Probability and Stochastic Processes I - Lecture 15

Michael Evans
University of Toronto
http://www.utstat.utoronto.ca/mikevans/stac62/STAC622023.html

2023

III.2 Convergence With Probability 1

Definition III.2.1 The sequence of r.v.'s $\{X_n\}$ *converges with probability 1* to r.v. X if

$$P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$$

and write $X_n \stackrel{wp1}{\to} X$. ■

note

$$\begin{aligned} \{\omega : & \lim_{n\to\infty} X_n(\omega) = X(\omega)\} \\ = & \cap_{m=1}^{\infty} \liminf_n \{\omega : |X_n(\omega) - X(\omega)| < 1/m\} \\ = & \cap_{m=1}^{\infty} \cup_{n=1}^{\infty} \cap_{i=n}^{\infty} \{\omega : |X_i(\omega) - X(\omega)| < 1/m\} \in \mathcal{A} \end{aligned}$$

Example III.2.1

$(\Omega, \mathcal{A}, P) = (R^1, \mathcal{B}^1, P)$ where P is the uniform distribution on $[0, 1]$ so

$$P(B) = \int_{B \cap [0,1]} dx$$

and let $X_n(\omega) = \frac{n}{n+1}\omega^2$ and $X(\omega) = \omega^2$

- then $\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = R^1$ and $P(R^1) = \int_{[0.1]} dx = 1$ so $X_n \overset{wp1}{\to} X$

- let

$$X_*(\omega) = \begin{cases} \omega^2 & \text{if } \omega \neq 1/2 \\ 1 & \text{if } \omega = 1/2 \end{cases}$$

then $\{\omega : \lim_{n\to\infty} X_n(\omega) = X_*(\omega)\} = R^1 \setminus \{1/2\}$ and

$$P(R^1 \setminus \{1/2\}) = \int_{[0,1/2)} dx + \int_{(1/2,1]} dx = 1/2 + 1/2 = 1$$

and so $X_n \overset{wp1}{\to} X_*$ too

- we could change X at every rational $q \in \mathbb{Q}$ to obtain X_{**} and since $P(\mathbb{Q}) = 0$ we still have $X_n \overset{wp1}{\to} X_{**}$ ■

- a *measure* ν defined on $(\Omega, \mathcal{A})$ is a function $\nu : \mathcal{A} \rightarrow [0, \infty]$ that satisfies $\nu(\phi) = 0$ and $\nu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \nu(A_i)$ whenever $A_1, A_2, \ldots \in \mathcal{A}$ are mutually disjoint

Example III.2.2

- a probability measure ν defined on $(\Omega, \mathcal{A})$ is a measure

- counting measure defined by $\nu(A) = \#(A)$ is a measure

- if $(\Omega, \mathcal{A}) = (R^k, \mathcal{B}^k)$ and $\nu(A) = Vol(A)$ is a measure ■

- now suppose $h : (\Omega, \mathcal{A}) \rightarrow (R^1, \mathcal{B}^1)$ ($h : \Omega \rightarrow R^1$ and $h^{-1}B \in \mathcal{A}$ for every $B \in \mathcal{B}^1$)

- then just as we did for r.v. X and P we can define a kind of average of h with respect to ν (simple functions h, nonnegative functions h, general functions $h = h_+ - h_-$) which, when it exists, is denoted

$$\int_{\Omega} h(\omega)\, \nu(d\omega)$$

called the *integral of* h *with respect to* ν

- so, for example, the expectation of r.v. X can also be written as the integral of X with respect to $\nu = P$, namely,

$$E(X) = \int_{\Omega} X(\omega)\, P(d\omega)$$

- with $\nu =$ counting measure on $(\Omega, \mathcal{A})$ (fact)

$$\int_{\Omega} h(\omega)\, \nu(d\omega) = \sum_{\omega \in \Omega} h(\omega)$$

and with $\nu =$ volume measure on $(R^k, \mathcal{B}^k)$ (fact)

$$\int_{\Omega} h(\omega)\, \nu(d\omega) = \int_{R^k} h(\mathbf{x})\, d\mathbf{x}$$

- if $\{h_n\}$ is a sequence of such functions, then we say the sequence *converges almost surely* ν *to* h if

$$\nu(\{\omega : \lim_{n\to\infty} h_n(\omega) \neq h(\omega)\}) = 0$$

and write $h_n \stackrel{a.s.\,\nu}{\to} h$

- so convergence almost surely P to h is convergence with probability 1

- we need the following results

Proposition III.2.1 Suppose $h_n \stackrel{a.s.\,\nu}{\to} h$.

(i) (*Monotone Convergence MCT*) If $0 \leq h_1 \leq h_2 \leq \cdots$, then $\int_\Omega h_n(\omega)\,\nu(d\omega) \uparrow \int_\Omega h(\omega)\,\nu(d\omega)$.

(ii) (*Dominated Convergence DCT*) If there exists $g : (\Omega, \mathcal{A}) \to (R^1, \mathcal{B}^1)$ such that $\int_\Omega |g(\omega)|\,\nu(d\omega) < \infty$ and $|h_n| \leq |g|$ for very n, then $\int_\Omega h_n(\omega)\,\nu(d\omega) \to \int_\Omega h(\omega)\,\nu(d\omega)$.

Proof: Accept. ■

Corollary III.2.1 Suppose $X_n \stackrel{wp1}{\to} X$.

(i) If $0 \le X_1 \le X_2 \le \cdots$, then $E(X_n) \uparrow E(X)$.

(ii) If there exists r.v. Y such that $E(|Y|) < \infty$ and $|X_n| \le |Y|$ for very n, then $E(X_n) \to E(X)$.

Example III.2.1 (continued)

- then $X_n(\omega) = \frac{n}{n+1}\omega^2 \uparrow X(\omega) = \omega^2$ and so by MCT $E(X_n) \uparrow E(X)$ and $E(X_n) \uparrow E(X_*)$ ■

Example III.2.2

- suppose X is s.t. $E(X)$ is finite and let $X_n = XI_{\{|X| \le n\}}$

- then $X_n \stackrel{wp1}{\to} X$ and $|X_n| \le |X|$ so by DCT $E(X_n) \to E(X)$ ■

Lemma III.3.1 If X is a r.v. with respect to $(\Omega, \mathcal{A}, P)$ and $h : (R^1, \mathcal{B}^1) \to (R^1, \mathcal{B}^1)$, then $Y = h(X)$ is a r.v. with respect to $(\Omega, \mathcal{A}, P)$.

Proof: Let $B \in \mathcal{B}^1$. Then

$$\begin{aligned} Y^{-1}B &= \{\omega : Y(\omega) = h(X(\omega)) \in B\} \\ &= \{\omega : X(\omega) \in h^{-1}B\} = X^{-1}h^{-1}B \in \mathcal{A} \end{aligned}$$

since $h^{-1}B \in \mathcal{B}^1$ and X is a r.v. ■

- when h is a r.v. with respect to $(R^1, \mathcal{B}^1, P_X)$ does $E(Y) = E_{P_X}(h)$?

Proposition III.3.2 If X is a r.v. with respect to $(\Omega, \mathcal{A}, P)$ and $h : (R^1, \mathcal{B}^1) \to (R^1, \mathcal{B}^1)$, then $E(Y) = E_{P_X}(h)$ when it exists.

Proof: Suppose $h = \sum_{i=1}^k b_i I_{B_i}$ is a simple function. Then

$$Y(\omega) = h(X(\omega)) = \sum_{i=1}^k b_i I_{B_i}(X(\omega)) = \sum_{i=1}^k b_i I_{X^{-1}B_i}(\omega)$$

is a simple function on Ω and so

$$E(Y) = \sum_{i=1}^k b_i P(X^{-1}B_i) = \sum_{i=1}^k b_i P_X(B_i) = E_{P_X}(h).$$

If $h \geq 0$ so $Y = h(X) \geq 0$, then there exist nonnegative simple $W_n \uparrow h$ which implies $W_n(X) \uparrow h(X) = Y$. So using definition of expectation for nonnegative r.v.'s,

$$E_{P_X}(h) = \lim_{n\to\infty} E_{P_X}(W_n) = \lim_{n\to\infty} E(W_n(X)) = E(Y).$$

In general write $h = h_+ - h_-$ so $h(X) = h_+(X) - h_-(X)$ and apply the above result to both parts. ■

Proposition III.3.3 Suppose X is a r.v. with respect to $(\Omega, \mathcal{A}, P)$, $h : (R^1, \mathcal{B}^1) \longrightarrow (R^1, \mathcal{B}^1)$ and $E_{P_X}(h)$ exists.

(i) If P_x is discrete with prob. fn p_X, then $E_{P_X}(h) = \sum_{x \in R^1} h(x) p_X(x)$.

(ii) If P_x is a.c. with density fn f_X, then $E_{P_X}(h) = \int_{-\infty}^{\infty} h(x) f_X(x)\, dx$.

Proof: Suppose $h(x) = \sum_{i=1}^k b_i I_{B_i}(x)$ is a simple function in canonical form. Then

$$\begin{aligned} E_{P_X}(h) &= \sum_{i=1}^k b_i P_X(B_i) = \begin{cases} \sum_{i=1}^k b_i \sum_{x \in B_i} p_X(x), & \text{if } X \text{ discrete} \\ \sum_{i=1}^k b_i \int_{B_i} f_X(x)\, dx, & \text{if } X \text{ a.c.} \end{cases} \\ &= \begin{cases} \sum_{x \in R^1} h(x) p_X(x), & \text{if } X \text{ discrete} \\ \int_{-\infty}^{\infty} h(x) f_X(x)\, dx, & \text{if } X \text{ a.c.} \end{cases} \\ &= \begin{cases} \int_{-\infty}^{\infty} h(x) p_X(x)\, \nu(dx), & \nu = \text{counting measure} \\ \int_{-\infty}^{\infty} h(x) f_X(x)\, \nu(dx), & \nu = \text{volume measure} \end{cases} \end{aligned}$$

which proves the result for simple h.

If $h \geq 0$ and nonnegative simple $h_n \uparrow h$ then (i) $h_n p_X \uparrow hp_X$ (ii) $h_n f_X \uparrow hf_X$ and the result follows by MCT. The result follows for general h via the decomposition $h = h_+ - h_-$. ■

Example III.2.3 $X \sim N(\mu, \sigma^2)$

- then with $h(x) = x$

$$
\begin{aligned}
E(X) \; &= \; \int_0^\infty x \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx \; - \\
&\quad\; \int_{-\infty}^0 (-x) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx
\end{aligned}
$$

and making the change of variable $t = T(x) = (x - \mu)/\sigma$ in both integrals (with $J_T(x) = \sigma$ and $T^{-1}(t) = \mu + \sigma t$) and putting

$$\varphi(t) = (2\pi)^{-1/2} \exp(-t^2/2)$$

$$
\begin{aligned}
E(X) &= \int_0^\infty (\mu+\sigma t)\varphi(t)\,dt + \int_{-\infty}^0 (\mu+\sigma t)\varphi(t)\,dt \\
&= \mu\int_{-\infty}^\infty \varphi(t)\,dt + \sigma\left(\int_0^\infty t\varphi(t)\,dt + \int_{-\infty}^0 t\varphi(t)\,dt\right) = \mu
\end{aligned}
$$

since $\int_{-\infty}^0 t\varphi(t)\,dt = -\int_0^\infty t\varphi(t)\,dt$ on putting $w=-t$

- also, with t as before, $h(x) = (x-\mu)^2$

$$
\begin{aligned}
E\left((X-\mu)^2\right) &= \int_{-\infty}^\infty (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)dx \\
&= \sigma^2\int_{-\infty}^\infty t^2\varphi(t)\,dt
\end{aligned}
$$

- using integration by parts with $u=t, dv=t\varphi(t)$, then $du=dt, v=-\varphi(t)$

$$
\int_{-\infty}^\infty t^2\varphi(t)\,dt = -t\varphi(t)\big|_{t=-\infty}^{t=\infty} + \int_{-\infty}^\infty \varphi(t)\,dt = 0+1=1
$$

- so $E\left((X-\mu)^2\right) = \sigma^2$ ■

Definition III.3.1 The *k-th moment* of a r.v. X is given by $\mu_k = E(X^k)$ when this exists. When the first moment exists, the *k-th central moment* of a r.v. X is given by $\bar{\mu}_k = E((X - \mu_1)^k)$ when it exists. The *mean* of X is given by $\mu_X = E(X)$ and the variance of X is given by $\sigma_X^2 = Var(X) = E((X - \mu_X)^2)$ when μ_X exists. ■

Proposition III.3.4 If μ_k is finite then μ_l is finite for $l = 1, 2, \ldots, k$.

Proof: Note μ_k is finite iff $E(|X|^k)$ is finite and putting $h(x) = |x|^l$

$$0 \le E(|X|^l) = E_{P_X}(h) = \int_{-\infty}^{\infty} |x|^l \, P_X(dx)$$

$$\begin{aligned} &= \int_{-\infty}^{-1} |x|^l \, P_X(dx) + \int_{-1}^{1} |x|^l \, P_X(dx) + \int_{1}^{\infty} |x|^l \, P_X(dx) \\ &\le \int_{-\infty}^{-1} |x|^k \, P_X(dx) + \int_{-1}^{1} 1 \, P_X(dx) + \int_{1}^{\infty} |x|^k \, P_X(dx) \\ &\le \int_{-\infty}^{\infty} |x|^k \, P_X(dx) + P_X([-1, 1]) < \infty. \; ■ \end{aligned}$$

Exercise III.3.1 When $X \sim N(\mu, \sigma^2)$ compute $E(X^3)$ and $E(X^4)$.

Exercise III.3.2 When $X \sim$ Standard Cauchy, namely, X has density $f_X(x) = 1/\pi(1+x^2)$ for $-\infty < x < \infty$, show that μ_1 doesn't exist.

Exercise III.3.3 E&R 3.1.17.

Exercise III.3.4 E&R 3.1.22, E&R 3.3.18 and E&R 3.3.19.

Exercise III.3.5 E&R 3.2.16 and E&R 3.3.20.

Exercise III.3.6 E&R 3.2.22 and E&R 3.3.24.

Example III.2.4 *Monte Carlo Approximations*

- suppose $Y = h(X)$ for some $h : (R^1, \mathcal{B}^1) \to (R^1, \mathcal{B}^1)$ and we want to compute $E(Y)$

- often this can be very difficult unless P_Y is easy to work with

- but if we can generate $X_1, X_2, \ldots \overset{i.i.d.}{\sim} P_X$ then $Y_1, Y_2, \ldots \overset{i.i.d.}{\sim} P_Y$

- then a very natural estimator of $E(Y)$ is

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$$

and we will show (later) that this converges to $E(Y)$ as $n \to \infty$

- how accurate is this estimate for some specific n?

- the Central Limit Theorem (later) says, for large n,

$$\frac{\bar{Y} - E(Y)}{\sqrt{Var(Y)/n}} \approx N(0, 1)$$

provided $Var(Y) < \infty$

- $Var(Y) = E((Y - E(Y))^2) = E(Y^2) - (E(Y))^2$ can be estimated (later) by

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \bar{Y}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

and indeed (later)

$$\frac{\bar{Y} - E(Y)}{\sqrt{s^2/n}} \approx N(0,1)$$

- if $Z \sim N(0,1)$ then $P(-3 < Z < 3) = 0.9973002 \approx 1$

- combining these statements we can say that the true value of $E(Y)$ lies in the interval

$$[\bar{Y} - 3s/\sqrt{n}, \bar{Y} + 3s/\sqrt{n}]$$

with "virtual certainty" and the length of the interval assesses the accuracy of the estimate

note when $Y = I_A$ then $\bar{Y} =$ the relative frequency of A in $X_1, X_2, \ldots, X_n$ and $Y_i^2 = Y_i$ so $s^2 = \bar{Y}(1 - \bar{Y})$ and this is the same estimation procedure as previously discussed for estimating $P_X(A)$