Lecture 3: Convergences in Probability Theory

Michael Bishop
University of Arizona
Analysis Working Group

May 18, 2010

1 Introduction

The purpose of this lecture is to discuss various forms of convergence that exist in Probability Theory. I used the following works for background. [1, 2, 3, 4]

2 Borel-Cantelli Lemmas

First, limits of sequences of sets must be defined and motivated.

Definition 1. *Let $\{A_n\}$ be a countable collection of sets.*

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} A_m$$

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m$$

The liminf of events is referred to as the event "A_n eventually." That is, for any element in the liminf, there is an N such that the element is in every set in the rest of the limit. Conversely, the limsup of events is referred to as the event "A_n infinitely often (i.o.)." For every element in the limsup, for every N, there exists an A_n with $n \geq N$ that has the element. If the liminf equals the limsup, then that is the limiting set of the sequence. Sequences of set do not generally converge.

Example: Let $A_n = \{0, (-1)^n\}$. Then $\liminf A_n = \{0\}$ and $\limsup A_n = \{0, -1, +1\}$. So A_n has no limit in this case.

Now we have the means to define and show the Borel Cantelli Lemmas.

Theorem 2.1 (Borel-Cantelli 1)**.** *In $(\Omega, \mathcal{F}, P)$, given a sequence $\{A_n\}$ subset $\mathcal{F}$, if $\sum_{n=1}^{\infty} P[A_n] < \infty$, then $P[A_n$ i.o.$] = 0$*

Proof. Noting that $\bigcup_{m \geq 1} A_m \supseteq \bigcup_{m \geq 2} A_m \supseteq \bigcup_{m \geq 3} A_m \supseteq ...$ is a decreasing sequence, subadditivity of the measure can be applied such that

$$P[\limsup A_n] = \lim_{n \to \infty} P\left[\bigcup_{m \geq n} A_m\right] \leq \lim_{n \to \infty} \sum_{m \geq n} P[A_m]$$

Since $\sum_{m\geq n} P[A_m] < \infty$, the limit of the tail must go to zero. Thus $P[A_n i.o.] = 0$. □

Theorem 2.2 (Borel-Cantelli 2)**.** *In $(\Omega, \mathcal{F}, P)$, let $\{A_n\} \subset \mathcal{F}$ be a sequence of independent events. If $\Sigma_{n=1}^{\infty} P[A_n] = \infty$, then $P[A_n i.o.] = 1$*

To prove this theorem, note that A_n i.o. $= A_n^c$ eventually and $1 - x \leq e^{-x}$. I leave it as an exercise. The Borel-Cantelli Lemmas allow us to grasp limiting behaviors.

Example 1: Let X_i be Bernoulli random variables (coin flips), then Borel-Cantelli 2 tells us that $P[X_i = 1 \text{ i.o.}] = 1$ since $\sum P[X_i = 1] = \sum p = \infty$. So heads and tails will happen infinitely often if a coin flipped infinitely many times.

Example 2: Let X_n be independent identically distributed random variables with distribution $P[X_n > x] = e^{-x}$ for $x > 0$. Then $\forall \alpha > 0$, $P[X_n > \alpha \log n] = n^{-\alpha}$. Applying the Borel-Cantelli Lemmas,

$$P[X_n > \alpha \log(n) \text{ i.o.}] = \begin{cases} 0 & \alpha > 1 \\ 1 & \alpha \leq 1 \end{cases}$$

Let $L = \limsup \left(\frac{X_n}{\log n}\right)$. Then

$$P[L \geq 1] \geq P[X_n > \log n \text{ i.o.}] = 1$$

$$P\left[L \geq 1 + \frac{2}{k}\right] \leq P\left[X_n > (1 + \frac{1}{k}) \log n \text{ i.o.}\right] = 0$$

$$\Rightarrow P[L > 1] = 0 \Rightarrow P[L = 1] = 1.$$

3 Almost Sure Convergence

Almost sure convergence or (probability 1 convergence) is one of the strongest forms of convergence in probability theory.

Definition 2. *A sequence of random variables X_n converge to a random variable X almost surely (almost everywhwere or with probability 1) if $\{\omega \in \Omega : X_n(\omega) \nrightarrow X(\omega)\}$ is measure zero (probability zero).*
Equivalently: 1. $P[\lim X_n = X] = 1$
2. $\forall \epsilon > 0$, $P[|X_n - X| > \epsilon \text{ i.o.}] = 0$

Remark, this is the strongest convergence I will introduce. However, there are stronger forms of convergence like L^2 convergence of random variables. The associated theorem is the Strong Law of Large Numbers.

Theorem 3.1 (Strong Law of Large Numbers)**.** *Let X_i be independent random variables such that $\mathbb{E}[X_k] = 0$ and $\mathbb{E}\left[X_k^4\right] \leq K$ $\forall k$. Let $S_n = X_1 + ... + X_n$. Then $P\left[\frac{S_n}{n} \to 0\right] = 1$.*

Proof.

$$\mathbb{E}\left[S_n^4\right] = \mathbb{E}\left[\Sigma_k X_k^4 + 6 \sum_{i<j} X_i^2 X_j^2\right]$$

Since the cross terms with a factor of X_k must equal 0. Note $\mathbb{E}\left[X_i^2 X_j^2\right] = \mathbb{E}\left[X_i^2\right]^2 \leq \mathbb{E}\left[X_i^4\right] \leq K$. Then

$$\mathbb{E}\left[S_n^4\right] \leq nK + 3n(n-1)K \leq 3Kn^2$$
$$\mathbb{E}\left[\sum\left(\frac{S_n}{n}\right)^4\right] \leq 3K\sum n^{-2} < \infty$$
$$\Rightarrow P\left[\sum\left(\frac{S_n}{n}\right)^4 < \infty\right] = 1$$
$$\Rightarrow P\left[\frac{S_n}{n} \to 0\right] = 1$$

□

The Strong Law confirms our intuition on expectation: weighing each outcome with its probability is exactly the right formula for describiing long term behavior of a repeated random variable.

4 Convergence in Probability

A weaker form of convergence is convergence in probability (or in measure). This convergence is much easier to control and use.

Definition 3. *X_n converges to X in probability if* $\forall \epsilon > 0$

$$\lim_{n\to\infty} P[|X_n - X| > \epsilon] = 0$$

To show that probability in convergence is weaker than almost sure convergence. Assume that $X_n \to X$ almost surely. Then

$$0 = P[|X_m - X| > \epsilon \text{ i.o.}]$$

By continuity from above

$$P[|X_m - X| > \epsilon \text{ i.o.}] = \lim_{n\to\infty} P\left[\bigcup_{m\geq n}\{|X_m - X| > \epsilon\}\right]$$

By Reverse Fatou's Lemma

$$P\left[\bigcup_{m\geq n}\{|X_m - X| > \epsilon\}\right] \geq \limsup P[|X_m - X| > \epsilon] \geq \lim P[|X_m - X| > \epsilon] \geq 0$$

So $X_n \to X$ in probability.

Convergence almost everywhere is stronger than convergence in probability. If variables converge in probability, then to show that the variables converge almost everywhere the condition that $\sum P[|X_n - X| > \epsilon] < \infty$ must be satisfied. Then by Borel-Cantelli 1, $P[|X_n - X| > \epsilon \text{ i.o.}] = 0$, which is almost sure convergence.

To prove the Weak Law of Large Numbers, we need the Chebyshev inequality.

Theorem 4.1 (Chebyshev Inequality)**.** *Let g be a non-negative and increasing function. Then*

$$P[|X - \mathbb{E}[X]| \geq \epsilon] \leq \frac{\mathbb{E}[g(X - \mathbb{E}[X])]}{g(\epsilon)}$$

This is a powerful all purpose theorem for showing convergence and convergence rates.

Theorem 4.2 (Weak Law of Large Numbers)**.** *Let X_i be independent identically distributed random variables with bounded variance (=M). Then if $\bar{X}_n$ is the sample mean,*

$$\lim_{n\to\infty} P[|\bar{X}_n - \mathbb{E}[X_i]| < \epsilon] = 1$$

Proof. Using the Chebyshev Inequality with $g(x) = x^2$:

$$P[|\bar{X}_n - \mathbb{E}[X_i]| < \epsilon] \leq \frac{Var(\bar{X}_n)}{\epsilon^2} \leq \frac{M}{n^2\epsilon^2} \to 0 \text{ as } n \to \infty$$

□

5 Convergence in Distribution

Convergence in distribution is weaker than the above distributions, but it does give a means to estimate limiting probabilities in a clear way.

Definition 4. *Probability measures P_n converge in distribution (or weakly) to P if ∀ continuous and bounded f*

$$\int_\Omega f dP_n \to \int_\Omega f dP \text{ as } n \to \infty$$

Random variables X_n converge in distribution if

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$$

If $F_n(x) = P[X_n \leq x]$, then X_n converges in distribution to X if for all x such that $F(x) = P[X \leq x]$ is continuous,

$$F_n(x) \to F(x) \text{ as } n \to \infty$$

These are all equivalent definitions. The main theorem for distributional convergence is one of the most powerful:

Theorem 5.1 (Central Limit Theorem)**.** *If X_i are i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\sigma^2 = Var(X_i) < \infty$, then $\frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}}$ converges in distribution to the normal distribution with mean 0 and variance 1. Equivalently,*

$$P\left[\frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}} \leq x\right] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy \text{ as } n \to \infty$$

The idea is to take a fourier transform or laplace transform and show all terms disappear except the ones giving the normal distribution.

References

[1] Rabi Bhattacharya and Edward C. Waymire, *A basic course in probability theory*, Universitext, Springer, New York, 2007. MR MR2331066 (2009e:60001)

[2] George Casella and Roger L. Berger, *Statistical inference*, The Wadsworth & Brooks/Cole Statistics/Probability Series, Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990. MR MR1051420 (91j:62002)

[3] Gerald B. Folland, *Real analysis*, second ed., Pure and Applied Mathematics (New York), John Wiley & Sons Inc., New York, 1999, Modern techniques and their applications, A Wiley-Interscience Publication. MR MR1681462 (2000c:00001)

[4] David Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991. MR MR1155402 (93d:60002)