CONVERGENCE CONCEPTS & LAWS OF LARGE NUMBERS

Before discussing the Central Limit Theorem (CLT), *Weak Law of Large Numbers* (WLLN) and *Strong Law of Large Numbers* (SLLN) it helps to know some different convergence concepts that exist in probability (and measure theory).

We begin with two results that help us bound probabilities when only the mean is known:

Markov Inequality: For any non-negative valued r.v. Y with $E(Y) = \mu$, then for $a > 0$

$$P(Y \geq a) \leq \frac{E(Y)}{a}.$$

Proof (finite-variance, continuous case):

$$\frac{E(Y)}{a} = \frac{1}{a}\int_0^\infty y\, f(y)\, dy \geq \frac{1}{a}\int_a^\infty y\, f(y)\, dy \geq \frac{1}{a}\int_a^\infty a\, f(y)\, dy = P(Y \geq a) \qquad \blacksquare$$

Chebychev Inequality: For r.v. X with $E(X) = \mu$ and $Var(X) = \sigma^2 < \infty$, then for any $k > 0$ the probability that X deviates more than k from the mean is bounded by

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Sketch of Proof: Apply the Markov Inequality using $Y = (X - \mu)^2$ and $a = k^2$.

CONVERGENCE CONCEPTS IN PROBABILITY

Definition: The r.v.s X_n **converge in distribution** to r.v. X ($X_n \overset{D}{\to} X$) if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

for all x where $F_X(x)$ is continuous. This is point-wise convergence of cdfs.

Definition: The r.v.s X_n **converge in probability** to r.v. X ($X_n \overset{P}{\to} X$) if, for all $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0 \quad \Leftrightarrow \quad \lim_{n\to\infty} P(|X_n - X| \leq \epsilon) = 1.$$

This convergence of probability values is in measure theory called *convergence in measure.*

Definition: The r.v.s X_n **converges almost surely** to r.v. X ($X_n \overset{a.s.}{\to} X$) if for all $\epsilon > 0$

$$P\left(\lim_{n\to\infty} |X_n - X| \leq \epsilon\right) \equiv P\left(\{\text{all } \omega \in S \text{ such that } \lim_{n\to\infty} |X_n(\omega) - X(\omega)| \leq \epsilon\}\right) = 1$$

In measure theory, *almost everywhere* means a statement holds true for all but a set of measure zero. Thinking of random variables as functions on our sample space, this is just *pointwise convergence of the random variables* except perhaps on some set of measure zero.

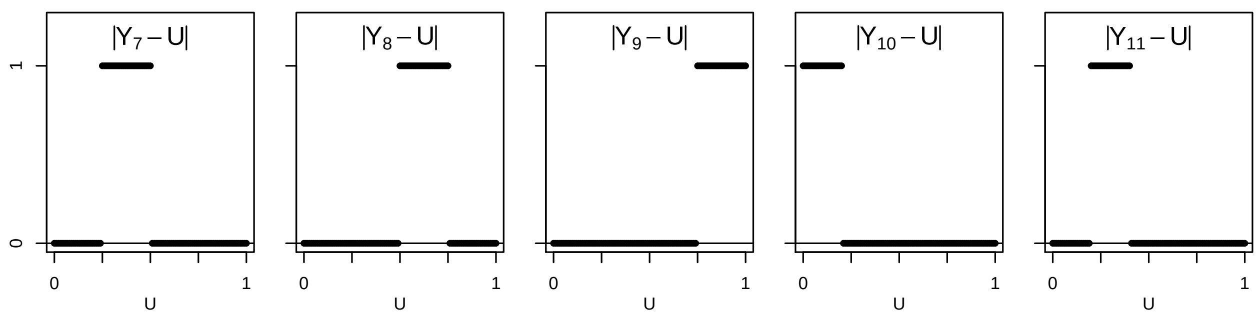
Theorem: If X_n *converges almost surely* to X, then it also *converges in probability.* If X_n *converges in probability* to X, the it also *converges in distribution.*

Example 1 (Convergence in probability, but not almost surely.)

Let U be uniform on [0,1], and define the sequence of random variables Y_n to all depend directly on U according to $Y_n = U + \mathbb{1}_{A_n}(U)$ where intervals A_n are defined as the n^{th} interval in the sequence [0,1/2], [1/2,1], [0,1/3], [1/3,2/3], [2/3,1], [0,1/4],... That is, for observation $U = u$, $Y_n = u + 1$ if $u \in A_n$, otherwise $Y_n = u$. Note these r.v.s Y_n are not independent, since each depends directly on U! Observing that, as $n \to \infty$, the width of interval $A_n \to 0$, it follows that Y_n converges in probability to U since

$$\lim_{n\to\infty} P(|Y_n - U| \geq \epsilon) = \lim_{n\to\infty} P(U \in A_n) = 0.$$

But for a given outcome $U = u$, $Y_n(u)$ never converges since for any $N > 0$ there is always some $k > N$ where $Y_k(u) = 1 + u$.



Since $|Y_n - U|$ converges nowhere on [0,1],

$$P\left(\lim_{n\to\infty} |Y_n - U| \leq \epsilon\right) = P\left(\emptyset\right) = 0 \neq 1$$

That is, there is no almost sure convergence.

Example 2 (Convergence in distribution, but not in probability.)

Let X be a standard Normal r.v. ($E(X) = 0$, $Var(X) = 1$). Let $X_n = -X$ for all n. Then all X_n and X have the same distribution (i.e., $F_{X_n}(x) = F_X(x)$ for all x and n), so trivially X_n *converges in distribution* to X. However, for $\epsilon > 0$, symmetry gives that

$$P(|X_n - X| \geq \epsilon) = P(|2X| \geq \epsilon) = P(|X| \geq \epsilon/2) = P\left(X \notin \left[\frac{-\epsilon}{2}, \frac{\epsilon}{2}\right]\right).$$

Since $P\left(X \notin \left[\frac{-\epsilon}{2}, \frac{\epsilon}{2}\right]\right) > 0$ for all $\epsilon > 0$, it follows that

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = P\left(X \notin \left[\frac{-\epsilon}{2}, \frac{\epsilon}{2}\right]\right) > 0.$$

Therefore X_n does not converge in probability to X.

Weak Law of Large Numbers (WLLN): Let X_i be iid with mean μ. Then $\overline{X_n} = \sum_{i=1}^n X_i$ *converges in probability* to μ, i.e., $\overline{X_n} \xrightarrow{P} \mu$. That is, for all positive ϵ near zero,

$$\lim_{n\to\infty} P(|\overline{X_n} - \mu| \geq \epsilon) = 0 \quad \Leftrightarrow \quad \lim_{n\to\infty} P(|\overline{X_n} - \mu| \leq \epsilon) = 1.$$

Proof (when $Var(X) = \sigma^2 < \infty$): Apply the Chebychev Inequality. This was first proven in the 1700s by Bernoulli, and incrementally generalized by Markov then Chebychev.

Strong Law of Large Numbers (SLLN): Let X_i be iid with mean μ, and let $\overline{X_n} = \sum_{i=1}^n X_i$. Then $\overline{X_n}$ *converges almost surely* to μ. That is, for all positive ϵ near zero,

$$P\left(\lim_{n\to\infty} |\overline{X_n} - \mu| \geq \epsilon\right) = 0 \quad \Leftrightarrow \quad P\left(\lim_{n\to\infty} |\overline{X_n} - \mu| \leq \epsilon\right) = 1.$$

NOTE: Borel gave the first proof of the SLLN, 200 years later, in 1909. It was incrementally improved by Cantelli, Khintchine (who named it the SLLN) and Kolmogorov (in the 1930s).

Weak vs Strong: Accordingly, *almost sure convergence* is called a *stronger* form of convergence than *convergence in probability*, and *convergence in distribution* is even more *weak*.

NOTE: The WLLN and SLLN basically both state that the average of n *iid* random variables (with mean $\mu < \infty$) converges to μ as $n \to \infty$. The Weak LLN states this in the *weaker* form ($X_n \xrightarrow{P} \mu$), while the Strong LLN states this in the (stronger) form ($X_n \xrightarrow{a.s.} \mu$).

CENTRAL LIMIT THEOREM

Classic CLT (Lindberg-Levy): Suppose random variables X_n are *iid* with finite mean μ and finite variance σ^2. Then the quantity

$$S_n = \frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\left(\sum_{i=1}^n X_i\right) - \mu\right)$$

converges *in distribution* to a standard Normal r.v., i.e., $S_n \xrightarrow{D} \mathcal{N}(0,1)$.

NOTE: Other CLTs relax the *iid* assumptions, but require additional conditions that must be met. The Lyapunov CLT, for example, relaxes the assumption of identical distributions:

CLT (Lyapunov): For independent random variables X_n, each with finite mean $E(X_n) = \mu_n$ and variance $Var(X_n) = \sigma_n^2$, define $s_n = \sqrt{\sum_{i=1}^n \sigma_i^2}$. The quantity

$$S_n = \frac{1}{s_n}\sum_{i=1}^n (X_i - \mu_i)$$

converges in distribution a standard Normal *if* the following condition holds for some $\delta > 0$ (usually checking $\delta = 1$ is all it takes):

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E\left(|X_i - \mu_i|^{2+\delta}\right) = 0.$$

NORMAL (GAUSSIAN) DISTRIBUTION

Normal (Gaussian) distribution. Continuous random variable X has a normal distribution with mean μ and variance σ^2 if its pdf is of the form:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

where μ and σ^2 are real valued constants. If X has pdf as above, we denote it: $X \sim N(\mu, \sigma^2)$. The mgf of X is $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$, for any real t.

The normal pdf is bell shaped and centered around the mean μ. There is a special Normal distribution with mean 0 and variance 1, called standard normal distribution, and denoted by $Z \sim N(0, 1)$. The standard normal pdf is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

The values of the standard normal cdf are tabulated. To find probabilities related to general normal random variables, use the following fact:

Theorem. If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Theorem: Linear combinations of independent normal r.v.s are themselves normal.

1. Let $X_1 \sim N(\mu_1, \sigma_1^2)$, and $X_2 \sim N(\mu_2, \sigma_2^2)$, with X_1 and X_2 independent. Then $X_1 \pm X_2 \sim N(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$, and more generally:
2. Let $X_i \sim N(\mu_i, \sigma_i^2)$, for $i = 1, \ldots, n$, and X_i's ind. Then $Y = \sum_{i=1}^n X_i \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$, and
3. For any real numbers $a_1, a_2, \ldots a_n$, $Y = \sum_{i=1}^n a_i X_i \sim N(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$.
4. Let $X_i \sim N(\mu, \sigma^2)$ iid for $i = 1, \ldots, n$. Then $\bar{X} \sim N(\mu, \sigma^2/n)$.

Normal Approximation to Binomial. Let $X \sim Bin(n, p)$ and $Y \sim N(np, np(1-p))$. Then for large n

$$P(a \leq X \leq b) \approx P(a \leq Y \leq b).$$

Continuity correction for the normal approximation to binomial. To "correct" for the fact that binomial is discrete and normal is a continuous distribution, we do the following *correction for continuity*: $P(X = x) \approx P(x - 0.5 < Y < x + 0.5)$.