Convergence almost surely $X_n \xrightarrow{a.s.} X$

$$\mathbb{P}(\{s \in S \mid X_n(s) \to X(s)\}) = 1$$

convergence in probability $X_n \xrightarrow{P} X$

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

distributional convergence $X_n \xrightarrow{d} X$

$$\forall x \quad F_{X_n}(x) \longrightarrow F_X(x)$$

Theorem: Algebraic Properties of Convergence

For seq of numbers $x_n, y_n \in \mathbb{R}$; $x_n \to x$, $y_n \to y$

$\to x_n + y_n \longrightarrow x + y$

$\to x_n y_n \longrightarrow xy$

$\to x_n / y_n \longrightarrow x/y$

$\to a x_n + b y_n \to ax + by$

If $X_n \to X$, $Y_n \to Y$ then for $a, b \in \mathbb{R}$

① $aX_n + bY_n \to aX + bY$

② $X_nY_n \to XY$ $\quad (X_n/Y_n \to X/Y)$

true for convergence a.s. or convergence i.p.

not generally true for convergence in dist.

---

note: treat a seq of $c_n \in \mathbb{R}$ as a seq. of degenerate r.v.s.

So if $c_n \to c$ (as numbers)

then $c_n \xrightarrow{a.s.} c$ (as RVs)

(hence $c_n \xrightarrow{p} c$, $c_n \xrightarrow{d} c$)

---

punchline: $c_n \to c$ and $X_n \to X$ (a.s., p)

then ① $aX_n + bc_n \to aX_n + bc$

② $c_nX_n \to cX$.

What about convergence in dist?

Slutsky's Theorem If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ const.

then (1) $X_n + Y_n \xrightarrow{d} X + c$

(2) $X_n Y_n \xrightarrow{d} Xc$

$(X_n / Y_n \xrightarrow{d} X/c)$

Theorem: Continuous Mapping Theorem

If $g: \mathbb{R} \to \mathbb{R}$ is a continuous function and $X_n \longrightarrow X$ (for any type of convergence) then

$$g(X_n) \longrightarrow g(X).$$

---

proof? Aside: in real analysis we define a cts function as a function that preserves limits: $x_n \to x$ then $g(x_n) \to g(x)$.

Ex. If $Y_n \to Y$ and $Y_n > 0$ and $Y > 0$ then cts mapping theorem says $g(y) = 1/y$ is continuous for $y > 0$ hence

$$g(Y_n) \to g(Y)$$

i.e. $1/Y_n \to 1/Y$.

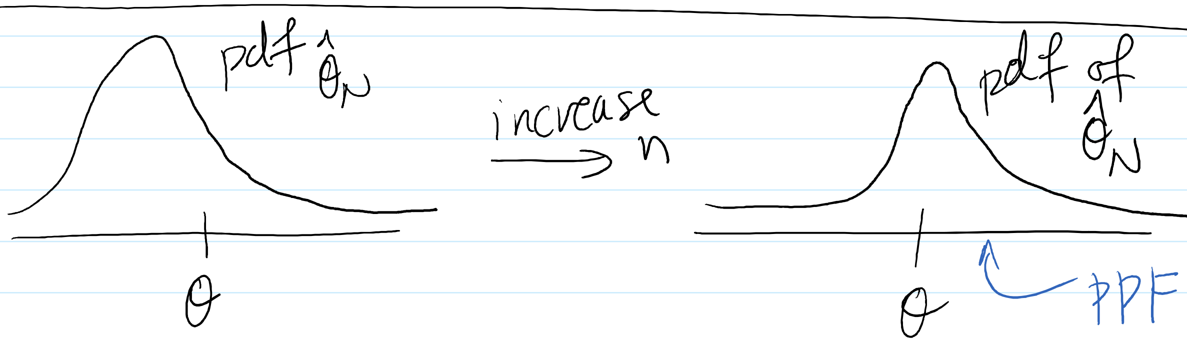
means: If $X_n \to X$ then $X_n / Y_n \to X/Y$.

Defn: Consistent Estimator

We say an estimator $\hat{\theta}_N$ is consistent for θ if $\hat{\theta}_N \xrightarrow{P} \theta$.

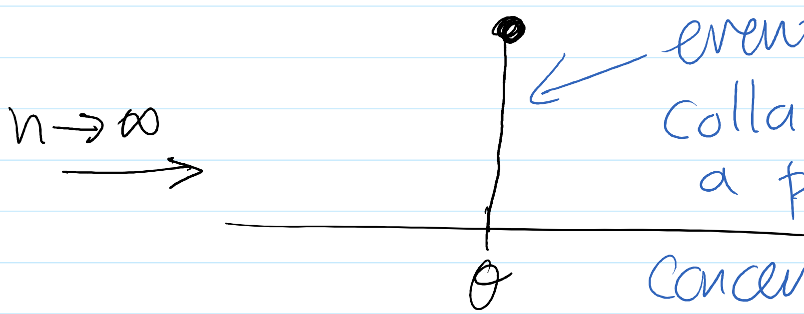
based on N samples

a constant.



PDF tighter around θ

$$\lim_{n\to\infty} \mathbb{P}(|\hat{\theta}_N - \theta| \geq \varepsilon) = 0$$



eventually collapse to a point dist concentrated on θ.

Another way: consistency = asymptotically unbiased

(($*$) for regular ...)

($\smile$ 'enough dists)

$$S^2 = \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \bar{X})^2$$, Know $\mathbb{E}[S^2] = \sigma^2$

(unbiased)

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}(X_n - \bar{X})^2$$, $E[\hat{\sigma}^2] = \frac{N-1}{N}\sigma^2$

(not unbiased)

however $\mathbb{E}[\hat{\sigma}^2] \xrightarrow{N} \sigma^2$

Theorem: MSE $\to 0$ then $\hat{\theta}$ consistent

If $\mathrm{MSE}(\hat{\theta}) \xrightarrow{N} 0$

then $\hat{\theta} \xrightarrow{P} \theta$. (i.e. $\hat{\theta}$ consistent for θ)

---

pf Show consistency:

$\forall \varepsilon \quad \mathbb{P}(|\hat{\theta} - \theta| \geq \varepsilon) \xrightarrow{N} 0$

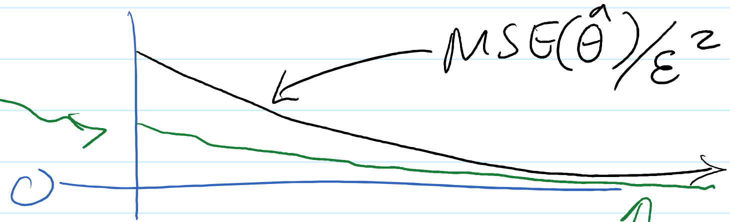
Markov's Ineq: $X \geq 0 \quad \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$.

$0 \leq \mathbb{P}(|\hat{\theta} - \theta| \geq \varepsilon) = \mathbb{P}(\underbrace{(\hat{\theta} - \theta)^2}_{X} \geq \underbrace{\varepsilon^2}_{a})$

$$0 \le \underbrace{\mathbb{P}(|\hat{\theta}-\theta| \geq \varepsilon)}_{(*)} = \mathbb{P}(\overbrace{(\hat{\theta}-\theta)^2}^{X} \geq \overbrace{\varepsilon^2}^{a})$$

$$\leq \frac{\mathbb{E}[(\hat{\theta}-\theta)^2]}{\varepsilon^2} = \frac{MSE(\hat{\theta})}{\varepsilon^2}$$

if $MSE(\hat{\theta}) \xrightarrow{N} 0$ then



squeeze theorem says...

$$\mathbb{P}(|\hat{\theta}-\theta| \geq \varepsilon) \xrightarrow{N} 0$$

i.e. $\hat{\theta} \xrightarrow{P} \theta$.

Intuition: $\bar{X}_N = \frac{1}{N}\sum_{n=1}^{N} X_n$ this should be a good approx of $\mu = E[X_n]$.

Theorem: Weak Law of Large Numbers (WLLN)

We have some X_ns that are uncorrelated and

(1) $E[X_n] = \mu$

(2) $Var(X_n) = \sigma^2 < \infty$

then $\bar{X} \quad \frac{1}{N} \sum^{N} X \quad P$

then $\boxed{\bar{X}_N = \frac{1}{N}\sum_{n=1}^{N} X_n \xrightarrow{P} \mu.}$

Called "weak" b/c its convergence in prob.

also has weaker requirements.

$\boxed{\text{Summary: WLLN says } \bar{X}_N \xrightarrow{P} \mu.}$

Pf: Let X_n are iid and $E[X_n] = \mu$,

$Var(X_n) = \sigma^2 < \infty$.

Notice that $E[\bar{X}_N] = \mu$ and $Var(\bar{X}_N) = \sigma^2/N$

by Chebyshev's Ineq.

$$\boxed{\mathbb{P}(|X-\mu| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}}$$

$$0 \leq \underbrace{P(|\bar{X}-\mu| \geq \varepsilon)}_{} \leq \frac{Var(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{N\varepsilon^2} \xrightarrow{N} 0$$

hence by Squeeze theorem

$$P(|\bar{X}-\mu| \geq \varepsilon) \xrightarrow{N} 0$$

hence $\bar{X} \xrightarrow{P} \mu$.

Ex. Let $X_n \overset{iid}{\sim} \text{Pois}(\lambda)$

$E[X_n] = \lambda$ and $\text{Var}(X_n) = \lambda < \infty$.

WLLN: $\bar{X}_N \xrightarrow{P} \lambda$.

More interesting relaxation

Assume that $\text{Var}(X_n) = \sigma_n^2$ but

$$\underline{\frac{1}{N}\sum_{n=1}^{N} \sigma_n^2 < \infty} \quad \forall N$$

can derive a similar rule b/c key Chebyshev's

$$P(|\bar{X}-\mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X})}{\varepsilon^2}$$

now: $\text{Var}(\bar{X}) = \frac{1}{N^2}\sum_{n=1}^{N} \sigma_n^2$

$$= \frac{1}{N} \frac{\left(\frac{1}{N}\sum_{n=1}^{N} \sigma_n^2\right)}{\varepsilon^2}$$

$< \infty$

$$= \frac{1}{N} \frac{\left(\frac{1}{N} \sum_{n=1} \sigma_n\right)}{\varepsilon^2} \xrightarrow{N} 0$$

const

here $\bar{X}_N \xrightarrow{P} \mu$.

<u>Consider:</u> $S^2 = \frac{1}{N-1} \sum_{n=1}^{N} (X_n - \bar{X})^2$

if X_n are independent and $E[X_n] = \mu$

$Var(X_n) = \sigma^2$

<u>Can show $S^2 \xrightarrow{P} \sigma^2$?</u> [$S^2$ consistent for σ^2]

<u>Chebyshev's:</u> $\boxed{P(|Y - E[Y]| \geq \varepsilon) \leq \frac{Var(Y)}{\varepsilon^2}}$

$$P(|S^2 - \sigma^2| \geq \varepsilon) \leq \frac{Var(S^2)}{\varepsilon^2}$$

<u>$E[S^2] = \sigma^2$</u>

If <u>$Var(S^2) \xrightarrow{n} 0$</u>

then we have $S^2 \xrightarrow{P} \sigma^2$

<u>Ex.</u> <u>$X_n \overset{iid}{\sim} N(\mu, \sigma^2)$</u> then we can show

$$Var(S^2) = \frac{2\sigma^4}{N-1} \xrightarrow{N} 0$$

hence $S^2 \xrightarrow{P} \sigma^2$.

Add on: Cts mapping theorem say $\sqrt{\cdot}$ is continuous so if $s^2 \xrightarrow{P} \sigma^2$ then

$$\sqrt{s^2} \xrightarrow{P} \sqrt{\sigma^2}$$

i.e. $S \xrightarrow{P} \sigma$

sample s.d. pop. s.d.

---

what about $\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^{N} (X_n - \bar{X})^2$?

let $c_n = \frac{N-1}{N}$. Then $c_n \to 1$

by our algebraic properties since

$$S^2 \xrightarrow{P} \sigma^2$$

and

$$\hat{\sigma}^2 = c_n S^2 \xrightarrow{P} 1 \cdot \sigma^2 = \sigma^2$$

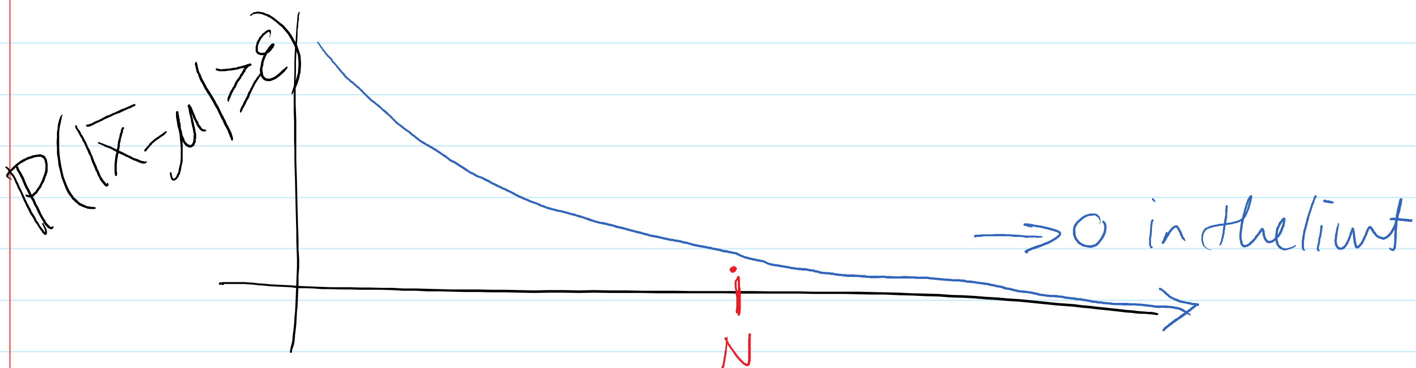
Theorem: Strong Law of Large Numbers (SLLN)

If $X_n \overset{iid}{\sim}$ w/ $E[X_n] = \mu$ and $Var(X_n) = \sigma^2 < \infty$.

Then

$$\bar{X}_N \xrightarrow{a.s.} \mu.$$

WLLN: $\bar{X}$ gets close to μ as N increases but possible w/ non-zero prob. that $\bar{X}$ can be arbitrarily far from μ.



SLLN: $\bar{X}$ can't be arb. far from μ beyond some point

Ex. We saw that $S \xrightarrow{P} \sigma$ and so $\sigma / S \xrightarrow{P} 1$.

Assum... $\bar{X} - \mu$ ... d ...

Assume $\boxed{Y_n = \frac{\bar{X}-\mu}{\sigma/\sqrt{N}} \xrightarrow{d} N(0,1)}$ (*)

let $t_N = \frac{\bar{X}-\mu}{S/\sqrt{N}}$

$$= \frac{\sigma}{\sigma}\frac{1}{S}\frac{\bar{X}-\mu}{1/\sqrt{N}} = \underbrace{\left(\frac{\sigma}{S}\right)}_{\xrightarrow{P} 1}\underbrace{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{N}}\right)}_{Y_n \xrightarrow{d} N(0,1)}$$

by Slutsky's theorem says

$$\xrightarrow{d} 1 \cdot N(0,1)$$

$$= N(0,1)$$