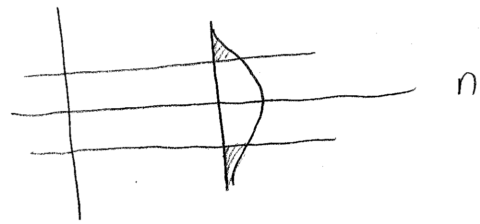
$E_n = \{ |X_n - X| > \varepsilon \}$



$\lim\limits_{n\to\infty} E_n = \limsup\limits_{n\to\infty} E_n = \liminf\limits_{n\to\infty} E_n$

$$= \bigcap_{I=1}^{\infty} \bigcup_{i=I}^{\infty} E_i$$

Let $x \in E_i$. Suppose x not in infinitely many E_i's. Then $\exists\ I^*$ beyond which $x \notin E_i \ \forall i \geq I^*$.

$$\Rightarrow x \notin \bigcup_{i=I^*}^{\infty} E_i \Rightarrow x \notin \bigcap_{I=1}^{\infty} \bigcup_{i=I}^{\infty} E_i = \limsup_{i\to\infty} E_i$$

We sometimes refer to $\limsup E_i$ as "E_i i.o."

For a.s. convergence $\Pr[\lim\limits_{n\to\infty} |X_n - X| > \varepsilon] = 0 \quad \forall \varepsilon > 0$

$\Leftrightarrow \Pr[\limsup |X_n - X| > \varepsilon] = 0 \quad \forall \varepsilon > 0$

$\Leftrightarrow \Pr[|X_n - X| > \varepsilon \text{ i.o.}] = 0 \quad \forall \varepsilon > 0$

Lemma: (Borel-Cantelli)

1] For an arbitrary sequence of events E_i, if $\sum\limits_{i=1}^{\infty} \Pr[E_i] < +\infty$, then $\Pr[E_i \text{ i.o.}] = 0$

2] For E_i sequence of independent events, if $\sum\limits_{i=1}^{\infty} \Pr[E_i] = +\infty$, then $\Pr[E_i \text{ i.o.}] = 1$

Remarks:

1. Definition allows for lack of convergence, but only on sets of measure zero.

e.g. Let $S \sim U[0,1]$

Let $X_n(S) = S + S^n$, $X = S$

Does $X_n \xrightarrow{as} X$? $\forall s \in [0,1)$, $X_n \to X$, but for $s=1$, there is no convergence. Since $Pr[S=1]=0$, we still have $X_n \xrightarrow{as} X$.

2. $X_n \xrightarrow{as} X \Rightarrow X_n \xrightarrow{p} X$

$X_n \xrightarrow{p} X \not\Rightarrow X_n \xrightarrow{as} X$

3. When applied to sample mean, we get the strong law of large numbers. (SLLN)

e.g. Kolmogorov SLLN: If $X_1, X_2, \ldots$ are iid with $E[|X_i|] < +\infty$, then $\bar{X}_n \xrightarrow{as} \mu \equiv E[X_i]$

4. Can extend these results to vectors of random variables.

$\underset{k\times 1}{X_n} \xrightarrow{a.s.} \underset{k\times 1}{X}$ iff $X_n^i \xrightarrow{a.s.} X^i$ $\forall i = 1, \ldots, k$

where $X_n = \begin{bmatrix} X_n^1 \\ \vdots \\ X_n^k \end{bmatrix}$ and $X^i = \begin{bmatrix} X^1 \\ \vdots \\ X^k \end{bmatrix}$

Defn 1.5.8 Two random variables X and Y are identically distributed if $\forall A$, $Pr[X \in A] = Pr[Y \in A]$

Defn 5.5.10 (Convergence in Distribution) a sequence of random variables $X_1, X_2, \ldots$ converges in distribution to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \text{for all } x \text{ s.t.}$$

$F_X(x)$ is continuous at x

Remarks: 1] It is cdfs that are getting close here.

2] "at all continuity points"

Thm 5.5.14: (Central Limit Thm) Let $X_1, X_2, \ldots$ be a sequence of iid random variables. Suppose $E[X_i] < +\infty$ and $0 < Var(X_i) = \sigma^2 < +\infty$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $G_n(\bar{X})$ be the cdf of $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$. Then

$$\lim_{n\to\infty} \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy \quad \text{ie } N(0,1).$$

eg (de Moivre): $X_i \sim$ Bernoulli (p)

$\Rightarrow E[X_i] = p$, $Var(X_i) = p(1-p) < +\infty$

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow{d} N(0,1)$$

Remarks: 1] Spectacular result.

2] Terminology: If $X_n \to X$ in distribution, we write $X_n \xrightarrow{d} X$ or $X_n \xrightarrow{law} X$.

3] Heuristic of proof

Define $Z_n = \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) = \sqrt{n}\left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mu}{\sigma}\right)$

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

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$$

By thm 4.6.3, $M_{Z_n}(t) = \left[M_{Y_i}\left(\frac{t}{\sqrt{n}}\right)\right]^n$. Using Taylor Series expansion around zero,

$$M_{Z_n}(t) \approx \left[\sum_{k=0}^{\infty} M_Y^{(k)}(0) \frac{(t/\sqrt{n})^k}{k!}\right]^n$$

(a) at $k=0$, $M_Y^{(k)}(0) = E[e^{0 \cdot Y}] = E[1] = 1$

(b) $k=1$, $M_Y^{(1)}(0) = \frac{d}{dt} M_Y(t)\Big|_{t=0} = E[Y] = 0$

(c) $k=2$, $M_Y^{(2)}(0) = \frac{d^2}{dt^2} M_Y(t)\Big|_{t=0} = Var(Y) = 1$

$$\Rightarrow M_{Z_n}(t) \approx \left[1 + 0 + \frac{(t/\sqrt{n})^2}{2!} + R_Y\left(\frac{t}{n}\right)\right]^n$$

Consider $n R_Y\left(\frac{t}{\sqrt{n}}\right)$. $n R_Y\left(\frac{t}{\sqrt{n}}\right) \to 0$ as $n \to \infty$

$$\lim_{n\to\infty}\left[M_X\left(\tfrac{t}{\sqrt{n}}\right)\right]^n = \lim_{n\to\infty}\left[1+\frac{t^2}{2n}+R_X\left(\tfrac{t}{\sqrt{n}}\right)\right]^n$$

$$= \lim_{n\to\infty}\left[1+\frac{1}{n}\left\{\frac{t^2}{2}+nR_X\left(\tfrac{t}{\sqrt{n}}\right)\right\}\right]^n$$

but $\lim_{n\to\infty}\left[1+\frac{a_n}{n}\right]^n = \exp\{\lim_{n\to\infty} a_n\}$

$$\Rightarrow \lim_{n\to\infty}\left[M_X\left(\tfrac{t}{\sqrt{n}}\right)\right]^n = \exp\left\{\frac{t^2}{2}\right\}$$

$\Rightarrow M_{Z_n}(t) \to \exp\left\{\frac{t^2}{2}\right\}$ and thus $Z_n \to N(0,1)$ "□"

4] Just an approximation! $X_i \sim N(0,1)$. Define $s^2 = \frac{\sum X_i^2}{n}$

Since $X_i \sim N(0,1)$, $X_i^2 \sim \chi^2(1)$. Thus $E[X_i^2]=1$

and $Var(X_i) = 2 < +\infty$. Using CLT,

$$\sqrt{n}\left(\frac{s^2-1}{\sqrt{2}}\right) \xrightarrow{d} N(0,1).$$

$$0 = Pr[s^2<0] = Pr\left[\sqrt{n}\left(\frac{s_n^2-1}{\sqrt{2}}\right) < -\frac{\sqrt{n}}{2}\right]$$

$$\overset{a}{=} \underbrace{Pr\left[Z < -\frac{\sqrt{n}}{\sqrt{2}}\right]}_{>0 \text{ for finite } n}, \quad Z\sim N(0,1)$$

For finite n, this is not a good approximation!

5] When is approximation good or bad?

$n M_X^{(3)}(0)\frac{t^3}{3!\,n^{3/2}}$ (*) If data is highly skewed, need a larger n.

- Restricted support helps
- Symmetry helps.

6 Multivariate central limit thms:

$\sqrt{n}(\underset{k\times 1}{\bar{X}_n} - \underset{k\times 1}{\mu}) \xrightarrow{d} N(0, \underset{k\times k}{\Sigma})$ jointly converge to normal

- Cramer-Wold device

  joint convergence iff $\forall$ vectors a,

  $a'\bar{X}_n \xrightarrow{d} N(a'\mu, a'\Sigma a)$