Key Concepts in Advanced Probability

Peter Glynn (glynn@stanford.edu)

April 8, 2007

1 Expectations

The expectation $\mathrm{E}[X]$ of any non-negative random variable X can always be defined uniquely; the expectation may be either finite or infinite. If X is of arbitrary sign, put

$$\begin{aligned} X^+ &= \max(X, 0) \\ X^- &= \max(-X, 0) \end{aligned}$$

$\mathrm{E}[X]$ is said to *exist* if at least one of $\mathrm{E}[X^+]$ and $\mathrm{E}[X^-]$ are finite, in which case we put $\mathrm{E}[X] \triangleq \mathrm{E}[X^+] - \mathrm{E}[X^-]$. If $\mathrm{E}[|X|] < \infty$ then X is said to be *integrable*.

If $X = g(Y)$ (for $g : \mathbb{R} \to \mathbb{R}$ and Y a real-valued r.v.), then $\mathrm{E}[X]$ can be computed as a Stieljes integral:

$$\mathrm{E}[X] = \int_{\mathbb{R}} g(y)F(dy),$$

where F is the distribution function of the r.v. Y given by $F(y) = \mathrm{P}\{Y \leq y\}$. Note that if Y has a density f, then

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} g(y)f(y)dy,$$

whereas if Y has a probability mass function p then

$$\mathrm{E}[X] = \sum_y g(y)p(y).$$

2 Useful Inequalities

For $p \geq 1$, put

$$\|X\|_p \triangleq \mathrm{E}[|X|^p]^{1/p}.$$

Then, we have

$$\|X_1 + \ldots + X_n\|_p \leq \|X_1\|_p + \ldots + \|X_n\|_p$$

(Minkowski's inequality), and

$$\|X_1 X_2\|_1 \leq \|X_1\|_p \, \|X_2\|_q$$

for $1/p + 1/q = 1$ with $p, q \geq 1$ (Hölder's inequality). The special case with $p = q = 2$, namely

$$\|X_1 X_2\|_1 \leq \|X_1\|_2 \|X_2\|_2,$$

is called the Cauchy-Schwarz inequality.

If X is nonnegative,

$$P\{X > x\} \le x^{-1} E[X]$$

for $x > 0$; this is called Markov's inequality. If $E\left[W^2\right] < \infty$, then we can set $X = (W - E[W])^2$ to yield

$$P\{|W - E[W]| > w\} \le \text{var}(W)/w^2;$$

this special case is called Chebyshev's inequality. If $E[\exp(\theta W)] < \infty$, put $X = \exp(\theta W)$; this yields the inequality

$$P\{W > w\} \le \exp(-\theta w) E[\exp(\theta W)].$$

Hence, we obtain the exponential inequality

$$P\{W > w\} \le \inf_{\theta \ge 0} \exp(-\theta w) E[\exp(\theta W)].$$

3 Weak Convergence

Let $(X_n : 1 \le n \le \infty)$ be a sequence of finite real-valued random variables. Then, X_n *converges weakly* to X_∞ (also known as *convergence in distribution*) if

$$P\{X_n \le x\} \to P\{X_\infty \le x\}$$

as $n \to \infty$ at each x at which $P\{X_\infty \le \cdot\}$ is continuous. We use the notation

$$X_n \Longrightarrow X_\infty$$

(or, equivalently, $P_n \Longrightarrow P_\infty$ where $P_n(\cdot) = P\{X_n \in \cdot\}$) to denote weak convergence.

Weak convergence can be re-formalized in several equivalent ways.

Theorem 1. *Let $(X_n : 1 \le n \le \infty)$ be a sequence of finite real-valued random variables. Then, the following are equivalent:*

i.) $X_n \Longrightarrow X_\infty$ as $n \to \infty$;

ii.) For each bounded and continuous $f : I\!R \to I\!R$, $E[f(X_n)] \to E[f(X_\infty)]$ as $n \to \infty$;

iii.) For each bounded and continuously differentiable $f : I\!R \to I\!R$, $E[f(X_n)] \to E[f(X_\infty)]$ as $n \to \infty$;

iv.) There exists a probability space supporting a sequence of r.v.'s $(X_n^* : 1 \le n \le \infty)$ for which $X_n^* \stackrel{\mathcal{D}}{=} X_n$ for $1 \le n \le \infty$ (where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution) and on which $X_n^* \to X_\infty^*$ a.s. as $n \to \infty$.

A key result in the theory of weak convergence is the fact that weak convergence is preserved under continuous mappings (This result is known as the "continuous mapping principle".).

Proposition 1: For $g : I\!R \to I\!R$, let $D(g) = \{x : g(\cdot) \text{is discontinuous at } x\}$. Suppose that $X_n \Longrightarrow X_\infty$ as $n \to \infty$, where $P\{X_\infty \in D(g)\} = 0$. Then,

$$g(X_n) \Longrightarrow g(X_\infty)$$

as $n \to \infty$.

4 Convergence in Probability

Let $(X_n : 1 \leq n \leq \infty)$ be a sequence of finite real-valued random variables. Then, X_n *converges in probability* to X_∞ if for each $\varepsilon > 0$,

$$\mathrm{P}\{|X_n - X_\infty| > \varepsilon\} \to 0$$

as $n \to \infty$. We use the notation

$$X_n \xrightarrow{p} X_\infty$$

to denote convergence in probability. Note that convergence in probability involves the joint distribution of (X_n, X_∞) whereas weak convergence concerns only the (marginal) distribution of X_n.

If $X_n \xrightarrow{p} X_\infty$ as $n \to \infty$, then $X_n \Longrightarrow X_\infty$ as $n \to \infty$. A partial converse also exists. If $X_n \Longrightarrow X_\infty$ as $n \to \infty$ where X_∞ is deterministic (i.e. $\mathrm{P}\{X_\infty = c\} = 1$ for some $c \in \mathbb{R}$), then $X_n \xrightarrow{p} X_\infty$ as $n \to \infty$.

Suppose that $X_n \Longrightarrow X_\infty$ as $n \to \infty$ and $Y_n \Longrightarrow Y_\infty$ as $n \to \infty$. Weak convergence and convergence in probability extend in the obvious way to $\mathbb{R}^d$ valued random variables. One might hope that $(X_n, Y_n) \Longrightarrow (X_\infty, Y_\infty)$ as $n \to \infty$. This generally is false. However:

- If $X_n \xrightarrow{p} X_\infty$ as $n \to \infty$ and $Y_n \xrightarrow{p} Y_\infty$ as $n \to \infty$ (with the X_n's and the Y_n's defined on a common probability space) then $(X_n, Y_n) \xrightarrow{p} (X_\infty, Y_\infty)$ as $n \to \infty$.
- If $X_n \Longrightarrow X_\infty$ as $n \to \infty$ and $Y_n \xrightarrow{p} c$ as $n \to \infty$ (for $c \in \mathbb{R}$) then $(X_n, Y_n) \Longrightarrow (X_\infty, c)$.

It follows from the continuous mapping principle for $\mathbb{R}^2$-valued random variables that if $X_n \Longrightarrow X_\infty$ as $n \to \infty$ and $Y_n \xrightarrow{p} c$ as $n \to \infty$, then

$$\begin{aligned} X_n + Y_n &\Longrightarrow X_\infty + c\,, \\ X_n Y_n &\Longrightarrow cX_\infty\,, \end{aligned}$$

as $n \to \infty$.

5 Convergence in p'th Mean

For $p > 0$ put

$$\|X\|_p = E^{1/p}|X|^p.$$

We say that X_n *converges to* X_∞ *in p'th mean* if

$$\|X_n - X_\infty\|_p \to 0$$

as $n \to \infty$. Markov's inequality implies that if X_n converges to X_∞ in p'th mean, then $X_n \xrightarrow{p} X_\infty$ as $n \to \infty$,
For $p \geq 1$, the vector space $L^p = \{X : \|X\| < \infty\}$ is a Banach space equipped with norm $\|\cdot\|_p$. (Minkowski's inequality is a statement of the "triangle inequality" in L^p.) For $p = 2$, L^2 is a Hilbert space equipped with inner product $\langle X, Y\rangle \triangleq \mathrm{E}[XY]$.

6 Almost Sure Convergence

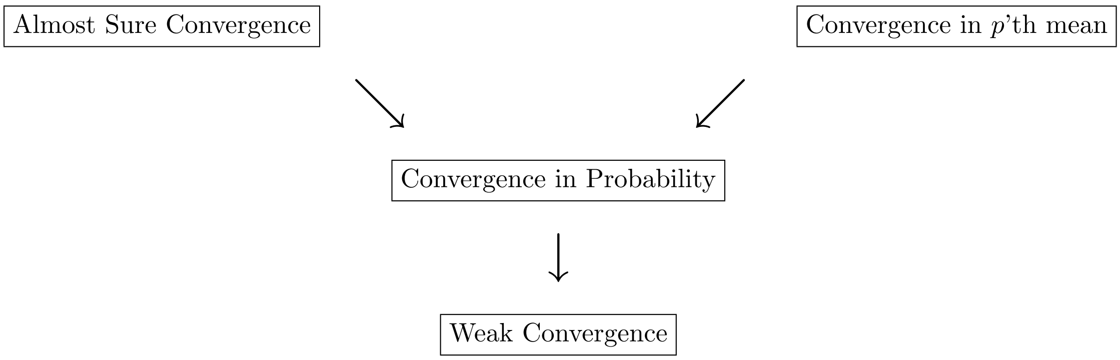
Let $(X_n : 1 \leq n \leq \infty)$ be a sequence of real-valued random variables defined on a common probability space. We say that X_n *converges almost surely* to X_∞ as $n \to \infty$ if $\mathrm{P}\{A\} = 1$, where

$$A = \{\omega : X_n(\omega) \to X_\infty(\omega) \text{ as } n \to \infty\}.$$

Hence, almost sure convergence is a statement about the "infinite dimensional" event A. Note that A is the collection of sample outcomes ω on which the numerical sequence $X_n(\omega)$ converges to $X_\infty(\omega)$ as $n \to \infty$.

Almost sure convergence has alternative names: Convergence with probability one, Convergence almost everywhere, Convergence almost certainly. Almost sure convergence implies convergence in probability.

7 Relationship between Types of Convergence



8 Interchanging Limits and Expectations

If $X_n \to X_\infty$ a.s. (i.e. X_n converges to X_∞ almost surely), it need not follow that $\mathrm{E}[X_n] \to \mathrm{E}[X_\infty]$ as $n \to \infty$. For example, if U is uniform on $[0,1]$ and

$$X_n = nI(U \leq 1/n),$$

then $X_n \to 0$ a.s. as $n \to \infty$, but $\mathrm{E}[X_n] = 1$ for $n \geq 1$.
The interchange of limit and expectation can be verified under certain conditions:

Dominated Convergence Theorem (DCT)
Suppose that $X_n \to X_\infty$ a.s. as $n \to \infty$. If there exists an integrable r.v. Y for which

$$|X_n(\omega)| \leq Y(\omega)$$

for $n \geq 1$, then $\mathrm{E}[X_n] \to \mathrm{E}[X_\infty]$ as $n \to \infty$(where $\mathrm{E}[X_\infty]$ is necessarily finite).
The special case where $Y = c$ a.s. (for some constant $c \in \mathbb{R}$) is known as the **Bounded Convergence Theorem**.

Monotone Convergence Theorem (MCT)
Suppose that $(X_n : 1 \leq n \leq \infty)$ is a sequence of non-negative r.v.'s for which $X_n \to X_\infty$ a.s. as $n \to \infty$. If the sequence $(X_n : 1 \leq n \leq \infty)$ is monotone in the sense that

$$X_n(\omega) \leq X_{n+1}(\omega)$$

for $n \geq 0$ and $\omega \in \Omega$, then $\mathrm{E}[X_n] \to \mathrm{E}[X_\infty]$ as $n \to \infty$(where $\mathrm{E}[X_\infty]$ may be finite or infinite).

Another useful result is the following: **Fatou's Lemma**
Suppose that $(X_n : 1 \leq n \leq \infty)$ is a sequence of non-negative r.v.'s. Then

$$\mathrm{E}\left[\varliminf_{n\to\infty} X_n\right] \leq \varliminf_{n\to\infty} \mathrm{E}[X_n].$$

The following extension to the Dominated Convergence Theorem is also sometimes useful. We say that a collection $(X_\lambda : \lambda \in \Lambda)$ of random variables is *uniformly integrable* if for each $\varepsilon > 0$, there exists $x = x(\varepsilon)$ such that

$$\sup_{\lambda \in \Lambda} \mathrm{E}\left[|X_\lambda| I\left(|X_\lambda| > x\right)\right] < \varepsilon$$

If $X_n \Longrightarrow X_\infty$ as $n \to \infty$where $(X_n : 1 \leq n < \infty)$ is uniformly integrable, then $\mathrm{E}[X_n] \to \mathrm{E}[X_\infty]$ as $n \to \infty$.

Finally, one key property of expectations is its linearity. In particular, if $X_1, \ldots, X_m$ are integrable r.v.'s, then

$$\mathrm{E}\left[\sum_{i=1}^{m} X_i\right] = \sum_{i=1}^{m} \mathrm{E}[X_i] \ .$$

Extending this linearity to the case where $m = \infty$ involves taking a limit. As for the case of interchanging a limit and an expectation, the extension to $m = \infty$ is not always valid.
Fubini's theorem provides a sufficient condition.

Theorem 2. Fubini's Theorem *Let $(X_\lambda : \lambda \in \Lambda)$ be a collection of r.v.'s and let $\mu(\cdot)$ be a non-negative measure on Λ.*

i.) If $X(\lambda)$ is non-negative for $\lambda \in \Lambda$, then

$$E \int_\Lambda X(\lambda)\mu(d\lambda) = \int_\Lambda EX(\lambda)\mu(d\lambda).$$

ii.) If $X(\lambda)$ is of mixed sign and

$$\int_\Lambda \mathrm{E}\left[|X(\lambda)|\right] \mu(d\lambda) < \infty \,,$$

then

$$\mathrm{E}\left[\int_\Lambda X(\lambda)\mu(d\lambda)\right] = \int_\Lambda \mathrm{E}\left[X(\lambda)\right] \mu(d\lambda).$$

Putting μ = counting measure on the integers yields the conclusion that

$$\mathrm{E}\left[\sum_{i=1}^{\infty} X_i\right] = \sum_{i=1}^{\infty} \mathrm{E}[X_i]$$

if the X_i's are either non-negative or satisfy

$$\sum_{i=1}^{\infty} \mathrm{E}\left[|X_i|\right] < \infty.$$

9 Transforms

Transform methods can be a useful approach for establishing weak convergence of real-valued random variables, as well as for dealing with sums of independent random variables (i.e. convolutions).
The *characteristic function* of a random variable X is the function defined by

$$\begin{aligned} c(t) &= \mathrm{E}\left[\exp(itX)\right] \\ &= \int_{\boldsymbol{R}} e^{itx}\, \mathrm{P}\{X \in dx\} \end{aligned}$$

The characteristic function exists and is finite-valued for every r.v. X. When X has a density f, the characteristic function is just the Fourier transform of f (up to a constant multiple).

Inversion Theorem Let $c(\cdot)$ be the characteristic function of a r.v. X. Then, for $a < b$,

$$\mathrm{P}\{X \in (a,b)\} + \frac{1}{2}\mathrm{P}\{X = a\} + \frac{1}{2}\mathrm{P}\{X = b\} = \lim_{u \to \infty} \frac{1}{2\pi} \int_{-u}^{u} \frac{e^{-ita} - e^{-itb}}{it} c(t) dt\,.$$

If $c(\cdot)$ is integrable over $(-\infty, \infty)$, then X has a continuous density f given by

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} c(t) dt\,.$$

It follows from the inversion theorem that if two r.v.'s X and Y have the same characteristic function, then $X \stackrel{\mathcal{D}}{=} Y$.

The following result establishes the connection between weak convergence and convergence of characteristic functions.

Theorem 3. *Let $(X_n : 1 \leq n \leq \infty)$ be a sequence of finite real-valued random variables, and put $c_n(t) =$ $\mathrm{E}[\exp(itX_n)]$ for $1 \leq n \leq \infty$.*

i.) If $X_n \Longrightarrow X_\infty$ as $n \to \infty$, then $c_n(\cdot) \to c_\infty(\cdot)$ uniformly in every finite interval.

ii.) Suppose that for each $t \in I\!R$, $c_n(t) \to c(t)$ as $n \to \infty$, where $c(\cdot)$ is continuous at $t = 0$. Then, $c(\cdot)$ is the characteristic function of a r.v. Y and $X_n \Longrightarrow Y$ as $n \to \infty$.