Lecture 11 / Week 7

Convergence

OUTLINE
1.) In Probability
2.) Almost Surely

Convergence in Probability

Definition Suppose we have a sequence $(\mathrm{X}_n)_{n=1}^{\infty}$ of random variables. Let X be a random variable. We say that $\mathbf{X}_n$ **converges to X in probability** if

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1 \ , \forall \varepsilon > 0.$$

For instance take an event $\{\omega : |X_n - X| < \varepsilon\}$. This event says that X_n is near X. Here the distance is expressed as absolute value. Then for n large enough

$$P(|X_n - X| < \varepsilon) \approx 1$$

Example Toss a coin infinetely many times. Call $F_n :=$ the frequency of heads in the first n trials. It can be proved that F_n converges in probablity to $\frac{1}{2}$. (It can also be shown with a computer simulation.) Fix an $\varepsilon > 0$.(it can be as small as we want) and then take the event

$$\{|F_n - \frac{1}{2}| < \varepsilon\}$$

what the above definition says that

$$P\{\frac{1}{2} - \varepsilon \le F_n \le \frac{1}{2} + \varepsilon\} \ge 0.9999...$$

if n is large enough.

The above definition can be expressed equivalentyly in terms of complement event, i.e.

$$\begin{aligned} \lim_{n\to\infty} P(|X_n - X| &< \varepsilon) = 1 \ \ , \forall \varepsilon > 0 \\ (|X_n - X| < \varepsilon)^c &= (|X_n - X| \ \ge \varepsilon) \end{aligned}$$

Then the equivalent definition will be

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0 \quad ,\forall\varepsilon > 0$$
$$or$$
$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0 \quad ,\forall\varepsilon > 0$$

There are different notation for expressing this concept such as

$$X_n \to^{P} X$$
$$\text{plim}_{n\to\infty} X_n = X$$

Weak Laws of Large Numbers

Suppose we have X_1,X_2........ Then the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ follows

$$\bar{X}_n \to^{P} E(X)$$

under certain assumptions. The following theorem tells us under which assumptions it holds.

Theorem (WLLN for uncorrelated r.v) Let X_1,X_2........ be a sequence of random variables such that $E(X^2) < \infty$ (i.e. second moment is *finite*) for every n, then $E(X_n) = \mu, var(X_n) = \sigma^2$, where μ and σ^2 *do not depend on n.* Moreover let $\mathbf{cov}(\mathbf{X}_n, \mathbf{X}_m) = \mathbf{0}$ $\forall n, m$. Then

$$\bar{X}_n \to^{P} \mu \quad n \to \infty$$

Note that this theorem makes the following assumptions: X_n are uncorrelated and $E(X_n), var(X_n)$ are constant w.r.t n.(independent of n)

Proof The theorem says

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0 \quad ,\forall\varepsilon > 0$$

$$E(\bar{X}_n) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu$$

We will use Chebishev Inequality to prove the result, i.e.

$$\lim_{n\to\infty} P(|\bar{X}_n - E(\bar{X}_n)| > \varepsilon) \leq^{Cheb.} \frac{var(\bar{X}_n)}{\varepsilon^2} = \frac{\frac{\sigma^2}{n}}{\varepsilon^2} \to 0 \quad ,\forall\varepsilon > 0$$

$$var(\bar{X}_n) = var(\frac{1}{n}\sum_{i=1}^{n} X_i) =^{uncorr.} \frac{1}{n^2}\sum_{i=1}^{n} var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}$$

Example Let F_n := the frequency of heads , $F_n := \frac{\text{the number of heads in the first n trials}}{n}$, $F_n = \frac{X_1+X_2+.....+X_n}{n} = \bar{X}_n$.

$$X_i \left\{ \begin{array}{cc} 1 & \text{head at toss i} \\ 0 & \text{tail at toss i} \end{array} \right\}$$

Note that X_i are independent and independence implies no correlation.

$$\begin{aligned}
P(X_i &= 1) = \frac{1}{2}, P(X_i = 0) = \frac{1}{2} \\
E(X_i) &= 1.\frac{1}{2} + 0.\frac{1}{2} = \frac{1}{2} \\
var(X_i) &= \frac{1}{2}(1-\frac{1}{2})^2 + \frac{1}{2}(0-\frac{1}{2})^2 = \frac{1}{4} \\
E(X_n) &= \mu = \frac{1}{2} \\
var(X_n) &= \sigma^2 = \frac{1}{4} \\
& \text{we can apply the theorem, hence} \\
\bar{X}_n &\rightarrow^P \; E(X) = \frac{1}{2}
\end{aligned}$$

Suppose we have $X_1, X_2....$ i.i.d random variables. $E(X_n^2) < \infty$. Then we can apply the previous theorem $\Rightarrow \bar{X}_n \rightarrow^P \mu.(n \rightarrow \infty).$, since X_n satisfy the hypothesis of the theorem. The indepence imply no correlation and since X_n are identically distributed $\Rightarrow E(X_n), var(X_n)$ do not depend on n. In the following theorem we will strive for a stronger result, in the sense that we will not require a finite second moment as an assumption.

Theorem (WLLN for i.i.d r.v) Let X_n be a sequence of *independent and identically distributed* random variables such that $E(|X_n|) < \infty.(integrable)$. This theorem guarantees convergence even though the random variable has infinite variance, i.e.

$$\begin{aligned}
Let \;\; \mu &= E(X_n) \\
Then \;\; \bar{X}_n &\rightarrow {}^P\mu \quad (n \rightarrow \infty)
\end{aligned}$$

Suppose $\bar{X}_n \rightarrow E(X_1)$. $E(X)$ is the limit of the mean of X_n where X_n is a sequence of i.i.d random variables distributed as X. We have shown that the above result holds for arithmetic mean.

Example Dice score E(X)=3.5

We wonder whether we can make any conclusion about the convergence of geometric mean. This is an important consideration since in economics and finance one often needs to use geometric averages. (interest rates, compunded returns in the long term.) Suppose we have a sequence of nonnegative random variables $X_1, X_2, ..X_n \geq 0$ and we consider the geometric mean

$$\begin{aligned} Geometric\ Mean &: (X_1.X_2....X_n)^{\frac{1}{n}} \to ? \\ (X_1.X_2....X_n)^{\frac{1}{n}} &= (e^{\log X_1}.e^{\log X_2}....e^{\log X_n})^{\frac{1}{n}} \end{aligned}$$

Note that we can use this transformation since we have nonnegative random variables.

$$\begin{aligned} (e^{\log X_1}.e^{\log X_2}....e^{\log X_n})^{\frac{1}{n}} &= \left(e^{\sum_{i=1}^n \log X_i}\right)^{\frac{1}{n}} = e^{\frac{1}{n}\sum_{i=1}^n \log X_i} \\ \text{Suppose } \mathrm{X}_i \text{ i.i.d} &\Leftrightarrow \log X_i \text{ i.i.d} \\ \frac{1}{n}\sum_{i=1}^n \log X_i &\to^P E(\log X_1). \\ e^{\frac{1}{n}\sum_{i=1}^n \log X_i} &\to^P e^{E(\log X_1)}. \end{aligned}$$

So we have shown that the result of the theorem also holds in case of geometric mean.

Theorem (Slutsky) Let X_n be a sequence of random variables converging in probability to a nonrandom constant c. Let ψ be a *continuous* function, then

$$\begin{aligned} X_n &\to^P c \\ \psi(X_n) &\to^P \psi(c) \quad n \to \infty \end{aligned}$$

Example $\psi(x) = e^x$. Since the *exponential function* is continuous, then

$$\psi(\frac{1}{n}\sum \log X_i) \to^P \psi(E(\log X_1))$$

Example $\psi(x) = \sqrt{x}$, $x > 0$, since square root is continuous, then

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n X_i^2 &\to^P E(X_1^2) = \mu \\ \sqrt{\frac{1}{2}(X_1^2 + X_1^2 + ...X_n^2)} &\to^P \sqrt{\mu} = \sqrt{E(X_1^2)} \end{aligned}$$

Almost Sure Convergence

Definition Suppose we have a sequence $(X_n)_{n=1}^{\infty}$ of random variables. Let X be a random variable. We say that $\mathbf{X}_n$ **converges almost surely to X** if

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

Since

$$P(\lim_{n\to\infty} \sup |X_n - X| = 0) = 1 \Leftrightarrow P(\lim_{n\to\infty} \sup_{m\geq n} |X_m - X| < \varepsilon) = 1 \text{ }, \forall\varepsilon > 0.$$

In the following we will show the relationship between the convergence in probability and a.s convergence. The following holds

$$\begin{aligned} \text{a.s convergence} \quad &\Rightarrow \quad \text{p.convergence} \\ P(\lim_{n\to\infty} \sup_{m\geq n} |X_m - X| \quad &< \quad \varepsilon) = 1 \text{ }, \forall\varepsilon > 0. \Rightarrow \lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1 \text{ }, \forall\varepsilon > 0. \end{aligned}$$

To see this, consider

$$\sup_{m\geq n} |X_m - X| \geq |X_n - X|$$

then the event

$$\begin{aligned} \sup_{m\geq n} |X_m - X| \quad &< \quad \varepsilon \subset |X_m - X| < \varepsilon \\ P(|X_m - X| \quad &< \quad \varepsilon) \geq P(\sup_{m\geq n} |X_m - X| < \varepsilon) \end{aligned}$$

then we can see that the fact that the supremum is smaller than ε implies that all the distances are smaller than ε, therefore a.s convergence⇒p.convergence, but the reverse is not true. This explains the fact that the strong law of large numbers(**SLLN**)) is based on a.s convergence, whereas weak law of large(**WLLN**) numbers is based on convergence in probability.

Theorem (Kolmogorov SLLN) Let $\{X_n\}$ be a sequence of independent and identically distributed random variables with $E(|X_n|) < \infty$. Let $E(X_1) = \mu$ and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then

$$\bar{X}_n \to \mu \quad \textbf{a.s}$$

Proof Complicated, hence omitted.
Example Let $F_n :=$ the frequency of heads, then

$$\begin{aligned} F_n \quad &\to \quad \frac{1}{2} \text{ a.s} \\ F_n \quad &= \quad \frac{1}{n}\sum_{i=1}^{n} X_i, \text{ } X_i \text{ i.i.d} \end{aligned}$$

In the following thereom we will see that Slutsky theorem carries over to a.s convergence, it is even stronger since $X_n \to c$ is not required.

Theorem (Slutsky for a.s convergence) Let $X_n \to X$ **a.s** and let ψ be a continuous function. Then

$$\psi(X_n) \to \psi(X) \quad a.s$$

Proof Take a set M=$\{\omega : X_n(\omega) \to X(\omega)\}$. By a.s. hypothesis $P(M) = 1$.

$$\begin{array}{rcl} \omega & \in & M : X_n(\omega) \to X(\omega) \quad n \to \infty \\ \psi - cont. & : & \psi(X_n(\omega)) \to \psi(X(\omega)) \;\; n \to \infty \\ \omega & \in & M \ then \ \psi(X_n(\omega)) \to \psi(X(\omega)) \;\; n \to \infty \\ \psi(X_n(\omega)) & \to & \psi(X(\omega)) \;\; (n \to \infty) \text{ a.s} \end{array}$$

Generalizations

Within the WLLN we saw to cases:

1. $\mathbf{X}'_n s$ **are uncorrelated**: This case can be generalized to *weakly correlated* $X'_n s$. $\Rightarrow$ **weakly stationary processe**s
2. $\mathbf{X}'_n s$ **are i.i.d**: This case can also be generalized to *weakly dependent* $X'_n s$. $\Rightarrow$ **strictly stationary processes**.

Definition A **time series process** is a sequence of random variables $X_t(t = 1, 2, 3....)$ where the random variables are indexed by time.

Definition A sequence of random variables X_t is called **strictly stationary** if *for every* **m**, the probability distribution of $(X_t, X_{t+1}, ...X_{t+n})$ does not depend on t.

$$\begin{array}{rcl} m & = & 1 : prob.\ dist\ X_1 = p.d\ X_2 = p.d\ X_3 \\ m & = & 2 : prob.\ dist\ (X_1, X_2) = p.d\ of\ (X_2, X_3) = p.d\ of\ (X_3, X_4) \end{array}$$

Example X_n are i.id. $\Rightarrow X_n$ is strictly stationary

$$f_{X_t, X_{t+1}....X_{t+m}}(x_0.....x_m) \;=\; \overset{(X_t\ X_{t+1}.....X_{t+m})}{\prod_{i=0}^{m} f(x_i)}$$

so the the distribution does not depend on t. What matters is the time lag.(m)

Example Suppose Z_n are i.i.d and $X_n = Z_n + Z_{n+1}$

X_n is strictly stationary $(X_t, X_{t+1}, ...X_{t+m})$=$(Z_t$+$Z_{t+1}, Z_{t+1}$+$Z_{t+2}....., Z_{t+n}$+$Z_{t+n+1})$

If $\{X_n\}$ is strictly stationary, than $X_1, X_2....$ are identically distributed. The probability distribution of X_t does not depend on t.

$$\begin{aligned} i.i.d &\Rightarrow strict\ stationarity \\ strict\ stationarity &\Rightarrow identical\ distribution \end{aligned}$$

So note that, strict stationarity does not imply independence.

Definition A sequence X_t of random variables is **weakly stationary** if $E(X_t^2) < \infty\ \forall t$ and $E(X_t) = \mu$, $cov(X_t, X_{t+m}) = \gamma(m)(m = 0, 1, ...)$.(Note that m=0$\Rightarrow$ *variance*)
when μ and $\gamma(m)$ do not depend on t.

Example Suppose we have weakly stationary process, then

$$\begin{aligned} m &= 1 : E(X_1) = E(X_2) = E(X_3)... \\ m &= 1 : var(X_1) = Var(X_2) = Var(X_3)..... \\ m &= 1 : cov(X_1, X_2) = cov(X_2, X_3).... = cov(X_{17}, X_{18}) \\ m &= 0 : cov(X_t, X_t) = var(X_t) \\ m &= 2 : cov(X_1, X_3) = cov(X_{14,} X_{16}) \end{aligned}$$