Donsker's Theorem

Statistics 710

October 19, 2006

Prologue: The Arc Sine Law. Let $X_1, X_2, \cdots$ be i.i.d. random variables taking the values ± 1 with probability one-half each; let $S_n = X_1 + \cdots + X_n$ and

$$N_n = \#\{k \le n : S_k > 0 \text{ or } S_k = 0, S_{k-1} > 0\}.$$

Then

$$\lim_{n\to\infty} P[N_n \le x] = \frac{2}{\pi}\arcsin(\sqrt{x}) \tag{1}$$

for $0 < x < 1$. In the proof it is first shown by purely combinatorial arguments, that

$$P[N_{2n} = 2k] = \binom{2k}{k}\binom{2n-2k}{n-k}2^{-2n}$$

for $k = 1, \cdots, n$. Relation (1) then follows from Stirlings formula. See [3], Chapter 3, for the details. Observe that the limiting distribution is $\beta(\frac{1}{2}, \frac{1}{2})$ with a U-shaped density

$$\frac{1}{\pi\sqrt{x(1-x)}}.$$

Question: Is this a general phenomenon, or special to the case $X = \pm 1$ with probability 1/2 each? Certainly, the proof is specific to the special case.

Brownian Motion. Let $\mathbb{B}_t,\ 0 \le t < \infty$ denote a standard Brownian motion. Thus, $\mathbb{B}_t$ is a stochastic process with continuous samples paths and stationary independent increments, and $\mathbb{B}_t = \text{Normal}[0, t]$ for all $0 < t < \infty$. It is easy to check that if $\mathbb{B}$ is a standard Brownian motion and $0 < c, s < \infty$ then the processes (defined by)

$$\frac{1}{\sqrt{c}}\mathbb{B}_{ct} \qquad \text{and} \qquad \mathbb{B}_{s+t} - \mathbb{B}_s$$

are again standard Brownian motions. The *Strong Markov Property* asserts that the second process is a standard Brownian even when s is replaced by a stopping time.

Brownian paths are continuous, but not smooth. They satisfy a Holder condition of the form

$$\sup_{0\le s<t\le 1} \frac{|\mathbb{B}_t - \mathbb{B}_s|}{\sqrt{|s-t|\log[\frac{1}{|s-t|}]}} < \infty \ w.p.1, \tag{2}$$

but are not differentiable.

Problem 1 Show that if $f : [0,1] \to \mathbb{R}$ is absolutely continuous, then

$$\lim_{n\to\infty} \sum_{k=1}^{2^n} \left[f(\frac{k}{2^n}) - f(\frac{k-1}{2^n}) \right]^2 = 0.$$

Then show that if $\mathbb{B}$ is standard Brownian motion, then

$$\lim_{n\to\infty} \sum_{k=1}^{2^n} \left[\mathbb{B}(\frac{k}{2^n}) - \mathbb{B}(\frac{k-1}{2^n}) \right]^2 = 1 \ w.p.1,$$

where $\mathbb{B}(t)$ is written for $\mathbb{B}_t$ when t is complicated.

The Distribution of $\mathbb{B}$. Let $C[0,1]$ be the space of continuous functions on $[0,1]$ endowed with the metric $d(f,g) = \sup_{0\le t\le 1} |f(t) - g(t)|$ for $f, g \in C[0,1]$; and let $\mathcal{B}$ be the class of Borel subsets of $C[0,1]$, the smallest sigma-algebra containing all open sets. If $\mathbb{B}$ is a standard Brownian motion, defined on a probability space $(\Omega, \mathcal{A}, P)$, then it can be shown that the function $\tilde{\mathbb{B}}$ which maps $\omega \in \Omega$ into the function (defined by) $\mathbb{B}_t(\omega),\ 0 \le t \le 1$, is a measurable mapping. For example, if $B = \{g : d(f,g) \le c\}$ is a closed ball in $C[0,1]$, then

$$\{\omega : \tilde{\mathbb{B}}(\omega) \in B\} = \{\omega : \sup_{0\le t\le 1} |\mathbb{B}_t(\omega) - f(t)| \le c\} = \{\omega : \sup_{0\le r\le 1} |\mathbb{B}_r(\omega) - f(r)| \le c\},$$

where r is restricted to rationals, and the right side is measurable, since each $\mathbb{B}_t$ is. Thus,

$$\Psi\{B\} = \{\omega : \tilde{\mathbb{B}}(\omega) \in B\} \tag{3}$$

defines a probability measure on $\mathcal{B}$, and Ψ is called the distribution of $\mathbb{B}$.

Broken Lines. Now let $X_1, X_2, \cdots \sim^{\text{ind}} F$ be i.i.d. random variables with mean 0 and variance 1; let $S_n = X_1 + \cdots + X_n$; and let $\mathbb{S}_n$ be a continuous piecewise linear function for which

$$\mathbb{S}_n(\frac{k}{n}) = \frac{S_k}{\sqrt{n}}$$

for $0 \le k \le n$. Then the function $\tilde{\mathbb{S}}_n$ which maps $\omega \in \Omega$ into the function (defined by) $\mathbb{S}_n(\omega, t),\ 0 \le t \le 1$, is also a measurable mapping; and the distribution of $\mathbb{S}_n$ is defined by

$$\Psi_n\{B\} = \{\omega : \mathbb{S}_n(\omega) \in B\} \tag{4}$$

for $B \in \mathcal{B}$.

Weak Convergence. Let μ_n and μ be probability measures defined on the Borel subsets of a metric space and recall that μ_n converges weakly to μ iff

$$\lim_{n\to\infty} \int_{\mathcal{X}} g d\mu_n = \int_{\mathcal{X}} g d\mu \tag{5}$$

for all bounded continuous functions $g : \mathcal{X} \to I\!R$. The Portmantau Theorem provides several equivalent conditions. In particular, (5) holds for all bounded continuous functions, if it holds for all bounded uniformly continuous functions. Recall too the continuous mapping theorem: if μ_n converges weakly to μ and $h : \mathcal{X} \to \mathcal{Y}$, a second metric space, is continuous almost everywhere (μ), then $\mu_n \circ h^{-1}$ converges weakly to $\mu \circ h^{-1}$.

If X_n and X are random elements with distributions μ_n and μ, then X_n converges in distribution to X iff μ_n converges weakly to μ. These relations are denoted by $X_n \Rightarrow X$ and $\mu_n \Rightarrow \mu$. With this notation, Slutsky's Theorem may be stated: *If* $X_n \Rightarrow X$ *and* $d(X_n, Y_n) \to^p 0$, *then* $Y_n \Rightarrow X$. Billingsley [1] is recommended for background on weak convergence.

Donsker's Theorem. $I\!\!S_n \Rightarrow I\!\!B$; *that is,* Ψ_n *converges weakly to* Ψ. As a corollary: if $h : C[0,1] \to I\!R$ is continuous almost everywhere (Ψ), then $h(I\!\!B_n)$ converges in distribution to $h(I\!\!B)$.

Example 1 Let λ denote Lebesgue measure. Then the function h defined on $C[0,1]$ by

$$h(f) = \lambda\{t : f(t) > 0\}$$

is continuous at every f for which $\lambda\{t : f(t) = 0\} = 0$ and, therefore, almost everywhere (Ψ). So, $N_n = h(I\!\!B_n)$ converges in distribution to $h(I\!\!B)$. It follows

$$P[h(I\!\!B) \le x] = \frac{2}{\pi} \arcsin(\sqrt{x})$$

for $0 < x < 1$ and then that relation (1) holds for any F with mean 0 and variance 1.

Some Details. Let $h_c(f) = \lambda\{t : f(t) > c\}$. If $d(f,g) \le \epsilon$, then clearly $\{t : f(t) > \epsilon\} \subseteq \{t : g(t) > 0\} \subseteq \{t : f(t) > -\epsilon\}$ and, therefore, $h_\epsilon(f) \le h(g) \le h_{-\epsilon}(f)$. So, if $d(f, f_n) \to 0$, then $h_\epsilon(f) \le \liminf_{n\to\infty} h(f_n) \le \limsup_{n\to\infty} h(f_n) \le h_{-\epsilon}(f)$ for all $\epsilon > 0$. Letting $\epsilon \downarrow 0$,

$$h(f) \le \liminf_{n\to\infty} h(f_n) \le \limsup_{n\to\infty} h(f_n) \le \lambda\{t : f(t) \ge 0\}$$

from which the continuity follows. It remains to show that $\lambda\{t : I\!\!B_t(\omega) = 0\} = 0$ for almost every ω; and this follows from

$$E\left[\lambda\{t : I\!\!B_t(\omega) = 0\}\right] = \int_0^1 P[I\!\!B_t = 0]\lambda\{dt\} = 0.$$

Skorohod Embedding. If F is any distribution function with mean 0 and finite variance $0 < \sigma^2 < \infty$, then, on a suitably rich probability space, there are a standard Brownian motion and stopping times $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$ for which

$$[\tau_k - \tau_{k-1}, \mathbb{B}(\tau_k) - \mathbb{B}(\tau_{k-1})] \text{ are i.i.d.},$$
$$E[\tau_k - \tau_{k-1}] = \sigma^2,$$
$$\mathbb{B}(\tau_k) - \mathbb{B}(\tau_{k-1}) \sim F$$

It is easy to describe the proof in the special case that F is a two point distribution, assigning masses $b/(a+b)$ and $a/(a+b)$ to $-a$ and b, where $0 < a, b < \infty$. Let

$$\tau_1 = \inf\{t > 0 : \mathbb{B}_t \notin [-a, b]\}.$$

Then $\mathbb{B}(\tau_1) = -a$ or b with probability $b/(a+b)$ or $a/(a+b)$, since $E[\mathbb{B}(\tau_1)] = 0$ (by a continous version of Wald's Lemma). Now, let

$$\tau_2 = \inf\{t > \tau_1 : \mathbb{B}(t + \tau_1) - \mathbb{B}(\tau_1) \notin [-a, b]\},$$

etc. $\cdots$, and use the Strong Markov Property. To get from two point distributions to a general F requires showing that any F with mean 0 can be represented as an average of two point distributions.

Skorohod Embedding and Donsker. Let

$$\mathbb{B}_n(t) = \frac{1}{\sqrt{n}}\mathbb{B}(nt).$$

Then $\mathbb{B}_n$ is again a standard Brownian motion and, so, has distribution Ψ. Next, since Donsker's Theorem refers only to the distribution of $\mathbb{S}_n$, there is no loss of generality is suppose that $S_k = \mathbb{B}(\tau_k)$ for all k. With this assumption, it suffices to show that $d[\mathbb{B}_n, \mathbb{S}_n] \to^p 0$. To see this, first observe that $\tau_n/n \to \sigma^2$ $w.p.1$, by the Strong Law of Large Numbers, and then that

$$\begin{aligned} d[\mathbb{B}_n, \mathbb{S}_n] &\leq \max_{k \leq n} |\mathbb{B}_n(\frac{\tau_k}{n}) - \mathbb{B}_n(\frac{k}{n})| + \max_{|s-t| \leq 1/n} |\mathbb{B}_n(s) - \mathbb{B}_n(t)| \\ &=\leq C \max_{k \leq n} \left[\sqrt{|\frac{\tau_k}{n} - \frac{k}{n}| \log[|\frac{\tau_k}{n} - \frac{k}{n}|]^{-1}} \right] + C\sqrt{\frac{1}{n}\log(n)}, \end{aligned} \qquad (*)$$

by Levy's Inequality, and this approaches 0 in probability, by a simple application of Levy's Theorem.

Rates. It is possible to get rates of convergence in this argument. Suppose for example that $E(X_1^4) < \infty$. Then, (it can be shown that) $E(\tau_1^2) < \infty$, in which case

$$|\tau_n - n| = O[\sqrt{n \log\log(n)}] \ w.p.1,$$

by the Law of the Iterated Logarithm. Then $\max_{k\leq n}|\tau_k - n| = O_p[\sqrt{n\log\log(n)}]$ too, and the last line in (*) is of order

$$\sqrt{\sqrt{n}\log^2(n)} = n^{-\frac{1}{4}}\log(n).$$

Skorohod embedding encounters a barrier at this point. Even if there are higher moments, the argument does not give better rates of convergence.

Strong Approximation.

...

The Least Concave Majorant (or Greatest Convex Minorant). A function f (defined on a convex subset C of a Hilbert space $\mathcal{H}$ is said to be concave if $-f$ is convex; that is,

$$f[\alpha x + (1-\alpha)y] \geq \alpha f(x) + (1-\alpha)f(y)$$

for all $x, y \in C$ and $0 \leq \alpha \leq 1$. It is easily seen that the infimum, $f_* = \inf_{i\in I} f_i$, of any collection $f_i, i \in I$, of concave functions is again concave. For if $i \in I$, $x, y \in C$, and $0 \leq \alpha \leq 1$, then

$$f_i[\alpha x + (1-\alpha)y] \geq \alpha f_i(x) + (1-\alpha)f_i(y) \geq \alpha f_*(x) + (1-\alpha)f_*(y)$$

and, therefore,

$$f_*[\alpha x + (1-\alpha)y] \geq \alpha f_*(x) + (1-\alpha)f_*(y).$$

It follows that if f is any bounded function on C, then there is a unique function $\tilde{f}$ for which $\tilde{f} \geq f$; $\tilde{f}$ is concave; if h is any concave function for which $h \geq f$, then $h \geq \tilde{f}$.

For existence, it suffices to let $\tilde{f}$ be the infimum of all concave functions that majorize f. There is at least one, since f is bounded. Uniqueness is clear. The function $\tilde{f}$ is called the least concave majorant of f and may also be denoted by $L(f)$.

Marshal's Lemma. Now let $\|f\| = \sup_{x\in C}|f(x)|$ for bounded functions f on C. If f, g are bounded functions, then $f(x) \leq g(x) + [f(x) - g(x)] \leq \tilde{g}(x) + \|f - g\|$ for all $x \in C$. Since the right side is concave, $\tilde{f} \leq \tilde{g} + \|f - g\|$. Then reversing the roles of f and g, yields

$$\|\tilde{f} - \tilde{g}\| \leq \|f - g\|. \tag{6}$$

That is, L is a contraction. It follows that *if g is any concave function, then*

$$\|\tilde{f} - g\| \leq \|f - g\|. \tag{7}$$

More on Derivatives of Concave Functions. Suppose now that C is a subinterval of $I\!R$. Let f and g are concave functions and $\epsilon^2 = \|f - g\|$. If $t_0 \in C^o$ and $d(t_0, C') \geq \epsilon$, then

$$\frac{g(t_0+\epsilon)-g(t_0)}{\epsilon} - 2\epsilon \leq f'_r(t_0) \leq f'_\ell(t_0) \leq \frac{g(t_0)-g(t_0-\epsilon)}{\epsilon} + 2\epsilon. \tag{8}$$

To establish the left inequality, simply observe that

$$\begin{aligned} f'_r(t_0) &\geq \frac{f(t_0)-f(t_0-\epsilon)}{\epsilon} \\ &\geq \frac{g(t_0)-g(t_0-\epsilon)}{\epsilon} - \frac{2\|f-g\|}{\epsilon} \geq \frac{g(t_0)-g(t_0-\epsilon)}{\epsilon} - 2\epsilon. \end{aligned}$$

The right side of (8) may be established similarly. As a corollary: if $\|f_n - f\| \to 0$ as $n \to \infty$, $t_0 \in C^o$, and $f'_r(t_0) = f'_\ell(t_0)$, then $f'_{nr}(t_0) \to f'_r(t_0)$.

Back to Broken Lines and Brownian Motion. Now (re)consider the broken line $\mathbb{S}_n$; regard $\mathbb{S}_n$ as a random element taking values in $C[0,1]$; and recall Donsker's Theorem, $\mathbb{S}_n \Rightarrow I\!B$. Next, let $\tilde{\mathbb{S}}_n$ and $\tilde{B}$ denote the least concave majorants of $\mathbb{S}_n$ and $I\!B$. Then $\mathbb{S}_n \Rightarrow I\!B$, by the Continuous Mapping Theorem.

...

The Space D[0,1]. Let $D[0,1]$ be the space of all functions on $[0,1]$ that are right continuous on $[0,1)$ and have (finite) left limits on $(0,1]$. Further, let Λ be the set of all continuous, strictly increasing functions $\lambda : [0,1] \to^{\text{onto}} [0,1]$, and let

$$d(f,g) = \inf_{\lambda \in \Lambda} \sup_{0 \leq t \leq 1} \left[|f(t) - g[\lambda(t)]| + |\lambda(t) - t| \right].$$

Then it is easily verfied that d is a metric. For example, if $f = \mathbf{1}_{[0,\frac{1}{2})}$ and $g = \mathbf{1}_{[0,\frac{1}{2}+\epsilon)}$, where $0 < \epsilon < 1/2$, then $d(f,g) \leq \epsilon$. It can be shown that $D[0,1]$ is a topologically complete metric space.

Problem 2 *Show that addition is not a continuous function with respect to the product topology on* $D[0,1]^2$.

References

[1] Billingsley, Patrick (1968). *Convergence of Probability Measures*. Wiley.

[2] Carolan, Chris and Richard Dykstra (2005). *Characterization of the least concave majorant of Brownian motion, conditional on a vertex point, with applications to construction.*

[3] Feller, William (1968). *An Introduction to Probability Theory and its Applications*, **1**, Wiley.

[4] Groeneboom, Piet (1983). The concave majorant of Brownian motion. *Ann. Prob.*, **11**, 1016-1027.

[5] Komlos, J., P. Major, and G. Tusnaldy (1975). An approximation of partial sums of independent random variables and the sample DF, I and II. *Zeit. Wahr. verw. Geb.*, **32**, 111-131, and **34**, 33-58.