid \mathcal{G}_m] = \frac{Y_m^{(j)}(m - Y_m)^{(n-j)}}{m^{(n)}} \quad (17.6.21)$$

For fixed n, j , and $(x_1, x_2, \dots, x_n) \in B_j^n$, the conditional expected value in the middle of the displayed equation, as a function of m , is a [Doob backward martingale](#) with respect to \mathcal{G} and hence converges to $\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid \mathcal{G}_\infty)$ as $m \rightarrow \infty$.

Next we show that \mathbf{M} is a backwards martingale with respect to \mathcal{G} . Trivially M_n is measurable with respect to \mathcal{G}_n and $\mathbb{E}(M_n) \leq 1$ for each $n \in \mathbb{N}$. Thus we need to show that $\mathbb{E}(M_n \mid \mathcal{G}_m) = M_m$ for $m, n \in \mathbb{N}_+$ with $n \leq m$. From our previous work with (X_1, X_2, \dots, X_n) we know that the conditional distribution of Y_n given $Y_m = k$ is hypergeometric with parameters m, k , and n :

$$\mathbb{P}(Y_n = j \mid Y_m = k) = \binom{n}{j} \frac{k^{(j)}(m-k)^{(n-j)}}{m^{(n)}}, \quad j \in \{0, 1, \dots, n\} \quad (17.6.22)$$

Recall that the mean of the hypergeometric distribution is the sample size times the proportion of type 1 objects in the population. Thus,

$$\mathbb{E}(M_n \mid Y_m = k) = \frac{1}{n} \mathbb{E}(Y_n \mid Y_m = k) = \frac{1}{n} n \frac{k}{m} = \frac{k}{m} \quad (17.6.23)$$

Or equivalently, $\mathbb{E}(M_n \mid Y_m) = Y_m/m = M_m$. Once again, given Y_m , the variables Y_{m+1}, Y_{m+2} give no additional information and so $\mathbb{E}(Y_n \mid \mathcal{G}_m) = Y_m$. Hence \mathbf{M} is a backwards martingale with respect to \mathcal{G} . From the [backwards martingale convergence theorem](#), there exists a random variable P such that $M_n \rightarrow P$ as $n \rightarrow \infty$ with probability 1.

It just remains to connect the dots. Suppose now that $n \in \mathbb{N}_+$ and $j \in \{0, 1, \dots, n\}$ and that $m \in \mathbb{N}_+$ and $k_m \in \{0, 1, \dots, m\}$. From simple calculus, if n and j are fixed and $k_m/m \rightarrow p \in [0, 1]$ as $m \rightarrow \infty$ then

$$\frac{k_m^{(j)}(m - k_m)^{(n-j)}}{m^{(n)}} \rightarrow p^j(1-p)^{n-j} \text{ as } m \rightarrow \infty \quad (17.6.24)$$

(You may recall that this computation is used in the proof of the convergence of the hypergeometric distribution to the binomial.) Returning to the joint distribution, recall that if $(x_1, x_2, \dots, x_n) \in B_j^n$ then

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid \mathcal{G}_m) = \frac{Y_m^{(j)}(m - Y_m)^{(n-j)}}{m^{(n)}} \quad (17.6.25)$$

Let $m \rightarrow \infty$. Since $Y_m/m \rightarrow P$ as $m \rightarrow \infty$ we get

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid \mathcal{G}_\infty) = P^j(1-P)^{n-j} \quad (17.6.26)$$

Random variable P is measurable with respect to \mathcal{G}_∞ so

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid P) = P^j(1-P)^{n-j} \text{ as } m \rightarrow \infty \quad (17.6.27)$$

Given $P = p \in [0, 1]$, \mathbf{X} is a sequence of Bernoulli trials with success parameter p .

De Finetti's theorem has been extended to much more general sequences of exchangeable variables. Basically, if $\mathbf{X} = (X_1, X_2, \dots)$ is an exchangeable sequence of random variables, each taking values in a significantly nice measurable space (S, \mathcal{S}) then there exists a random variable Θ such that \mathbf{X} is independent and identically distributed given Θ . In the proof, the result that $M_n \rightarrow P$ as $n \rightarrow \infty$ with probability 1, where $M_n = \frac{1}{n} \sum_{i=1}^n X_i$, is known as *de Finetti's strong law of large numbers*. De Finetti's theorem, and its generalizations are important in Bayesian statistical inference. For an exchangeable sequence of random variables (our observations in a statistical experiment), there is a hidden, random parameter Θ . Given $\Theta = \theta$, the variables are independent and identically distributed. We gain information about Θ by imposing a *prior distribution* on Θ and then updating this, based on our observations and using Baye's theorem, to a *posterior distribution*.

Stated more in terms of distributions, de Finetti's theorem states that the distribution of n distinct variables in the exchangeable sequence is a *mixture* of product measures. That is, if μ_θ is the distribution of a generic X on (S, \mathcal{S}) given $\Theta = \theta$, and ν is the distribution of Θ on (T, \mathcal{T}) , then the distribution of n of the variables on (S^n, \mathcal{S}^n) is

$$B \mapsto \int_T \mu_\theta^n(B) d\nu(\theta) \quad (17.6.28)$$

This page titled [17.6: Backwards Martingales](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.