Martingale convergence theorems

Janko Gravner

MAT 235B

February 2, 2024

Martingale convergence theorem

Theorem

Assume that X_n is a submartingale with $\sup_n EX_n^+ < \infty$. Then $X_n \to X$ a.s. for some r.v. X with $E|X| < \infty$.

Corollary

If $X_n \geq 0$ is a supermartingale, then $X_n \to X$ a.s. for some r.v. X with $EX \leq EX_0$.

Proof.

Observe that $-X_n$ is a submartingale with $(-X_n)_+ = 0$, so the theorem applies to $-X_n$. As $EX_0 \geq EX_n$, Fatou's lemma implies that

$$EX = E \liminf X_n \leq \liminf EX_n \leq EX_0.$$

□

Martingale convergence theorem

Example. In general, a martingale may go to ∞ a.s.

Let $S_n = \xi_1 + \ldots + \xi_n$, where ξ_n are independent and

$$P(\xi_n = 1) = 1 - 2^{-n}, P(\xi_n = -a_n) = 2^{-n},$$

where a_n are chosen so that $E\xi_n = 0$, namely, $a_n = 2^n - 1$.

Consider the event

$$\{\xi_n = 1 \text{ ev.}\} = \{\xi_n \neq 1 \text{ i.o.}\}^c$$

that $\xi_n = 1$ except for finitely many n. By Borel-Cantelli, $P(\xi_n = 1 \text{ ev.}) = 1$. On the event $\{\xi_n = 1 \text{ ev.}\}$, $S_n \to \infty$.

Any asymptotic behavior can be arranged this way.

Martingale convergence theorem

Example. Under the assumptions of the theorem, it is not necessary that $EX_n \to EX$.

Consider the product martingale $X_n = \xi_1 \cdots \xi_n$, where ξ_n are i.i.d., with $P(\xi_n = 0) = P(\xi_n = 2) = 1/2$.

Then $X_n \geq 0$, $EX_n = 1$, and $X_n \to 0$ a.s. (as $P(X_n = 0 \text{ ev.}) = 1$).

Martingale convergence theorem

We now define *upcrossings* and *downcrossings* of an interval $[a, b]$.
Define

$$
\begin{aligned}
N_0 &= 0 \\
N_1 &= \inf\{m \geq N_0 : X_m \leq a\} \\
N_2 &= \inf\{m \geq N_1 : X_m \geq b\} \\
N_3 &= \inf\{m \geq N_2 : X_m \leq a\} \\
&\dots
\end{aligned}
$$

Martingale convergence theorem

These are all stopping times, e.g.,

$$\{N_{2k} = n\}$$
$$= \bigcup_{0=n_0 \le n_1 < \cdots < n_{2k-1} < n} \Big[\{N_0 = n_0, N_1 = n_1, \ldots, N_{2k-1} = n_{2k-1}\}$$
$$\cap \bigcap_{m=n_{2k-1}+1}^{n-1} \{X_m < b\} \cap \{X_n \ge b\} \Big] \in \mathcal{F}_n,$$

by induction.

Martingale convergence theorem

Note also that the indicator of upcrossings

$$H_m = \begin{cases} 1 & \text{if } N_{2k-1} < m \leq N_{2k} \text{ for some } k = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

is predictable, as

$$\begin{aligned} \{H_m = 1\} &= \bigcup_{k=1}^{\infty} \{N_{2k-1} < m \leq N_{2k}\} \\ &= \bigcup_{k=1}^{\infty} \left(\{N_{2k-1} \leq m-1\} \cap \{N_{2k} \leq m-1\}^c\right) \in \mathcal{F}_{m-1}. \end{aligned}$$

We also introduce the number of completed upcrossings by time n:

$$U_n = \sup\{k : N_{2k} \leq n\}.$$

Martingale convergence theorem

Lemma

If X_n is a submartingale, then

$$(b-a)EU_n \leq E(X_n - a)_+ - E(X_0 - a)_+$$

Martingale convergence theorem

Proof.

The sequence $Y_n = (X_n - a)_+$ is a submartingale and upcrosses $[0, b-a]$ the same number of times as X_n upcrosses $[a, b]$. We have

$$(b-a)U_n \le \sum_{m=1}^{n} H_m(Y_m - Y_{m-1}) = (H \bullet Y)_n,$$

as we have a contribution at least $b-a$ for every completed upcrossing and a nonnegative contribution at the end.

Let $K_m = 1 - H_m$. This is predictable, so $K \bullet Y$ is a submartingale and

$$E(K \bullet Y)_n \ge E(K \bullet Y)_0 = 0.$$

As $Y_n - Y_0 = (H \bullet Y)_n + (K \bullet Y)_n$, $E(H \bullet Y)_n \le E(Y_n - Y_0)$. □

Martingale convergence theorem

Proof of MCT.

Easy to show: $(x-a)_+ \leq x_+ + |a|$. So,

$$EU_n \leq \frac{EX_n^+ + |a|}{b-a}$$

Now, $U_n \uparrow U$, the number of upcrossings of the entire sequence. As $\sup_n EX_n^+ < \infty$, MCT implies that $EU < \infty$, which in turn implies $P(U < \infty) = 1$. Thus the event

$$\cup_{a,b\in\mathbb{Q}}\{\liminf X_n \leq a < b \leq \limsup X_n\}$$

has probability 0. So, $P(\liminf X_n = \limsup X_n) = 1$, but we do not yet know that the two are finite.

Martingale convergence theorem

Proof of MCT, continued.

Let $X = \liminf X_n$, so that $X_n \to X$ a.s.

By Fatou, $EX^+ \leq \liminf EX_n^+ < \infty$, so $EX^+ < \infty$.

Also, $EX_n^- = EX_n^+ - EX_n \leq EX_n^+ - EX_0$, and so

$$EX^- \leq \liminf EX_n^- \leq \sup EX_n^+ - EX_0 < \infty$$

and so $E|X| < \infty$. □

Pólya's urn

Example. Pólya's urn. Start with an urn containing 1 red and 1 green ball. At each time, select a ball at random and return it to the urn, together with an additional ball of the same color.
Let G_n be the number of of green balls after n draws and additions (e.g., $G_0 = 1$), and $X_n = G_n/(n+2)$ the fraction of green balls at the same time.
We claim that X_n is a martingale w.r.t. $\mathcal{F}_n = \sigma\{G_1, \ldots, G_n\}$.
First we observe that

$$G_{n+1} = \begin{cases} G_n + 1 & \text{w.p. } \frac{G_n}{n+2} \\ G_n & \text{w.p. } 1 - \frac{G_n}{n+2} \end{cases}$$

Pólya's urn

Then,

$$\begin{aligned} E[X_{n+1} \mid \mathcal{F}_n] &= \frac{1}{n+3} E[G_{n+1} \mid \mathcal{F}_n] \\ &= \frac{1}{n+3}\left[(G_n+1)\frac{G_n}{n+2} + G_n\left(1-\frac{G_n}{n+2}\right)\right] \\ &= \frac{1}{n+3}\left[G_n \frac{n+3}{n+2}\right] = X_n \end{aligned}$$

As $X_n \geq 0$, martingale CT implies that $X_n \to X_\infty$ for some r.v. X_∞ (of course with values in $[0,1]$). What is the distribution of X_∞?

Pólya's urn

We have

P(draw r red balls on first r draws, g green balls on next g draws)

$$= \frac{1}{2} \cdot \frac{2}{3} \cdots \frac{r}{r+1} \cdot \frac{1}{r+2} \cdot \frac{2}{r+3} \cdots \frac{g}{g+r+1}$$

and the probability of any other configuration of r red and g green balls on the first $r+g$ draws is the same.
So, G_n is uniform: for $k = 0, \ldots, n$,

$$P(G_n = k+1) = \binom{n}{k} \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1},$$

and so, for $x \in (0,1)$,

$$P(X_n \leq x) = P(G_n \leq (n+2)x) = \frac{\lfloor (n+2)x \rfloor}{n+1} \to x$$

so that X_∞ is uniform on $[0,1]$.

Branching processes

Example. Branching process with offspring distribution ξ, with $p_k = P(\xi = k)$ and $\mu = E\xi < \infty$. If Z_n is the population size at generation n, then $M_n = Z_n/\mu^n$ is a nonnegative martingale, which therefore converges a.s.

In the *subcritical* case $\mu < 1$, then $P(Z_n = 0 \text{ for some } n) = 1$, because

$$P(Z_n \geq 1) \leq EZ_n = \mu^n$$

and so $P(Z_n \geq 1 \text{ i.o.}) = 0$.

Same holds in the *critical* case $\mu = 1$, when $P(\xi = 1) < 1$:

By martingale CT, $Z_n \to Z_\infty$ a.s., but Z_n is integer valued, so $Z_n = Z_\infty$ for some $n \geq n_0(\omega)$. For all natural numbers $k \geq 1$, $N \geq 0$, $P(Z_n = k \text{ for all } n \geq N) = 0$, and so for all $k \geq 1$,

$$P(Z_\infty = k) = P\left(\cup_{N=0}^{\infty}\{Z_n = k \text{ for all } n \geq N\}\right) = 0.$$

Branching processes

We now deal with the *supercritical* case $\mu > 1$.

Theorem

If $\mu > 1$, then $P(Z_n \geq 1$ for all $n) > 0$.

Proof.

Let $\theta_m = P(Z_m = 0) \uparrow P(Z_n = 0 \text{ for some n})$. By conditioning on Z_1,

$$\theta_m = \sum_{k=0}^{\infty} p_k \theta_{m-1}^k.$$

For $s \in [0, 1]$, let $\varphi(s) = \sum_{k=0}^{\infty} p_k s^k = E(s^{\xi})$, (a version of) the *moment generating function*.

For $s \in [0, 1)$, $\varphi'(s) = \sum_{k=0}^{\infty} k p_k s^{k-1} \to \mu$ as $s \to 1-$.

As $\varphi'' \geq 0$ on $[0, 1)$, φ is convex.

Proof, continued.

We now claim that there exists a unique $\rho < 1$ so that $\varphi(\rho) = \rho$.

For small $\epsilon > 0$: $\varphi(1-\epsilon) = \varphi(1) - \mu\epsilon + o(\epsilon) < 1 - \epsilon$, so $\varphi(x) < x$ when x is close to 1 and $\varphi(0) = p_0 \geq 0$. So there is a $\rho \in [0, 1)$ such that $\phi(\rho) = \rho$, and it must be unique as $\varphi'' > 0$ on $(0, 1)$ (otherwise, there are 3 points in $[0, 1]$ at which $\varphi(x) - x$ vanishes, so $(\varphi(x) - x)'' = \varphi''(x) = 0$ for some $x \in (0, 1)$).

Note that $x \leq \varphi(x) \leq \phi(\rho) = \rho$ for $x \in [0, \rho]$.

We now claim that $\theta_m \uparrow \rho$ as $m \to \infty$.

We have $\theta_0 = 0$, $\theta_m = \varphi(\theta_{m-1})$. Then $\theta_m \leq \rho$ and so $\theta_m \uparrow \theta_\infty \leq \rho$. By continuity, $\theta_\infty = \varphi(\theta_\infty)$. By uniqueness, $\theta_\infty = \rho$. □

For example, if $p_k = (1-p)^k p$, so that $P(\xi \geq k) = (1-p)^k$, $k \geq 0$, then $\mu = \sum_{k\geq 1} P(\xi \geq k) = (1-p)/p$, so $\mu > 1$ when $p < 1/2$. We have

$$\varphi(s) = \sum_{k=0}^{\infty} p(1-p)^k s^k = \frac{p}{1-(1-p)s},$$

and $\varphi(s) = s$ when $(1-p)s^2 - s + p = 0$, so

$$\rho = \frac{p}{1-p} = \frac{1}{\mu}$$

in *this* case.

Doob's inequality

Theorem (Doob's inequality)

Assume that X_n is a submartingale and $A = \{\max_{0 \le m \le n} X_m^+ \ge \lambda\}$. Then for $\lambda > 0$,

$$\lambda P(A) \le E[X_n 1_A] \le EX_n^+.$$

The idea is that we can "control the maximum of a finite submartingale sequence by its endpoint."

Doob's inequality

Corollary (Kolmogorov's inequality)

Let $S_n = \xi_1 + \cdots + \xi_n$, ξ_i *independent,* $E\xi_i = 0$, $E\xi_i^2 < \infty$. *Then, for* $x > 0$,

$$P(\max_{1 \le m \le n} |S_m| \ge x) \le \frac{1}{x^2} \mathrm{Var}(S_n).$$

Proof of Kolmogorov's inequality.

Use Doob's inequality for the submartingale $X_n = S_n^2$ and $\lambda = x^2$. □

Note the similarity to Chebyshev's inequality, but this one has $\max$ in it!

Doob's inequality

Lemma

Assume that X_n is a submartingale, $a \in \mathbb{N}$, and N a stopping time so that $P(N \leq a) = 1$. Then $EX_N \leq EX_a$.

Proof.

As $K_n = 1_{\{N \leq n-1\}}$ is predictable,

$$(K \bullet X)_n = \sum_{m=1}^{n} K_m(X_m - X_{m-1}) = X_n - X_{n \wedge N}$$

is a submartingale. So, $E(X_n - X_{n \wedge N}) \geq 0$. Take $n = a$. □

Proof of Doob's inequality.

The stopping time

$$N = \inf\{m : X_m \geq \lambda\} \wedge n$$

is bounded by n. On A, $X_N \geq \lambda$, that is, $X_N 1_A \geq \lambda 1_A$. So, $\lambda P(A) \leq E[X_N 1_A]$.
Also, $X_N 1_{A^c} = X_n 1_{A^c}$ and, by the lemma, $EX_N \leq EX_n$. So,

$$\begin{aligned}\lambda P(A) &\leq E[X_N 1_A] \\ &= EX_N - E[X_N 1_{A^c}] \leq EX_n - E[X_n 1_{A^c}] = E[X_n 1_A].\end{aligned}$$

Finally, $X_n 1_A \leq X_n^+$. □

L^p-maximum inequality

Theorem (L^p-maximum inequality)

Assume $p > 1$. *Assume* X_n *is a submartingale. Let* $\overline{X}_n = \max_{0 \le m \le n} X_m^+$. *Then*

$$E\overline{X}_n^p \le \left(\frac{p}{p-1}\right)^p E(X_n^+)^p.$$

Corollary

If X_n *is a martingale, then*

$$E\left[\max_{0 \le m \le n} X_m^2\right] \le 4EX_n^2$$

Proof.

Apply the L^p-maximum inequality, with $p = 2$, to the submartingale $|X_n|$. □

L^p-maximum inequality

Example. There is no L^1-maximum inequality: there is no constant C such that, for any nonnegative martingale X_n,

$$E\left(\sup_{0\le m\le n} X_m\right) \le C\, EX_n = C\, EX_0,$$

which would imply, by MCT, $E\left(\sup_{n\ge 0} X_n\right) \le C\, EX_0$.

Indeed, we know that such a martingale converges a.s. to some r.v. X, and then the above would (by DCT) imply that $EX_n \to EX$. But we already know that this is not necessarily true.

L^p-maximum inequality

Proof of L^p-maximum inequality.

Recall Hölder:

$$E|XY| \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}$$

if $p > 1$ and $q = p/(p-1)$. Also note that Doob's inequality implies, for $\lambda > 0$,

$$P(\overline{X}_n \geq \lambda) \leq \frac{1}{\lambda} E[X_n^+ 1_{\{\overline{X}_n \geq \lambda\}}].$$

Fix $M > 0$. Then also

$$P(\overline{X}_n \wedge M \geq \lambda) \leq \frac{1}{\lambda} E[X_n^+ 1_{\{\overline{X}_n \wedge M \geq \lambda\}}],$$

as both sides are 0 if $M < \lambda$.

L^p-maximum inequality

Proof of L^p-maximum inequality, continued.

$$\begin{aligned}
E(\overline{X}_n \wedge M)^p &= \int_0^\infty p\lambda^{p-1} P(\overline{X}_n \wedge M \geq \lambda)\, d\lambda \\
&\leq \int_0^\infty p\lambda^{p-1} \frac{1}{\lambda} E[X_n^+ 1_{\{\overline{X}_n \wedge M \geq \lambda\}}]\, d\lambda \\
&= \int_0^\infty p\lambda^{p-2}\, d\lambda \int_\Omega X_n^+ 1_{\{\overline{X}_n \wedge M \geq \lambda\}}\, dP \\
&= \int_\Omega X_n^+\, dP \int_0^{\overline{X}_n \wedge M} p\lambda^{p-2}\, d\lambda \\
&= \frac{p}{p-1} E[X_n^+ (\overline{X}_n \wedge M)^{p-1}] \\
&\leq \frac{p}{p-1} \left(E(X_n^+)^p\right)^{1/p} \left(E(\overline{X}_n \wedge M)^p\right)^{(p-1)/p}
\end{aligned}$$

L^p-maximum inequality

Proof of L^p-maximum inequality, continued.

Solving the inequality,

$$E(\overline{X}_n \wedge M)^p \leq \left(\frac{p}{p-1}\right)^p E(X_n^+)^p$$

Now send $M \to \infty$ and use MCT. □

Question: why did we need M in the proof?

L^p-maximum inequality

Proof of L^p-maximum inequality, continued.

Solving the inequality,

$$E(\overline{X}_n \wedge M)^p \leq \left(\frac{p}{p-1}\right)^p E(X_n^+)^p$$

Now send $M \to \infty$ and use MCT. □

Question: why did we need M in the proof?

We do not know, a priori, that $E\overline{X}_n^p$ is finite. Neither do we need to assume that the RHS in the L^p-maximum inequality is finite as the theorem trivially hold when it is infinite.

L^p-convergence of martingales

Recall that $X_n \to X$ in L^p if $E|X_n - X|^p \to 0$.

Theorem (Martingale L^p-convergence theorem)

Assume $p > 1$. Assume X_n is a martingale or a positive submartingale, with $\sup_n E|X_n|^p < \infty$. Then there exists a r.v. $X \in L^p$ so that $X_n \to X$ a.s. and in L^p.

Example. This is not true for general submartingales or for positive supermartingales. Let $\xi_i \geq 0$ be i.i.d. with $E\xi_i < 1$ and $E\xi_i^2 = 1$. (E.g., $P(\xi_0 = 0) = 3/4$ and $P(\xi_0 = 2) = 1/4$.)

Then $X_n = \xi_1 \cdots \xi_n$ is a supermartingale, $EX_n \to 0$, so $X_n \to 0$ in L^1. If X_n converges in L^2, it can only converge to 0. But $EX_n^2 = 1$ so $X_n \not\to 0$ in L^2.

L^p-convergence of martingales

Proof of martingale L^p-convergence theorem.

As $EX_n^+ \le E|X_n| \le (E|X_n|^p)^{1/p}$, a.s. convergence to some r.v. X follows from martingale CT. As $|X_n|$ is a submartingale, by L^p-maximum inequality,

$$E\left[\sup_{0\le m\le n} |X_m|^p\right] \le \left(\frac{p}{p-1}\right)^p E|X_n|^p \le \left(\frac{p}{p-1}\right)^p \sup_n E|X_n|^p$$

Then, by MCT, $\sup_{0\le m} |X_m| \in L^p$.
So, $X \in L^p$, and then by DCT $E|X_n - X|^p \to 0$. □

Supercritical branching processes

Example. Consider the supercritical ($E\xi > 1$) branching process with offspring distribution ξ, such that $\mu = E\xi > 1$ and $\sigma^2 = \mathrm{Var}(\xi) \in (0, \infty)$. We continue studying the martingale $M_n = Z_n/\mu^n$, which converges a.s. to a r.v. M.
We have

$$\begin{aligned} E[Z_n^2 \mid \mathcal{F}_{n-1}] &= \sum_{k=1}^{\infty} E[Z_n^2 1_{\{Z_{n-1}=k\}} \mid \mathcal{F}_{n-1}] \\ &= \sum_{k=1}^{\infty} 1_{\{Z_{n-1}=k\}} E[(\xi_1 + \cdots + \xi_k)^2] \\ &= \sum_{k=1}^{\infty} 1_{\{Z_{n-1}=k\}} (k\sigma^2 + k^2\mu^2) \\ &= \sigma^2 Z_{n-1} + \mu^2 Z_{n-1}^2 \end{aligned}$$

So, $EM_n^2 = \dfrac{\sigma^2}{\mu^{n+1}} + EM_{n-1}^2$.

It follows that

$$EM_n^2 = 1 + \sigma^2 \sum_{k=1}^{n} \mu^{-k-1}$$

and so $\sup_n EM_n^2 < \infty$. By the L^p-convergence theorem, $M_n \to M$ in L^2 and so $EM_n \to EM = 1$ and $EM_n^2 \to EM^2 = 1 + \frac{\sigma^2}{\mu(\mu-1)}$. In particular, $\theta = P(M = 0) < 1$. By conditioning on the first generation,

$$\theta = \sum_{k=0}^{\infty} p_k \theta^k,$$

so $\theta = \varphi(\theta)$, and, as $\theta < 1$, $\theta = \rho = P(Z_n = 0 \text{ for some } n)$.

Supercritical branching processes

As

$$\{Z_n = 0 \text{ for some } n\} \subset \{M = 0\},$$

the set-difference of these two sets has measure 0, and

$$P(\text{either: } Z_n = 0 \text{ for some } n; \text{ or } M > 0) = 1.$$

A supercritical branching process with $E\xi^2 < \infty$ either dies out or its population increases exponentially as μ^n (up to a random factor). This was proved by Kolmogorov in the 1930s.

A 1966 theorem of Kesten and Stigum says that the necessary and sufficient condition for the above to be true is that $E(\xi(\log \xi)_+) = \sum_{k \geq 2} k \log k \, p_k < \infty$, weaker than $E\xi^2 < \infty$.
A 1968 theorem by Seneta shows that if $\mu \in (1, \infty)$, one can normalize Z_n by some (deterministic) sequence C_n and get a nontrivial limit.

The classic on this subject is the book by Athreya and Ney.