Continue with Convergence in Probability

We say that $X_n \to_p X$, if

for any given $\varepsilon > 0$,

$$\lim_{n \to \infty} P(|X_n - X| > \varepsilon) = 0$$

The use of this notion in statistics is primarily to ensure that as sample size (n) increases, your estimate T_n (a function of the available data) approaches the true unknown parameter θ.

What we want as a fundamental property of an estimator is

$$\to P(|\underbrace{T_n}_{\text{estimator}} - \underbrace{\theta}_{\text{parameter}}| > \varepsilon) \to 0 \quad \text{as } n \to \infty$$

$\{X_n\}$]

WLLN (weak law of large numbers)

If $X_1, X_2, X_3, \ldots$ are i.i.d random variables with finite mean μ and finite variance $\sigma^2 > 0$

then $\overline{X}_n = \dfrac{X_1 + \ldots + X_n}{n} \xrightarrow[p]{} \mu$.

i.e $P(|\overline{X}_n - \mu| > \varepsilon) \to 0$, whatever be ε

so $\overline{X}_n \to_p \mu \to$ population average

$\overline{X}_n$: empirical average

i.i.d: independent and identically distributed.

[But this can be relaxed]

Markov's Inequality:

$X \geq 0$

We are interested in bounds of the form $P(X > t)$, t is a +ve number

Tail events

Take any $g : [0,\infty) \to [0,\infty)$.

$$\mathbb{E}[g(X)] \geq \mathbb{E}[g(X)\,\mathbb{1}(X \in A)]$$

for any event A

$$\mathbb{E}[g(X)\,\mathbb{1}(X \in A)] \geq \left(\inf_A g(u)\right) P(X \in A)$$

$$g(X)\,\mathbb{1}(X \in A) \geq \inf_{u \in A} g(u)\,\mathbb{1}(X \in A)$$

taking expectation of either side gives

So:

$$P(X \in A) \leq \boxed{\frac{\mathbb{E}[g(X)]}{\inf g(A)}} \quad (*)$$

$$\inf g(A) = \inf\{g(x) : x \in A\}$$

$\boxed{\text{range of } A \text{ under } g.}$

$(*)$ is true for any $g : [0,\infty) \to [0,\infty)$

and any Borel set $A \subseteq \mathbb{R}$.

Important application of $(*)$ is when g is non-decreasing i.e

$x \leq y \Rightarrow g(x) \leq g(y)$ and

$A = [t, \infty)$ for some $t > 0$

$$P[X \in A] = \underline{P[X \geq t]} \leq \frac{\mathbb{E}[g(X)]}{\inf_{u \geq t} g(u)}$$

so $P[X \geq t] \leq \dfrac{\mathbb{E}[g(X)]}{g(t)}$

Markov's inequality

Chebyshev's inequality:

Bounds $P(|Y - \xi| \geq \eta)$ for some $\eta > 0$, where $\mathbb{E}Y = \xi$

$\mathrm{Var}(Y) = \sigma^2$.

Take your X in Markov's inequality to be $|Y - \xi|$.

Take: $g(t)$: t^2 from $[0,\infty)$ to $[0,\infty)$.

Then: (by Markov)

$$P(|Y - \xi| \geq \eta) \leq \frac{\mathbb{E}[(Y-\xi)^2]}{\eta^2}$$

Chebyshev's Inequality $= \dfrac{\sigma^2}{\eta^2}$

Proof of WLLN:

$X_1, \dots, X_n$ are i.i.d
$\mathbb{E}X_i = \mu$
$\mathrm{Var}\, X_i = \tau^2$

look at:

$$P[\,|\bar{X}_n - \mu| > \varepsilon\,]$$

Take $Y = \bar{X}_n$ in Chebyshev.
$\eta = \varepsilon$.

Now $\xi = \mathbb{E}\bar{X}_n = \mu$.

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\tau^2}{n}.$$

So $\sigma^2 = \dfrac{\tau^2}{n}$

By C.I:

$$P(|\bar{X}_n - \mu| > \varepsilon) \le \frac{\mathrm{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\tau^2}{n\varepsilon^2} \to 0$$

So $\bar{X}_n \xrightarrow[P]{} \mu$.

In general,

suppose you have a sequence

$$\left(X_1, X_2, X_3, \ldots \ldots \right).$$

look at $\left\{ \bar{X}_n = \dfrac{X_1 + \cdots + X_n}{n} \right\}$.

$$\mathbb{E}(\bar{X}_n) = \frac{\mathbb{E}X_1 + \cdots + \mathbb{E}X_n}{n} = \bar{\mu}_n.$$

look at:

$$P(|\bar{X}_n - \bar{\mu}_n| > \varepsilon)$$

$$\leq \frac{Var(\bar{X}_n)}{\varepsilon^2}$$

$$= \frac{Var(X_1 + \cdots + X_n)}{n^2 \varepsilon^2}$$

$$= \frac{\sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j)}{n^2 \varepsilon^2}.$$

So:

$$P(|\bar{X}_n - \bar{\mu}_n| > \varepsilon)$$

$$\leq \frac{\sum_{i=1}^{n} Var(X_i)}{n^2 \varepsilon^2} + \frac{\sum_{i \neq j} Cov(X_i, X_j)}{n^2 \varepsilon^2}$$

In order for $\bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0$

I want the right side of the inequality to go to 0.

Now assume $Var(X_i) \leq \xi$ for all i

Then: $$\frac{\sum_{i=1}^{n} Var(X_i)}{n^2 \varepsilon^2}$$

$$\leq \frac{n\xi}{n^2 \varepsilon^2} = \frac{\xi}{n \varepsilon^2} \longrightarrow 0$$

$$\frac{\sum_{i \neq j} \underline{Cov(x_i, x_j)}}{n^2 \varepsilon^2}$$

$$\underline{|Cov(x_i, x_j)|} \le \sqrt{Var(x_i) Var(x_j)}$$

$$\le \xi$$

What does the above bound give?

$$\frac{\sum_{i \neq j} Cov(x_i, x_j)}{n^2 \varepsilon^2} \le \frac{\xi n(n-1)}{n^2 \varepsilon^2} \not\to 0.$$

However if

$$\sum_{\substack{i \neq j \\ 1 \le i, j \le n}} Cov(x_i, x_j)$$

grows at a rate slower than n^2, then

$$\frac{\sum_{i \neq j} Cov(x_i, x_j)}{n^2 \varepsilon^2} \xrightarrow{d} 0.$$

If I tell you that $\sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$ is of the order n^{λ} where $\lambda < 2$, then of course the second term $\longrightarrow 0$!

and $\bar{X}_n - \bar{\mu}_n \longrightarrow_p 0$.

Suppose X_i is only dependent on X_{i-1} and X_{i+1} and $\mathrm{Var}(X_i) \leq \xi$ for all i

Then check ~~that~~:

$$\frac{\left.\sum_{i \neq j} \mathrm{Cov}(X_i, X_j)\right\}}{n^2} \longrightarrow 0$$

$$\left.\begin{array}{l} \underline{X_1 \longleftrightarrow X_2} \\ X_1 \longleftarrow X_2 \longrightarrow X_3 \\ X_2 \longleftarrow X_3 \longrightarrow X_4 \end{array}\right\}$$

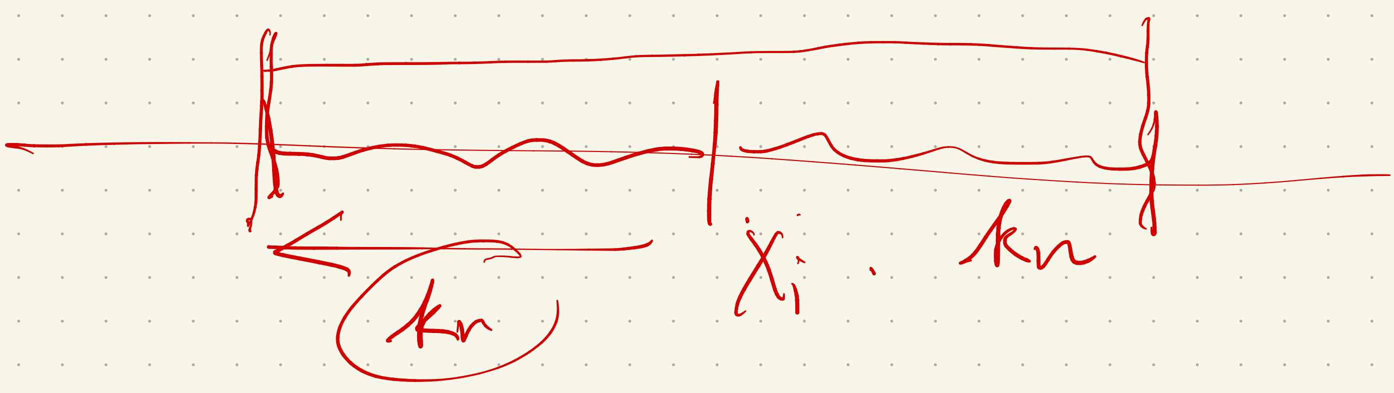
X_4 uncorrelated to $\underline{X_1}, X_2, \underline{X_6}, \ldots$

The same thing happens if X_i just depends on k neighbors to the left and k to the right, k fixed.

In fact, check that

$$\frac{\sum\limits_{1 \le i \ne j \le n} \mathrm{Cov}(X_i, X_j)}{n^2 \varepsilon^2} \to 0$$

even if X_i only depends on k_n neighbors to the left and k_n neighbors to the right, provided $\frac{k_n}{n} \to 0$



Check that the WLLN shows that $\hat{p}_D$: the estimated proportion of D's in SRSWR does converge in probability to p_D, the proportion of D's in the population!

So we now know:

$$\bar{X}_n - \mu \xrightarrow{P} 0 \; !$$

Can we say anything as to how the fluctuation $\bar{X}_n - \mu$ behaves?

Yes, we know this is approximately normal! Central Limit Theorem tells us this.

Useful for constructing confidence intervals for μ and other parameters in statistical problems.

CLT Statement:

Let $X_1, X_2, X_3 \ldots$ be iid random variables with $\mathbb{E}X_i = \mu$, $\mathrm{Var}(X_i) = \sigma^2 > 0$.

Let $S_n = X_1 + X_2 + \cdots + X_n$

$$\bar{X}_n = \frac{S_n}{n}.$$

Then:
$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n - \mathbb{E}S_n}{\sqrt{\mathrm{Var}\,S_n}}$$

$$\parallel$$

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - \mathbb{E}\bar{X}_n}{\sqrt{\mathrm{Var}(\bar{X}_n)}}$$

and $\dfrac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z$ where

$Z \sim N(0,1)$!

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) - \Phi(t) \longrightarrow 0 \text{ for every } t.$$

Restricted proof: We'll assume that the mgf of X_1 (which is also the mgf of all other X_i's) exists finitely in a nbhd $(-h,h)$ of 0.

We're going to show that the mgf of $\frac{S_n - n\mu}{\sigma\sqrt{n}} \equiv \widetilde{S}_n$ converges to the mgf of Z.

$$M_Z(t) = \mathbb{E}\left[e^{tZ}\right] = e^{\frac{1}{2}t^2}$$

$$\widetilde{S}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i, \text{ where } Y_i = \frac{X_i - \mu}{\sigma}.$$

Notice $Y_1, Y_2, \ldots$ are i.i.d

$\mathbb{E}Y_i = 0$, $\mathrm{Var}(Y_i) = 1$. $\mathrm{Var}\,Y_i = \mathbb{E}Y_i^2$

and $M_{Y_1}(t) = \mathbb{E}\left[e^{tY_1}\right]$ also exists for all $t \in (-h,h)$.

So I only need to show that:

$$M_{\tilde{S}_n}(t) \to e^{\frac{1}{2}t^2}.$$

$$M_{\tilde{S}_n}(t) = \mathbb{E}\left[\exp\left(\frac{t\cdot(Y_1+\dots+Y_n)}{\sqrt{n}}\right)\right]$$

$$= \mathbb{E}\left[\prod_{i=1}^{n} \exp\left(\frac{tY_i}{\sqrt{n}}\right)\right]$$

$$= \prod_{i=1}^{n} \mathbb{E}\left(e^{\frac{tY_i}{\sqrt{n}}}\right)$$

$$= \left(\mathbb{E}\left(e^{\frac{tY_1}{\sqrt{n}}}\right)\right)^n = \underline{M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)}^n$$

Expand $M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)$ around 0 up to quadratic with remainder.

$$M_{Y_1}\left(\frac{t}{\sqrt{n}}\right) = M_{Y_1}(0) + \frac{t}{\sqrt{n}} M_{Y_1}'(0) + \frac{t^2}{2n} M_Y''(0) + \delta_n.$$

where $\delta_n = O\left(\left(\frac{t}{\sqrt{n}}\right)^3\right)$

$$= O\left(\frac{1}{n\sqrt{n}}\right) = o\left(\frac{1}{n}\right)$$

i.e $n\delta_n \to 0$.

so: $M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)^n$

$$= \Big[1 + \underbrace{\frac{t}{\sqrt{n}} M_{Y_1}'(0)}_{=0} + \frac{t^2}{2n} M_{Y_1}''(0) + \delta_n\Big]^n$$

$M_{Y_1}'(0) = \mathbb{E}Y_1 = 0$

$M_{Y_1}''(0) = \mathbb{E}Y_1^2 = 1$

$$= \left[1 + \frac{t^2}{2n} + \delta_n\right]^n$$

$$= \Big[1 + \frac{1}{n}\underbrace{\left(\frac{t^2}{2} + n\delta_n\right)}_{= x_n}\Big]^n \to e^{t^2/2}$$

So $x_n \to \frac{t^2}{2}$

$\boxed{\text{if } x_n \to x \quad \left(1 + \frac{x_n}{n}\right)^n \to e^x}$