The Glivenko-Cantelli theorem

Jordan Bell

April 12, 2015

1 Narrow topology

Let X be a metrizable space and let $C_b(X)$ be the Banach space of bounded continuous functions $X \to \mathbb{R}$, with the norm $\|f\| = \sup_{x \in X} |f(x)|$. If X is metrizable with the metric d, let $U_d(X)$ be the collection of bounded d-uniformly continuous functions $X \to \mathbb{R}$. This is a vector space and is a closed subset of $C_b(X)$, thus is itself a Banach space.

Let X be a metrizable space and denote by $\mathscr{P}(X)$ the collection of Borel probability measures on X. The **narrow topology on** $\mathscr{P}(X)$ is the coarsest topology on $\mathscr{P}(X)$ such that for every $f \in C_b(X)$, the mapping $\mu \mapsto \int_X f d\mu$ is continuous $\mathscr{P}(X) \to \mathbb{R}$. It can be proved that if X is metrizable with a metric d and D is a dense subset of $U_d(X)$, then the narrow topology is equal to the coarsest topology such that for each $f \in U_d(X)$, the mapping $\mu \mapsto \int_X f d\mu$ is continuous $\mathscr{P}(X) \to \mathbb{R}$.[1]

If X is a separable metrizable space, then it is metrizable by a metric d such that the metric space (X, d) is totally bounded. It is a fact that if (X, d) is a totally bounded metric space, then $U_d(X)$ is separable.[2]

Theorem 1. *If X is a separable metrizable space, then X is metrizable by a metric d for which there is a countable dense subset D of $U_d(X)$ such that μ_n converges narrowly to μ if and only if*

$$\int_X f d\mu_n \to \int_X f d\mu, \qquad f \in D.$$

2 Independent and identically distributed random variables

Let (Ω, S, P) be a probability space and let X be a separable metric space, with the Borel σ-algebra $\mathscr{B}_X$. We say that a finite collection measurable functions

[1] Charalambos D. Aliprantis and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third ed., p. 507, Theorem 15.2.

[2] Daniel W. Stroock, *Probability Theory: An Analytic View*, p. 371, Lemma 9.1.4.

$\xi_i : \Omega \to X$, $1 \leq i \leq n$, is **independent** if

$$P\left(\bigcap_{i=1}^{n} \xi_i^{-1}(A_i)\right) = \prod_{i=1}^{n} P(\xi_i^{-1}(A_i)), \qquad A_1, \ldots, A_n \in \mathscr{B}_X,$$

i.e.

$$P(\xi_1 \in A_1, \ldots, \xi_n \in A_n) = P(\xi_1 \in A_1) \cdots P(\xi_n \in A_n), \qquad A_1, \ldots, A_n \in \mathscr{B}_X.$$

We say that a family of measurable functions is independent if every finite subset of it is independent.

We say that two measurable functions $f, g : \Omega \to X$ are **identically distributed** if the pushforward f_*P of P by f is equal to the pushforward g_*P of P by g, i.e. $P(f^{-1}(A)) = P(g^{-1}(A))$ for every $A \in \mathscr{B}_X$. We say that a family of measurable functions is identically distributed if any two of them are identically distributed.

3 Strong law of large numbers

If $\zeta \in L^1(P)$, the **expectation of** ζ is

$$E(\zeta) = \int_\Omega \zeta dP,$$

and by the change of variables theorem,

$$\int_\Omega \zeta(\omega) dP(\omega) = \int_{\mathbb{R}} x d(\zeta_* P)(x).$$

The **strong law of large numbers**[3] states that if $\zeta_1, \zeta_2, \ldots \in L^1(P)$ are independent and identically distributed, with common expectation E_0, then

$$P\left(\left\{\omega \in \Omega : \sum_{i=1}^{n} \frac{\zeta_i(\omega)}{n} \to E_0\right\}\right) = 1.$$

4 Sample distributions

Let X be a separable metrizable space and let $\xi_1, \xi_2, \ldots$ be independent and identically distributed measurable functions $\Omega \to X$. For $\omega \in \Omega$, define μ_n^ω on $\mathscr{B}_X$ by

$$\mu_n^\omega = \sum_{i=1}^{n} \frac{1}{n} \delta_{\xi_i(\omega)},$$

which is a probability measure. We call the sequence μ_n^ω the **sample distribution of** ω.

[3]M. Loève, *Probability Theory I*, 4th ed., p. 251, 17.B.

The following is the **Glivenko-Cantelli theorem**, which shows that the sample distributions of a sequence of independent and identically distributed measurable functions converge narrowly almost everywhere to the common pushforward measure.[4]

Theorem 2 (Glivenko-Cantelli theorem)**.** *Let (Ω, S, P) be a probability space, let X be a separable metrizable space and let $\xi_1, \xi_2, \dots$ be independent and identically distributed measurable functions $\Omega \to X$, with common pushforward measure μ. Then*

$$P\left(\{\omega \in \Omega : \mu_n^\omega \to \mu \text{ narrowly}\}\right) = 1.$$

Proof. For $g \in C_b(X)$, $g \circ \xi_i : \Omega \to \mathbb{R}$ is measurable bounded, hence belongs to $L^1(P)$. Also, $(g \circ \xi_i)_* P = g_* \mu$, so the sequence $g \circ \xi_i$ are identically distributed. We now check that the sequence is independent. Let $A_1, \dots, A_n \in \mathscr{B}_\mathbb{R}$. Then $g^{-1}(A_1), \dots, g^{-1}(A_n) \in \mathscr{B}_X$, and because $\xi_1, \xi_2, \dots$ are independent,

$$P\left(\bigcap_{i=1}^n \xi_i^{-1}(g^{-1}(A_i))\right) = \prod_{i=1}^n P(\xi_i^{-1}(g^{-1}(A_i))),$$

i.e.,

$$P\left(\bigcap_{i=1}^n (g \circ \xi_i)^{-1}(A_i)\right) = \prod_{i=1}^n P((g \circ \xi_i)^{-1}(A_i)),$$

showing that $(g \circ \xi_1), (g \circ \xi_2), \dots$ are independent. For any i, by the change of variables theorem

$$E(g \circ \xi_i) = \int_\Omega g \circ \xi_i dP = \int_X gd((\xi_i)_* P) = \int_X gd\mu,$$

so the strong law of large numbers tells us that there is a set $N_g \in S$ with $P(N_g) = 0$ such that for all $\omega \in \Omega \setminus N_g$,

$$\sum_{i=1}^n \frac{(g \circ \xi_i)(\omega)}{n} \to \int_X gd\mu.$$

But

$$\sum_{i=1}^n \frac{(g \circ \xi_i)(\omega)}{n} = \sum_{i=1}^n \frac{1}{n} \int_X gd\delta_{\xi_i(\omega)} = \int_X gd\mu_n^\omega,$$

so for all $\omega \in \Omega \setminus N_g$,

$$\int_X gd\mu_n^\omega \to \int_X gd\mu.$$

Because X is separable, Theorem 1 tells us that there is a metric d that induces the topology of X and some countable dense subset G of $U_d(X)$ such that a sequence ν_n in $\mathscr{P}(X)$ converges narrowly to ν if and only if

$$\int_X gd\nu_n \to \int_X gd\nu, \qquad g \in G.$$

[4] K. R. Parthasarathy, *Probability Measures on Metric Spaces*, p. 53, Theorem 7.1.

Now let $N = \bigcup_{g \in G} N_g$, which satisfies $P(N) = 0$, and if $\omega \in \Omega \setminus N$ then for each $g \in G$,

$$\int_X g d\mu_n^\omega \to \int_X g d\mu.$$

This implies that for all μ_n^ω converges narrowly to μ. That is, there is a set $N \in S$ with $P(N) = 0$ such that for all $\omega \in \Omega \setminus N$, the sample distribution μ_n^ω converges narrowly to the common pushforward measure μ, proving the claim. □