Theorem (0-1 Law)

Let $A_1, A_2, \ldots$ be a countably # of events.

Set $\mathcal{A} = \sigma(A_1, A_2, \ldots)$.

If $A \in \mathcal{A}$ is independent of $A_1, \ldots, A_n$ for all $n \geq 1$ then either $P(A) = 0$ or $P(A) = 1$.

Proof: Let $\mathcal{A}_n = \sigma(A_1, \ldots, A_n)$.

Then $\exists\, C_n \in \mathcal{A}_n \ni P(A \triangle C_n) \to 0$

essentially approximating A by C_n

so that

$$P(AC_n) \to P(A)$$

$$P(C_n) \to P(A)$$

But A & C_n are independent & so

$$P(AC_n) = P(A)\,P(C_n)$$

$$\downarrow \qquad\qquad \downarrow$$

$$P(A) = P(A)^2$$

$\Rightarrow P(A) = 0$ or 1

Kolmogorov 0-1 Law

Let $X_1, X_2, \cdots$ be ind. Then all tail events have probability 0 or 1 & hence tail rv's are constants wp1.

We see this result in the Borel-Cantelli Lemma (in the independent case) & the SLLN.

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho & 0 \\ -\rho & 1+\rho^2 & -\rho \\ 0 & -\rho & 1 \end{pmatrix}$$

$\vdots$

$$\begin{pmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & \ddots & & \vdots \\ \vdots & & \ddots & \\ \rho^{n-1} & \cdots & & \end{pmatrix}^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho & & 0 & 0 \\ -\rho & 1+\rho^2 & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & 1+\rho^2 & -\rho \\ 0 & & & -\rho & 1 \end{pmatrix}$$

Let $X_0, X_1, \ldots$ be rv's such that

$$\begin{pmatrix} X_0 \\ \vdots \\ X_n \end{pmatrix} \sim N(\underset{\sim}{\mu}, \Sigma),$$

where $\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho^n \\ \rho & \ddots & & \\ \vdots & & \ddots & \\ \rho^n & & & \ddots \end{pmatrix}$

This is a discrete time Gaussian stochastic process. If $\underset{\sim}{\mu} = \mu \underset{\sim}{1}$ then

$$E(X_n) = \mu \quad , \forall n$$

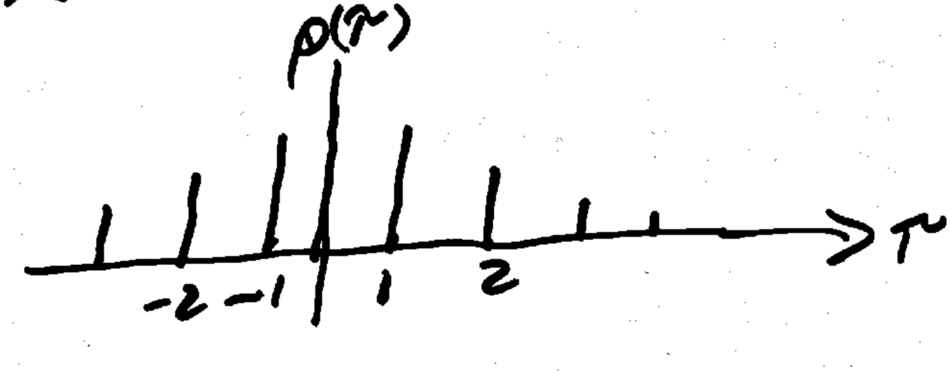
$$\text{cov}(X_n, X_{n+\tau}) = \sigma^2 \rho^\tau \quad , \tau = 0, 1, \cdots$$

(define $\text{cov}(X_n, X_{n-\tau})$ + $\text{cov}(X_n, X_{n+\tau})$ to be the autocovariance function

So $\text{cov}(X_n, X_{n+s}) = \sigma^2 \rho^{|s|} \quad , \forall s$

Set $\rho(\tau) = \rho^{|\tau|} \quad , \forall \tau$

This is the autocorrelation function.



The $\text{cov}(X_n, X_{n+s})$ does not depend on n (for each s) & neither does $E(X_n)$. This is the <u>weakly stationary</u> property.

In this case since everything is normal we have for any

$t_1 < t_2 < \cdots < t_k$

$$\begin{pmatrix} X_{t_1} \\ \vdots \\ X_{t_k} \end{pmatrix} \overset{d}{=} \begin{pmatrix} X_{t_1+s} \\ \vdots \\ X_{t_k+s} \end{pmatrix}, \quad \forall s$$

so that the process is <u>strictly stationary</u>.

Assume $\mu = 0$

$$f(\underbrace{x_0, \ldots, x_n}_{\underset{\sim}{x}'}) \propto e^{-\frac{1}{2\sigma^2} \underset{\sim}{x}' \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & & 0 \\ -\rho & 1+\rho^2 & \ddots & \\ & \ddots & \ddots & \\ 0 & & & 1+\rho^2 \end{pmatrix} \underset{\sim}{x}}$$

$$\propto f_N(x_0)\, f_N(x_0 + x_1) \cdots f_N(x_{n-1} + x_n)$$

$$\| $$

$$f(x_0)\, f(x_1 | x_0) \cdots f(x_n | x_{n-1})$$

Conditional Expectation

Y rv

$\underset{\sim}{X}$ rvec

$\underbrace{E(Y \mid \underset{\sim}{X})}_{\hat{Y}}$ — rv which is a f'n of $\underset{\sim}{X}$.

$\hat{Y}$ is that f'n of $\underset{\sim}{X}$ s.t.

(*) $E(Y h(\underset{\sim}{X})) = E(\hat{Y} h(\underset{\sim}{X}))$, "$\forall$" h

or if $E(Y^2) < \infty$ then $\hat{Y}$ is that f'n of $\underset{\sim}{X}$ which minimizes

(**) $E(Y - \text{f'n}(\underset{\sim}{X}))^2$

If $E(Y^2) < \infty$ then (*) + (**) are $\Leftrightarrow$.

We have yet to show that $E(Y \mid \underset{\sim}{X})$ exists.

Convergence in mean-square

$X_n \xrightarrow{ms} X$ if $E(X_n - X)^2 \to 0$

Remarks

- This is also called convergence in L_2 & is a special case of $X_n \xrightarrow{r} X$ or $X_n \xrightarrow{L_r} X$ ($E(|X_n - X|^r) \to 0$)

- Set $\|X\| = \sqrt{E(X^2)}$. $\{X : E(X^2) < \infty\}$ is called L_2 (it's a linear space — note that $X_1, \ldots, X_k \in L_2 \Rightarrow c_1 X_1 + \cdots + c_k X_k \in L_2$). $\|X\|$ is a norm & satisfies

$$\left\| \sum_{i=1}^{K} X_i \right\| \le \sum_{i=1}^{K} \| X_i \|$$ Δ inequality

$$\left\| \sum_{i=1}^{K} X_i \right\|^2 = \sum_{i=1}^{K} \| X_i \|^2 \text{ if } E(X_i X_j) = 0 \text{ if } i \neq j$$

$$\| X + Y \|^2 + \| X - Y \|^2 = 2 \| X \|^2 + 2 \| Y \|^2$$

↑ parallelogram property

$L_2 + \| \ \|$ is a normed linear space

<u>Some other inequalities</u>

① $E(g(X)) \ge g(E(X))$, g is convex

② $|x+y|^r \le c_r (|x|^r + |y|^r)$, where

$$c_r = 1, \quad 0 < r \le 1$$
$$= 2^{r-1}, \quad r > 1$$

③ $\left\| \sum_{i=1}^{K} X_i \right\|_r \le \sum_{i=1}^{K} \| X_i \|_r$, $r \ge 1$

where $\|X\|_r = (E(|X|^r))^{1/r}$

④ $E|XY| \le \|X\|_r \, \|Y\|_s$

$r, s > 1$ + $\frac{1}{r} + \frac{1}{s} = 1$

($r = s = 2$ gives Cauchy Schwartz)

Hölder's Inequality

⑤ $X_n \xrightarrow{ms} X \Rightarrow \exists \, X_{n_k} \xrightarrow{as} X$

$X_n \xrightarrow{ms} X \Rightarrow X \in L_2$ + $E(X_n^2) \to E(X^2)$

(Whittle p288)

$X_n \xrightarrow{ms} X \Leftrightarrow \sqrt{E(X_n - X)^2} \to 0$

$\Leftrightarrow \|X_n - X\| \to 0$

Look at

$$\| X_n - X_m \|$$

Suppose $\|X_n - X_m\| \to 0$ as $n, m \to \infty$

then $\exists$ X st $X_n \xrightarrow{ms} X$

__Note__ $X_n \xrightarrow{ms} X$ then

$$\| X_n - X_m \| = \|(X_n - X) + (X - X_m)\|$$

$$\leq \underbrace{\| X_n - X\| + \| X - X_m \|}$$

$\downarrow$ as $n, m \to \infty$

0

So $\xrightarrow{ms}$ of $\{X_n\}$

Yes? $\Leftarrow$ $\Rightarrow$ Cauchy property

Some more background

$X \in L_2$ then $\|X\| = \sqrt{E(X^2)}$

$X_n \xrightarrow{ms} X$ means $\|\underbrace{X_n - X}_{\in L_2}\| \to 0$

X_n ↑ $\in L_2$

$E(X_n - X)^2 \to 0$

Also use the notation

$X_n \xrightarrow{ms} X$ or $X_n \xrightarrow{L_2} X$ or $X_n \xrightarrow{2} X$

Is $X \in L_2$?

$$\|X\| = \|X - X_n + X_n\|$$
$$\leq \|X - X_n\| + \|X_n\| < \infty$$

so that $X \in L_2$.

$X_n \xrightarrow{ms} X$ & $X_n \xrightarrow{ms} X' \Rightarrow \|X - X'\| = 0$

i.e. $X \overset{ms}{=} X'$

$E(X - X')^2 = 0$

$\Rightarrow X \overset{wp1}{=} X'$

Let $H = \{ h(X) : E(h(X)^2) < \infty \}$

$\subset L_2$

H is in fact a linear space (subspace of L_2) which is closed wrt $\| \ \|$ in the sense that

if $\| Z_n - Z_m \| \to 0$ as $n, m \to \infty$

($Z_n, Z_m \in H$)

then $\exists\ Z \in H$ st $Z_n \overset{ms}{\to} Z$.

$X_n \xrightarrow{ms} X \Rightarrow X_n \xrightarrow{P} X$ (Markov)

$\Rightarrow \exists$ subsequence $X_{n_k} \xrightarrow{as} X$ (BCL part a)

So if $\{Z_n\}$ is a "Cauchy" sequence in H (i.e $\|Z_n - Z_m\| \to 0$ as $n, m \to \infty$ w.p.1 Z),

then $\exists$ a subsequence Z_{n_k}

where $Z_n \xrightarrow{ms} Z$ } this is true $\therefore$

$Z_n \in H \subset L_2$, $Z \in L_2$

$\xrightarrow{ms}$ has the Cauchy property.

Is Z in H? Yes $\because$ it is in L_2 and it is the pointwise limit of f'ns of X + hence is a f'n of X.

Let B = set of #'s (ie a subset of $\mathbb{R}$) & suppose $\sup B = 5$. Is there a sequence $b_n \in B$ st $b_n \to 5$? Yes, almost by def'n of sup.

<u>Theorem</u> Let H be a closed linear space of rv's & $H \subset L_2$. If $Y \in L_2$ then $\exists\ \hat{Y} \in H$ st

$$\|Y - \hat{Y}\| \le \|Y - Z\|, \ \forall Z \in H$$

<u>Proof</u> Let $d = \inf_{Z \in H} \|Y - Z\|$

$$(= \inf \{\|Y - Z\| : Z \in H\})$$

Then $\exists\ Z_n \in H$ st $\|Z_n - Y\| \to d$.

Now use the // property to get

$$\|Z_m - Z_n\|^2 + \|2Y - (Z_n + Z_m)\|^2 = 2\|Y - Z_n\|^2 + 2\|Z_m - Y\|^2$$

$$\Rightarrow \| Z_m - Z_n \|^2 = 2\|Y - Z_n\|^2 + 2\|Z_m - Y\|^2 - 4\underbrace{\left\| Y - \frac{Z_n + Z_m}{2} \right\|^2}_{\geq d^2 \text{ since } \frac{Z_n + Z_m}{2} \in H}$$

$$\leq 2\|Y - Z_n\|^2 + 2\|Z_m - Y\|^2 - 4d^2$$

$$\to 0 \quad \text{as } n, m \to \infty$$

So $0 \leq \| Z_m - Z_n \|^2 \leq \underbrace{(\quad)}_{\to 0}$

$$\Rightarrow \| Z_m - Z_n \|^2 \to 0 \quad \text{as } n, m \to \infty$$

$$\Rightarrow \exists \underbrace{Z}_{\hat{Y}} \in H \text{ st } Z_m \xrightarrow{ms} \underbrace{Z}_{\hat{Y}}$$

Now

$$\| Y - \hat{Y} \| \leq \| Y - Z_n \| + \| Z_n - \hat{Y} \| \to d$$

so that $\| Y - \hat{Y} \| \leq d \leq \| Y - Z \|, \; \forall Z \in H$

qed

Let $B \subset \mathbb{R}$. B is <u>bounded</u> if $\exists$ $M > 0$ st $x \in B \Rightarrow |x| \leq M$.

In this case there exists a smallest finite upper bound (called $\sup B$) + a largest finite lower bound (called $\inf B$).
or lub / glb

<u>BTW</u> $a_n \to a \Leftrightarrow \overline{\lim}\, a_n = \underline{\lim}\, a_n = a$

<u>Note</u> – $\underline{\lim} \leq \overline{\lim}$

– if $a_n \geq 0$ + $\overline{\lim}\, a_n = 0$ $\Rightarrow a_n \to 0$ } try to show

<u>Fatou's Lemma</u> If $Y_n \geq 0$ then
$$E(\underline{\lim}\, Y_n) \leq \underline{\lim}\, E(Y_n)$$

Proof: $E(\underline{\lim}\, Y_n)$

$$= E\left(\lim_{N\to\infty} \inf_{n\geq N} Y_n\right)$$

$$= \lim_{N\to\infty} \underbrace{E\left(\inf_{n\geq N} Y_n\right)}$$ by MCT

Now $\inf_{n\geq N} Y_n \leq Y_n$, $\forall n \geq N$

$$\Rightarrow E\left(\inf_{n\geq N} Y_n\right) \leq E(Y_n), \ \forall n \geq N$$

$$\Rightarrow E\left(\inf_{n\geq N} Y_n\right) \leq \inf_{n\geq N} E(Y_n)$$

$$\therefore E(\underline{\lim}\, Y_n) \leq \lim_{N\to\infty} \inf_{n\geq N} E(Y_n)$$

$$= \underline{\lim}\, E(Y_n)$$

qed

MCT $X_n \geq 0$ & $X_n \overset{as}{\uparrow} X \Rightarrow E(X_n) \to E(X)$

DCT $X_n \overset{as}{\to} X$ & $|X_n| \overset{as}{\leq} W$ with $E(W) < \infty \Rightarrow E(X_n) \to E(X)$

PDCT $X_n \overset{P}{\to} X$ " .

Note MCT & DCT work for integrals of type $\int X(\omega)\,\mu(d\omega)$ or $\int X\,d\mu$

μ ↑ σ-finite measure

We will sometimes use the notation $E_\mu(X)$ in place of the integral notation. E_μ has the same properties as E except that it is not normed (i.e. $E(1)$ need not be 1).

Counting rv's, generating functions

A rv X taking on possible values $\{0, 1, 2, \ldots\}$ is called a counting rv. It is, of course, a special case of a discrete rv (and an example of a lattice rv).

The pgf of X is defined as

$$G(s) = E(s^X) = \sum_{i=0}^{\infty} s^i P(X=i)$$

Notes

- G is the generating function of the sequence $\{ P(X=i) : i = 0, 1, \ldots \}$
- G is a power series with radius of convergence ≥ 1. This follows as
$$|G(s)| \leq E(|s|^X) \leq 1 \quad \text{for } |s| \leq 1$$
- G determines the distribution of X & in fact
$$P(X=k) = G^{(k)}(0)/k!$$

- $G^{(k)}(s)$ is also a power series with radius of convergence ≥ 1 so that

$$\lim_{s \uparrow 1} G^{(k)}(s) = E\left[\lim_{s \uparrow 1} X(X-1)\cdots(X-k+1)s^{X-k}\right]$$

$$= E\left[X(X-1)\cdots(X-k+1)\right]$$

which is the kth factorial moment of X. Unless otherwise stated we denote $\lim_{s \uparrow 1} G^{(k)}(s)$ by $G^{(k)}(1)$

For a rvec $\underset{\sim}{X} = (X_1, \cdots, X_n)'$ with counting rv components we define the (joint) pgf

$$G(\underset{\sim}{s}) = E(\underset{\sim}{s}^{\underset{\sim}{X}}) = E(s_1^{X_1} \cdots s_n^{X_n})$$

In this case independence of $X_1, \cdots, X_n$ is equivalent to

$$G(\underset{\sim}{s}) = G_1(s_1) \cdots G_n(s_n)$$

If N is a counting rv independent of $X_1, X_2, \ldots$ which are iid with pgf G_X, then

$$S = X_1 + \cdots + X_N$$

has pgf

$$\begin{aligned} G_S(s) &= E(s^S) \\ &= E(s^{X_1 + \cdots + X_N}) \\ &= E\{ E[s^{(X_1 + \cdots + X_N)} \mid N] \} \\ &= E[G_X(s)^N] \end{aligned}$$

(Note $X_0 \equiv 0$ for these formulae)

Hence if G_N is the pgf of N then

$$G_S(s) = G_N[G_X(s)]$$

<u>example</u> If $X \sim$ Bernoulli (p) then it has pgf

$$G_X(s) = q + ps \quad ; \quad q = 1-p$$

Now let $X_1, \dots, X_n$ be iid X & set

$$Y = X_1 + \cdots + X_n$$

Then

$$\begin{aligned} G_Y(s) &= E\left(s^{X_1 + \cdots + X_n}\right) \\ &= E(s^{X_1}) \cdots E(s^{X_n}) \\ &= (q + ps)^n \\ &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} s^k \end{aligned}$$

$$\Rightarrow \quad P(Y = k) = \binom{n}{k} p^k q^{n-k} \quad \text{— binomial probabilities}$$

Now take $np = \lambda$ fixed with $n \to \infty$ (so that $p \to 0$). This yields

$$\lim_{n \to \infty} G_Y(s) = \lim_{n \to \infty} (q + ps)^n$$

$$= \lim_{n \to \infty} \left[1 + \frac{\lambda}{n}(s-1)\right]^n$$

$$= e^{\lambda(s-1)} = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} s^k$$

and $\left\{ \frac{e^{-\lambda}\lambda^k}{k!} ; k = 0, 1, \cdots \right\}$ are the Poisson(λ) probabilities.

<u>example</u> Let $U \sim$ Poisson(λ_1), $V \sim$ Poisson(λ_2) and $W \sim$ Poisson(λ_3) be ind.

Set

$$X = U + V$$

$$Y = V + W$$

The pgf of $\underset{\sim}{X} = (X, Y)'$ is

$$G(\underset{\sim}{s}) = E(s_1^{U+V} s_2^{V+W})$$
$$= E(s_1^{U}) E[(s_1 s_2)^{V}] E(s_2^{W})$$
$$= \exp[\lambda_1 (s_1 - 1) + \lambda_2 (s_1 s_2 - 1) + \lambda_3 (s_2 - 1)]$$

which shows that X & Y are independent iff

$$\lambda_2 = \text{cov}(X, Y) = 0$$

$\underset{\sim}{X}$ is called a bivariate Poisson.

Notice

$$Y | X \overset{d}{=} \text{binomial rv} + \text{Poisson rv}$$

(ind)

since $V | X$ is binomial.