Different notions of convergence for Random Variables

Let $X_1, X_2, X_3, \ldots, X_n, \ldots$ be a sequence of random variables, with cumulative distrib. fct $F_1, F_2, F_3, \ldots, F_n, \ldots$, i.e.

$$F_i(x) = P(X_i \le x).$$

1. Convergence in distribution

$$X_n \xrightarrow{d} X_\infty \quad \text{if} \quad \lim_{n\to\infty} F_n(x) = F_\infty(x) \quad \text{at all } x$$

where the cumulative distr. fct $F_\infty(x)$ of X_∞ is continuous.

i.e. $P(X_n \le x) \xrightarrow{n\to\infty} P(X_\infty \le x)$ for all x where F_∞ is cont.

This is the convergence in the statement of CLT:

$$\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim \text{Normal}(0,1)$$

2. Convergence in probability

$$X_n \xrightarrow{P} X_\infty \text{ if for all } \varepsilon > 0, \quad \lim_{n\to\infty} P(|X_n - X_\infty| \ge \varepsilon) = 0.$$

This is the convergence in the statement of Weak LLN:

$$\overline{X_n} \xrightarrow{P} \mu$$

3. Almost sure convergence

$$X_n \xrightarrow{a.s.} X_\infty \quad \text{if} \quad P\left(\lim_{n\to\infty} X_n = X_\infty\right) = 1.$$

This is the convergence in the (upcoming) statement of Strong LLN.

Q: How do they compare?

$$X_n \xrightarrow{a.s.} X_\infty \implies X_n \xrightarrow{P} X_\infty \implies X_n \xrightarrow{d} X_\infty.$$

(strongest) (weakest)

Strong Law of Large Numbers (Kolmogorov)

Let $X_1, X_2, X_3, \ldots$ be a sequence of iid random variables, with common mean $E(X_i) = \mu < \infty$. Then, as $n \to \infty$,

$$\overline{X_n} := \frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{a.s.} \mu$$

i.e., $P\left(\lim_{n\to\infty} \overline{X_n} = \mu\right) = 1.$

Pf: Let us make the extra assumption that $E(X_i^4) < \infty$; let $K := E(X_i^4)$. Define $S_n = X_1 + X_2 + \cdots + X_n$, i.e.,

$$S_n = n \cdot \overline{X_n}, \text{ or } \overline{X_n} = \frac{S_n}{n}.$$

Multinomial Thm

$$E(S_n^4) = E\left((X_1 + \cdots + X_n)^4\right) =$$ Sum of terms of the form

(*) $X_i^4,\ X_i^3 X_j,\ X_i^2 X_j^2,\ X_i^2 X_j X_k,\ X_i X_j X_k X_\ell$

$1 \le i, j, k, \ell \le n$; different

Suppose $\mu = E(X_i) = 0$. Then, since $E(\cdot)$ is linear and, $E(XY) = E(X)E(Y)$ if X and Y are independent, we have:

$$E(X_i^3 X_j) = E(X_i^3)\underbrace{E(X_j)}_{=\mu=0} = 0$$

$$E(X_i^2 X_j X_k) = E(X_i^2)\underbrace{E(X_j)}_{=\mu=0}\underbrace{E(X_k)}_{=\mu=0} = 0.$$

$$E(X_i X_j X_k X_\ell) = \underbrace{E(X_i)}_{=\mu=0}\underbrace{E(X_j)}_{=\mu=0}\underbrace{E(X_k)}_{=\mu=0}\underbrace{E(X_\ell)}_{=\mu=0} = 0.$$

However $E(X_i^4)$ and $E(X_i^2 X_j^2)$ need not vanish. There are

n terms of the form $E(X_i^4)$ in $(*)$

$\underbrace{\binom{n}{2}\binom{4}{2}}_{= 6\binom{n}{2} = 6\frac{n(n-1)}{2} = 3n(n-1)}$ ——— " ——— $E(X_i^2 X_j^2)$ in $(*)$

$$(*) = E(S_n^4) = n.\underbrace{E(X_i^4)}_{K} + 3n(n-1)E(X_i^2 X_j^2)$$

$$= nK + 3n(n-1)E(X_i^2 X_j^2).$$

Since $0 \le Var(X_i^2) = E((X_i^2)^2) - (E(X_i^2))^2 = E(X_i^4) - E(X_i^2)^2$

we have $E(X_i^2)^2 \le E(X_i^4) = K$, thus

$$\textcircled{*} = E(S_n^4) \overset{\text{indep.}}{=} nK + 3n(n-1)\underbrace{E(X_i^2)E(X_j^2)}_{\le K} \le (n + 3n(n-1))K$$

$$\le (n+3n^2)K$$

$$E\left(\frac{S_n^4}{n^4}\right) \le \frac{n+3n^2}{n^4}K = \left(\frac{1}{n^3} + \frac{3}{n^2}\right)K$$

$$\| $$

$$E\left(\left(\frac{S_n}{n}\right)^4\right) = E\left(\overline{X}_n^4\right)$$

Sum over all $n \in \mathbb{N}$:

This series converges b/c it is a p-Series w/ $p > 1$.

$$E\left(\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4\right) = \sum_{n=1}^{\infty} E\left(\left(\frac{S_n}{n}\right)^4\right) \le \sum_{n=1}^{\infty}\left(\frac{1}{n^3} + \frac{3}{n^2}\right)K < \infty$$

Thus, $\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4 < \infty$ with probability 1, since if

$P\left(\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4 < \infty\right) < 1$, then $P\left(\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4 = +\infty\right) > 0$ and

hence $E\left(\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4\right) = +\infty$. In particular, then also

$\left(\frac{S_n}{n}\right)^4 \longrightarrow 0$ with probability 1, i.e. $\overline{X}_n = \frac{S_n}{n} \longrightarrow \underset{\mu}{\overset{\|}{0}}$ w/ prob. 1.

Thus, $P\left(\lim_{n\to\infty} \overline{X_n} = 0\right) = 1$. For the case of $\mu \neq 0$, consider $\tilde{X}_n := X_n - \mu$ and apply the above reasoning to $\tilde{X}_n$ to conclude, since $\overline{\tilde{X}_n} = \overline{X_n - \mu} = \overline{X_n} - \mu$ that

$$P\left(\lim_{n\to\infty} \overline{X_n} = \mu\right) = P\left(\lim_{n\to\infty} (\overline{X_n} - \mu) = 0\right) = P\left(\lim_{n\to\infty} \overline{\tilde{X}_n} = 0\right) = 1.$$

$\square$

Main differences between <u>weak</u> and <u>strong</u> LLN:

<u>Weak LLN</u>	<u>Strong LLN</u>
$\overline{X_n} \xrightarrow{P} \mu$, i.e.	$\overline{X_n} \xrightarrow{a.s.} \mu$, i.e.,
$\forall \varepsilon > 0$, $\lim_{n\to\infty} P\left(\lvert \overline{X_n} - \mu \rvert \geq \varepsilon\right) = 0$	$P\left(\lim_{n\to\infty} \overline{X_n} = \mu\right) = 1$.
The event $\lvert \overline{X_n} - \mu \rvert \geq \varepsilon$ could happen an infinite number of times, although at infrequent intervals	The event $\lvert \overline{X_n} - \mu \rvert \geq \varepsilon$ can only happen a finite number of times.

Recall:

Markov's inequality: if $X \geq 0$, then for all $a > 0$, $P(X \geq a) \leq \frac{E(X)}{a}$.

Chebyshev's inequality: if $\mu = E(X)$, $\sigma^2 = Var(X)$ are finite, then

$$P(X - \mu \geq a \text{ or } X - \mu \leq -a) = P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

Suppose we want to find an upper bound for $P(X - \mu \geq a)$.

$$P(X - \mu \geq a) \leq P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}. \qquad (\text{if } a > 0)$$

But we can actually find a much better bound:

One-sided Chebyshev's inequality: $\sigma^2 = Var(X) = E(X^2) - \cancel{E(X)^2}^{\,0} = E(X^2)$

Suppose X has $\mu = E(X) = 0$ and $\sigma^2 = Var(X) < \infty$. Then for all $a > 0$,

$$P(X \geq a) \leq \frac{\sigma^2}{a^2 + \sigma^2}.$$

Pf. If $a > 0$, then $X \geq a \Leftrightarrow X + b \geq a + b$, so if $b > 0$,

$$P(X \geq a) = P(X + b \geq a + b) \leq P\left((X+b)^2 \geq (a+b)^2\right) \overset{\text{Markov}}{\leq} \frac{E((X+b)^2)}{(a+b)^2} =$$

$$= \frac{E(X^2 + 2bX + b^2)}{(a+b)^2} = \frac{\cancel{E(X^2)}^{\,\sigma^2} + 2b\,\cancel{E(X)}^{\,0} + b^2}{(a+b)^2} = \frac{\sigma^2 + b^2}{a^2 + b^2}.$$

Let $b = \sigma^2/a$, we obtain the claimed inequality.

Ex: If an apiary produces, on average, 100 pounds of honey per year, with variance 400, find an upper bound for the probability that next year's production excedes 120 pounds.

$X =$ production (in pounds) $\qquad \mu_X = E(X) = 100$

$\sigma_X^2 = Var(X) = 400$

$$P(X \geq 120) = P(\underbrace{X - 100}_{\tilde{X} = X - 100} \geq \underbrace{120 - 100}_{20}) \leq \frac{(\sigma_{\tilde{X}})^2}{(20)^2 + (\sigma_{\tilde{X}})^2} = \frac{400}{800} = \frac{1}{2}.$$

now $E(\tilde{X}) = 0$.

One-sided Chebyshev to $\tilde{X}$

b/c: $\sigma_{\tilde{X}}^2 = Var(\tilde{X}) = Var(X - 100) = Var(X) = 400$

Thus: $\boxed{P(X \geq 120) \leq \frac{1}{2}}$ $= 50\%$

Note: Compare with what Markov's inequality would give:

$$P(X \geq 120) \leq \frac{E(X)}{120} = \frac{100}{120} = \frac{5}{6} = 83.33\%$$

Much worse than $\leq 50\%$!

There are other useful bounds involving Moment generating functions, known as Chernoff bounds.