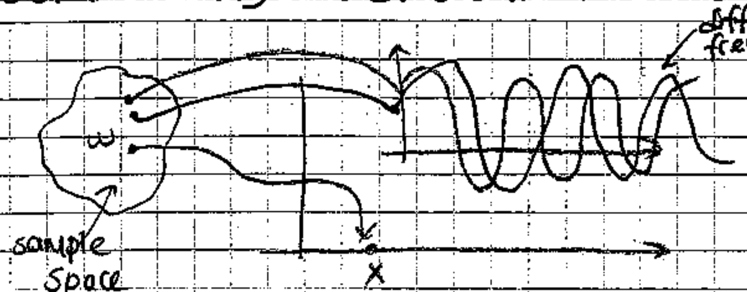
Hande Karadoğan, EE 531, Fall 2014-15, METU, Ankara (1)

EE 531: Probability And Stochastic Processes 22/09/2014

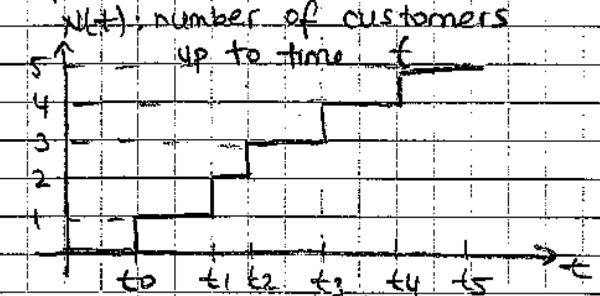


r.v → mapping from sample space (Ω) → real line

$x(\omega) = x$

r.p → mapping from sample space (Ω) to functions

If r.p involves "continuous functions", that process is called continuous r.p



discrete / point r.p

Definitions:

Probability space: σ-algebra ⟶ random variables ↓ random processes

(Borel field)

(σ-field)

Probability space:

Random experiment: An experiment whose outcome is not known in advance

Outcome: An experiment result

Sample space (Ω): set of all outcomes

Event: A subset of sample space

EX: Coin Toss

$\Omega = \{H, T\}$ $A = \{H\}$ "Finite sample space"

event

EX: "Countably Infinite Sample Space"

exp: random integer pick

$\Omega = \{-\infty, \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots, \infty\}$

$A = \{0, 2, 4, \ldots, \infty\}$

EX: "Uncountably Infinite Sample Space"

Exp: pick a real number in $[0,1]$

$\Omega = \{x : 0 \leq x \leq 1\}$

$A = \{x : 0 \leq x \leq 1/2\}$

EX: Gender and height

$\Omega = \{(M,h) : 150 \leq h \leq 250\} \cup \{(F,h) : 130 \leq h \leq 225\}$

$A = \{(M,h) : h \geq 200\}$

24.09.2014

Q: How to assign probabilities to Events.

Probabilities are assigned to sets and probability as a function is from sets to real number

The assignment procedure should satisfy following:

A1. $P\{\Omega\} = 1$

A2. $P\{A\} \geq 0$

A3. $A_1, A_2, \ldots$ are disjoint events then $P\{\bigcup_{n=1}^{\infty} A_n\} = \sum_{n=1}^{\infty} P\{A_n\}$
$A_k \cap A_\ell = \phi \quad (k \neq \ell)$

Kolmogorov's Axioms (~1930)

Ex: 2-coin tosses

$\Omega = \{HH, HT, TH, TT\}$

$P\{HH\} = \alpha_1$

$P\{HT\} = \alpha_2$

$P\{TH\} = \alpha_3$

$P\{TT\} = \alpha_4$

A3 $P\{HH \cup HT \cup TH \cup TT\} = 1$

$\hookrightarrow P\{HH\} + P\{HT\} + P\{TH\} + P\{TT\} \overset{A3}{=} 1$

$= \sum_{k=1}^{4} \alpha_k = 1 \quad \alpha_k \geq 0$

Any other event

$A = \{HH, TT\}$

$P\{A\} \overset{A3}{=} P\{HH\} + P\{TT\}$

$= \alpha_1 + \alpha_4$

To do some algebra with sets, we need to have consistency in our operations. The algebra for sets is done by union and intersection operations.

We consider the set events satisfying the following conditions as a valid field:

Borel field (σ-algebra)

1. Ω is an event
2. $A_1, A_2 \ldots$ are events
   $\bigcup_{k=1}^{\infty} A_k$ is also an event.
3. A_i is an event, A_i^c is also an event.

Ex: $\Omega = \{HH, HT, TH, TT\}$

$F_1 = \{\Omega, \phi, HH, HT, TH, TT, \{HH, HT\}, \{HH, HT, TH\}$
$, \{HH, TH\}, \{HH, HT, TT\}$
$, \{HH, TT\}, \{HT, TH, TT\}$
$, \{HT, TH\}, \{HH, TH, TT\}$
$, \{HT, TT\}$
$, \{TH, TT\}$

$2^{|\Omega|} = 2^4 = 16$ subsets
power set.

$F_2 = \{\Omega, \phi, \{HH, TT\}, \{TH, HT\}\}$

$F_3 = \{\Omega, \phi\}$

EX: $\Omega = \{ x : 0 < x \leq 1 \} \quad x \in \mathbb{R}$

We construct events from half-open intervals, $(a,b]$

$F_1 = \{ \Omega, \phi, (0, \frac{1}{2}], (\frac{1}{2}, 1] \}$

A field can be constructed using all half open intervals in $(0,1]$

Note: Event of $\{\frac{1}{2}\}$, $[\frac{1}{4}, \frac{1}{2}]$ is also in the field constructed.

$$\{ \tfrac{1}{2} \} = \bigcap_{k=2}^{\infty} \left(\tfrac{1}{2} - 1/k , \tfrac{1}{2} \right]$$

since $A \cap B = (A^c \cup B^c)^c$

$$\left[\tfrac{1}{4}, \tfrac{1}{2} \right] = \{ \tfrac{1}{4} \} \cup \left(\tfrac{1}{4}, \tfrac{1}{2} \right]$$

Consequences of Kolmogorov's Axioms

① $P\{\phi\} = 0$

② $P\{ \bigcup_{n=1}^{m} A_n \} = \sum_{n=1}^{M} P\{A_n\}$ A_n: disjoint

③ $P\{A^c\} = 1 - P\{A\}$

④ $P\{A\} \leq P\{B\}$, $A \subset B$

⑤ $\sum_n P\{A_n\} \leq 1$; A_n: disjoint

⑥ $P\{ \bigcup_{n=1}^{\infty} A_n \} = \lim_{m \to \infty} P\{ \bigcup_{n=1}^{m} A_n \}$

⑦ $P\{ \bigcup_{n=1}^{\infty} A_n \} = \lim_{m \to \infty} P\{A_n\}$, $A_1 \subset A_2 \subset A_3 \cdots$

⑧ $P\{ \bigcap_{n=1}^{\infty} A_n \} = \lim_{m \to \infty} P\{A_n\}$, $A_1 \supset A_2 \supset A_3 \cdots$

Sec 1.2.2

1. $P\{\phi\} \stackrel{?}{=} 0$.

$A_1 = A_2 = A_3 = \cdots = A_n = \phi$

Apply (A3)

$$P\{ \bigcup_{n=1}^{\infty} A_k \} = \lim_{m \to \infty} \sum_{n=1}^{M} P\{A_n\}$$

$$P\{\phi\} = \lim_{m \to \infty} m \cdot P\{\phi\}$$

$$\Downarrow$$

$$P\{\phi\} = 0 \checkmark$$

2. $A_{m+1} = A_{m+2} = A_{m+3} = \cdots = \phi$

Apply (A3) and we $P\{\phi\} = 0$ ✓

3. $P\{A^c\} = 1 - P\{A\}$

(A1) $P\{\underbrace{\Omega}_{A \cup A^c}\} = 1 \stackrel{A_3}{=} P\{A\} + P\{A^c\}$ ✓

4. $A \subset B \rightarrow P\{A\} \leq P\{B\}$

$B = A \cup \underbrace{(B-A)}_{B \cap A^c}$

Apply A3

$P\{B\} = P\{A\} + \underbrace{P\{B \cap A^c\}}_{\geq 0 \rightarrow A2}$

$P\{B\} \geq P\{A\}$

5. $\sum P\{A_n\} \leq 1 \quad A_n$: disjoint

$\bigcup_{n=1}^{\infty} A_n \subset \Omega$

from previous statement

$\underbrace{P\{\bigcup_{n=1}^{\infty} A_n\}}_{\text{A3}} \leq \underbrace{P\{\Omega\}}_{1} \ \text{A3}$

$\sum_{n=1}^{\infty} P\{A_n\}$

6. $P\{\bigcup_{n=1}^{\infty} A_n\} = \lim_{m \to \infty} P\{\bigcup_{n=1}^{m} A_n\}$

not disjoint

$B_1 = A_1, \ B_2 = A_2 \cap B_1^c \longrightarrow B_1$ and B_2 are disjoint

$B_1 \cup B_2 = A_1 \cup (A_2 \cap A_1^c) = (A_1 \cup A_2) \cap (A_1 \cup A_1^c)$

$= (A_1 \cup A_2) \cap \Omega = A_1 \cup A_2$

$B_3 = A_3 \cap B_1^c \cap B_2^c \rightarrow$ disjoint

B_k's are disjoint, $\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} B_k$

Then, $P\{\bigcup_{n=1}^{\infty} A_n\} \overset{A3}{=} \lim_{m \to \infty} \sum_{n=1}^{m} P\{B_n\}$

$\overset{A3}{=} \lim_{m \to \infty} P\{\bigcup_{n=1}^{m} B_n\}$

$= \lim_{m \to \infty} P\{\bigcup_{n=1}^{\infty} A_n\}$

7. Apply ⑥ ✓

8. $A_1 \to A_1^c$

$A_2 \to A_2^c$ then apply ⑥ ✓

Union Bound:

$$P\left(\bigcup_{k=1}^{N} A_k\right) \le \sum_{k=1}^{N} P(A_k)$$

A_k : Event k

$$P(A_1 \cup A_2) = p(A_1 \cup \underbrace{(A_2 \cap A_1^c)}_{A_2 - A_1}) \overset{A3}{=} p(A_1) + \underbrace{p(A_2 \cap A_1^c)} \le p(A_1) + p(A_2)$$

since $(A_2 \cap A_1^c) \subset A_2$

$\downarrow$

$p(A_2 \cap A_1^c) \le P(A_2)$

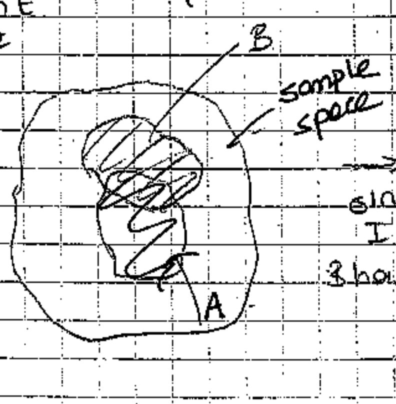
Using this fact

$$(A_1 \cup \bigcup_{k=2}^{N} A_k) \le p(A_1) + p\left(\bigcup_{k=2}^{N} A_k\right) \le p(A_1) + p(A_2) + p\left(\bigcup_{k=3}^{N} A_k\right)$$

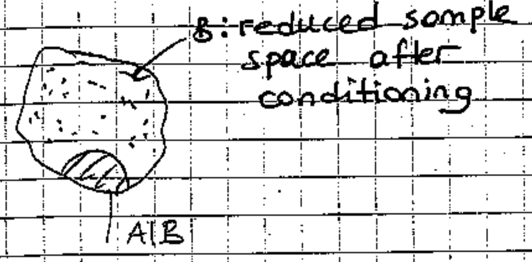
Conditional Probability

$$\underset{\substack{\uparrow \\ \text{event } A}}{P(A} \mid \underset{\substack{\uparrow \\ \text{event } B}}{B)} \triangleq \frac{P(A \cap B) \rightarrow \text{event A and B occur at the same time}}{P(B) \rightarrow \text{event B occurred}}$$

$$\underset{\substack{\uparrow \\ \text{event } z}}{z} = A \mid B = \begin{cases} \text{A occurs given B occurs} \\ \text{A does not occur given B occurs.} \end{cases}$$



→ since I know B has occurred.



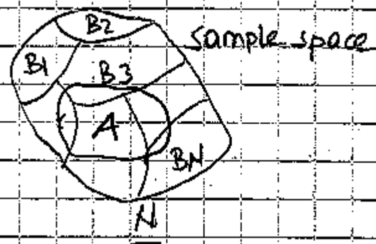
Baye's Rule

$$p(A|B) = \frac{p(B|A)\,p(A)}{p(B)} \quad \text{since} \quad \underbrace{p(A|B)\,p(B)}_{\text{from def.}} = p(B|A)\,p(A)$$

$$p(A \cap B) = p(B \cap A)$$

Note: A|B event is related to B|A event by Baye's Rule.

Total Probability Theorem: B_i disjoint sets covering sample space (partition)



① $B_k \cap B_\ell = \Phi \quad k \neq \ell$

② $\bigcup_k B_k = \Omega$

$$p(A) = \sum_{k=1}^{N} \underbrace{P(A|B_k)\,P(B_k)}_{p(A \cap B_k)}$$

Independence: If $p(A \cap B) \overset{\triangle}{=} p(A)\,p(B)$, A and B are independent

Also equivalent to $P(A|B) = P(A)$
$P(B|A) \overset{\text{or}}{=} P(B)$

If $A_1, A_2, A_3, \ldots$ are independent

① If they are pairwise independent, that is

$p(A_k \cap A_\ell) = p(A_k) \cdot p(A_\ell) \quad k \neq \ell \quad \forall k, \forall \ell$

② They should independent in triplets, that is

$p(A_k \cap A_\ell \cap A_m) = p(A_k) \cdot p(A_\ell) \cdot p(A_m) \quad k \neq \ell \neq m \quad \forall k, \forall \ell, \forall m$

③ Independent in quartets (and so on)

Conditional Independence:

$$p(A \cap B \mid C) = p(A \mid C)\, p(B \mid C)$$

Borel - Contelli Lemma (Papulis 4^{th} Edition)

① $A_1, A_2, A_3, \ldots$ are a sequence of events

$$p_k = p(A_k)$$

$$\sum_{k=1}^{\infty} p_k < \infty \longrightarrow \underbrace{\sum_{k=1}^{\infty} 1_{A_k} < \infty}_{\text{only finite number of } A_k \text{ happens}} \quad \text{with prob. } 1$$

$$1_{A_k} = \begin{cases} 1 & \text{if } A_k \text{ occurs} \\ 0 & A_k \text{ does not occur} \end{cases}$$

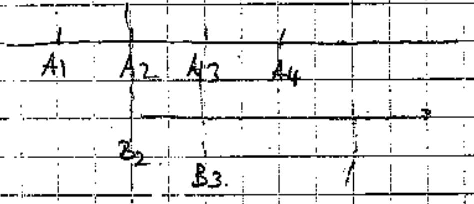
indicator function

Proof: $B_n = \bigcup_{k \geq n} A_k$

at least one of A_n, A_{n+1}, A_{n+2} occurs

B = infinitely many $A_1, A_2, \ldots$ occur

$B_1 \supset B_2 \supset B_3 \supset \ldots$



infinitely many A_k occurs.

then, $B = \bigcap_{n=1}^{\infty} B_n$, | B occurs $\Longleftrightarrow$ B_n occurs of $n = \{1, 2, 3, \ldots\}$

since, if $\xi \in B \longrightarrow \xi \in B_n \quad \forall n$ and then $\xi \in \bigcap_{n=1}^{\infty} B_n$

an event occuring infinite many times

$X \in \bigcap_{n=1}^{\infty} B_n \longrightarrow X$ belongs to infinite number of A_k events

outcome

$$P(B) = P\Big\{\bigcap_{n=1}^{\infty} \underbrace{\Big(\bigcup_{k\geq n}^{\infty} A_k\Big)}_{B_n}\Big\} \quad , \quad B_1 \supset B_2 \supset B_3 \cdots \quad (*)$$

$$= P\Big\{\bigcap_{n=1}^{\infty} B_n\Big\}$$

$$= P\Big\{\lim_{N\to\infty} \bigcap_{n=1}^{N} B_n\Big\} \quad \text{using } (*)$$

$$= \lim_{N\to\infty} P\{B_n\}$$

$$P\{B_n\} \underbrace{\leq}_{\text{union bound}} \sum_{k\geq n} P(A_k) \to 0 \text{ as } n\to\infty.$$

since claim is

$$\sum_{k=1}^{\infty} P(A_k) < \infty$$

$$P\{B\} = \lim_{n\to\infty} P\{B_n\} = 0$$

$P\{B^c\} = 1 \rightarrow$ Finitely many A_k occurs.

B.C Lemma 2:

If $A_1, A_2, \ldots$ are independent events

and if $\sum_{k=1}^{\infty} P(A_k)$ diverges $\longrightarrow$ infinitely many A_k events occur

Proof: Papoulis 4^{th} ed

EX: $P(\text{success}) = p$ You repeat success/fail trials infinitely many times. Each trial is independent

$P(\text{fail}) = 1 - p$

Q: Can there be infinite number n-successes in a row?

$\underset{\text{try\#1}}{S}\,F\,S\,\underset{4}{F} \ldots \overbrace{|\underset{n_0+1}{S}\,S \ldots \underset{n_0+n}{S}|}^{n}$

01.10.2014

Soln: | F_1 | F_2 | --- | F_k |

first n-trials   second n-trials

$P\{F_k = n\text{-successes}\} = p^n \quad k = \{1, 2, 3, --\}$ and

each frame n-success probability is independent from others.

then by B.C lemma 2:

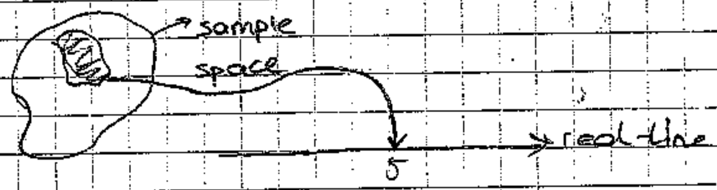
$$\sum_{k=1}^{\infty} \underbrace{P(F_k = \text{success})}_{p^n} \longrightarrow \infty$$

→ so, there are infinitely many n-successes in a row.

Random Variables:

r.v is a mapping from Ω to $\mathbb{R}$ ← real number (real line).

$X = X(w)$
$\forall w \in \Omega$



Required Condition For Valid R.V:

$\{w : X(w) \leq x\}$ should be a valid event.

Ex: $\Omega = \{a_1, a_2, a_3, a_4\}$

$F = \{\phi, \Omega, \{a_1, a_2\}, \{a_3, a_4\}\}$

$\underset{rv}{X} = X(w) = i \leftarrow w = a_i$

Q: Is X a valid random variable?

A: $\{w : X(w) \leq 1^-\} = \phi \in F$ ✓

$\{w : X(w) \leq 1^+\} = \{a_1\} \notin F$ not a valid r.v cdf: cumulative distribution function

$F_X(x) \triangleq P\{\omega : X(\omega) < x\}$

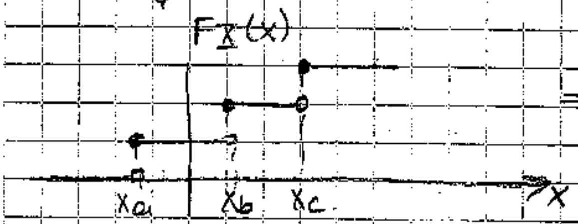
Properties:

1. $\lim_{x\to\infty} F_X(x) = 1$, $\lim_{x\to-\infty} F_X(x) = 0$

2. $F_X(x)$ is non-decreasing.

3. $F_X(x) = F_X(x^+)$; i.e cdf is cont. from right.

(see Exercise 1.5 Textbook)



$\Rightarrow$ ex. for cont. from right.

pdf: prob. density function

$f_X(x) \triangleq \frac{\partial}{\partial x} F_X(x)$

or

$F_X(x) \triangleq \int_{\infty}^{x} f_X(x')\,dx'$

2-r.v:

$F_{X,Y}(x,y) \triangleq P\{X \leq x, Y < y\}$

$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$

Ex: X and Y are ind. rv's

Let

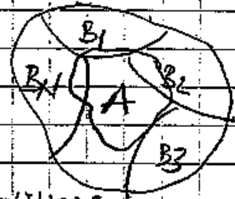
$Z = X + Y$; $f_Z(z) = ?$

Soln: $f_Z(z) = \frac{\partial}{\partial z} F_Z(z)$

$$F_Z(z) = P\{w : Z(w) \leq z\} = P\{Z \leq z\}$$
$$= P\{X + Y \leq z\}$$

$$\int_{-\infty}^{\infty} P\{\underset{\text{a set}}{A} \mid X = x\} f_X(x) = P\{A\}$$

(Remember



$$P\{A\} = \sum_{k=1}^{\#\text{ of partitions}} P(A|B_k)P(B_k) \quad \text{total prob. thm.}$$

Then;

$$P\{X + Y \leq z \mid Y = y\} = P\{X + y \leq z \mid Y = y\}$$
$$= P\{X \leq z - y \mid Y = y\}$$

since X and Y are independent.

$$= P\{X \leq z - y\}$$
$$= F_X(z-y)$$

$$P\{X + Y \leq z\} \underset{\text{total prob thm}}{=} \int_{-\infty}^{\infty} P\{X + Y \leq z \mid Y\} f_Y(y)\,dy$$

$$\boxed{F_Z(z) = \int_{-\infty}^{\infty} F_X(z-y) f_Y(y)\,dy}$$

$\partial/\partial z$

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\,dy$$

$$= f_X(z) * f_Y(z)$$

independent and identical $\hookrightarrow f_{X_k}(x_k) = f_X(x_k) \quad k=1,\dots,N$

Ex: $X_1, X_2, \dots, X_N$: i.i.d r.v. with pdf $f_{X_N}(x)$.

a) $Z = \max(X_1, X_2, \dots, X_N)$, find $f_Z(z)$

b) $P\{Z = X_1\} = ?$

Soln:

a) $F_Z(z) = P\{\overbrace{\max(X_1, \dots, X_N) \le z}^{A}\}$ $\quad A \subset B,\ B \subset A \Big\} A = B$

$= P\{\underbrace{X_1 \le z, X_2 \le z, \dots, X_N \le z}_{B}\}$

$= P\{X_1 \le z\}\, P\{X_2 \le z\} \dots P\{X_N \le z\}$

$= [F_X(z)]^N$

$f_Z(z) = \frac{\partial}{\partial z} F_Z(z) = N[F_X(z)]^{N-1} \cdot f_X(z)$

b) $P\{Z = X_1 \mid X_1 = x_1\} = P\{X_2 \le x_1, X_3 \le x_1, \dots X_N \le x_1 \mid \cancel{X_1 = x_1}\}$ independent

$= [F_X(x_1)]^{N-1}$

$P\{Z = X_1\} = \int_{-\infty}^{\infty} P(Z = X_1 \mid X_1 = x_1)\, f_X(x_1)\, dx_1$

$= \int_{-\infty}^{\infty} [\underbrace{F_X(x_1)}_{u}]^{N-1} \underbrace{f_X(x_1)\, dx}_{dv}$

$= \underbrace{F_X(x_1)}_{u} \underbrace{F_X(x_1)}_{v} \Big|_{x_1=-\infty}^{x_1=\infty} - \int \underbrace{(N-1)[F_X(x_1)]^{N-2} f_X(x_1)}_{du}\, \underbrace{F_X(x_1)}_{v} dx_1$

$= 1 - (N-1) \underbrace{\int [F_X(x_1)]^{N-1} f_X(x_1)\, dx_1}_{P\{Z = X_1\}}$

$\therefore P(Z = X_1) = 1/N$

08.10.2014

Expectation:

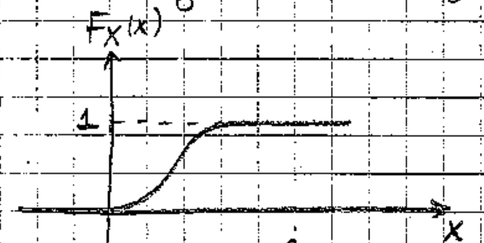
$$E\{x\} \triangleq \int_{-\infty}^{\infty} x\, f_X(x)\, dx$$

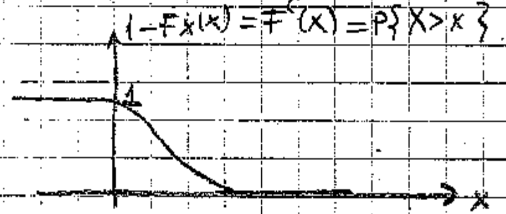
Note ①: If $x>0$, that is $f_X(x)=0$ for $x<0$,

then

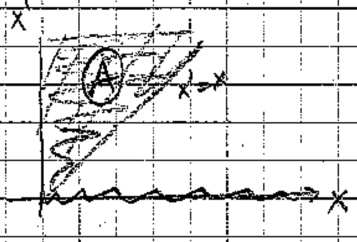
$$E\{x\} = \int_0^{\infty} \underbrace{F^c(x)}_{\text{Complementary cdf } = P\{X>x\} = 1-F_X(x)}\, dx$$

Proof ①: $\displaystyle\int_0^{\infty} F^c(x)\,dx = \int_0^{\infty} P\{X>x\}\,dx = \int_0^{\infty}\left(\int_x^{\infty} f_X(x')\,dx'\right)dx$





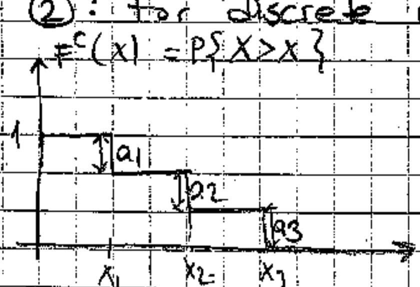
$$= \int_{x'=0}^{\infty}\left(\int_{x=0}^{x'} dx\right)dx'$$

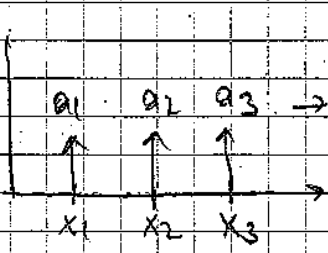


$$\iint_A f_X(x')\,dx'\,dx$$

$$\int_{x'=0}^{\infty} f_X(x')\left(\int_{x=0}^{x'} dx\right)dx' = \int_0^{\infty} x' f_X(x')\,dx' = E\{X\}$$

Proof ②: For discrete r.v.





$\rightarrow E(X) = x_1 P(X=x_1)$

$+ x_2 P(X=x_2)$

$+ x_3 P(X=x_3)$

$= x_1 \cdot a_1 + x_2 \cdot a_2 + x_3 \cdot a_3$

Note ②: If X takes both (+)ve and (-)ve values, then,

$$E\{X\} = -\int_{-\infty}^{0} \underbrace{F_X(x)}_{\text{cdf}}dx + \int_{0}^{\infty} \underbrace{F_X^c(x)}_{\text{complementary cdf}}dx$$

See the picture in Fig. 1.4 from book and convince yourself!

Note ③: The original def for $E\{X\}$ is still very valuable; the "new" relation for $E\{X\}$ is also useful in some cases

Note ④: $E\{\underbrace{g(x)}_{\text{function of single r.v}}\} = \int g(x) f_X(x)dx$

Note ⑤: $E\{X^k\} = M_k \rightarrow k^{th}$ moment

$$\begin{cases} E\{X\} = \bar{X} \rightarrow \text{mean} \\ var(X) = E\{(X-\bar{X})^2\} = E\{X^2\} - (\bar{X})^2 \rightarrow \text{variance} \end{cases}$$

When we have 2 or more r.v's

$$E\{g(x,y)\} = \int_{-\infty}^{\infty}\int g(x,y) f_{X,Y}(x,y)dxdy$$

$$\underbrace{cov(X,Y)}_{\text{covariance of } X \text{ and } Y} = E\{(X-\bar{X})(Y-\bar{Y})\} = E\{XY\} - \bar{X}\bar{Y}$$

Note that $var(X) = cov(X,X)$

Definition:

$\mathrm{Cov}(X,Y) = 0 \longrightarrow$ X and Y are uncorrelated rv's

Note ⑥: If X and Y are independent

then, $E\{g(x)h(y)\} = E\{g(x)\}\, E\{h(y)\}$

$$\iint g(x)h(y)\underbrace{f_{X,Y}(x,y)}_{\substack{\text{independent}\\ \text{so,}\\ f_X(x)\cdot f_Y(y)}}dx\,dy = \underbrace{\left[\int g(x) f_X(x)\,dx\right]}_{E\{g(x)\}}\underbrace{\left[\int h(y) f_Y(y)\,dy\right]}_{E\{h(y)\}}$$

Note again that if X,Y are independent $\rightarrow$ X,Y are also uncorrelated

Uncorrelatedness check $\longrightarrow \mathrm{cov}(x,y) = 0 \longrightarrow E\{(X-\bar{X})(Y-\bar{Y})\} \overset{?}{=} 0$

$$\downarrow$$

$$E\{\underbrace{(X-\bar{X})}_{=0}\}\, E\{\underbrace{(Y-\bar{Y})}_{=0}\} = 0 \checkmark$$

Note: The converse is not true in general

X, Y uncorrelated

$\not\Downarrow$

X, Y independent

Note ⑦: $E\{X_1 + X_2 + \cdots + X_N\} = E\{X_1\} + E\{X_2\} + \cdots + E\{X_N\}$

This relation is valid when X_k's are independent or dependent

correlated or uncorrelated

etc.

Note ⑧: $I_A = \begin{cases} 1 & \text{event } A \text{ occurs} \\ 0 & \text{event } A \text{ does not occur} \end{cases}$

indicator function.

$E\{I_A\} = 1 \cdot P(A \text{ happening}) + 0 \cdot P(A \text{ does not happening})$

$= P(A \text{ occuring})$

Note (9): Iterated Expectation

$$E_{XY}\{g(X,Y)\} = E_Y\{\underbrace{E_{X|Y}\{g(X,Y)|Y\}}_{\text{func}(Y)}\}$$

$$= \int \left(E_{X|Y}\{g(X,Y)|Y\}\right) f_Y(y)\,dy$$

$$= \int \left(\int g(x,y) f_{X|Y}(x|y)\,dx \right) f_Y(y)\,dy$$

$$= \int\int g(x,y) \underbrace{f_{X|Y}(x|y) f_Y(y)}_{f_{X,Y}(x,y)}\,dy$$

note 7 & 8

Ex: In data communications, some bit patterns are reserved for signaling, for example, 011111 can denote end of transmission and once receivers decodes this sequence of bits, it stops listening.

If transmitter tries to send $\underbrace{01100111111100}_{\text{payload}}$ bits to the receiver

then,

$$011001111\,0\,11100 \mid 011111$$

additional 0 inserted to avoid false termination in payload

Q If I am sending n-bits and let $P\{"1"\} = p$,

then what is the expected number of stuffed bits?

Soln.

$\underbrace{b_1 \; b_2 \; b_3 \; ---- \; b_n}_{\text{payload}}$

A_k: Event of bit stuffing after b_k

$$I_{A_k} = \begin{cases} 1 & A_k \text{ occurs} \\ 0 & \text{other} \end{cases}$$

$$\#\text{bits stuffed} = \sum_{k=1}^{n} I_{A_k} \longrightarrow E\left\{\#\text{bits stuffed}\right\} = \sum_{k=1}^{n} \underbrace{E\{I_{A_k}\}}_{p(A_k \text{ happening})}$$

$p(A_1) = 0$

$p(A_2) = 0$

$p(A_3) = 0$

$p(A_4) = 0$

$p(A_5) = (1-p)p^4 + 0$

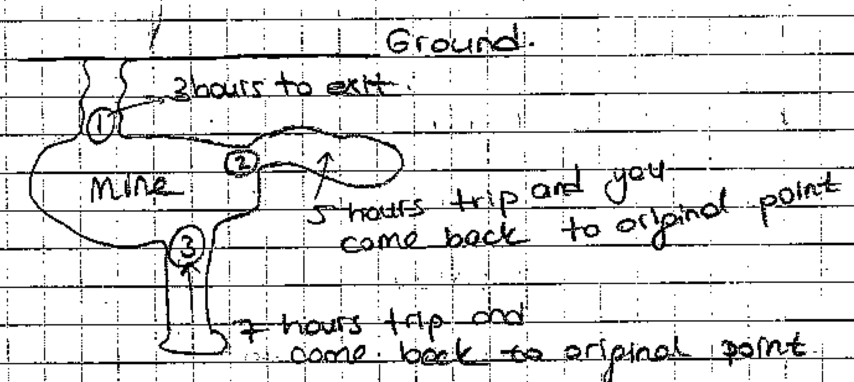
$p(A_6) = (1-p)p^4$

$\vdots$

$p(A_n) = (1-p)p^4$

$$= \sum_{k=5}^{n} (1-p)p^4 = (n-4)(1-p)p^4$$

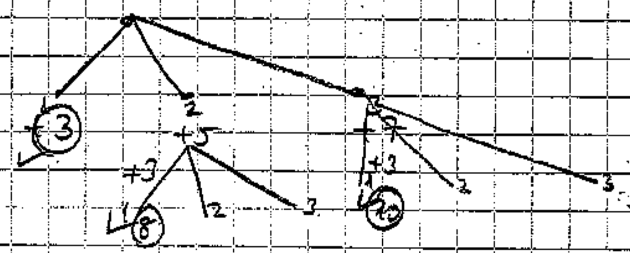
EX: Note 9:

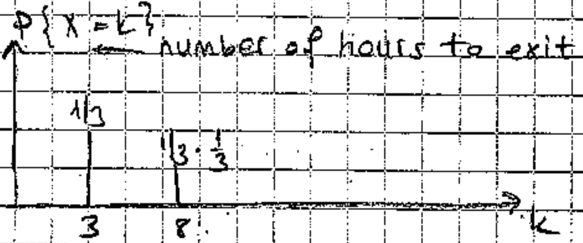


Miners always select one of doors with equal prob. of $1/3$.

Q: What is expected time for mine to go out.

A1:





$$E\{X\} = \sum_{k=0}^{\infty} k \; P(X=k)$$

$E\{E\{X|Y\}\}$

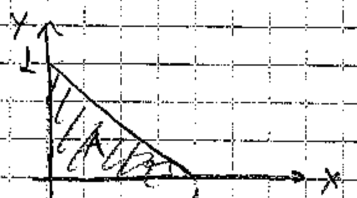
A2: Y be his first choice

$$E\{X\} = E\{X|Y=1\}P\{Y=1\} + E\{X|Y=2\}P\{Y=2\} + E\{X|Y=3\}P\{Y=3\}$$

$$E\{X\} = 3 \cdot \frac{1}{3} + \left(E\{X\}+5\right) \cdot \frac{1}{3} + \left(E\{X\}+7\right) \cdot \frac{1}{3}$$

$$E\{X\} = 15$$

EX:



X, Y are uniform distr. over region A

a) $f_{X,Y}(x,y) = ?$

b) $f_X(x)$, $f_Y(y) = ?$

c) $f_{X|Y}(x|y)$, $f_{Y|X}(y|x)$

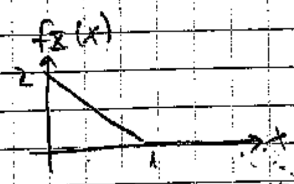
d) $E\{X|Y\}$, $E\{Y|X\}$, $E\{X\}$, $E\{Y\}$

e) Are X and Y ind? $\mathrm{cov}(X,Y) = ?$

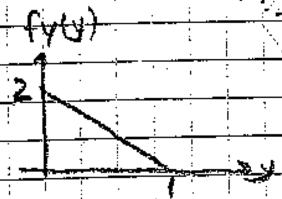
$y \geq 0, \; 0 \leq x \leq 1-y$ (11)

a) $f_{X,Y}(x,y) = \begin{cases} 2 & (x,y) \in A \rightarrow \; x \geq 0 \;\; 0 \leq y \leq 1-x \leftarrow \text{A region} \\ 0 & \text{other} \end{cases}$ same A region

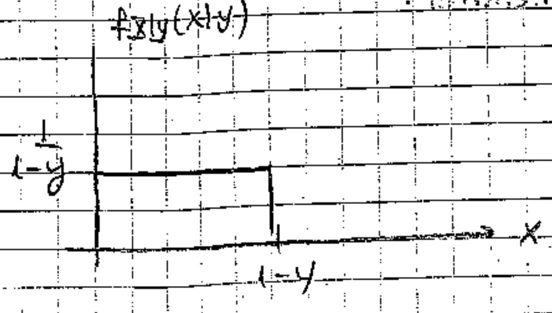
b) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \begin{cases} \int_0^{1-x} 2\,dy = 2(1-x) & 0 \leq x \leq 1 \\ 0 & \text{other} \end{cases}$



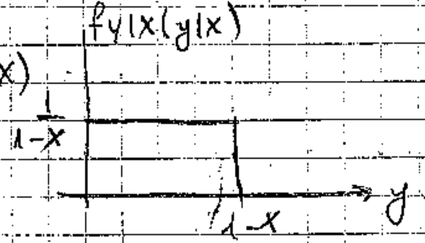
$f_Y(y) = \int f_{X,Y}(x,y)\,dx = \begin{cases} \int_0^{1-y} 2\,dx = 2(1-y) & 0 \leq y \leq 1 \\ 0 & \text{other} \end{cases}$



c) $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \dfrac{\cancel{2}}{\cancel{2}(1-y)} & 0 \leq x \leq 1-y, \; y \geq 0 \\ 0 & \text{other} \end{cases}$



$f_{Y|X}(y|x)$



$\int_0^{1-y} x \frac{1}{1-y}\,dx = \frac{1}{1-y} \cdot \frac{x^2}{2}\Big|_0^{1-y} = \frac{1-y}{2}$

d) $E\{X|Y=y\} = \dfrac{1-y}{2}$

$E\{Y|X\} = \dfrac{1-x}{2}$

$E\{X\} = E\{E\{X|Y\}\} = E_Y\left\{\dfrac{1-Y}{2}\right\} = \dfrac{1}{2} - \dfrac{E\{Y\}}{2} = \int_0^1 y \cdot 2(1-y)\,dy$

$= 1/3 = E\{Y\}$

e) $f_{X,Y}(x,y) \overset{?}{=} f_X(x) f_Y(y)$ not possible

they are not independent

$\text{cov}(x,y) = E\{XY\} - E\{X\}E\{Y\} = E\{XY\} - \frac{1}{3}\cdot\frac{1}{3} = 1/12 - \frac{1}{9} = -1/36$

$E\{XY\} = \iint_A xy f_{X,Y}(x,y)\,dx\,dy = \int_{x=0}\int_{y=0}^{1-x} xy \cdot 2 \cdot dy\,dx = \int_x 2x \frac{y^2}{2}\Big|_{y=0}^{1-x} dx = \int_{x=0}^{1} x(1-x)^2 dx = 1/12$

13.10.2014

Moment Generating Functions:

cont. r.v

$$g_X(r) = E_X\{e^{rX}\} = \int_{-\infty}^{\infty} e^{rX} f_X(x)\,dx \quad , r \in ROC$$

laplace transform
$x \rightarrow s = -r$

dist. r.v

$$g_X(r) = E\{e^{rX}\} = \sum_{k=-\infty}^{\infty} e^{rk} P\{X = k\}$$

z-transform

$z\{a_k\}$

$a_k \leftarrow P\{X = k\}$

$k \rightarrow z = e^{-r}$

$$z\{a_n\} = \sum_{n=-\infty}^{\infty} a_n z^{-n}$$

Remember:

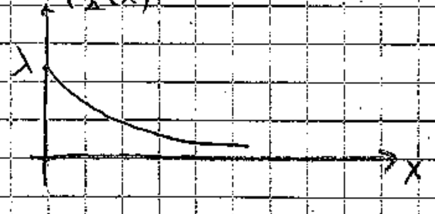
① $g_X(0) = 1 \longrightarrow r = 0 \in ROC$

② $\underbrace{E\{x^k\}}_{m_k} = \dfrac{\partial^k}{\partial r^k} g_X(r)\Big|_{r=0} = g_X^{(k)}(0) \longleftarrow$ assuming that an interval around $r=0$ is in ROC

$m_k \uparrow$ k^{th} moment

$$= E\left\{\frac{\partial^k}{\partial r^k} e^{rX}\right\} = E\left\{X^k e^{rX}\right\} \xrightarrow[r=0]{\text{substitute}} E\{X^k\}$$

Ex: $f_X(x) = \lambda e^{-\lambda x} u(x)$ exponential dist.



$$g_X(r) = \frac{\lambda}{s+\lambda}\Bigg|_{s=-r} = \frac{\lambda}{\lambda - r}$$

$$E\{X\} = g_X^{(1)}(0) = \left.\frac{\lambda}{(\lambda-r)^2}\right|_{r=0} = \frac{1}{\lambda}$$

$$E\{X^2\} = g_X^{(2)}(0) = \left.\frac{2\lambda}{(\lambda-r)^3}\right|_{r=0} = \frac{2}{\lambda^2}$$

Then, exp. distribution with parameter λ has

$$E\{X\} = 1/\lambda$$

$$var\{X\} = E\{X^2\} - \left(E\{X\}\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Probability Inequalities: non-negative r.v (like exp. dist)

① Markov Inequality $\longrightarrow$ $Y \geq 0$, $P\{Y > a\} \leq E\{Y\}/a$

② Chebyshev Inequality $\rightarrow$ $P\{|Y - \underset{E\{Y\}}{\mu_Y}| \geq a\} \leq \overset{var\{Y\}}{\sigma_Y^2}/a^2$

③ One-sided Chebyshev Inequality $\rightarrow$ $P\{Y - \mu_Y \geq a\} \leq \sigma_Y^2 + b^2/(a+b)^2$, for any $b > 0$

④ Chernoff Bound

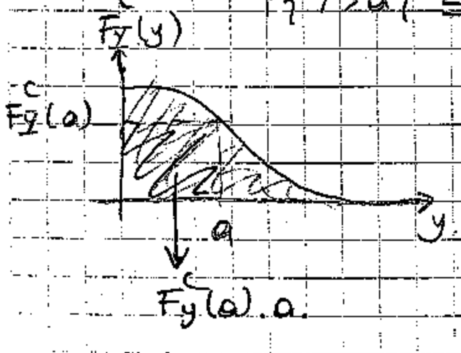
① Markov Inequality

For $Y \geq 0$ r.v $\quad P\{Y > a\} \leq \dfrac{E\{Y\}}{a}$

$$E\{Y\} = \int_0^{\infty} \underbrace{F^c(y)}_{P\{Y>y\}}\,dy \geq \int_0^a \underbrace{P\{Y>y\}}_{\geq P\{Y>a\}}\,dy \geq P\{Y>a\}\cdot a$$

$$\downarrow$$

$$P\{Y > a\} \leq \frac{E\{Y\}}{a}$$



② Chebyshev Inequality

Let $z = (Y-\mu_Y)^2$ and apply Markov Inequality.
$z \geq 0$

$$P\{z > a^2\} \leq \frac{E\{z\}}{a^2}$$

$$P\{(Y-\mu_Y)^2 > a^2\} \leq \frac{\overbrace{E\{(Y-\mu_Y)^2\}}^{var(Y)}}{a^2}$$

$$P\{(Y-\mu_Y)^2 > a^2\} = \frac{var(Y)}{a^2}$$

$$\{Y: (Y-\mu_Y)^2 > a^2\} \underset{\text{equivalent}}{\longleftrightarrow} \{Y: |Y-\mu_Y| > a\}$$

$$\downarrow$$

$$P\{|Y-\mu_Y| \geq a\} \leq \frac{\sigma_Y^2}{a^2}$$

③ One-sided Chebyshev Inequality

$a > 0$

$$Y-\mu_Y \geq a \underset{b>0}{\longleftrightarrow} (Y-\mu_Y)+b \geq \underbrace{a+b}_{>0} \rightarrow \left((Y-\mu_Y)+b\right)^2 \geq (a+b)^2$$

$$\left(z > w \xrightarrow[z > w > 0]{} z^2 > zw > ww = w^2 \right.$$

$$P\{Y-\mu_Y > a\} = P\{(Y-\mu_Y)+b > a+b\}$$
$$\leq P\{[(Y-\mu_Y)+b]^2 > (a+b)^2\}$$
$$\overset{\text{markov}}{\leq} \frac{E\{[(Y-\mu_Y)+b]^2\}}{(a+b)^2} \quad \leftarrow \sigma_Y^2 + b^2 + 2E\{\overset{0}{\cancel{(Y-\mu_Y)}}\}b$$

$$P\{Y-\mu_Y \geq a\} \leq \frac{\sigma_Y^2+b^2}{(a+b)^2} \quad \text{for any } b>0$$

I can take derivative of $\frac{\sigma_y^2+b^2}{(a+b)^2}$ and select b such that $\frac{\sigma_y^2+b^2}{(a+b)^2}$ is minimized

$\rightarrow$ If I do that $b_* = \frac{\sigma_y^2}{a} \longrightarrow \boxed{P\{Y-\mu_Y \geq a\} \leq \frac{\sigma_y^2}{\sigma_y^2+a^2}}$

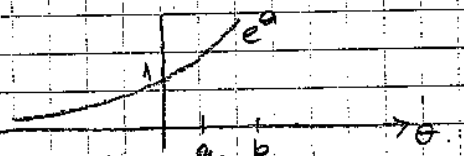
Chernoff Bound:

$P\{Y \geq a\} \leq e^{-ra} g_Y(r)$, $r>0$, $r \in ROC$

moment generating func. of Y.

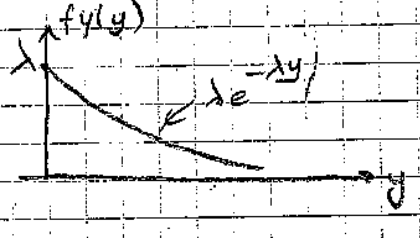
$P\{Y \leq a\} \leq e^{-ra} g_Y(r)$, $r<0$, $r \in ROC$

Proof: $Y \geq a \underset{(r\geq 0)}{\longleftrightarrow} rY \geq ra \rightarrow e^{rY} \geq e^{ra}$



$P\{Y \geq a\} = P\{e^{rY} \geq e^{ra}\} \underset{\text{Markov}}{\leq} \frac{E\{e^{rY}\}}{e^{ra}}$ $e^{-ra} g_Y(r)$ $\Rightarrow$ Note: This is valid for any $r \geq 0$ in ROC.

Ex: exp. dist $f_Y(y) = \lambda e^{-\lambda y} u(y)$



$\mu_y = 1/\lambda$

$\sigma_y^2 = 1/\lambda^2$

$g_Y(r) = \frac{\lambda}{\lambda - r}$

$P\{Y \geq k\mu_y\} = ?$ $(k \geq 2)$

Exact Result: $P\{Y \geq k\mu_y\} = \int_{k/\lambda}^{\infty} \lambda e^{-\lambda y} dy = \int_{y'=k}^{\infty} e^{-y'} dy' = \left.\frac{e^{-y'}}{-1}\right|_{y'=k}^{\infty}$

$= e^{-k}$

1. Markov Inequality

$$P\{Y > k\mu_y\} \le \frac{E\{Y\}}{k\mu_y} = \frac{1}{k}$$

$\swarrow \mu_y$

2. Chebyshev Inequality

$$P\{\underbrace{|Y-\mu_y| > (k-1)\mu_y}_{\equiv}\} \le \frac{\sigma_y^2}{(k-1)^2\mu_y^2} \swarrow 1/(k-1)^2$$

$(Y \ge k\mu_y)$

or

$y \le (2-k)\mu_y$

3. One sided Chebyshev Inequality

$$P\{Y-\mu_y > (k-1)\mu_y\} \le \frac{\sigma_y^2}{\sigma_y^2+(k-1)\mu_y^2} \swarrow \frac{1}{1+(k-1)^2}$$

4. Chernoff

$$P\{Y \ge k\mu_y\} \le e^{-k\mu_y r}\, g_Y(r) \quad \text{Let's minimize RHS wrt "r"}$$

$$\le \underbrace{e^{-k\mu_y r}\frac{\lambda}{\lambda-r}}_{RHS}$$

$$\frac{\partial}{\partial r}\left(e^{-\frac{k}{\lambda}r}\cdot\frac{\lambda}{\lambda-r}\right) = 0 \quad (*)$$

$(*)$ is satisfied at $r^* = \lambda\left(1-\frac{1}{k}\right)$

by substituting r^* in RHS

$$P\{Y \ge k\mu_y\} \underset{\text{chernoff}}{\le} k\, e^{-(k-1)}$$

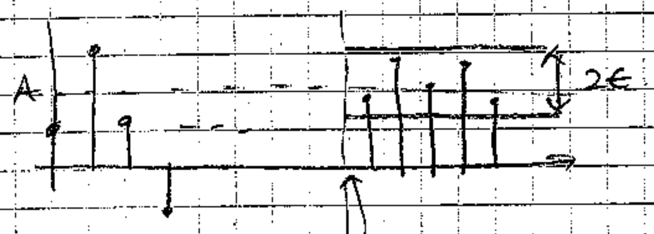
STOCHASTIC CONVERGENCE:

$X_1, X_2, \ldots$ a sequence of r.v

Q: Does X_k as $k \to \infty$ converge in some sense to a r.v?

Remember: ① Convergence in real numbers:

$\lim_{n\to\infty} a_n = A \longleftrightarrow$ for any ϵ, there exists an $N(\epsilon)$ such that $|a_n - A| < \epsilon$, $n > N(\epsilon)$



Ex: $\lim_{n\to\infty} \frac{1}{n} = 0$, Given ϵ $\left|\frac{1}{n} - 0\right| < \epsilon \rightarrow n > 1/\epsilon$

$N(\epsilon) = \underbrace{\left\lceil \frac{1}{\epsilon} \right\rceil}_{\text{ceiling function}}$

$\lceil 2,1 \rceil = 3$

② Convergence of Functions

i) Pointwise Convergence

If $f_k(x) \xrightarrow[\text{pointwise}]{} f(x)$,

then for a given "x", the sequence of numbers $f_k(x) \xrightarrow[k\to\infty]{} f(x)$, that is the convergence of real numbers for a fixed x.

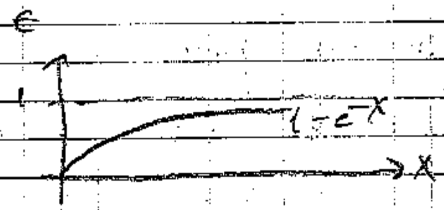
ii) Uniform Convergence

$f_k(x) \xrightarrow[u]{} f(x)$, Given ϵ, there exists an $N(\epsilon)$ such that

$\sup |f_k(x) - f(x)| < \epsilon$

$k > N(\epsilon)$

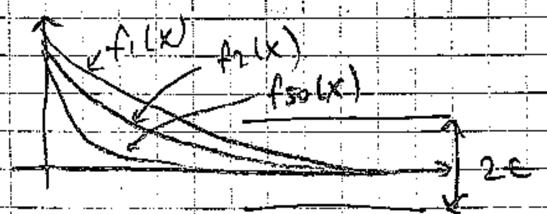
sup $\rightarrow$ supremum (maximum)



$\sup_{x>0} (1 - e^{-x}) = 1$

(lowest upper bound)

Ex: $f_k(x) = e^{-kx}$, $x \geq 0$



claim: $f_k(x) \xrightarrow[k\to\infty]{u} 0$

$f_k(x) = e^{-kx}$

Choose $k > \ln(1/\epsilon) \rightarrow f_k(x) = e^{-kx} < e^{-k}$ for $x > 1$

and $e^{-k} < \epsilon$ if $k > \ln(1/\epsilon)$

so, by choosing $k > \ln(1/\epsilon)$

we satisfy uniform convergence condition for a given ϵ.

STOCHASTIC CONVERGENCE

1. Convergence in Distribution

$$\underbrace{F_{Z_k}(z)}_{CDF} \longrightarrow \underbrace{F_Z(z)}_{CDF}$$

$Z_1, Z_2, \ldots$ r.v's
Z is the limiting r.v

$\Bigg\}\quad Z_k \xrightarrow[k\to\infty]{} Z$ in distribution

if $F_{Z_k}(z) \rightarrow F_Z(z)$ at each z for which $F_Z(z)$ is continuous.

Ex: Central Limit theorem:

Let $X_1, X_2, \ldots$ i.i.d with finite mean $\overline{X}$ and variance σ_X^2

Then,

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

CDF of S_n

$$\lim_{n\to\infty} P\left\{ \frac{S_n - \overline{X}}{\sigma_X \sqrt{n}} \le z \right\} = \underbrace{\Phi(z)}_{\text{CDF of } N(0,1)}$$

where $\Phi(z) = \displaystyle\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-(z')^2/2} dz'$

NOTE: $\theta = \dfrac{S_n - n\overline{X}}{\sigma_X \sqrt{n}}$

$E(\theta) = 0$, since $E\{S_n\} = n\overline{X}$

$var(\theta) = 1$, since $var\{S_n\} = n\sigma_X^2$

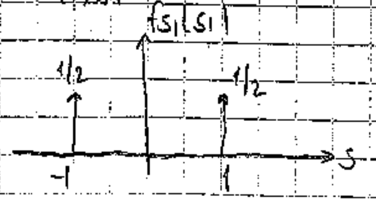
and $var\{\theta\} = \dfrac{var\{S_n\}}{\sigma_X^2\, n} = 1$

Ex: $X = \begin{cases} 1 & \text{with prob. } 1/2 \\ -1 & \text{with prob. } 1/2 \end{cases}$, X_k i.i.d

$E\{X\} = 0$, $\dfrac{E\{X^2\}}{\sigma_X^2} = 1$

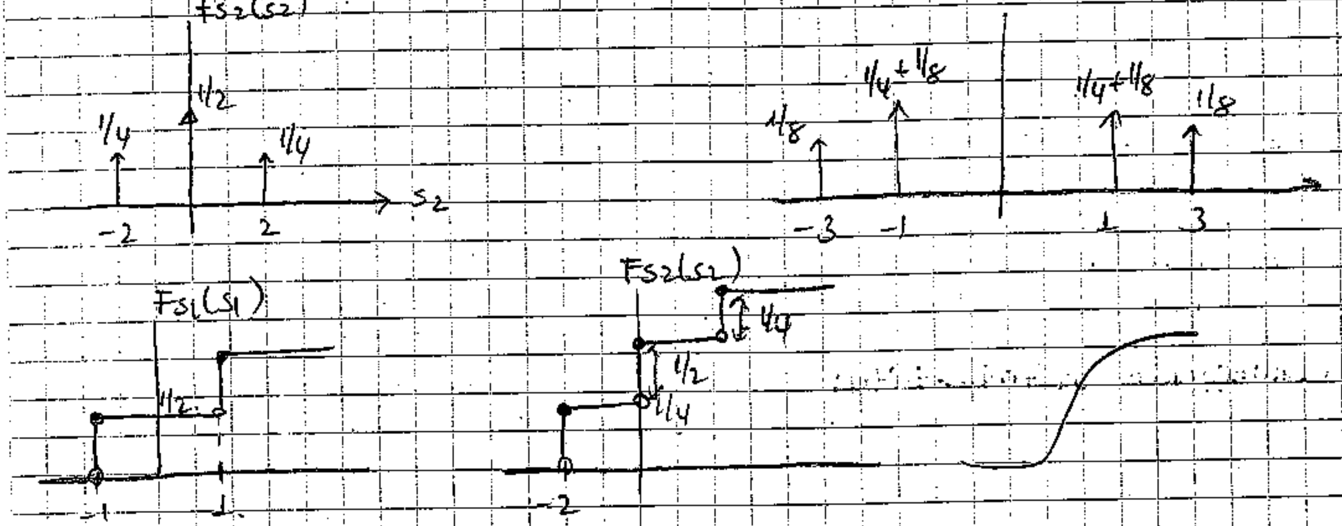
$S_n = X_1 + X_2 + \cdots + X_n$

$\boxed{n=1 \quad S_1 = X_1}$



$S_2 = X_1 + X_2$

Remember $f_{S_2}(s_2) = f_X(s_2) * f_X(s_2)$ $\quad S_3 = X_1 + X_2 + X_3$



② Convergence in Probability

A sequence of rv's converges to z in probability

if $\lim_{n\to\infty} P\{|z_n - z| > \epsilon\} = 0$ for any $\epsilon > 0$

Notes: ① Let's call

$\hat{z}_n = z_n - z$
↑
another r.v

We can show that if $\hat{z}_n \to 0$ in probability

then $z_n \to z$ in probability

② call $a_k = P\{|\hat{z}_k| > \epsilon\}$ then

convergence in probability is $\lim_{k\to\infty} a_k = 0$ for every ϵ

20/10/2014

STOCHASTIC CONVERGENCE

(1) Conv. In Distribution ← EX: central Limit Theorem

(2) Conv. In Probability

$Z_1, Z_2, Z_3, \ldots$ a sequence of r.v.

$$Z_k \xrightarrow[k\to\infty]{?} Z$$

if converges, in what sense?

Convergence In Probability:

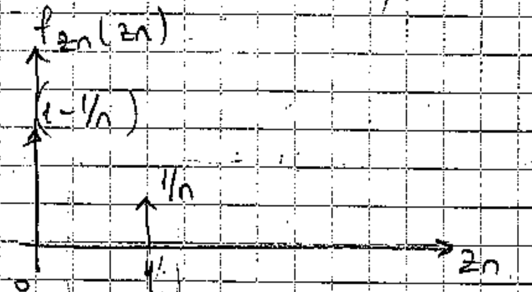
$$\lim_{n\to\infty} P\{\underbrace{|Z_n - Z|}_{error_n} > \epsilon\} = 0 \quad \forall \epsilon \rightarrow Z_n \xrightarrow[P]{} Z \text{ in probability.}$$

$a_n = P\{|Z_n - Z| > \epsilon\}$, then conv. In prob. is as simple as

$$\lim_{n\to\infty} a_n \overset{?}{=} 0$$

EX: $Z_n = \begin{cases} 1 & \text{with prob. } 1/n \\ 0 & \text{" prob. } 1-1/n \end{cases}$

Assume Z_n's are independent.



Let's show $Z_n \xrightarrow[P]{} 0$

$$P\{|Z_n - 0| > \epsilon\} = \begin{cases} 0 & \epsilon > 1 \\ 1/n & 0 < \epsilon \leq 1 \end{cases}$$

$$\lim_{n\to\infty} P\{\underbrace{|z_n - 0| > \epsilon}_{1/n}\} = 0 \quad \text{elect}$$

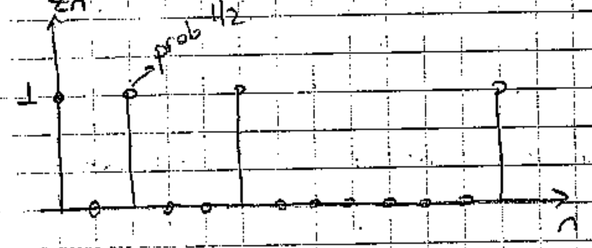
↓

converges in probability

z_1, z_2, z_3, z_4

Sample Path: $1, \{0,1\}, \{0,1\}, \{0,1\}$

$1/2$ prob, $1/3$ prob, $1/4$ prob.



③ Convergence In Mean Square:

If $\lim_{n\to\infty} E\{(z_n - z)^2\} = 0$, Then $z_n \xrightarrow[m.s]{} z$ in mean-square.

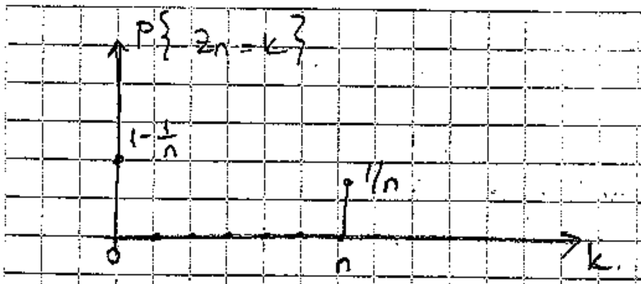
Ex: Let's apply earlier exaple to check conv. in M.S.

$$E\{(z_n - 0)^2\} = 1^2 . P\{z_n = 1\} + 0^2 P\{z_n = 0\}$$

$= \frac{1}{n}$ → So $z_n \xrightarrow[m.s]{} 0$ in mean-square sense also.

Ex: Let's modify earlier exaple as;

$$z_n = \begin{cases} n & \text{with prob } 1/n \\ 0 & \text{with prob } 1 - 1/n \end{cases}$$



so $z_n \xrightarrow[P]{} 0.$

but $z_n \not\xrightarrow[m.s]{} 0$

$$E\{(z_n - 0)^2\} = n^2 \underbrace{P\{z_n = n\}}_{1/n} + 0^2 \cdot P\{z_n = 0\}$$

$$= n$$

$$\lim_{n\to\infty} E\{(z_n - 0)^2\} = \infty !$$

<u>Claim:</u> If $z_n \xrightarrow[m.s]{} z$ in mean square sense,

then $z_n \xrightarrow[P]{} z$ in probability.

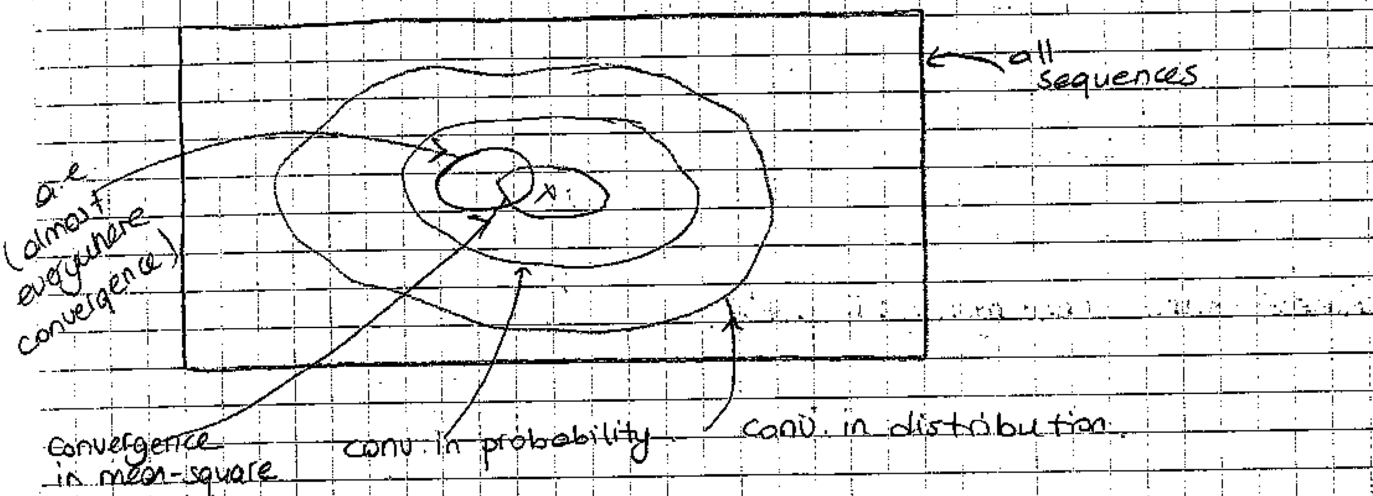
since; $P\{|z_n - z| > \epsilon\} \leq \dfrac{E\{(z_n - z)^2\}}{\epsilon^2}$

Chebyshev inequality.

So, taking limit as $n \to \infty$

$$\lim_{n\to\infty} P\{|z_n - z| > \epsilon\} \leq \lim_{n\to\infty} \left\{ \frac{E\{(z_n - z)^2\}}{\epsilon^2} \right\}$$

Relationships Between Convergence Modes



④ Almost Everywhere convergence / convergence with prob. 1 / Almost sure convergence

A random experiment is said to converge with prob. 1

if $P\{w \in \Omega : \underbrace{\lim_{n\to\infty} z_n(w) = z(w)}\} = 1$ → def. of r.v $z(w)$

Ω ← sample space

for a fixed w this is an ordinary limit operation

definition of r.v. z_n

Ex: $\Omega = [0, 1]$

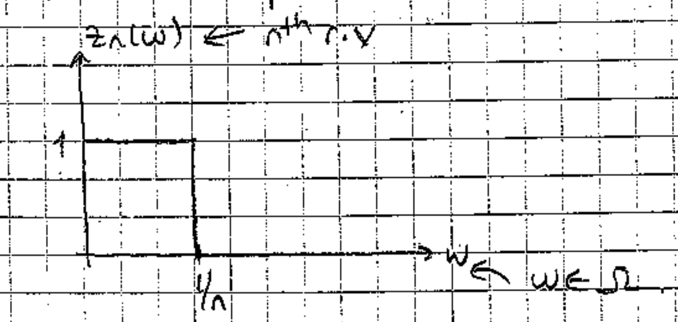
F = measurable sets in $[0,1]$

P = prob. assignment associated with length of interval

} sample space

$P\{w \in [a,b]\} > b-a$

$z_n(w)$ ← n^{th} r.v



z_n is two-valued r.v

$z_n = \{0, 1\}$

22/10/2014

$P\{z_n(w) = 0\} = 1 - \frac{1}{n}$

$P\{z_n(w) = 1\} = 1/n$

Q: Do I have $z_n \to 0$ with prob. 1?

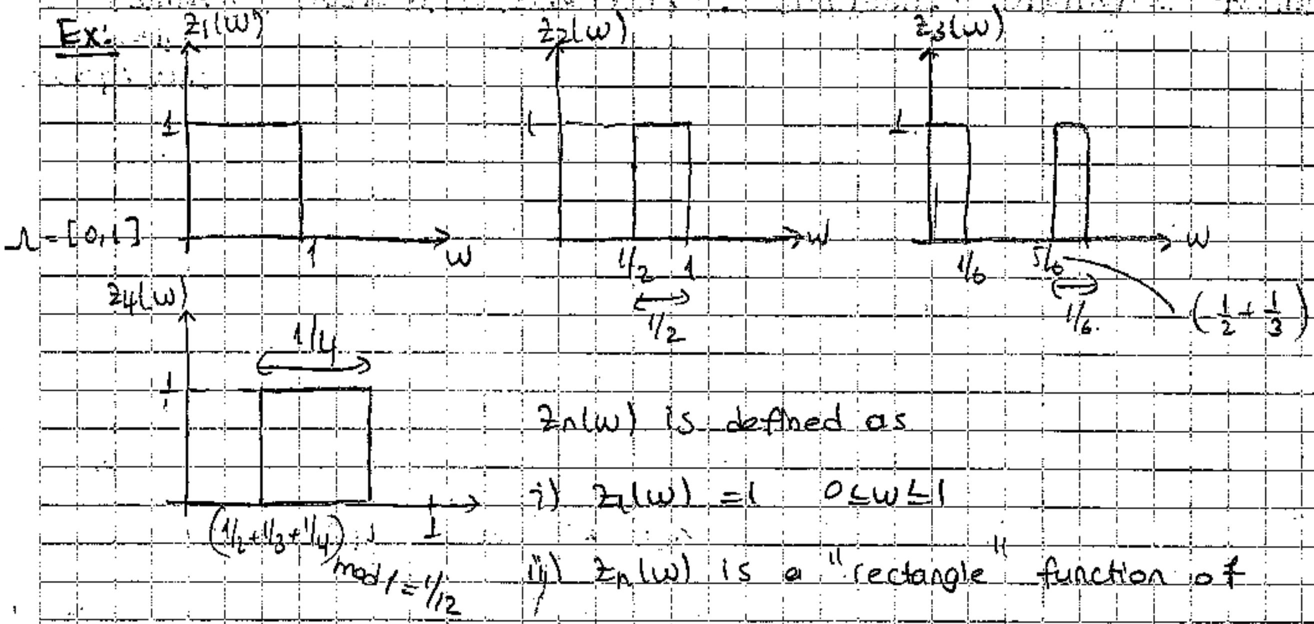
$\downarrow$ $\quad p(w=0) = 0$

Yes, apart from $w=0$, all outcomes $z_n(w)$ become exactly 0 for sufficiently large "n"

Almost sure convergence: (cont'd)

$$z_n \xrightarrow[as]{} z \iff P\{w \in \Omega : \underbrace{\lim_{n\to\infty} z_n(w) = z(w)}_{\text{ordinary limit for a fixed } w}\} = 1$$

Ex:



$z_n(w)$ is defined as

i) $z_1(w) = 1 \quad 0 \leq w \leq 1$

ii) $z_n(w)$ is a "rectangle" function of length $1/n$ and starting at

$\left(\frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}\right)$ mod 1, and rect. func. may do a circular shift

$(5.25 \text{ mod } 1 \triangleq 0.25)$

$P\{z_2(\omega)=1\}=1/2$

$P\{z_2(\omega)=0\}=1/2$

$P\{\underbrace{z_2(\omega)=1,\ z_3(\omega)=0}_{\omega\in[\frac{1}{2},\frac{5}{6}]}\}=1/3$

$P\{z_3=1\}=1/3$

$P\{z_3=0\}=2/3$

$P\{z_n=1\}=1/n$

Q: $z_n \xrightarrow[a.s]{} 0$

$\omega=0.99 \to z_1(0.99),\ z_2(0.99),\ z_3(0.99)$

$[1,\ 1,\ 1,\ 0,$

Since $\sum 1/n$ diverges, the rectangular func. gets thinner, but it rotates infinite number of times in $[0,1]$ interval $\to$ so,

z_n does not converge to 0 with prob. 1.

Convergence with prob. 1 is difficult to check and it requires going back to the definition of r.v. There are some sufficient conditions which are simpler to use and if satisfied guarantees a.s convergence.

(S1) $\sum_n P\{|z_n-z|>\epsilon\}<\infty,\ \forall\epsilon \to z_n\to z$ a.s (Borel-Cantelli Lemma)

(S2) z_n's r.v's with finite expectation

$$\sum_{n=1}^{\infty} E\{|z_n-z|\}<\infty \to z_n \xrightarrow[a.s.]{} z$$ (Textbook p. 218 Lemma 5.2.1)

EX:
$$z_n = \begin{cases} 1 & \text{with prob } 1/n^2 \\ 0 & \text{with prob } 1-1/n^2 \end{cases}$$

Q: $z_n \xrightarrow[a.s]{} 0$?

Apply S_1: $P\{|z_n - 0| > \epsilon\} = \begin{cases} 0 & \epsilon > 1 \\ 1/n^2 & 0 < \epsilon < 1 \end{cases}$

$$\sum_{n=1}^{\infty} P\{|z_n - 0| > \epsilon\} = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} < \infty$$

$$\downarrow$$

$$z_n \xrightarrow[a.s]{} 0$$

Weak Law of Large Numbers:

Let $S_n = X_1 + X_2 + \dots + X_n$ where

X_k's are iid rv's with finite variance σ_X^2, then

$$\boxed{X_k \sim f_X(x_k)}$$

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - \bar{X}\right| > \epsilon\right\} = 0 \quad \forall \epsilon$$

; i.e. $\frac{S_n}{n} \xrightarrow[P]{} \bar{X}$ $\quad E\{X\} = \bar{X}$

Proof: Remember mean-square convergence guarantees conv. in prob.

Do I have M.S conv. ?

$$\lim_{n\to\infty} E\left\{\left(\frac{S_n}{n} - \bar{X}\right)^2\right\} \overset{?}{=} 0$$

$$\frac{S_n}{n} - \bar{X} = \frac{S_n - n\bar{X}}{n}$$

$$= \frac{(X_1 - \bar{X}) + (X_2 - \bar{X}) + \dots + (X_n - \bar{X})}{n}$$

$$E\left\{\left(\frac{S_n}{n}-\bar{X}\right)^2\right\} = \frac{E\left\{\left[(X_1-\bar{X})+(X_2-\bar{X})+\cdots+(X_n-\bar{X})\right]^2\right\}}{n^2}$$

$$= \frac{E\left\{(X_1-\bar{X})^2+(X_2-\bar{X})^2+\cdots+(X_n-\bar{X})^2+2(X_1-\bar{X})(X_2-\bar{X})+\cdots\right\}}{n^2}$$

$$= \frac{\sigma_X^2+\sigma_X^2+\cdots+\sigma_X^2+0+0+\cdots}{n} = \frac{n\cdot\sigma_X^2}{n^2} = \frac{\sigma_X^2}{n}$$

$$\lim_{n\to\infty} \underbrace{E\left\{\left(\frac{S_n}{n}-\bar{X}\right)^2\right\}}_{\frac{\sigma_X^2}{n}} = 0. \longrightarrow$$

conclusion: We have M.S conv. to $\bar{X}$ and therefore conv. in prob. to $\bar{X}$.

Comments:

① $\frac{S_n}{n}$ is nothing but sample mean $\left(\hat{\bar{X}} = \frac{1}{N}\sum_{n=1}^{N} x[n]\right)$ $\uparrow \simeq E\{x\}$

and $\frac{S_n}{n} \to E\{X\}$ is the basis of computer experiments or Monte-Carlo trials conducted to "calculate" $\bar{X}$.

$$\bar{X} = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

② $X_k = 1A_k = \begin{cases} 1 & A_k \text{ happens} \\ 0 & \text{otherwise} \end{cases} \rightarrow \frac{S_n}{n} = \underbrace{\frac{\#A_k \text{ happens}}{n}}_{\text{relative frequency of event } A}$

$\rightarrow E\{X\} = 1.P\{A \text{ happening}\} + 0.P\{A \text{ not happening}\}$

$= P\{A \text{ happening}\}$ ← relative frequency interpretation of probability

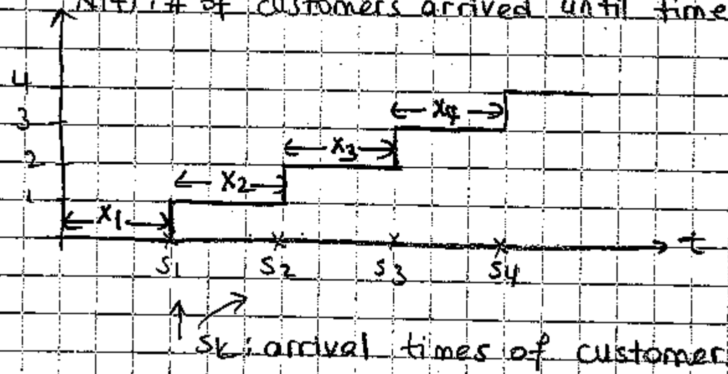
③ Weak law of large numbers can be extended to dependent X_k's and X_k's having infinite variance.

④ There is a strong form for law of Large Numbers, for X_k i.i.d and $E\{X_k^4\} < \infty$,

$\frac{S_n}{n} \xrightarrow[a.s]{} \bar{X}$ ← strong law of Large Numbers.

- Poisson Processes -

A member of "arrival" processes. (such as arrival times of customers to a shop.)



* S_k: arrival time / epoch

* $\left.\begin{array}{l} X_1 = S_1 \\ X_k = S_k - S_{k-1} \end{array}\right\}$ Interarrival time **Note:** $S_n = \sum_{k=1}^{\infty} X_k$

Hence,

$$\underbrace{S_1, S_2, --, S_n}_{\text{Arrival Times}} \longleftrightarrow \underbrace{X_1, X_2, --, X_n}_{\text{Interarrival Times}}$$

So, knowing either of one of them gives the other one. So Joint pdf of either one is should be sufficient to find joint pdf of other one.

* $N(t)$: Counting r.v., $N(t)$ starts from 0 at $t=0$ and incremented by 1 at different "t" values.

$N(t)$: # arrivals until time t and including time "t".

* $\underbrace{\{S_n \le t\}}_{\substack{n^{th} \text{ customer} \\ \text{has arrived before} \\ t}} = \underbrace{\{N(t) \ge n\}}_{\substack{\text{at time } t \\ \text{I have n or} \\ \text{more customers.}}}$

complement $\{S_n > t\} = \{N(t) < n\}$ *

Renewal Process: An arrival process with i.i.d interarrival times (X_i's) are called renewal process.

Poisson Process: A renewal process with exponential PDF i.e.;

$$f_{\bar{X}}(x) = \lambda e^{-\lambda x} u(x) \quad (\lambda\text{: rate of the process})$$

* Properties of Poisson Processes

1. Memoryless: A r.v is memoryless if

$$P\{\underbrace{\bar{X}}_{\text{waiting time}} > t+x\} = P\{\bar{X} > x\}\, P\{\bar{X} > t\}$$

or

$$P\{\bar{X} > t+x \mid \bar{X} > t\} = P\{\bar{X} > x\}$$

Exponential distribution is memoryless, since

$$P\{\underbrace{\bar{X}}_{\exp(\lambda)} > z\} = e^{-\lambda z}$$

Note that only exp. r.v satisfies memoryless property.

Poisson Process:

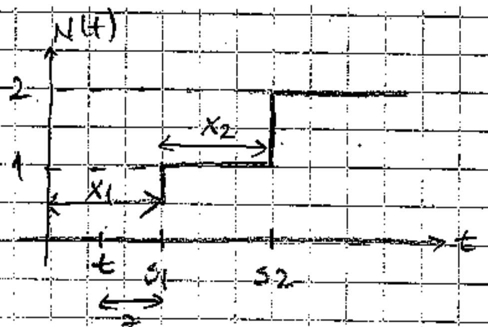
Remember, Poisson process is a renewal process with exponential inter arrival distribution (exp. dist is memoryless, i.e

$P\{ X > t+z \mid X > t \} = P\{ X > z \}$)

Theorem 2.2.5: For a poisson process at any time "t", the first arrival after "t" (waiting time) is independent of $N(t)$ and all arrival epochs before t. It's also independent of r.v's $N(t_1), N(t_2), \ldots N(t_k)$ s.t $t_k < t$

Proof: case

(a) $N(t) = 0$



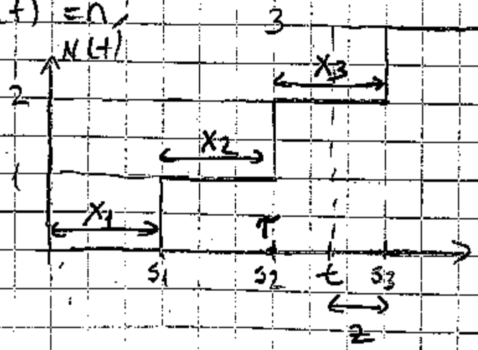
← additional waiting time until first arrival.

$$P\{ Z > z \mid N(t) = 0 \} = P\{ X_1 > t+z \mid N(t) = 0 \}$$

$$= P\{ X_1 > t+z \mid X_1 > t \}$$

$$= P\{ X_1 > z \} = e^{-\lambda z}$$

(b) $N(t) = n$,



(*) $P\{Z > z \mid N(t) = n, \underbrace{S_n = \tau}_{\text{last arrival time is } \tau}\} = P\{X_{n+1} > t - \tau + z \mid N(t) = n, S_n = \tau\}$

$= P\{X_{n+1} > t - \tau + z \mid X_{n+1} > t - \tau, S_n = \tau\}$

$= P\{X_{n+1} > t - \tau + z \mid X_{n+1} > t - \tau\}$

$= P\{X_{n+1} > z\} = e^{-\lambda z}$

Note: The same argument (*) holds when conditioning is not only on S_n, but also on $S_1, S_2, S_3, \ldots, S_n$.

The $S_1, S_2, \ldots, S_n$ information is equivalent to $N(t')$, $t' < t$

$P\{Z > z \mid \{N(\tau) : 0 < \tau \leq t\}\} = e^{-\lambda z}$

So, "additional waiting time" Z is independent of $N(\tau)$ for $0 < \tau \leq t$.

Definition: Stationary Increment / A counting process is called stationary increment if $N(t') - N(t) = N(t' - t) - N(0)^{0}$ for all $0 < t < t'$

Poisson process is stationary increments, since

\# arrivals in $t' - t$ seconds $= N(t') - N(t)$

is independent of t' and t, but depends only on $t' - t$ (waiting period)

Definition: Independent Increments

$\{N(t) : t > 0\}$ is independent increments

if for every k

$0 < t_1 < t_2 < \dots < t_k$

$N(t_1), \underbrace{N(t_1, t_2)}_{\substack{\#\text{ of arrivals} \\ \text{in} \\ (t_1, t_2] \\ \text{interval}}}, N(t_2, t_3), \dots, N(t_{k-1}, t_k)$

are independent from each other.

Conclusion: Poisson process is stationary and independent increments process.

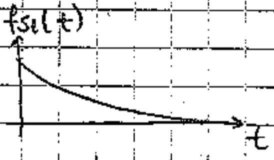
Probability Density of n^{th} arrival:

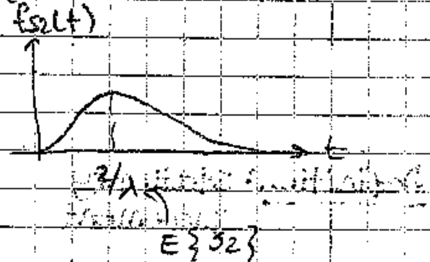
$S_n = X_1 + X_2 + \dots + X_n$

X_k's i.i.d and exp. dist.

$$f_{S_n}(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$$

n^{th} arrival time

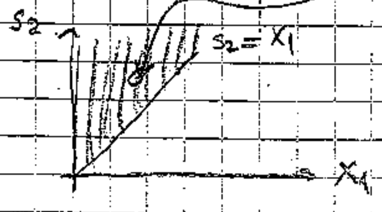




Note: $f_{X_1, S_2}(x_1, s_2) = f_{X_1}(x_1)\, f_{S_2 | X_1}(s_2 | x_1)$

$= \lambda e^{-\lambda x_1} \cdot \underbrace{\lambda e^{-\lambda(s_2 - x_1)}}_{f_{X_2}(s_2 - x_1)}$

$= \lambda^2 e^{-\lambda s_2}, \quad 0 < x_1 < s_2$



$$f_{s_2}(s_2) = \int_{-\infty}^{\infty} f_{X_1,s_2}(x_1, s_2)\, dx_1 = \int_{0}^{s_2} \lambda^2 e^{-\lambda s_2} dx_1$$

$$= s_2 \lambda^2 e^{-\lambda s_2} \quad , \quad s_2 > 0.$$

$s_2 = X_1 + X_2$

$s_2 = 5 + x_2$ given $f_{x_2}(x_2)$

Find $f_{s_2}(s_2)$

$f_{s_2}(s_2) = f_{X_2}(s_2 - 5)$

Review: * 1 func. of 1 r.v

* 1 func. of 2 r.v $\quad \rho^2 = x^2 + y^2$

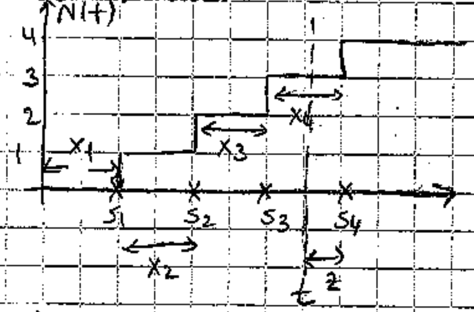
* 2 func. of 2 r.v $\quad \rho^2 = x^2 + y^2$

$\phi = \tan^{-1}(y/x)$

Review:

03/11/2014

Poisson Process:



X_k: i.i.d exp. distributed with param λ

$$f_{X_k}(x_k) = \lambda e^{-\lambda x_k}, \quad x_k > 0$$

$$P\{ Z > z \mid N(t) = 3 \} = \lambda e^{-\lambda z}$$

Properties: ① Ind. Increments

$\tilde{N}(t_1, t_2)$ and $\tilde{N}(t_3, t_4) \rightarrow$ equivalent to $\tilde{N}(0, t_4 - t_3)$

arrivals in $(t_1, t_2]$

arrivals in $(t_3, t_4]$

$N(t_4 - t_3) \leftarrow$ # arrivals $(0, t_4 - t_3]$

② Stationary Increments

$$f_{S_n}(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!} \leftarrow \text{Erlang distribution (Chi-square with even degrees of freedom)}$$

$$S_n = X_1 + X_2 + \dots + X_n$$

Probability Mass Function Of $N(t)$:

Th. 2.2.10 p.79 Textbook $\Rightarrow$ For a poisson process with rate λ, $N(t)$ (# arrivals in $(0, t]$) is given by Poisson r.v.

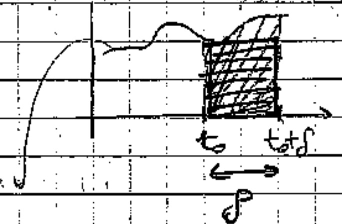
$$P_{N(t)}(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

03/11/2014

Proof: We know $f_{S_n}(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$

I will calculate $P\{t < S_{n+1} \le t+\delta\}$ in two different ways

① $P\{t < S_{n+1} \le t+\delta\} = \int_t^{t+\delta} f_{S_{n+1}}(z)\,dz = f_{S_{n+1}}(t)\,\delta + o(\delta)$, δ: small quantity



$f(t_0) + f'(t_0)(t-t_0) + \frac{f''(t_0)(t-t_0)^2}{2!}$

Remember!

* $o(\delta)$ represents functions of δ s.t. $g(\delta)$ with the property

$$\lim_{\delta \to 0} \frac{g(\delta)}{\delta} = 0$$

Ex. δ^2 is $o(\delta)$

Ex. $\sin(\delta)$ is not $o(\delta)$

② $P\{t < S_{n+1} \le t+\delta\} = P\{(N(t)=n) \text{ and } (1 \text{ arrival in } (t,t+\delta])\}$
$+ P\{(N(t)=n-1) \text{ and } (2 \text{ arrival in } (t,t+\delta])\}$
$+$
$\vdots$
$+$
$P\{(N(t)=0) \text{ and } (n+1 \text{ arrival in } (t,t+\delta])\}$

$P\{N(t)=n, 1 \text{ arrival in } (t,t+\delta]\} = P\{1 \text{ arrival} \mid N(t)=n\}\, P\{N(t)=n\}$ indep. increments property

$= P\{1 \text{ arrival in } (t,t+\delta]\} \cdot P\{N(t)=n\}$ stationary increments property

$= P\{1 \text{ arrival in } (0,\delta]\}\; P\{N(t)=n\}$

$$= \left(\int_0^\delta f_{S_1}(z)\,dz\right) P\{N(t)=n\}$$

$$= \left[f_{S_1}(0)\,\delta + o(\delta)\right] \cdot P\{N(t)=n\}$$

$$= \left[\lambda\delta + o(\delta)\right] \cdot P\{N(t)=n\}$$

$f_{S_n}(z) = \frac{\lambda^n z^{n-1} e^{-\lambda z}}{(n-1)!}$

$f_{S_1}(z) = \lambda e^{-\lambda z} = \lambda\left(1 - \lambda z + \frac{(\lambda z)^2}{2!} - \dots\right)$

$f_{S_2}(z) = \lambda^2 z e^{-\lambda z}$

$$P\{(N(t)=n-1) \text{ and } 2 \text{ arrivals}\} = \Big(\underbrace{\int_0^\delta f_{S_2}(\tau)\,d\tau}_{\cancel{f_{S_2}(\delta)}\,0\;\delta + o(\delta)}\Big) P\{N(t)=n-1\} = o(\delta)$$

Conclusion For (2)

(2) $P\{t < S_n \le t+\delta\} = (\lambda\delta)\,P\{N(t)=n\} + o(\delta)$

Equate (1) and (2)

$$f_{S_{n+1}}(t)\,\delta + o(\delta) = \lambda\delta\, P\{N(t)=n\} + o(\delta)$$

$\cdot\frac{1}{\delta}$

$$f_{S_{n+1}}(t) + \frac{o(\delta)}{\delta} = \lambda P\{N(t)=n\} + \frac{o(\delta)}{\delta}$$

$$\lim_{\delta\to 0} \qquad f_{S_{n+1}}(t) = \lambda\, P\{N(t)=n\}$$

$$P\{N(t)=n\} = \frac{f_{S_{n+1}}(t)}{\lambda} = \frac{\lambda^{n+1} t^n e^{-\lambda t}}{\lambda\, n!}$$

$$= \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

Also check textbook p.80 for a second proof. Poisson distr. !!!

Poisson R.V Properties:

PMF: $P_N\{N=n\} = \dfrac{\lambda^n e^{-\lambda}}{n!}$, $n=\{0,1,2,\ldots\}$

poisson

Mean: $E\{N\} = \lambda$

Var: $\mathrm{var}\{N\} = \lambda$

MGF: $g_N(r) = e^{\lambda(e^r - 1)}$

moment generating function.

Notes:

1. For poisson process with rate λ (λ unit is arrivals/sec)

then λt has the unit of arrivals

then $P\{N(t)=n\} = \frac{(\lambda t)^n \cdot e^{-\lambda t}}{n!}$

$E\{\underbrace{N(t)}_{\text{poisson process}}\} = \lambda t$ rate par of poisson process (arrival/sec)

(rate discussion, last lecture)

when two ind Poisson r.v are added the resultant r.v is also Poisson with rate $\lambda_1+\lambda_2$

[see MGF $\rightarrow g_n(r) = e^{\lambda(e^r-1)}$

λ: arrivals/sec $\longrightarrow$ $\lambda t \leftarrow$ # arrivals

$E\{N(t)\} = \lambda t$

$var\{N(t)\} = \lambda t$

$N = N_1 + N_2$

$\downarrow$

poisson has rate $\lambda_1+\lambda_2$

Poisson r.v is in some ways "discrete analog" of Gaussian r.v, that is under mild condition several counting processes (not necessarily Poisson), when summed approach to a Poisson process (similar to CLT for gaussian r.v)

Def #1

Renewal Process with Exp Interarrival times

$\longrightarrow$ ✓ $\qquad \longleftarrow$ ex 2.4

Def #2

$N(t)$: poisson r.v with rate λt

$N(t)$ ind and stationary increments

$\longrightarrow$ ✓ $\qquad \longleftarrow$ ex 2.7

Def #3

$P\{N(t,t+\delta)=0\} = 1-\lambda\delta+o(\delta)$

$P\{N(t,t+\delta)=1\} = \lambda\delta+o(\delta)$

$P\{N(t,t+\delta)\geq 2\} = o(\delta)$

$N(t)$ ind and stationary increments.

$N_1(t)$ and $N_2(t)$ are ind. Poisson processes.

(Two counting processes are ind if for every N

$0<t_1<t_2<\dots<t_N$

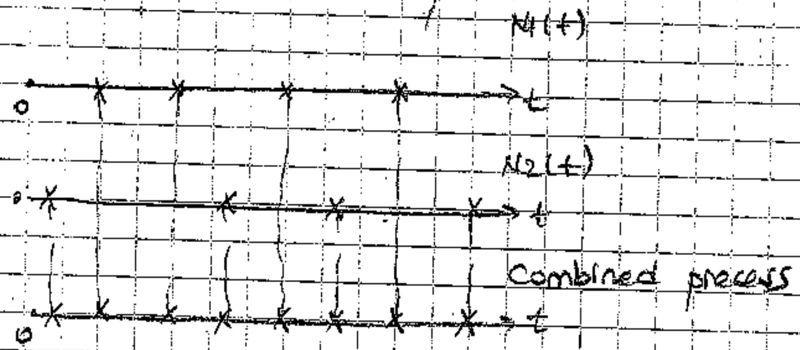
$N_1(t_1), N_1(t_2), \dots, N_1(t_N) \leftarrow$ Joint PMF

$\updownarrow$ independent

$N_2(t_1), N_2(t_2), \dots, N_2(t_N) \leftarrow$ Joint PMF

then $N(t) = N_1(t) + N_2(t)$ is Poisson with rate $\lambda_1+\lambda_2$

combined process



$$P\{N(t,t+\delta)=0\} = 1-\lambda\delta + o(\delta)$$

$$P\{N_1(t,t+\delta)=0,\ N_2(t,t+\delta)=0\} = P\{N_1(t,t+\delta)=0\}\, P\{N_2(t,t+\delta)=0\}$$

$$= [1-\lambda_1\delta + o(\delta)][1-\lambda_2\delta + o(\delta)]$$

$$1-\underbrace{(\lambda_1+\lambda_2)}_{\lambda new}\delta + o(\delta)$$

$$P\{N(t,t+\delta)=1\} = P\{N_1(t,t+\delta)=1,\ N_2(t,t+\delta)=0\}$$

$$+$$

$$P\{N_1(t,t+\delta)=0,\ N_2(t,t+\delta)=1\}$$

$$= [\lambda_1\delta + o(\delta)][1-\lambda_2\delta + o(\delta)]$$

$$+$$

$$[1-\lambda_1\delta + o(\delta)][\lambda_2\delta + o(\delta)]$$

$$= (\lambda_1\delta + o(\delta)) + (\lambda_2\delta + o(\delta))$$

$$= \underbrace{(\lambda_1+\lambda_2)}_{\lambda new}\delta + o(\delta)$$

$$P\{N(t,t+\delta)\geq 2\} = o(\delta)$$

since $N_1(t)$ and $N_2(t)$ are poisson dist. r.v and ind

→ their sum is also Poisson dist. r.v

Let X be the interarrival time of combined process.

$$P\{X > x\} = P\{N_1(t, t+x) = 0,\ N_2(t, t+x) = 0\}$$

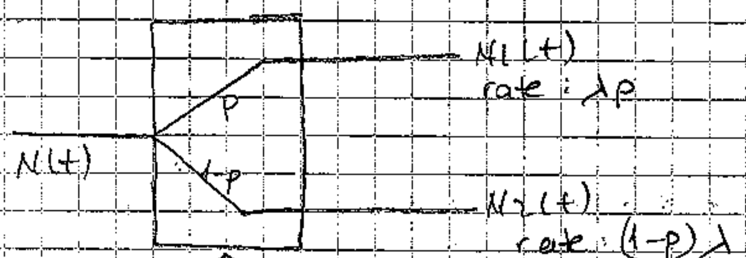
waiting time for next arrival for the combined process

$$= P\{N_1(t, t+x) = 0\}\, P\{N_2(t, t+x) = 0\}$$

$$= e^{-\lambda_1 x} \cdot e^{-\lambda_2 x}$$

$$= e^{-\underbrace{(\lambda_1+\lambda_2)}_{\lambda new}x}$$

So, waiting time is exp. distributed for combined process.



p: prob. of switching input to $N_1(t)$

(p: Prob. Head)

and a coin toss experiment operates the switch

ind. from $N(t)$

$N_1(t)$ and $N_2(t)$ are Poisson Process with rate $\lambda_1 = \lambda p$ and $\lambda_2 = (1-p)\lambda$.

Furthermore, $N_1(t)$ and $N_2(t)$ are independent.

Let's show $N_1(t)$ is Poisson

$$P\{N_1(t,t+\delta)=1\} = P\{N(t,t+\delta)=1, \text{ switched to } 1\}$$
$$+$$
$$P\{N(t,t+\delta)\geq 2, \text{ one of them switched to } 1\}$$

$$= [\lambda\delta + o(\delta)]\cdot p + o(\delta)$$

$$= (\lambda p)\delta + o(\delta)$$

$$P\{N_1(t,t+\delta)=0\} = P\{N(t,t+\delta)=1, \text{ switched to } 2\}$$
$$+$$
$$P\{N(t,t+\delta)\geq 2, \text{ All of them switched to } 2\}$$
$$+$$
$$P\{N(t,t+\delta)=0\}$$

$$= [\lambda\delta(1-p) + o(\delta)] + o(\delta) + (1-\lambda\delta + o(\delta))$$

$$= 1 - \underbrace{(\lambda p)}_{\lambda_1}\delta + o(\delta)$$

$$P\{N_1(t,t+\delta)\geq 2\} = o(\delta)$$ then $N_1(t)$ is Poisson with rate $\lambda_1 = \lambda p$ (similarly for $N_2(t)$)

Proof for $N_1(t)$ and $N_2(t)$ are independent

$$P\{N_1(t)=m, N_2(t)=k \mid N(t)=m+k\} = \binom{m+k}{m} p^m (1-p)^k$$

$$P\{N_1(t)=m, N_2(t)=k; N(t)=m+k\} = P\{N_1(t)=m, N_2(t)=k \mid N(t)=m+k\} \times P\{N(t)=m+k\}$$

$$= \frac{(m+k)!}{m!\,k!} p^m (1-p)^k \cdot \frac{(\lambda t)^{m+k} e^{-\lambda t}}{(m+k)!}$$

$$= \frac{(\lambda p t)^m}{m!} e^{-\lambda p t} \cdot \frac{(\lambda(1-p)t)^k}{k!} e^{-\lambda(1-p)t}$$

$$= P\{N_1(t)=m\}\, P\{N_2(t)=k\} \leftarrow \text{independent!}$$

30

$N_1(t)$ is also independent of $N_2(tx)$ $tx < t$

$(0,t] = (0,tx] \cup (tx,t]$

and from proof $N_1(tx)$ is ind. $N_2(tx)$.

$N_1(tx,t)$ is ind. from $N_2(tx)$, since N_1 is a Poisson process with ind. increments.

You are at car wash. there are two lines generating "clean cars" with rate λ_1 and λ_2.

The processes are Poisson and independent.

You join line 1. There are many cars in both lines.

Let $S_k^{(1)}$ is the departure time of k^{th} car from line 1.

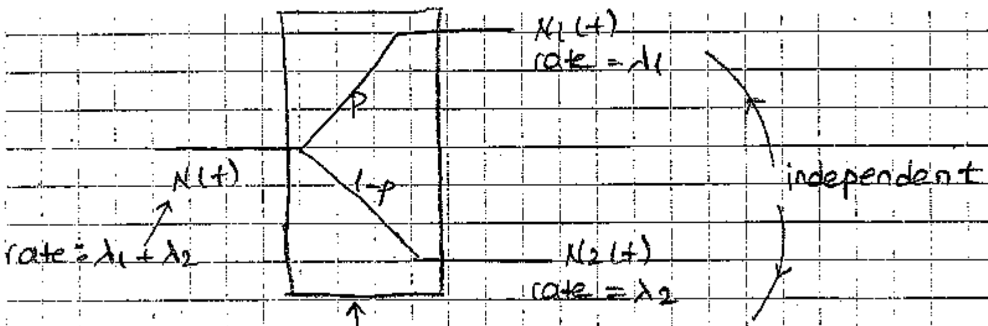
$$P\{S_1^{(2)} < S_1^{(1)}\} = ?$$

$$P\{\underset{\substack{\uparrow \\ \text{waiting time for} \\ \text{departure of } 1^{st} \text{ car}}}{X_1^{(2)}} < X_1^{(1)}\}$$

exp λ_1 exp λ_2

$$= \int P\{X_1^{(2)} < X_1^{(1)} = x_1 \mid X_1^{(1)} = x_1\} f_{X_1}(x_1)\, dx_1$$

$$= \frac{\lambda_2}{\lambda_1+\lambda_2}$$



$$p = \frac{\lambda_1}{\lambda_1+\lambda_2}$$

$$P\{X_1^{(2)} < X_1^{(1)}\} = P\{\text{First event is switched to } (2)\} = 1-p = \frac{\lambda_2}{\lambda_1+\lambda_2}$$

You arrive at the post office. Two clerks on busy and no other clients waiting. Clerks operate at rate λ_1 customers/hour and λ_2 customers/hour.

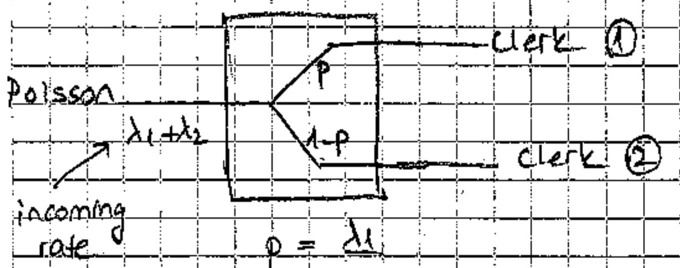
Processes are Poisson and independent.

Find the expected amount of time that you spent in post office until your task is completed.

$$T = W + P$$

- T: total task time
- W: waiting for clerk to be available
- P: processing time of your request.

$E\{T\} = ?$



$$p = \frac{\lambda_1}{\lambda_1+\lambda_2}$$

$$E\{T\} = E\{W\} + E\{P\}$$

$$= \frac{1}{\lambda_1+\lambda_2} + \underbrace{E\{P \mid \text{Clerk 1 is assisting}\}}_{1/\lambda_1} \underbrace{P\{\text{clerk 1}\}}_{\frac{\lambda_1}{\lambda_1+\lambda_2}} + \underbrace{E\{P \mid \text{clerk 2 is assisting}\}}_{1/\lambda_2} \underbrace{P\{\text{clerk 2}\}}_{\frac{\lambda_2}{\lambda_1+\lambda_2}}$$

$$= \frac{3}{\lambda_1+\lambda_2}$$

$$f_X(x) = \lambda e^{-\lambda x}$$

$$E\{X\} = \frac{1}{\lambda}$$

Also see Ross p. 305, 10^{th} Edition for other solutions.

$0 \quad s_1 \quad s_2 \quad s_3 \quad \cdots \quad s_n \quad t$

$N(t) = n, \quad f_{s_1, s_2, \ldots, s_n \mid N(t)=n}(s_1, s_2, \ldots, s_n \mid N(t) = n) = ?$

Let $Y_1, Y_2, \ldots, Y_n$ be n rv's iid dist by distribution $f_Y(y)$.

$Y_{(1)} = \min\{Y_1, \ldots, Y_n\}$

$Y_{(2)} = \text{second minimum}\{Y_1, \ldots, Y_n\} = \min\{\{Y_1, Y_2, \ldots, Y_n\} - Y_{(1)}\}$

set difference

$Y_{(3)} \Rightarrow 3^{rd}$ smallest in the list

$\vdots$

$Y_{(n)} = \max\{Y_1, \ldots, Y_n\}$

$f_{Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}}(y_{(1)}, y_{(2)}, \ldots, y_{(n)}) = ?$

Joint density of ordered rv's

Let's remember

$$f_{Y_1, Y_2, \ldots, Y_n}(y_1, \ldots, y_n) = \prod_{k=1}^{n} f_Y(y_k)$$

Y_k's are iid with density $f_Y(y)$

not ordered

$$\underline{Y} = \begin{bmatrix}2\\-2.1\\3\end{bmatrix} \longrightarrow \underline{Y}_{(\cdot)} = \begin{bmatrix}-2.1\\2\\3\end{bmatrix}$$

a realization for Y_1, Y_2, Y_3

ordered r.v realization

$$\underline{Y} = \begin{bmatrix}3\\2\\-2.1\end{bmatrix} \qquad Y = \begin{bmatrix}3\\-2.1\\2\end{bmatrix}$$

$3!$ orderings of $\begin{bmatrix}2\\-2.1\\3\end{bmatrix}$ realization gives the same ordered realization.

$$f_{Y(1),Y(2),Y(3)}(y(1),y(2),y(3)) = ?$$

$$P\left\{\begin{array}{l} y(1) < Y(1) \leq y(1)+\delta \\ y(2) < Y(2) < y(2)+\delta \\ y(3) < Y(3) \leq y(3)+\delta \end{array}\right\} = \left.\begin{array}{l} P\{y(1)<Y(1)<y(1)+\delta_1,\ y(2)<Y(2)<y(2)+\delta_2,\ y(3)<Y(3)<y(3)+\delta_3\} \\ + \\ P\{y(1)<Y(2)<y(1)+\delta,\ y(2)<Y(1)<y(2)+\delta,\ y(3)<Y(3)<y(3)+\delta\} \\ + \\ \vdots \end{array}\right\} \text{6 terms}$$

1	2	3
2	1	3
3	2	1
1	3	2
2	3	1
3	1	2

$$= \left.\begin{array}{l} f_Y(y_1)\, f_Y(y_2)\, f_Y(y_3)\, \delta_1 \delta_2 \delta_3 \\ + \\ f_Y(y_2)\, f_Y(y_1)\, f_Y(y_3)\, \delta_1 \delta_2 \delta_3 \\ + \\ \_\ \_\ \_ \end{array}\right\} \text{6 terms}$$

$$= \left(3! \prod_{k=1}^{3} f_Y(y_k)\right) \delta_1 \delta_2 \delta_3$$

$$P\{A\} = n! \prod_{k=1}^{n} f_Y(y_k)\, \delta_1 \delta_2 \delta_3 \quad (n=3)$$

$$f_{Y(1),Y(2),Y(3)}(y(1),y(2),y(3))\, \delta_1 \delta_2 \cancel{\delta_3} = n! \prod_{k=1}^{n} f_Y(y_k)\, \delta_1 \delta_2 \cancel{\delta_3} \quad (n=3)$$

$$\underbrace{f_{Y(1),\ldots,Y(n)}(y(1),y(2),\ldots,y(n))}_{\text{Joint density of the ordering}} = n! \underbrace{\prod_{k=1}^{n} f_Y(y_k)}_{\text{Joint density before ordering}}, \quad y(1) < y(2) < \ldots < y(n)$$

$Y_1, Y_2, \ldots Y_n$ be uniform in $(0,t]$

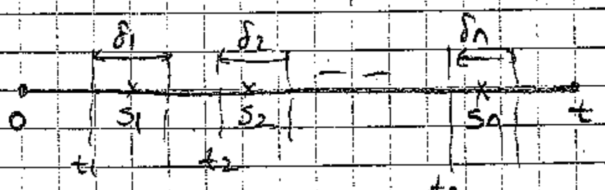
$\downarrow$

$$f_{Y(1),Y(2),\ldots,Y(n)}(y(1),y(2),\ldots,y(n)) = \frac{n!}{t^n}, \quad y(1) < y(2) < \ldots < y(n)$$

Conditional arrival joint density is

$$f_{S_1,S_2,\ldots S_n \mid N(t)=n}(s_1,s_2,\ldots,s_n) = \frac{n!}{t^n}, \quad s_1 < s_2 < \ldots < s_n$$

$$P\{t_k < S_k \le t_k + \delta_k,\ k = \{1,2,\ldots,n\} \mid N(t) = n\}$$



$$P\{A \mid N(t) = n\} = \frac{P\{\overbrace{(1 \text{ arrival in } (t_k, t_k+\delta],\ k=\{1,2,\ldots,n\}) \text{ and } (\text{no other arrivals})}^{A \text{ and } N(t)=n}\}}{P\{N(t)=n\}}$$

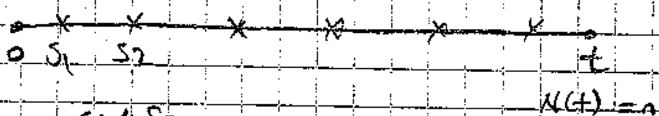
$$= \frac{(\lambda \delta_1 e^{-\lambda\delta_1})(\lambda \delta_2 e^{-\lambda\delta_2})\cdots(\lambda\delta_n e^{-\lambda\delta_n})\, \underbrace{\overbrace{e^{-\lambda(t-\delta_1-\delta_2-\cdots-\delta_n)}}^{\text{length of remaining intervals}}}_{0 \text{ arrival}}}{(\lambda t)^n e^{-\lambda t}/n!}$$

$$= \frac{n!}{t^n}\,\delta_1\delta_2\cdots\delta_n$$

$$P\{A \mid N(t)=n\} = f_{S_1,S_2,\ldots,S_n \mid N(t)=n}(s_1,s_2,\ldots,s_n)\,\delta_1\delta_2\cdots\delta_n$$

Proof is completed by cancelling $\delta_1 \cdot \delta_2 \cdots \delta_n$ from both parts.

Joint dist. of $S_1, S_2, \ldots S_n$ given $N(t)=n$ is nothing but ordered statistics of n i.i.d r.v's with unif dist. in $[0,t]$



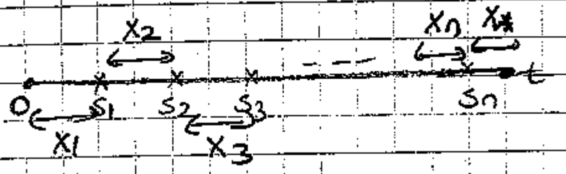
$s_1 < s_2$

the result is not surprising since Poisson process is stationary and independent increments.

12/11/2014

Conditional Waiting Times

Given $N(t)=n$,



$$f_{S_1,S_2,\dots,S_n|N(t)=n}(s_1,s_2,\dots,s_n|N(t)=n) = \frac{n!}{t^n}, \quad 0<s_1<s_2 \dots <s_{n-1}<s_n<t$$

ordered dist. of uniform picks in $(0,t]$ interval.

X_n: the waiting time from $(n-1)^{th}$ arrival to n^{th} arrival.

$$\left.\begin{aligned} X_1 &= S_1 \\ X_2 &= S_2 - S_1 \\ X_3 &= S_3 - S_2 \quad * \\ &\vdots \\ X_n &= S_n - S_{n-1} \end{aligned}\right\}$$

I need to find Joint distribution of X_k's given $N(t)=n$ (LHS of $*$) from joint dist. of S_k's given $N(t)=n$ (RHS of $*$)

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} 1 & & & & 0 \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & -1 & 1 & \\ 0 & & & -1 & 1 \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ \vdots \\ S_n \end{bmatrix}$$

$\Downarrow$

$$f_{X_1,X_2,\dots,X_n|N(t)=n}(x_1,x_2,\dots,x_n|N(t)=n) = \frac{f_{S_1,S_2,\dots,S_n}(M^{-1}x)}{\det(M)} = \frac{n!}{t^n}$$

provided that
$x_1 > 0$
$x_2 > 0$
$\vdots$
$x_n > 0$

$x_1 + x_2 + \dots + x_n < t$

$$f_{X_1}(x_1 \mid N(t) = n) = ?$$

$$= \int_{x_2=0}^{t-x_1} \int_{x_3=0}^{t-x_1-x_2} \int_{x_4=0}^{t-x_2-x_3-x_1} \dots \int_{x_n=0}^{t-x_2-x_3-\dots-x_{n-1}-x_1} (\text{joint density})\, dx_2\, dx_3\, dx_4 \dots dx_n$$

Second way to get $f_{X_1}(x_1 \mid N(t) = n)$

$$P\{X_1 > x_1 \mid N(t) = n\} = \frac{P\{X_1 > x_1 \text{ and } N(t) = n\}}{P\{N(t) = n\}}$$

$$= \frac{P\left\{\begin{matrix}\text{no arrivals} \\ \text{in } (0, x_1]\end{matrix} \text{ and } \begin{matrix}n \text{ arrivals} \\ \text{in } (x_1, t]\end{matrix}\right\}}{P\{N(t) = n\}} = \frac{P\{0 \text{ arrival in } (0, x_1]\} \cdot P\{n \text{ arrivals in } (x_1, t]\}}{P\{N(t) = n\}}$$

$$= e^{-\lambda x_1} \cdot e^{-\lambda(t-x_1)} \cdot \frac{(\lambda(t-x_1))^n}{n!} \Big/ \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

$$= \left(\frac{t - x_1}{t}\right)^n$$

$$f_{X_1 \mid N(t)=n}(x_1 \mid N(t) = n) = \frac{\partial}{\partial x_1} \underbrace{\left(1 - \left(\frac{t - x_1}{t}\right)^n\right)}_{\text{cdf of } X_1 \mid N(t) = n}$$

$$= n \cdot \frac{(t - x_1)^{n-1}}{t^n}$$

$$f_{X_1 | N(t)=1}(x_1 | N(t)=1) = \frac{1}{t}$$

Non-Homogenous Poisson Process:

If the rate λ of the Poisson process varies by time "t", i.e. $\lambda(t)$; then the resultant process is called non-homogenous Poisson process.

$$P\{\tilde{N}(t, t+\delta) = 0\} = 1 - \delta\lambda(t) + o(\delta)$$

$$P\{\tilde{N}(t, t+\delta) = 1\} = \delta\lambda(t) + o(\delta)$$

$$P\{\tilde{N}(t, t+\delta) \geq 2\} = o(\delta)$$

Equivalent of Def #3 we have given earlier.

Theorem 24.1 p.90 Let $\lambda(t)$ be rate of non-homogenous Poisson Process; then

$$P\{\tilde{N}(t,\tau) = n\} = \frac{(m(t,\tau))^n e^{-m(t,\tau)}}{n!}$$

$$m(t,\tau) = \int_t^\tau \lambda(t')dt'$$

equivalent of Def #2 defined earlier.

Ex: A barber shop operates as follows:

00:00 - 06:00 : Closed

06:00 - 12:00 : AM hours, typically 1 customer in morning hours.

12:00 - 18:00 : PM hours, typically 2 customers in afternoon hours.

18:00 - 24:00 : Closed

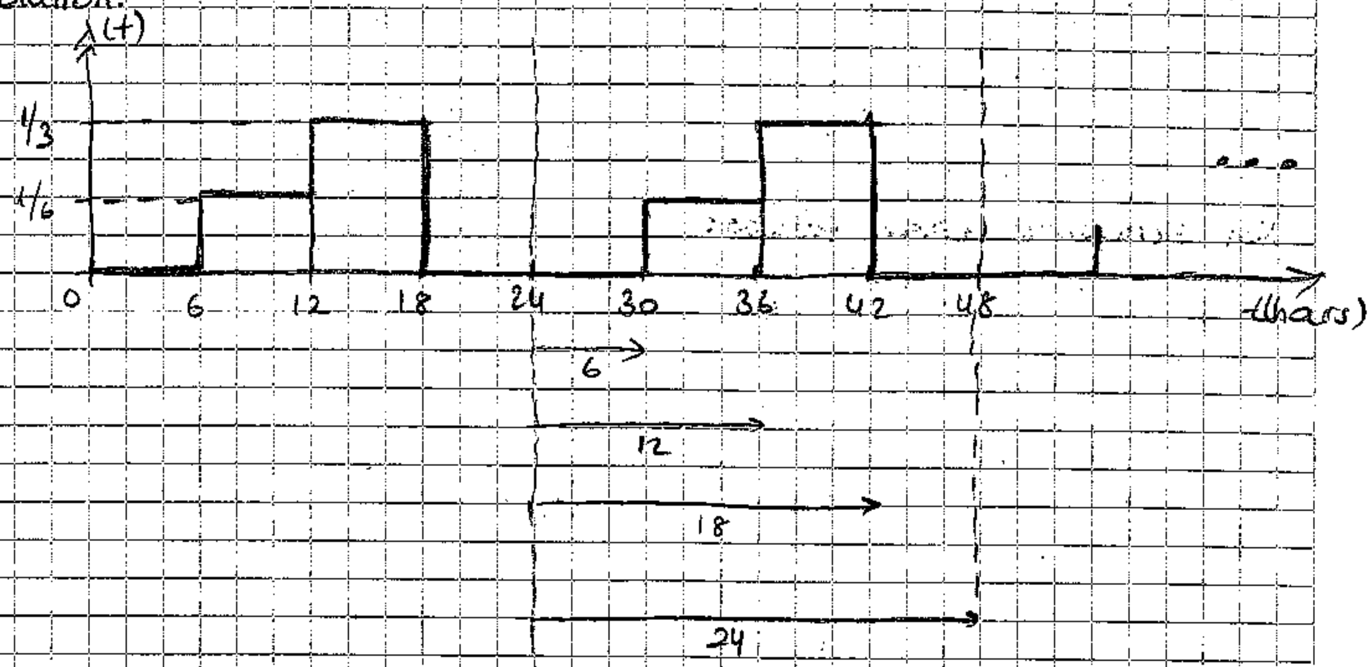
Assume arrivals are Poisson distributed

a) P{2 customers in AM hours} = ?

b) P{2 customers in 24 hours} = ?

c) P{2 customers in AM hours | 2 customers in 24 hours} = ?

Solution:



$E\{N(t)\} = \lambda t$

a) $P\{$ 2 customers in AM hours $\}$

$$m_{AM} = \int_0^{12} \lambda(t')\,dt' \rightarrow P\{N(0,12) = 2\} = \frac{e^{-m} \cdot m^2}{2!}$$

$$= \int_6^{12} \frac{1}{6}\,dt' \qquad\qquad = e^{-1}/2$$

$m_{AM} = 1$

b) $P\{$ 2 customers in 24 hours $\} = ?$ $\qquad m_{24} = \int_0^{24} \lambda(t)\,dt = 3$

$$P\{N(0,24) = 2\} = \frac{e^{-3} \cdot 9}{2}$$

Soln 2: $P\{2 \text{ customers in } 24 \text{ hours}\}$

$= P\{2 \text{ in AM}, 0 \text{ in PM}\}$

$+$

$P\{1 \text{ in AM} + 1 \text{ in PM}\}$

$+$

$P\{0 \text{ in AM} + 2 \text{ in PM}\}$

$$= \frac{e^{-1}}{2} \cdot e^{-2^{\,\mu_{PM}}} + \frac{e^{-1}}{1!} \cdot \frac{e^{-2} \cdot 2}{1!} + e^{-1} e^{-2} \frac{2^2}{2!}$$

$$= e^{-3}\left(\frac{1}{2} + 2 + 2\right) = e^{-3} \cdot \frac{9}{2}$$

c) $P\{2 \text{ AM customers} \mid 2 \text{ customers in } 24 \text{ hours}\}$

$$= \frac{P\{2 \text{ AM}, 0 \text{ PM}\}}{P\{2 \text{ in } 24 \text{ hours}\}} = \frac{e^{-3}/2}{e^{-3} \cdot 9/2} = \frac{1}{9}$$

Claim: $P\{\text{AM arrival} = 1 \mid 1 \text{ arrival in } 24 \text{ hours}\} = 1/3$

Since

$$\frac{P\{AM = 1, PM = 0\}}{P\{AM + PM = 1\}} = \frac{\left[e^{-\mu_{AM}} \frac{(\mu_{AM})^1}{1!}\right]\left[e^{-\mu_{PM}}\right]}{e^{-(\mu_{AM} + \mu_{PM})} \, / \, (\mu_{AM} + \mu_{PM})^1 / 1!}$$

$$= \frac{\mu_{AM}}{\underset{1}{\mu_{AM}} + \underset{2}{\mu_{PM}}} = \frac{1}{3}$$

P{Two arrivals in AM hours given only 2 arrivals in 24 hours}

is then $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

Compound Poisson Process:

$$X(t) = \sum_{i=1}^{N(t)} Y_i$$

$N(t)$: Homogenous Poisson process with rate λ.
Y_i : i.i.d dist. r.v $f_Y(y)$

Y_i's and $N(t)$ are independent.

① $E\{X(t)\} = ?$

$$= \sum_{n=0}^{\infty} E\{X(t) \mid N(t) = n\} P\{N(t) = n\}$$

$$= \sum_{n=0}^{\infty} E\{\sum_{i=1}^{n} Y_i\} P\{N(t) = n\}$$

$$= \sum_{n=0}^{\infty} n\overline{Y} P\{N(t) = n\}$$

$$= \overline{Y} \cdot E\{N(t)\}$$

$$= (\lambda t)\overline{Y}$$

$$E_{N(t),Y_i}\{X(t)\} = E_{N(t)}\{E_{Y}\{X(t) \mid N(t) = n\}\}$$

$$= E_{N(t)}\{N(t)\overline{Y}\}$$

$$= (\lambda t) \cdot \overline{Y}$$

② $\mathrm{var}\{x(t)\} = (\lambda t)\, E\{Y^2\}$ ← DIT
check wikipedia (Compound Poisson Process)

Ex:

$X = N(0,5)$

$Y = N(0,6)$

$N(0,t)$ is poisson process with rate λ.

$E\{N(0,5)\} = 5\lambda$.

$\mathrm{cov}(X,Y) = E\{(X-\bar{X})(Y-\bar{Y})\} = E\{XY\} - \underbrace{\bar{X}}_{5\lambda}\,\underbrace{\bar{Y}}_{6\lambda}$

$E\{N(0,5)\,N(0,6)\} = E\{N(0,5) \times [N(0,5) + N(5,6)]\}$ → arrival time not intersect they are independent

$= E\{N(0,5)^2\} + E\{N(0,5)\}\,E\{N(5,6)\}$

$= \lambda 5 + (\lambda 5)^2 + \lambda 5 \cdot \lambda 1$ $\qquad \mathrm{var}\{N\} = E\{N\} = \lambda t$

$= 5\lambda + 30\lambda^2$ $\qquad$ poisson rv

$\mathrm{cov}(X,Y) = 5\lambda + 30\lambda^2 - 30\lambda^2 = \underline{\underline{5\lambda}}$

17/11/2014

Random Vectors:

A random vector is completely defined by the joint pdf of its components

$$\underset{\substack{\uparrow \\ 2\times 1}}{\underline{X}} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad X_1, X_2 \text{ rv's with joint pdf } f_{\underline{X}}(x_1, x_2)$$

Ex: $\underline{X}$: $N\times 1$ vector whose entries are i.i.d $N(\mu_x, \sigma_x^2)$

Find joint pdf $\underline{X}$.

$$f_{\underline{X}}(x_1, x_2, \ldots x_N) = f_X(x_1) f_X(x_2) \ldots f_X(x_N)$$

$$= \prod_{k=1}^{N} f_X(x_k) = \frac{1}{(\sqrt{2\pi})^N \cdot \sigma_x^N} e^{\boxed{-\sum_{k=1}^{N} \frac{(x-\mu_x)^2}{2\sigma_x^2}}}$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-(x-\mu_x)^2/2\sigma_x^2}$$

$$\frac{\|\underline{x} - \mu_x \underline{1}\|^2}{2\sigma_x^2}$$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} - \mu_x \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\|z\|^2 = z^T \cdot z$$
(Euclidean norm)

$$= \frac{1}{(\sqrt{2\pi}\,\sigma_x)^N} e^{-\frac{1}{2\sigma_x^2}\|\underline{x} - \mu_x \underline{1}\|^2}$$

Gaussian Vector: A random vector $\underline{z}$ which can be expressed as

$\underline{z} = \underline{A}\,\underline{w}$ where $\underline{A}$ is a real valued matrix

and $f_{\underline{w}}(\underline{w})$ is $N(\underline{\mu_w}, \underline{I})$.

$$\hookrightarrow \frac{1}{(\sqrt{2\pi}\cdot 1)^N} e^{-\frac{1}{2}\|\underline{w} - \underline{\mu_w}\|^2}$$

Special case:

$\underline{\underline{A}}$ in the definition is invertible.

$$\underline{z} = \underline{\underline{A}}\,\underline{w}$$

$$f_{\underline{z}}(\underline{z}) = \frac{f_{\underline{w}}(A^{-1}\underline{z})}{|\det(A)|}$$

N function of N r.v mapping.

$$f_{\underline{z}}(\underline{z}) = \frac{1}{(\sqrt{2\pi})^N} \cdot \underbrace{\frac{1}{|\det(A)|}}_{(**)} \cdot e^{-\frac{1}{2}\underbrace{\|\underline{\underline{A}}^{-1}\underline{z} - \mu_w\|^2}_{(*)}}$$

(*) $\|\underline{\underline{A}}^{-1}\underline{z} - \underline{\mu_w}\|^2 = \|\underline{\underline{A}}^{-1}(\underline{z} - \underline{\underline{A}}\,\underline{\mu_w})\|^2$

$$= (\underline{z} - \underline{\underline{A}}\,\underline{\mu_w})^T (A^{-1})^T A^{-1} (\underline{z} - \underline{\underline{A}}\,\underline{\mu_w})$$

$$\underline{\underline{K}} \triangleq \underline{\underline{A}}\,\underline{\underline{A}}^T$$

covariance matrix

$$= (\underline{z} - \underline{\underline{A}}\,\underline{\mu_w})\, K^{-1}\, (\underline{z} - \underline{\underline{A}}\,\underline{\mu_w})$$

(**) $\det(K) = \det(\underline{\underline{A}})\det(\underline{\underline{A}}^T)$

(***) $\mu_z = \underline{\underline{A}}\,\mu_w$

$$f_{\underline{z}}(\underline{z}) = \frac{1}{(\sqrt{2\pi})^N \sqrt{|\det(K)|}} \exp\left(-\frac{1}{2}(\underline{z} - \mu_z)^T K^{-1} (\underline{z} - \underline{\mu_z})\right)$$

Joint pdf of Gaussian vector

Notes: ① Joint pdf only depends on μ_z and $\underline{\underline{K}}$.

mean vector

covariance matrix

② N=1, we get

$$f_z(z) = \frac{1}{\sqrt{2\pi}\,\sigma_z} e^{-\frac{1}{2\sigma_z^2}(z-\mu_z)^2}$$

2^{nd} Moment Descriptions of Random Vectors:

$\underline{X}$: a random vector

2^{nd} moment description (first two moments)

① Mean : $\underline{\mu_X} = E\{\underline{X}\}$

② Covariance : $\underline{\underline{K_X}} = E\{(\underline{X}-\underline{\mu_X})(\underline{X}-\underline{\mu_X})^T\}$

covariance matrix

$$\underline{\underline{K_X}} = E\left\{ \begin{bmatrix} X_1-\mu_{X_1} \\ X_2-\mu_{X_2} \\ \vdots \\ X_N-\mu_{X_N} \end{bmatrix} \begin{bmatrix} X_1-\mu_{X_1} & X_2-\mu_{X_2} & \cdots & X_N-\mu_{X_N} \end{bmatrix} \right\}$$

$$= \begin{bmatrix} \underbrace{var\{X_1\}}_{cov(X_1,X_1)} & cov(X_1,X_2) & \cdots & cov(X_1,X_N) \\ cov(X_2,X_1) & var\{X_2\} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_N,X_1) & cov(X_N,X_2) & \cdots & var(X_N) \end{bmatrix}$$

$cov(X,Y) = E\{(X-\mu_X)(Y-\mu_Y)\}$

Change in 2^{nd} order Descriptions After a Linear Mapping

$$\underline{X} \longrightarrow \boxed{A} \longrightarrow \underline{Y}$$

random vector $\underline{X}$: $\underline{\mu_X}$, $\underline{\underline{K_X}}$

"new" random vector $\underline{Y}$: $\underline{\mu_Y}$?, $\underline{\underline{K_Y}}$?

$$\underline{y} = \underline{\underline{A}}\,\underline{x} \rightarrow \text{①}\ \mu_y = E\{\underline{y}\} = E\{\underline{\underline{A}}\,\underline{x}\} = \underline{\underline{A}}\,E\{\underline{x}\} = \underline{\underline{A}}\,\underline{\mu_x}$$

$$\rightarrow \text{②}\ \underline{\underline{K_y}} = E\{(y-\mu_y)(y-\mu_y)^T\}$$

$$= E\{\underline{y}\,\underline{y}^T\} - \mu_y \mu_y^T$$

$$= E\{(\underline{\underline{A}}x)(x^T\underline{\underline{A}}^T)\} - \underline{\underline{A}}\,\mu_x\mu_x^T\,\underline{\underline{A}}^T$$

$$\underline{\underline{K_y}} = \underline{\underline{A}}\,E\{\underline{x}\,\underline{x}^T\}\,\underline{\underline{A}}^T - \underline{\underline{A}}\,\mu_x\mu_x^T\,\underline{\underline{A}}^T$$

$$= \underline{\underline{A}}\left(E\{\underline{x}\,\underline{x}^T\} - \mu_x\mu_x^T\right)\underline{\underline{A}}^T = \underline{\underline{A}}\,\underline{\underline{K_x}}\,\underline{\underline{A}}^T$$

$$[\underline{\underline{A}}\,\underline{x}]_j = \sum_{k=1}^{N} a_{jk} x_k$$

$$\begin{bmatrix} \ \\ \ \\ \ \end{bmatrix}_{N\times 1} \leftarrow j$$

After a linear mapping:

① Mean changes to $\underline{\mu_x} \rightarrow \underline{\underline{A}}\,\mu_x$

Cov changes to $\underline{\underline{K_x}} \rightarrow \underline{\underline{A}}\,\underline{\underline{K_x}}\,\underline{\underline{A}}^T$

<u>Note:</u> cov. of 2 r.v's X and Y are <u>not</u> a func. of μ_x and μ_y

$$\mathrm{Cov}(X,Y) = E\{\underbrace{(X-\mu_x)}_{\substack{\text{mean subtracted} \\ \text{r.v.} \\ X}}\underbrace{(Y-\mu_y)}_{\substack{\text{mean subtracted} \\ \text{r.v.} \\ Y}}\}.$$

Because of this in cov. matrix and some similar calculations there is no harm in assuming that all vectors have zero mean.

$$\underset{N\times 1}{\underline{X}} = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}$$ ← joint pdf of all rv's is sufficient to char. the random vector

$$r_k = s + n_k \quad , \quad k = \{1, \dots, N\}$$

r_k: k^{th} measurement
s: signal of interest
n_k: noise of k^{th} measurement

$n_k \sim N(0, \sigma_n^2)$
i.i.d. → indep from measurement to measurement

s: non-random parameter (not changing, fix)

$$\Rightarrow \underset{N\times 1}{\underline{r}} = \underline{1}\, s + \underline{n}, \quad \underline{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \underline{n} \sim N(\underline{0}, \sigma_n^2 \underline{I})$$

$$r \sim N(\underline{1} s, \sigma_n^2 \underline{I}) \rightarrow \hat{s}_{ML} = \frac{\underline{1}^T \underline{r}}{N} = \frac{\sum_{k=1}^{N} r_k}{N}$$

$\underline{\underline{A}}\,\underline{X} = \underline{b}$ ← more equation than unknown

$$\hat{X}_{LS} = (\underline{\underline{A}}^T \underline{\underline{A}})^{-1} \underline{\underline{A}}^T \underline{b}$$

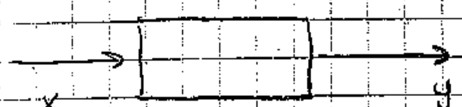
$$\hat{\mu} = \frac{1}{N} \sum_{k=1}^{N} x[k]$$

$$\hat{\sigma}_x^2 = \frac{1}{N-1} \sum_{k=1}^{N} (x[k] - \hat{\mu}_x)^2$$

1. Mean Vector $E\{\underline{X}\}$

2. Cov. Matrix $E\{\underbrace{(X - \mu_X)}_{\text{zero-mean vector}}(X - \mu_X)^T\}$

cov. matrix does not depend on μ_X and we can assume that $\mu_X = 0$ for cov. calculations $\longrightarrow E\{\underbrace{X_{zm}}_{\text{zero mean vector}} X_{zm}^T\}$



$\underline{x}$ r.vector $\qquad\qquad$ $\underline{y}$ new r.vector

$$E\{\underline{y}\} = \underline{A}\, E\{\underline{x}\}$$

$$K_y = \underline{A}\, K_x\, \underline{A}'$$

$\uparrow$ covariance matrix of $\underline{y}$

For Gaussian vectors, $\underline{\mu_x}$ and $\underline{K_x}$ is sufficient to write joint pdf.

$$\mathrm{cov}(X,Y) = E\{(X-\bar{X})(Y-\bar{Y})\}$$

$$\rho_{XY} = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}\{X\}\mathrm{var}\{Y\}}}$$

$\downarrow$ corr. coef.

$$\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$$

$$\mathrm{cov}(X+Y,Z) = \mathrm{cov}(X,Z) + \mathrm{cov}(Y,Z)$$

$$\mathrm{var}(X) = \mathrm{cov}(X,X)$$

$$\mathrm{cov}(\alpha X,Y) = \alpha\,\mathrm{cov}(X,Y)$$

$$\mathrm{var}\left(\sum_{k=1}^{N} X_k\right) = ?$$

$$\overset{(3)}{=} \mathrm{cov}\left(\sum_{k_1=1}^{N} X_{k_1}, \sum_{k_2=1}^{N} X_{k_2}\right)$$

$$\overset{(2)}{=} \sum_{k_1=1}^{N} \mathrm{cov}\left(X_{k_1}, \sum_{k_2=1}^{N} X_{k_2}\right)$$

$$\overset{(2)}{=} \sum_{k_1=1}^{N}\left(\sum_{\substack{k_2=1 \\ \text{and} \\ k_2\neq k_1}}^{N} \mathrm{cov}(X_{k_1}, X_{k_2})\right) + \sum_{k_1=1}^{N} \underbrace{\mathrm{cov}(X_{k_1}, X_{k_1})}_{\mathrm{var}\{X_{k_1}\}}$$

$$= \sum_{k=1}^{N} var(X_k) + \sum_{k_1=1}^{N} \sum_{\substack{k_2=1 \\ k_2 \neq k_1}}^{N} cov(X_{k_1}, X_{k_2})$$

$$\longrightarrow \triangleq 2 \times \sum_{k_1=1}^{N} \sum_{k_2=k_1+1}^{N} cov(X_{k_1}, X_{k_2})$$

$$\underset{(1\times 1)}{z} = \underset{(1\times N)}{\underline{1}^T} \underset{(N\times 1)}{X}$$

$$= \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}$$

$$\underset{\substack{cov(z) \\ \parallel \\ var(z)}}{K_z} = \underline{1}^T \underline{\underline{K}} \times \underline{1}$$

$$= \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} var(x_1) & cov(x_1,x_2) & \cdots \\ & \ddots & \\ cov(x_N,x_1) & & var(x_N) \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

(see document on the web with the same title)

$$\rho_{XY} = \frac{cov(X,Y)}{\sqrt{var(X)\,var(Y)}}$$

$|\rho_{XY}| \leq 1$, i.e. $-1 \leq \rho_{XY} \leq 1$

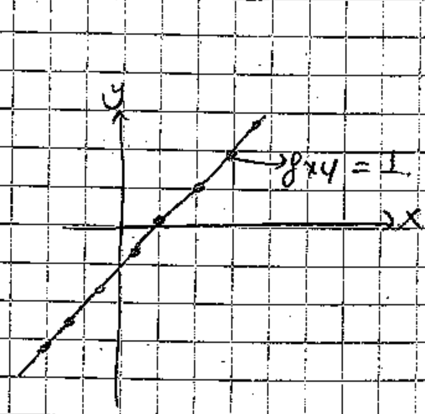
$\rho_{XY} = 0 \leftrightarrow cov(X,Y) = 0$, X and Y are uncorrelated

$\rho_{XY} = \mp 1 \iff Y = \alpha X + c$

fully correlated

random variable

non-random constants



$\underline{\underline{K_z}}$ is symetric matrix $\qquad \underline{\underline{K_z}} = E\{\underline{z}\,\underline{z}^T\}$

$(\underline{\underline{K_z}} = \underline{\underline{K_z}}^T)$ $\qquad$ zero mean vectors

Sol $\quad$ Eigendecomposition

$$\underline{\underline{K_z}} = \underline{\underline{\Theta}}\,\underline{\underline{\Omega}}\,\underline{\underline{\Theta}}^{-1} \quad \} \text{ eigendecomposition}$$

$\underline{\underline{\Theta}} = [\underbrace{e_1 \; - - - - \; e_N}_{\text{eigenvectors}}]$

$\underline{\underline{\Omega}} = \text{diag}(\underbrace{\lambda_1, \; - - , \lambda_N}_{\text{eigenvalue}})$

$$\underline{\underline{K_z}}\,\underline{e_k} = \underline{e_k}\,\lambda_k$$

$$\underline{\underline{K_z}}[e_1 \; -- \; e_5] = [e_1 \; -- \; e_5]\begin{bmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_5\end{bmatrix}$$

$$\downarrow$$

$$\underline{\underline{K}}\,\underline{\underline{\Theta}} = \underline{\underline{\Theta}}\,\underline{\underline{\Omega}}$$

$$\underline{\underline{K}} = \underline{\underline{\Theta}}\,\underline{\underline{\Omega}}\,\underline{\underline{\Theta}}^{-1}$$

Since $\underline{\underline{K_z}}$ is symetric

λ_k's are real-valued

$\underline{e_k} \perp \underline{e_\ell}$ for $\lambda_k \neq \lambda_\ell$

and

$\underline{e_{k1}}, \underline{e_{k2}}$ for $\lambda_{k1} = \lambda_{k2}$ can be also orthogonalized.

$$\underline{\underline{K}} = \underline{\underline{\Theta}}\,\underbrace{\underline{\underline{\Omega}}}_{\text{diagonal matrix of real numbers}}\,\underline{\underline{\Theta}}^{-1} \rightarrow \underline{\underline{\Theta}}\,\underline{\underline{\Theta}}^T = \underline{\underline{\Theta}}^T\underline{\underline{\Theta}} = \underline{\underline{I}}$$

$\underline{\underline{K_z}}$ is positive semi-definite

A symetric matrix A is positive semi-definite

if $\underline{x}^T \underline{\underline{A}}\, \underline{x} \geq 0 \quad \forall \underline{x} \in \mathbb{R}^N$

quadratic form

$$[x_1 \; --- \; x_N] \begin{bmatrix} a_{11} & - - & a_{31} \\ - - - & a_{55} & \\ & & \end{bmatrix} \begin{bmatrix} x_1 \\ | \\ x_N \end{bmatrix} \geq 0$$

$$a_{11} x_1^2 + a_{55} x_5^2 + a_{31} x_1 x_3 - -$$

$\underline{\underline{A}} \geq 0 \rightarrow A$ is positive semi-definite

$\underline{\underline{A}} > 0 \rightarrow A$ is positive definite

$A < 0 \rightarrow$ negative definite

Show

$\underline{\underline{K_z}} \geq 0$

$$\underline{x}^T \underline{\underline{K_z}}\, \underline{x} \overset{?}{\geq} 0$$

$$x^T E\{\underline{z}\,\underline{z}^T\}\, x \overset{?}{\geq} 0$$

$$E_z\Big\{ \underbrace{(\underline{x}^T \underline{z})}_{w} \underbrace{(\underline{z}^T \underline{x})}_{w} \Big\} \overset{?}{\geq} 0$$

$$E\{(x^T z)^2\} \geq 0 \quad \checkmark$$

If these two properties are satisfied by matrix $\underline{\underline{K}}$,

then can I be sure that $\underline{\underline{K}}$ is a covariance matrix?

(Is two properties sufficient to generate a valid covariance matrix?)

Yes!

I will construct a gaussian vector with given $\underline{\underline{K}}$ matrix

and covariance matrix of gaussian vector will be equal to $\underline{K}$ matrix since $\underline{K}$ is symetric, I can decompose it as follows:

$$\underline{K} = \underline{\underline{\Theta}}\,\underline{\Omega}\,\underline{\underline{\Theta}}^T$$

$$\underline{K}^{1/2} = \underline{\underline{\Theta}}\,\underline{\underline{\Omega}}^{1/2}\,\underline{\underline{\Theta}}^T \quad (\text{sqrtm}(\underline{\underline{K}}))$$

$$\begin{bmatrix} \sqrt{\lambda_1} & & & 0 \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ 0 & & & \sqrt{\lambda_n} \end{bmatrix}$$

eigenvalues of $\underline{\underline{K}}$ matrix.

$$\underline{K}^{1/2} \cdot \underline{K}^{1/2} = \Theta\,\Omega^{1/2}\,\underbrace{\Theta^T\,\Theta}_{\underline{\underline{I}}}\,\Omega^{1/2}\,\Theta^T = \Theta\,\Omega\,\Theta^T = \underline{\underline{K}}$$

$$\Omega = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

Then, Gaussian vector definition says that

$$\underline{z} = \underline{\underline{A}}\,\underline{w}$$

is gaussian vector. provided that $\underline{w} \sim N(0, \underline{\underline{I}})$

$$\underline{\underline{A}} = \underline{\underline{K}}^{1/2} \longrightarrow \underline{z} = \underline{\underline{K}}^{1/2}\,\underline{w}$$

gaussian vector

$$\rightarrow E\{\underline{z}\} = 0$$

$$\rightarrow \underline{\underline{K}}_z = \underline{\underline{A}}\,\underbrace{\underline{\underline{K}}_w}_{\underline{\underline{I}}}\,\underline{\underline{A}}^T = \underline{\underline{K}}^{1/2}\left(\underline{\underline{K}}^{1/2}\right)^T = \underline{\underline{K}}^{1/2} \cdot \underline{\underline{K}}^{1/2} = \underline{\underline{K}}$$

So, I can create a gaussian vector $\underline{z}$ with a given $\underline{\underline{K}}$ as a covariance matrix provided that $\underline{\underline{K}}$ is symmetric and positive-semidefinite

$$\begin{bmatrix} 1 & a \\ a & 2 \end{bmatrix} \rightarrow \begin{array}{l} tr\{\cdots\} \geq 0 \\ |A| \geq 0 \end{array} \quad 2-a^2>0 \quad |a|<\sqrt{2}$$

$$N=2 \quad , \quad z = \begin{bmatrix} x \\ y \end{bmatrix}$$

corr. coef

standart deviation

$$z \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} , \underbrace{\begin{bmatrix} \sigma_x^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{xy}\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}} \right)$$

$$\det(\;) = \sigma_x^2\sigma_y^2 - \rho_{xy}^2\sigma_x^2\sigma_y^2 \geq 0$$

$$|\rho_{xy}| \leq 1$$

$$\underline{\underline{K_z}} = E\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} x & y \end{bmatrix} \right\} = \begin{bmatrix} var\{x\} & \boxed{cov(x,y)} \\ cov(y,x) & var\{y\} \end{bmatrix} \rightarrow \frac{cov(x,y)}{var(x)\,var(y)} = \rho_{xy}$$

$$\downarrow$$

$$\frac{cov(x,y)}{\sigma_x\sigma_y} = \rho_{xy}$$

$$cov(x,y) = \rho_{xy}\sigma_x\sigma_y$$

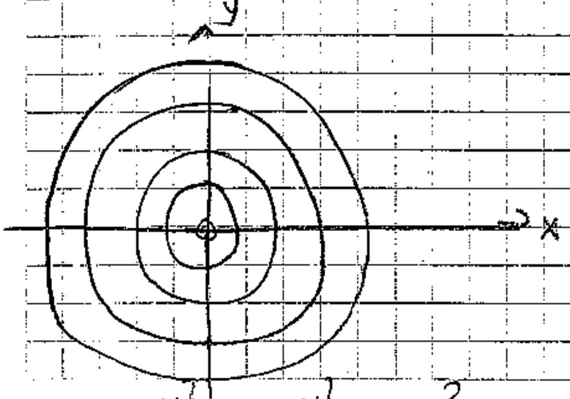
$$f_{\underline{z}}(z) = \frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{\det(K_z)}} e^{-\frac{1}{2}\underline{z}^T \underline{\underline{K_z}}^{-1} \underline{z}} \Bigg|_{N=2}$$

$$= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \exp\left(\frac{-\frac{x^2}{\sigma_x^2} + \frac{2\rho_{xy}\,xy}{\sigma_x\sigma_y} - \frac{y^2}{\sigma_y^2}}{2(1-\rho_{xy}^2)} \right)$$

$\rho_{xy} = 0$, $\sigma_x^2 = \sigma_y^2 = \sigma^2$ (uncorrelated x and y ($\rho_{xy} = 0$)
but

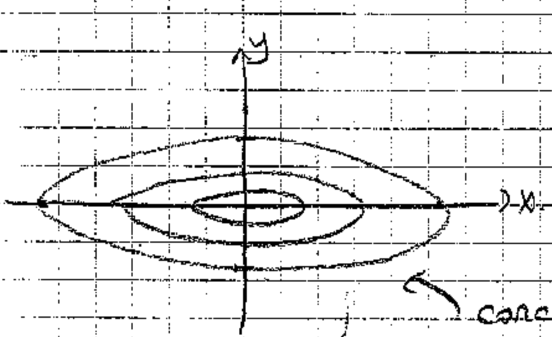
$f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$)

only for Gaussian r.v uncorrelatedness
implies ↓
independence



$$\frac{x^2}{\sigma^2} + \frac{y^2}{\sigma^2} = c^2$$

$\rho_{xy} = 0$, $\sigma_x^2 \neq \sigma_y^2$



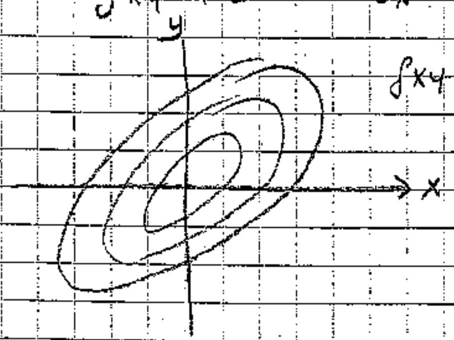
$\sigma_x^2 > \sigma_y^2$

$\Rightarrow$ along x direction → more variation

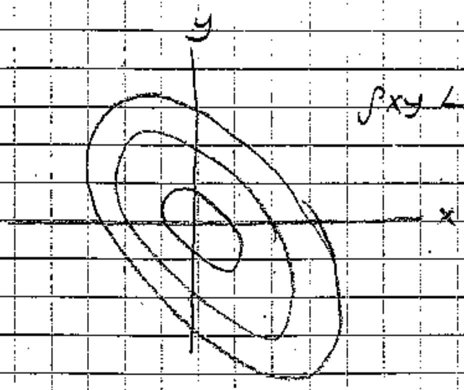
concentrated ellipses

$$\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} = c^2$$

$\rho_{xy} \neq 0$ $\sigma_x^2 \neq \sigma_y^2$



$\rho_{xy} > 0$



$\rho_{xy} < 0$

24/11/2014

$$\frac{k_1 x^2 + k_2 y^2 + k_{12} xy}{\exp\left(-\frac{1}{2} \underline{z}^T \underline{K_z}^{-1} \underline{z}\right)} = c^2$$

Gaussian Processes:

A stochastic process with process variable t is called Gaussian if its samples $t_1, t_2, \ldots, t_N$ is jointly Gaussian distributed for all N (N: number of samples), $\forall t_1, t_2, \ldots \forall t_N$

Def: $E\{x(t)\} = \mu_x(t) \leftarrow$ mean function

$$K_x(t,\tau) = E\{[x(t) - \underbrace{\bar{x}(t)}_{\mu_x(t)}][x(\tau) - \underbrace{\bar{x}(\tau)}_{\mu_x(\tau)}]\} \leftarrow \text{cov. function}$$

Note: Gaussian process is completely characterized by mean function and covariance function

Ex: $x(t)$: gaussian Process

$E\{x(t)\} = 2t+3 = \mu_x(t)$

$K_x(t,\tau) = 3e^{-|t-\tau|}$

a) Find pdf $X(5)$. $x(5) \sim N(\overset{13}{\mu_x(5)}, \overset{K_x(5,5)}{var\{x(5)\}}) = N(13, 3)$

b) Find joint pdf $x(5)$ and $x(10)$

$$\begin{bmatrix} x(5) \\ x(10) \end{bmatrix} \sim N\left(\underbrace{\begin{bmatrix} \mu_x(5) \\ \mu_x(10) \end{bmatrix}}_{\begin{bmatrix} 13 \\ 23 \end{bmatrix}}, \underbrace{\begin{bmatrix} K_x(5,5) & K_x(5,10) \\ K_x(10,5) & K_x(10,10) \end{bmatrix}}_{\begin{bmatrix} 3 & 3e^{-5} \\ 3e^{-5} & 3 \end{bmatrix}}\right)$$

c) $E\{(x(5) - x(6))^2\} = E\{(x(5))^2\} + E\{(x(6))^2\} - 2E\{(x(5)x(6))\}$

$$= (K_x(5,5) + \mu_x^2(5)) + (K_x(6,6) + \mu_x^2(6))$$

$$- 2(K_x(5,6) + \mu_x(5)\mu_x(6))$$

Ex: $X_n = \alpha X_{n-1} + w_n \quad , \quad n \geq 1$, filter is stable and causal $\rightarrow |\alpha| < 1$

$x_0 \sim N(\mu_0, \sigma_0^2)$

initial cond

w_n: i.i.d Gaussian with dist $N(0, \sigma_w^2)$ and w_n's are independent of X_0.

a) Find pdf of X_n, $n \geq 0$

$$X_n = \underbrace{X_0 \alpha^n}_{\text{zero-input solution}} + \underbrace{\sum_{k=1}^{n} \alpha^{n-k} w_k}_{\text{zero-stable solution}} \quad , \quad n \geq 0.$$

$X_0 = X_0$
$X_1 = \alpha X_0 + w_1$
$X_2 = \alpha(\alpha X_0 + w_1) + w_2$
$\quad = \alpha^2 X_0 + \alpha w_1 + w_2$

So, X_n is Gaussian distributed

$E\{x[n]\} = \mu_0 \alpha^n$

$$\text{var}\{x[n]\} = \text{var}\{\alpha^n X_0\} + \text{var}\left\{\sum_{k=1}^{n} \alpha^{n-k} w_k\right\} \quad \left(\text{since } X_0 \text{ and } w_k\text{'s are independent}\right)$$

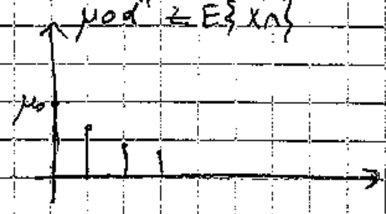
$$= \alpha^{2n} \sigma_0^2 + \sum_{k=1}^{n} \underbrace{\text{var}\{\alpha^{n-k} w_k\}}_{\alpha^{2(n-k)} \text{var}\{w_k\}} \quad \left(w_k\text{'s are independent}\right)$$

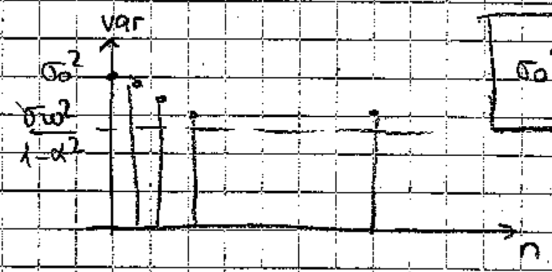
$$= \alpha^{2n} \sigma_0^2 + \underbrace{\left[\sum_{k=1}^{n} \alpha^{2(n-k)}\right]}_{\alpha^{2(n-1)} + \alpha^{2(n-2)} + \cdots + \alpha^2 + 1} \sigma_w^2$$

$$= \alpha^{2n} \sigma_0^2 + \frac{1 - \alpha^{2n}}{1 - \alpha^2} \cdot \sigma_w^2$$

$$= \alpha^{2n}\left(\sigma_0^2 - \frac{1}{1-\alpha^2}\sigma_w^2\right) + \frac{1}{1-\alpha^2}\sigma_w^2 \quad (*)$$

$$x_n \sim N(\mu_0 \alpha^n, (*))$$





$$\boxed{\sigma_0^2 > \frac{\sigma_w^2}{1-\alpha^2}}$$

So, $E\{x_n\} \to 0$.

$$\text{var}\{x_n\} \to \frac{\sigma_w^2}{1-\alpha^2} \quad \text{as } n \to \infty$$

b) $$K_x[n,k] = E\{(x_n - \underbrace{\bar{x}_n}_{\mu_0\alpha^n})(x_k - \underbrace{\bar{x}_k}_{\mu_0\alpha^k})\}$$

$$= E\left\{\left[(x_0-\mu_0)\alpha^n + \sum_{l_1=1}^{n}\alpha^{n-l_1}w_{l_1}\right]\left[(x_0-\mu_0)\alpha^k + \sum_{l_2=1}^{k}\alpha^{k-l_2}w_{l_2}\right]\right\}$$

$$= \text{var}\{x_0\}\,\alpha^{n+k} + E\left\{\sum_{l_1=1}^{n}\sum_{l_2=1}^{k}\alpha^{n-l_1}\alpha^{k-l_2}w_{l_1}w_{l_2}\right\}$$

$$= \alpha^{n+k}\sigma_0^2 + \sum_{l_1}\sum_{l_2}\alpha^{n-l_1}\alpha^{k-l_2}\underbrace{E\{w_{l_1}w_{l_2}\}}_{\sigma_w^2\delta[l_1-l_2]} = \begin{cases}\sigma_w^2 & l_1 = l_2\\ 0 & l_1 \neq l_2\end{cases}$$

$$= \alpha^{n+k}\sigma_0^2 + \sum_{l_1=1}^{k}\alpha^{n-l_1}\alpha^{k-l_1}\sigma_w^2$$

Assume that $k < n$, $\min(k,n) = k$

without any loss of generality

$$= \alpha^{n+k}\sigma_0^2 + \sigma_w^2\underbrace{\sum_{l_1=1}^{k}\alpha^{n-l_1}\alpha^{k-l_1}}_{\sum_{m=0}^{k-1}\alpha^{n-k+m}\cdot\alpha^{m}} \quad (k-l_1 = m)$$

$$= \alpha^{n+k}\sigma_0^2 + \sigma_w^2\alpha^{n-k}\sum_{m=0}^{k-1}\alpha^{2m}$$

$$= \alpha^{n+k}\sigma_0^2 + \sigma_w^2\alpha^{n-k}\frac{1-\alpha^{2k}}{1-\alpha^2}$$

01/12/2014

c) Joint pdf

$$\begin{bmatrix} X_3 \\ X_5 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_X(3) \\ \mu_X(5) \end{bmatrix}, \begin{bmatrix} K[3,3] & K[3,5] \\ K[5,3] & K[5,5] \end{bmatrix} \right)$$

Stationary Process:

1st order stationarity:

A process $X(t)$ is 1st order stationary

if $f_{X(t_1)}(x_1) = f_{X(\underbrace{t_1+\Delta}_{t_2=t_1+\Delta})}(x_1)$ $\forall t_1, \forall \Delta$

density for $X(t_1)$

2nd order Stationarity:

A process is 2nd order stationary

if $f_{X(t_1),X(t_2)}(x_1,x_2) = f_{X(t_1+\Delta),X(t_2+\Delta)}(x_1,x_2)$ $\forall t_1, \forall t_2, \forall \Delta$

joint pdf for $X(t_1)$ and $X(t_2)$

Nth order stationarity

$$f_{X(t_1),X(t_2),\dots,X(t_N)}(x_1,x_2,\dots,x_N) = f_{X(t_1+\Delta),X(t_2+\Delta),\dots,X(t_N+\Delta)}(x_1,\dots,x_N) \quad \forall t_1,t_2,\dots,t_N \;\; \forall \Delta$$

A process is strict sense stationary (SSS) if it is N^{th} order stationary for all N.

Let's focus on Gaussian Process and examine the conditions for which Gaussian process is stationary.

$x(t)$: gaussian process

1^{st} order:

$x(t_1) \sim N_{x_1}(\mu_x(t_1), \sigma^2_{x(t_1)})$

$x(t_2) \sim N_{x_2}(\mu_x(t_2), \sigma^2_{x(t_2)}) \qquad (t_2 \neq t_1)$

so, 1^{st} order stationarity requires $\rightarrow$ ① $\mu_x(t_1) = \mu_x(t_2)$ ② $\sigma^2_{x(t_1)} = \sigma^2_{x(t_2)}$ } since equalities ① and ② should be satisfied for all t_1 and t_2

$\Downarrow$

$\mu_x(t) = \text{constant}$

$\sigma^2_x(t) = \text{constant}$

2^{nd} order.

$$\begin{bmatrix} x(t_1) \\ x(t_2) \end{bmatrix} \sim N_{x_1,x_2}\left(\begin{bmatrix} \mu_x(t_1) \\ \mu_x(t_2) \end{bmatrix}, \underbrace{\begin{bmatrix} cov(x(t_1), x(t_1)) & cov(x(t_1), x(t_2)) \\ cov(x(t_2), x(t_1)) & cov(x(t_2), x(t_2)) \end{bmatrix}}_{C_{x(t_1), x(t_2)}} \right)$$

$$\begin{bmatrix} x(t_1+\Delta) \\ x(t_2+\Delta) \end{bmatrix} \sim N_{x_1,x_2}\left(\begin{bmatrix} \mu_x(t_1+\Delta) \\ \mu_x(t_2+\Delta) \end{bmatrix}, \begin{bmatrix} cov(x(t_1+\Delta), x(t_1+\Delta)) & cov(x(t_1+\Delta), x(t_2+\Delta)) \\ cov(x(t_2+\Delta), x(t_1+\Delta)) & cov(x(t_2+\Delta), x(t_2+\Delta)) \end{bmatrix} \right)$$

Then 2^{nd} order stationarity

① $\mu_x(t) = \text{constant} \quad \forall t$

② $cov(x(t_1), x(t_2)) = cov(x(t_1+\Delta), x(t_2+\Delta)) \quad \forall t_1 \; \forall t_2 \; \forall \Delta$

$\hookrightarrow$ then set $\Delta = -t_1$

$\Downarrow$

$$cov(x(t_1), x(t_2)) = cov(x(0), x(\underbrace{t_2 - t_1}_{u = t_2 - t_1}))$$

$\downarrow$

then cov. function $cov(x(t_1), x(t_2))$ can be written as a function of $u = t_2 - t_1$ (function of single variable)

N^{th} order case:

$$\begin{bmatrix} x(t_1) \\ \vdots \\ x(t_N) \end{bmatrix} \sim N\left(\underline{\mu}_{x(t_1), x(t_2), \ldots}, \underline{\underline{K}}_{x(t_1), \ldots, x(t_N)} \right)$$

$$\begin{bmatrix} x(t_1+\Delta) \\ \vdots \\ x(t_N+\Delta) \end{bmatrix} \sim N\left(\underline{\mu}_{x(t_1+\Delta), x(t_2+\Delta), \ldots}, \underline{\underline{K}}_{x(t_1+\Delta), x(t_2+\Delta), \ldots, x(t_N+\Delta)} \right)$$

so, for N^{th} order stationarity

① $\mu_x(t)$ = constant

② $\text{cov}(x(t_k), x(t_\ell)) = \text{cov}(x(t_k+\Delta), x(t_\ell+\Delta)) \quad \forall \Delta$

$= \text{cov}(x(0), x(t_\ell - t_k))$

conditions are the same as 2^{nd} order stationarity

Wide Sense Stationarity:

In many applications, stationarity in the pdf sense can not be checked or guaranteed; we use a relaxed form of stationarity which is more practical in many applications.

WSS:

① $E\{x(t)\}$ = constant $\forall t$,

mean function, $\mu_x(t)$

② $\text{cov}(x(t_1), x(t_2)) = \text{func}(t_2 - t_1) \quad \forall t_1, \forall t_2$

cov. func can be written as a function of $t_2 - t_1$, not t_2 and t_1.

if ① and ② are satisfied

$\downarrow$

$x(t)$: WSS ✓

It should be clear that WSS check does not require any knowledge of joint pdf; but $E\{\cdot\}$ calculations are sufficient for the WSS check

Note: ① A 1^{st} order stationary process is also stationary in the mean, $\mu_x(t)$ = constant. (①st check of WSS)

② A 2^{nd} order stationary process is also stationary in the covariance function,
(2^{nd} check of WSS)

So, from ① and ② we can say that $\boxed{2^{nd} \text{ order stationarity} \xrightarrow[\text{implies}]{} \text{WSS}}$

(A process that is 2^{nd} order stationary is guaranteed to be first order stationary, since

$$f_{x(t_1),x(t_2)}(x_1,x_2) = f_{x(t_1+\Delta),x(t_2+\Delta)}(x_1,x_2)$$

can be marginalized wrt to x_2 $\left(\int_{\infty}^{\infty} \text{--} dx_2 \right)$ then we have 1^{st} order stationarity)

Note:

$$\text{WSS} \underset{\text{does not imply}}{\not\longrightarrow} \text{even } 1^{st} \text{ order stationarity.}$$

Since WSS is about moments, but not joint pdf's.

Note: An important special case is Gaussian process;
stationarity requires for joint pdf sense coincides with WSS checks,
so, $x(t)$: gaussian process and WSS $\xrightarrow[\text{implies}]{}$ $x(t)$: Gaussian, SSS

03/12/2014

STATIONARY PROCESS

1) stationarity in joint-pdf (1^{st} order, 2^{nd} order, -- , N^{th} order, -- SSS)

2) wide sense stationarity (WSS) → ① $E\{x(t)\}$ = constant ← stationary in the mean
→ ② $E\{x(t)x(t-\tau)\}$ = func(τ) ← stationary in auto-correlation

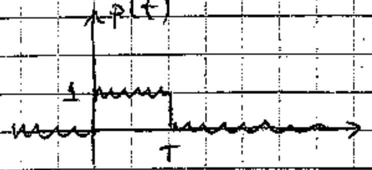
* 2^{nd} order stationarity → WSS

* $\underbrace{\text{WSS}}_{\text{about moments}} \not\longrightarrow$ even $\underbrace{1^{st} \text{ order stationarity}}_{\text{about pdf}}$

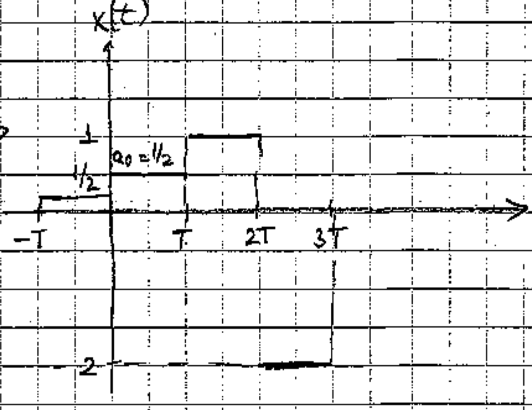
* If process is Gaussian → SSS
and $\underbrace{\text{WSS}}$

expected value & cov are invariant of the t

Ex:



$x(T) = \sum_{k=-\infty}^{\infty} a_k \, p(t-kT)$ ---->



a_k is i.i.d $N(0,\sigma^2)$

Q: Is $x(t)$ stationary?

1st order stationary:

$f_{x(t_1)}(x_1) \sim N_{x_1}(0,\sigma^2)$

$f_{x(t_2)}(x_1) \sim N_{x_1}(0,\sigma^2)$

So, $x(t)$ is 1st order stationary

$\lfloor \cdot \rfloor$ = floor func. of Matlab

$\lfloor x \rfloor$ = largest integer less than x $\lfloor 1.3 \rfloor = 1$.

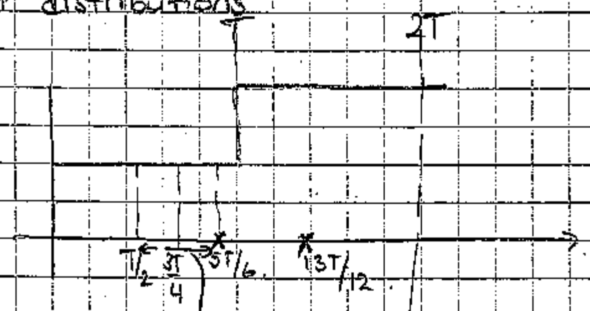
2nd order stationary:

$$f_{x(t_1)x(t_2)}(x_1,x_2) = \begin{cases} f_{x(t_1)}(x_1)\, f_{x(t_2)}(x_2), & \lfloor \frac{t_1}{T} \rfloor \neq \lfloor \frac{t_2}{T} \rfloor \\ f_{x(t_1)}(x_1)\, \underbrace{\delta(x_2 - x_1)}_{f_{x(t_2)|x(t_1)}(x_2|x_1)}, & \text{other} \end{cases}$$

Q: Do we have 2nd order stationary?

No, we do not; since by giving a shift of Δ, we can have two different distributions

Ex:



So, this example shows that $x(t)$ is not 2^{nd} order stationary.

No need to check 3^{rd} or higher order stationarities.

Q: Is $x(t)$ WSS?

1. $E\{x(t)\} = ?$ $\rightarrow E\{x(t)\} = 0$

2. $E\{x(t)\,x(t-\tau)\} \stackrel{?}{=} \text{func}(\tau)$

$E\{x(t_1)\,x(t_2)\} \stackrel{?}{=} \text{func}(t_1 - t_2)$

Ex: $t_2 - t_1 = T/4$ $\begin{cases} t_1 = T/2 \\ t_2 = 3T/4 \end{cases}$ 2 samples are in the same pulse $\rightarrow E\{\underbrace{x(t_1)}_{a_0}\underbrace{x(t_2)}_{a_0}\} = E\{a_0^2\} = \sigma^2$

$t_2 - t_1 = T/4$ $\begin{cases} t_1 = T/2 + \Delta = 5T/6 \\ t_2 = 3T/4 + \Delta = 13T/12 \end{cases}$ $\Delta = T/3$

$\rightarrow E\{x(t_1)\,x(t_2)\} = E\{a_0\}\,E\{a_1\}$
$= 0 \cdot 0 = 0$

$x(t)$ is stationary in the mean,
$x(t)$ is not stationary in the autocorrelation
$\}$ $x(t)$ is not WSS.

MT#2 ; (13:30 – 15:30) (21^{st} Dec. Sunday)

Properties of Auto-correlation Function for WSS Processes:

(0) For zero-mean processes, autocorrelation function $E\{x(t_1)\,x(t_2)\}$ is identical to covariance function $E\{(x(t_1)-0)(x(t_2)-0)\}$

$$\underset{\substack{\uparrow \\ \text{autocorrelation sequence}}}{r_x[k]} = E\{x[n]\,x^*[n-k]\}$$

(1) $r_x[k] = r_x^*[-k]$

Proof: $r_x[-k] = E\{x[n]\,x^*[n-(-k)]\}$
$= E\{x^*[n+k]\,x[n]\}$
$= E\{x[n+k]\,x^*[n]\}^* = r_x[k]^*$

So, if $x[n]$ is real-valued, $r_x[k]$ is an even sequence, $r_x[k] = r_x[-k]$

$x[n]$ is complex valued, $r_x[k]$ is hermitian symetric, $r_x[k] = r_x^*[-k]$

② $r_x[0] \geq |r_x[k]| \quad \forall k$

$$r_x[0] = E\{x[n]x[n]\} = E\{\underbrace{x^2[n]}_{\text{Ensemble power of the process } x[n]}\}$$

Proof: two rv's $z = x[n]$, $w = x[n-k]$ $\rightarrow |\underset{\text{corr coef}}{\rho_{zw}}| \leq 1 \rightarrow \rho_{zw} = \dfrac{E\{zw\}}{\sqrt{E\{z^2\}E\{w^2\}}}$ only valid for zero-mean z and w.

$$\rho_{zw} = \left|\frac{r_x[k]}{\sqrt{r_x[0]\,r_x[0]}}\right| \leq 1$$

$$r_x[0] \geq |r_x[k]| \checkmark$$

③ If $r_x[0] = r_x[N] \quad \exists N \neq 0$

$r_x[k]$ is periodic by N. $\left(r_x[k] = r_x[k+N] \quad \forall k\right)$

Proof: 2 rv's $z = x[n-k] - x[n-k-N]$, $w = x[n]$

$$|\rho_{zw}| \leq 1 \rightarrow (E\{zw\})^2 \leq E\{z^2\}E\{w^2\} \qquad 2(r_x[0] - r_x[N]) = 0$$

$$(r_x[k] - r_x[k-N])^2 \leq E\{z^2\}\,r_x[0]$$

$$\hookrightarrow E\{z^2\} = E\{(x[n-k] - x[n-k-N])^2\}$$

$$= E\{x^2[n-k] - 2x[n-k]x[n-k-N] + x^2[n-k-N]\}$$

$$= r_x[0] - 2r_x[N] + r_x[0]$$

So, this shows that $(r_x[k] - r_x[k-N])^2 \leq 0$

$$\downarrow$$

$$r_x[k] = r_x[k-N] \quad \forall k$$

④ Auto-corr Matrix for wss $x[n]$

$$\underset{3\times 1}{\underline{z}} = \begin{bmatrix} x[n] \\ x[n-1] \\ x[n-2] \end{bmatrix} \longrightarrow \underline{\underline{R}}_z = E\{\underline{z}\,\underline{z}^H\} = E\left\{ \begin{bmatrix} x[n] \\ x[n-1] \\ x[n-2] \end{bmatrix} \begin{bmatrix} x^*[n] & x^*[n-1] & x^*[n-2] \end{bmatrix} \right\}$$

$$= \begin{bmatrix} r_x[0] & r_x[1] & r_x[2] \\ r_x[-1] & r_x[0] & r_x[1] \\ r_x[-2] & r_x[-1] & r_x[0] \end{bmatrix}$$

$(r_x[-k] = r_x^*[k])$

$$= \begin{bmatrix} r_x[0] & r_x[1] & r_x[2] \\ r_x^*[1] & r_x[0] & r_x[1] \\ r_x^*[2] & r_x^*[1] & r_x[0] \end{bmatrix}$$

① Hermitian symmetric matrix

② $\underline{\underline{R}}_z \geq 0$

Valid for all R_z matrices (not specific for wss process samples)

③ $\underline{\underline{R}}_z$: Toeplitz structure

over diagonal and sub/super diagonals we have the same value in the matrix.

⑤ A sequence is a valid autocorrelation sequence

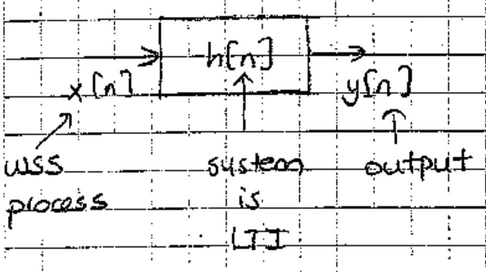
if and only if ($\Longleftrightarrow$)

$$\underline{\underline{R}}_x = E\left\{ \begin{bmatrix} x[n] \\ \vdots \\ x[n-(N-1)] \end{bmatrix} \begin{bmatrix} x^*[n] & \cdots & x^*[n-(N-1)] \end{bmatrix} \right\}$$

is $\underline{\underline{R}}_x \geq 0$ for all N. ($\underline{\underline{R}}_x$: $N\times N$)

08/12/2014

Filtering of WSS Processes:



$$y[n] = \sum_{k=-\infty}^{\infty} h[k]\, x[n-k]$$

Q: If $x[n]$ is WSS, what can we say about the stationarity of $y[n]$?

Let's check whether $y[n]$ satisfies WSS conditions:

1. $E\{y[n]\} \stackrel{?}{=}$ constant

$$E\{y[n]\} = \sum_{k=-\infty}^{\infty} h[k] \underbrace{E\{x[n-k]\}}_{\mu_x \text{ (since } x[n] \text{ is WSS)}} = \mu_x \sum_{k=-\infty}^{\infty} h[k] = \mu_x \, H(e^{j\omega})\Big|_{\omega=0}$$

$$\left(H(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h[n]\, e^{-j\omega n} \right)$$

provided that $H(z)\Big|_{z=1}$ is finite (i.e., no poles at $z=1$), then

$E\{y[n]\} =$ constant,

so, 1^{st} condition for WSS is satisfied.

2. $E\{y[\underbrace{n}_{n_1}]\, y[\underbrace{n-k}_{n_2}]\} \stackrel{?}{=}$ func(k) $\forall n$

i) $E\{y[n]\, x^*[n-k]\} = E\left\{ \sum_{k'=-\infty}^{\infty} h[k']\, x[n-k']\, x^*[n-k] \right\}$

since $x[n]$ is WSS

$$= \sum_{k'=-\infty}^{\infty} h[k']\, r_x[k-k'] = h[k] * r_x[k]$$

then $E\{y[n]\, x[n-k]\} = \underbrace{r_{yx}[n, n-k]}_{\text{depends only on "}k\text{"}} = h[k] * r_x[k]$

ii) $E\{y[n]\overset{*}{y}[n-k]\} \overset{?}{=} \text{func}(k)$ $\qquad y[n] = \sum_k h[k]\,x[n-k]$

$$E\{y[n]\overset{*}{y}[n-k]\} = E\Big\{y[n]\Big(\sum_{k'} h[k']\,x[n-k-k']\Big)^{*}\Big\}$$

$$= \sum_{k'} \overset{*}{h}[k']\,\underbrace{E\{y[n]\overset{*}{x}[n-k-k']\}}_{r_{yx}[k+k']}$$

$$= \sum_{k'=-\infty}^{\infty} \overset{*}{h}[k']\,r_{yx}[k+k']$$

$(k' = -\ell)$

$$= \sum_{\ell=-\infty}^{\infty} \overset{*}{h}[-\ell]\,r_{yx}[k-\ell] = \overset{*}{h}[-k] * r_{yx}[k]$$

then $E\{y[n]\overset{*}{y}[n-k]\} = \overset{*}{h}[-k] * r_{yx}[k]$

$$= \overset{*}{h}[-k] * h[k] * r_x[k]$$

$\boxed{r_y[k] = h[k] * \overset{*}{h}[-k] * r_x[k]}$ ← since RHS depends only on "k", but not "n", $y[n]$ satisfies 2^{nd} condition and it is WSS.

$x[n] \longrightarrow \boxed{h[n]} \longrightarrow y[n]$

$x[n]$ ↑ input WSS $\qquad$ $y[n]$ ↑ then output is also WSS!

Joint WSS: Two processes $(x[n], y[n])$ are called jointly WSS

if

① $x[n]$ is WSS
② $y[n]$ is WSS
} individually WSS

③ $\underbrace{E\{x[n]y[n-k]\}}_{r_{xy}[n,n-k]} = \text{func}(k)$

Comment: $\underset{\text{WSS}}{x[n]} \rightarrow \boxed{LTI} \rightarrow y[n]$, if $x[n]$ is WSS, then $x[n], y[n]$ are jointly WSS.

Power Spectral Density:

$r_y[k] \rightleftarrows S_y(e^{j\omega})$

autocorrelation of a WSS sequence $y[n]$

power spectral density

$e^{-\alpha t} \xrightarrow[F.T]{} \frac{1}{j\omega+\alpha}$

$$S_y(e^{j\omega}) \triangleq DTFT\{r_y[k]\} = \sum_{k=-\infty}^{\infty} r_y[k]\, e^{-j\omega k}$$

$$r_y[k] \triangleq IDTFT\{S_y(e^{j\omega})\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_y(e^{j\omega})\, e^{j\omega k}\, d\omega.$$

$$x[n] \longrightarrow \boxed{h[n]} \longrightarrow y[n]$$

$x[n]$	$y[n]$
$r_x[k]$	$r_y[k]$
$\updownarrow$ DTFT	$\updownarrow$ DTFT
$S_x(e^{j\omega})$	$S_y(e^{j\omega})$

$$S_y(e^{j\omega}) = H(e^{j\omega})\, H^*(e^{j\omega})\, S_x(e^{j\omega})$$

$$\boxed{S_y(e^{j\omega}) = |H(e^{j\omega})|^2\, S_x(e^{j\omega})}$$

PSD of the output

Note: Cross power spectral density

$$S_{yx}(e^{j\omega}) = DTFT\{r_{yx}[k]\}$$

10/12/2014

Properties of $S_y(e^{j\omega})$

1. $S_y(e^{j\omega})$ is real-valued.

(since $r_y[k] = r_y^*[-k]$)

2. $S_y(e^{j\omega}) \geq 0$,

Proof: (Assume $S_y(e^{j\omega})$ is the output of a LTI system with input $S_x(e^{j\omega}) = 1$ (white noise), white noise; $r_x[k] = \delta[k]$)

then $S_y(e^{j\omega}) = \underbrace{|H(e^{j\omega})|^2}_{\geq 0} \underbrace{S_x(e^{j\omega})}_{1}$

3. Q: Is any non-negative function say $S(e^{j\omega}) \geq 0$ a valid power spectral density?

A: Yes, For proof

see 503 notes or Papoulis's book on prob and random variables

POWER SPECTRAL DENSITY (CONT'D)

$$S_y(e^{j\omega}) = \text{DTFT}\{r_y[k]\}$$

Note:

$$r_y[0] = \text{IDFT}\{S_y(e^{j\omega})\}\Big|_{k=0}$$

$$E\{(y[n])^2\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_y(e^{j\omega})\, e^{j\omega k}\, d\omega \Big|_{k=0}$$

Ensemble power
power of random sequence $y[n]$

$$= \frac{1}{2\pi}\,(\text{Area under PSD})$$

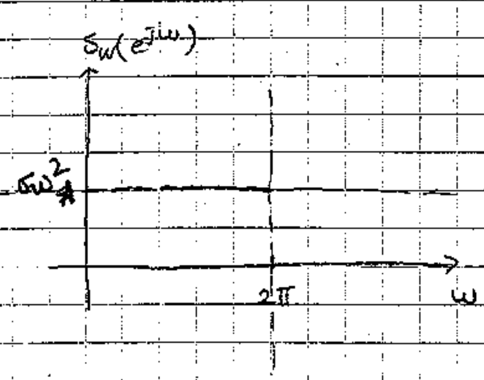
Properties:

1. $S_y(e^{j\omega})$: real valued

2. $S_y(e^{j\omega}) \geq 0$

3. Any non-negative func with finite "area" can be considered as a PSD of a process.

Ex:



$S_w(e^{j\omega}) = A \quad 0 \le \omega \le 2\pi$

$$\underbrace{E\{(w[n])^2\}}_{\text{power}} = \frac{1}{2\pi}\left(2\pi \cdot \sigma_w^2\right) = \sigma_w^2$$

white noise

If PSD is flat (i.e. constant) for all ω's, then the associated process is called white noise. (Implicitly in all definitions we assume "noise" is zero-mean)

$E\{w[n]w[n-k]\} = 0, \; k \ne 0$

or if $r_w[k] = \sigma_w^2 \delta[k]$ for the WSS process $\rightarrow w[n]$ is white noise.

i.e. white noise samples are uncorrelated ($r_w[k] = 0, \; k \ne 0$)

Ex:

$$x[n] \rightarrow \boxed{H(z) = \frac{1}{1-\alpha z^{-1}}} \rightarrow y[n]$$

$x[n]$: white noise with variance σ_w^2

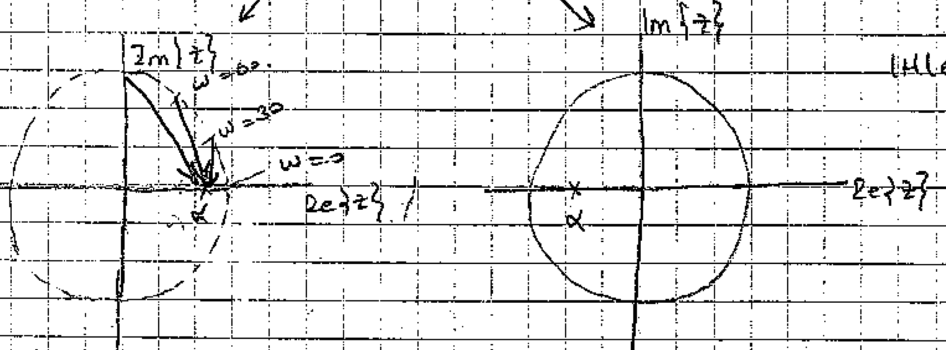
$H(z)$: single-pole LTI system

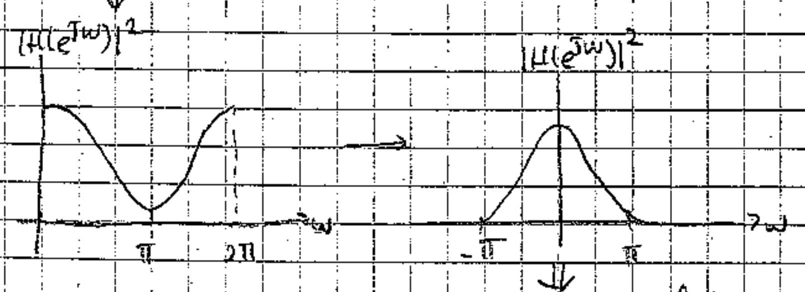
$$H(z)\Big|_{z=e^{j\omega}} = \frac{1}{1-\alpha z^{-1}}\Big|_{z=e^{j\omega}} = \frac{z}{z-\alpha}$$

$$H(e^{j\omega}) = H(z)\Big|_{z=e^{j\omega}} = \frac{e^{j\omega}}{e^{j\omega}-\alpha}$$

$$|H(e^{j\omega})| = \frac{1}{|e^{j\omega}-\alpha|} \quad \omega = 0:30^\circ:360^\circ$$

$$S_y(e^{j\omega}) = |H(e^{j\omega})|^2 \cdot S_x(e^{j\omega})$$



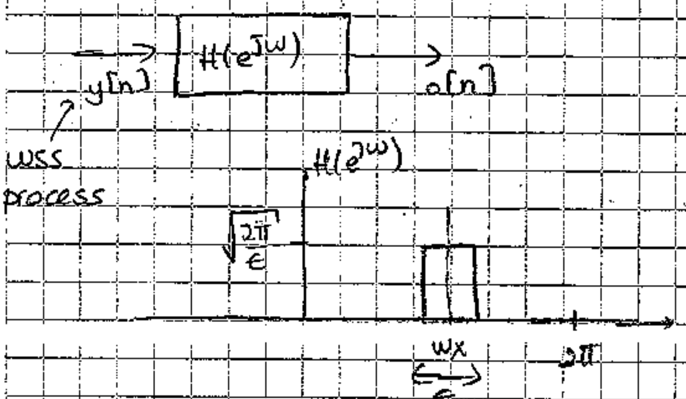


$$S_y(e^{jw}) = \frac{1}{|1-\alpha e^{-jw}|^2} \sigma_w^2$$

pole location

In the figure, $\alpha \approx 1$, so $H(z)$ is a low-pass filter $\Rightarrow y[n]$ is called a low pass process.

Interpretation for PSD:



$H(e^{jw})$ is a bandpass filter centered around $w = w_x$.

Q. What's output process variance?

$$E\{(o[n])^2\} = \frac{1}{2\pi}\int_0^{2\pi} S_o(e^{jw})\,dw$$

$$= \frac{1}{2\pi}\int_0^{2\pi} |H(e^{jw})|^2 S_y(e^{jw})\,dw$$

$$= \frac{1}{2\pi}\int_{|w-w_x|<\frac{\epsilon}{2}} \left(\sqrt{\frac{2\pi}{\epsilon}}\right)^2 S_y(e^{jw})\,dw = \frac{1}{\epsilon}\int_{|w-w_x|<\frac{\epsilon}{2}} S_y(e^{jw})\,dw \approx S_y(e^{jw_x})$$

(for small ϵ)

Markov Chains:

Let X_n, $n = \{0, 1, 2, \ldots\}$, be a random process taking finite or countable number of possible values,

$X_n \in \{1, 2, 3, \ldots\}$

We call X_n as the state of the process at time "n" and consider that process jumps from state to state with some probabilities at any time instant.

The process is said to be Markov, if

$$P\{X_{n+1} = j \mid X_n = i, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1, X_0 = x_0\}$$

$$=$$

$$P\{X_{n+1} = j \mid X_n = i\} = P_{ij}$$ ← state transition probability from state i to state j.

So, given the present state (X_n), the future state (X_{n+1}) is independent from states in history ($X_0, X_1, \ldots, X_{n-1}$)

↓

this kind of independence is called conditional independence

So, future is ind. from past given present sample

Q: $P\{X_{n-1} = A \mid X_n = B, X_{n+1} = x_{n+1}, \ldots, X_{n+m} = x_{n+m}\} = ?$

X_n: Markov Process

$$= \frac{P\{X_{n-1} = A, X_n = B, \overbrace{\boxed{X_{n+1} = x_{n+1}, \ldots, X_{n+m} = x_{n+m}}}^{F}\}}{P\{X_n = B, \underbrace{\boxed{X_{n+1} = x_{n+1}, \ldots, X_{n+m} = x_{n+m}}}_{F}\}}$$

$$= \frac{P\{X_{n-1}, X_n, F\}}{P\{X_n, F\}} = \frac{P\{X_{n-1}\} P\{X_n \mid X_{n-1}\} \cancel{P\{F \mid X_n, X_{n-1}\}}}{P\{X_n\} \cancel{P\{F \mid X_n\}}}$$

$$= \frac{P\{\bar{X}_{n-1}=A\}}{P\{\bar{X}_n=B\}} \cdot \underbrace{P\{\bar{X}_n=B \mid \bar{X}_{n-1}=A\}}_{P_{AB}}$$

So, $P\{\bar{X}_{n-1}=A \mid \bar{X}_n=B, \bar{X}_{n+1}=x_{n+1}, \ldots, \bar{X}_{n+M}=x_{n+M}\}$ does only depend $\bar{X}_n$ (current time) but not on future samples (F)

So, this says that a Markov chain in "reverse time" is also a Markov chain with different state transition probabilities

The Markov chains are denoted as

$$a \rightarrow b \rightarrow c$$ (these a, b, c are rv's)

$a = \bar{X}_{n-1}$, $b = \bar{X}_n$, $c = \bar{X}_{n+1}$

So, if I have $a \rightarrow b \rightarrow c$,

then, $a \leftarrow b \leftarrow c$ is also a Markov chain with different transition probabilities.

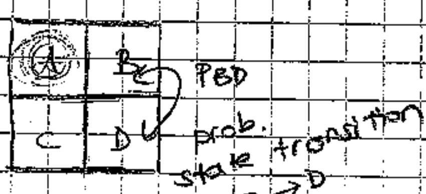
In some books: $a \leftrightarrow b \leftrightarrow c$

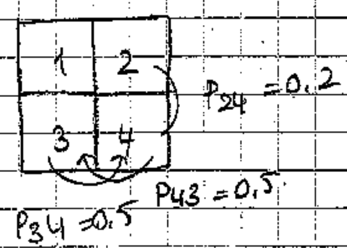
A Markov chain whose transition prob. does not change by time "n" are called "homogenous" Markov chains. We will mostly focus on homogenous Markov chains.

Ex: Spider And Fly

A spider is located at A, a fly moves randomly between A, B, C, D positions without any knowledge of spider

The moves are according to some assigned probabilities.





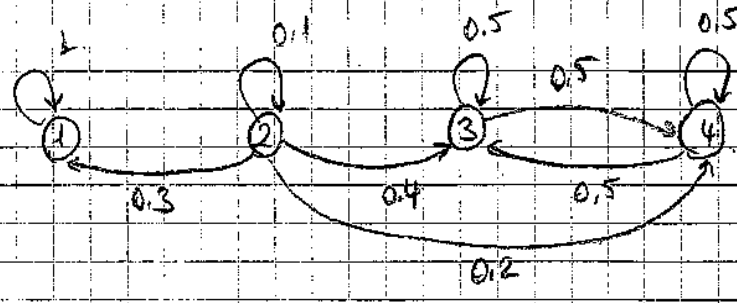
$[\underline{P}]_{ij} = P_{ij}$

$\underline{P}$: state trans. prob. matrix

P_{ij} : state trans. prob. $i \to j$

$$P = \begin{array}{c|cccc|c} & ① & ② & ③ & ④ & \\ \hline ① & 1 & 0 & 0 & 0 & 1 \\ ② & 0.3 & 0.1 & 0.4 & \textcircled{0.2} & 1 \\ ③ & 0 & 0 & 0.5 & 0.5 & 1 \\ ④ & 0 & 0 & 0.5 & 0.5 & 1 \end{array}$$

$P_{2\to 4} = P_{24}$



= State Transition Diagram

<u>Notes:</u>

① Row sums of $\underline{P}$ matrix is equal to 1 ($|S|$: cardinality of states = # of states)

$$\left(\text{since } \sum_{j=1}^{|S|} \underbrace{P\{X_{n+1}=j \mid X_n = i\}}_{P_{ij}} = 1 \right)$$

Such $\underline{P}$ matrices are called stochastic matrices.

② If both row and column sum is equal to 1

→ such matrices are called doubly stochastic matrices.

<u>Q:</u> What is Prob{ Fly captured | $X_0 = i_0$ } $i_0 = \{1, 2, 3, 4\}$

$$P\{\text{Fly captured} \mid X_0 = i_0\} = \begin{cases} 1 & i_0 = 1 \\ ? & i_0 = 2 \\ 0 & i_0 = 3 \\ 0 & i_0 = 4 \end{cases}$$

the value for ? is not uncertain, we need to calculate it, but we can see that reaching safety has prob. at least 0.6 since in the first transition fly can reach states 3 and 4 (reaches safety) with 0.6 prob.

Q: What is 2-step transition prob?

$$P\{X_{n+2}=A \mid X_n=B\} = ?$$

A: $$P\{X_{n+2}=A \mid X_n=B\} = \sum_{k=1}^{|S|} P\{X_{n+2}=A,\ X_{n+1}=s \mid X_n=B\}$$

$$= \sum_{s} P\{X_{n+1}=s \mid X_n=B\}\, P\{X_{n+2}=A \mid X_{n+1}=s, \cancel{X_n=B}\}$$

$$= \sum_{s=1}^{|S|} \underbrace{P_{BS}}_{\substack{\text{state trans prob}\\ B\to S}} P_{SA} = [\underline{P}\ \underline{P}]_{\underbrace{BA}_{\substack{B^{th}\text{ row and}\\ A^{th}\text{ column}}}}$$

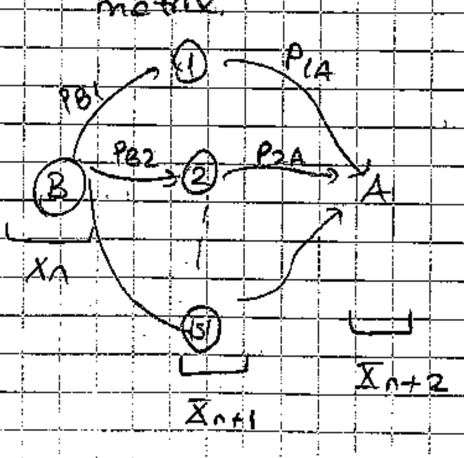
$$\boxed{[AB]_{ij} = \sum_{k=1}^{N} a_{ik} b_{kj}}$$

$$= [\underline{P}^2]_{BA}$$

So, two-step transition matrix is

$$\underline{P}^2 = \underline{P}\ \underline{P}$$

1-step transition matrix.



15/12/2014

So, 3-step transition matrix $\underline{\underline{P}}^3$

n-step transition matrix P^n

Chapman – Kolmogorov Equation:

$$\underline{\underline{P}}^{n+m} = \underline{\underline{P}}^{n} \cdot \underline{\underline{P}}^{m}$$

↓

$$P_{ij}^{(n+m)} = \sum_{k=1}^{|S|} P_{ik}^{(n)} \underbrace{P_{kj}^{(m)}}$$

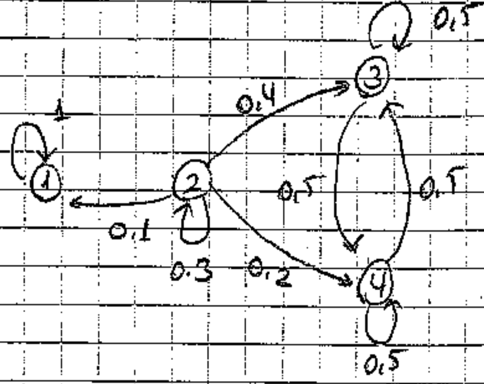
← Chapman – Kolmogorov Equation For discrete Markov chains

↑ m-step transition probability $k \to j$

Gaussian preceslerden soru sorcak
take-home verecek → Markov-chain computer assignment
ödev verecek soru alabilir

Markov Chains (cont'd)

Fly-spider



$$\underline{\underline{P}} = \begin{array}{c} ① \\ ② \\ ③ \\ ④ \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.1 & 0.3 & 0.4 & 0.2 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}$$

↑ probability transition matrix

p_{ij} = Prob{ moving $i \to j$ }

Q. P{ fly moves to state 1 (meets the spider) for the first time in time "n" | $X_0 = 2$ } = ?
↑ initial state

$= P\{X_n = 1, X_{n-1} \neq 1, X_{n-2} \neq 1, \ldots, X_1 \neq 1 \mid X_0 = 2\}$

$= P\{$ fly remains in state 2 for $\{1, 2, \ldots, n-1\}$ and $X_n = 1 \mid X_0 = 2\}$

$= (0.3)^{n-1} \cdot (0.1)$ ← $F_{21}^{(n)}$

Q. P{ fly meets spider at any time | $\bar{X}_0 = 2$ }

$$= \sum_{n=1}^{\infty} P\{\text{fly meets spider first time at instant } n \mid \bar{X}_0 = 2\}$$

$$= \sum_{n=1}^{\infty} (0.3)^{n-1} (0.1) = 0.1 \frac{1}{1-0.3} = \frac{1}{7} \leftarrow F_{21}$$

Q. P{ fly escapes to safety at time n | $\bar{X}_0 = 2$ }

$$= (0.3)^{n-1} \cdot 0.6 \leftarrow \triangleq \cancel{F_{23}^{(n)}} = F_{24}^{(n)}$$

Q. P{ Fly escapes at any time and remains safe }

$$= \sum_{n=1}^{\infty} (0.3)^{n-1} \cdot (0.6) = 0.6 \cdot \frac{1}{1-0.3} = \frac{6}{7} \leftarrow F_{23} = F_{24}$$

Q. P{ fly remains in state 2 at time n | $\bar{X}_0 = 2$ } $= (0.3)^n$

Q. $\underbrace{\text{P\{ fly returns back to state 2 at any time} \mid \bar{X}_0 = 2\}}_{\text{revisiting state 2 any time} \mid \bar{X}_0 = 2} = 0.3$

State 2: "transient" state

State 1: Danger state (fly remains there!)

States 3-4: Safety state (fly reaches safety states and remains there!)

Classification of states:

Recurrent / Transient: state j is <u>recurrent</u> if and only if (iff) starting at j, returning to j has the probability of 1. ($F_{jj} = 1$)

prob{ visiting state j at any time | $\bar{X}_0 = j$ }

state j is transient if $F_{jj} < 1$, that is not returning to state j (starting from state j) has a non-zero probability.

Absorbing State (Trapping State):

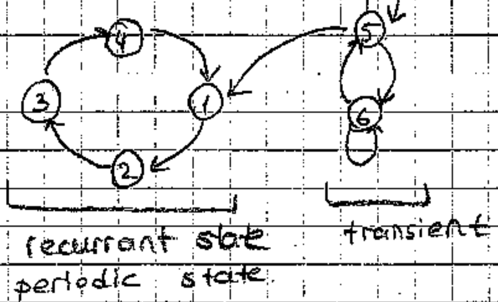
state j is absorbing if $P_{jj} = 1$ (no way out !)

Clearly, absorbing state is a recurrent state.

Periodic state:

state j is periodic if there exists an integer Δ ($\Delta \geq 2$) for which $P_{jj}^{(\Delta)} = 1$

Δ-step transition matrix



Definitions:

① $i \rightarrow j$: state j is accesible from state i

(there exists a sequence of transitions connecting state i to state j)

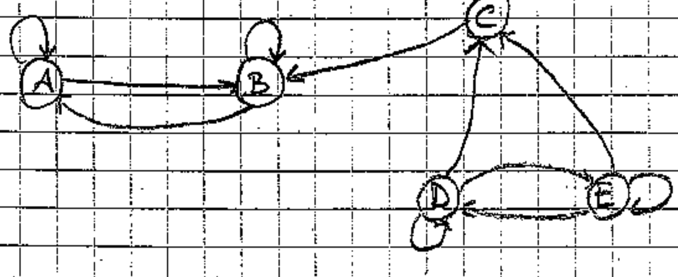
② $i \leftrightarrow j$: states i and j communicate, ie

$i \rightarrow j$ and $j \rightarrow i$

By definition every state communicates with itself

Class: A maximal set of communicating states

Ex:



$A \leftrightarrow B$

$D \leftrightarrow E$

C

Classes = {A, B}, {D, E}, {C}

17/12/2014

Observations: Classes portion the states into sets.

that is

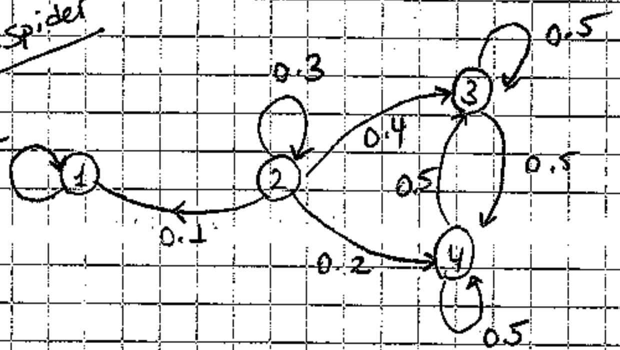
each state is in one and only one class

and there are no states without classes

Fact: states of a class are all recurrent or transient (all states of a class have the same property)

Proof: Theorem 4.2.6 p.105 of textbook

Fly-spider



$P\{\text{Ever reaching } ① \mid X_0 = 2\} = ?$

$= \underbrace{P\{\text{Ever reaching } ① \mid X_0 = 2\}}_{F_{21}} = \sum_{k=1}^{4} P\{\text{ever reaching } ①, X_1 = k \mid X_0 = 2\}$

Prob{ever reaching ① given that $X_0 = 2$}

$= \sum_{k=1}^{4} P\{\text{ever } ① \mid X_1 = k, \cancel{\underbrace{X_0 = 2}_{C}}\} P\{X_1 = k \mid \underbrace{X_0 = 2}_{C}\}$

$P(A,B) = P(A|B)\,P(B)$

$P(A,B|C) = P(A|B,C)\,P(B|C)$

$= \sum_{k=1}^{4} P\{\text{ever } ① \mid X_1 = k\} \cdot \underbrace{P_{2k}}$

1-step transition matrix 2^{nd} row k^{th} column entry.

$= \sum_{k=1}^{4} F_{k1} P_{2k} = \sum_{k=1}^{4} P_{2k} F_{k1}$

$$F_{21} = \underbrace{\begin{bmatrix} \overset{0.1}{P_{21}} & \overset{0.3}{P_{22}} & \overset{0.4}{P_{23}} & \overset{0.2}{P_{24}} \end{bmatrix}}_{2^{nd} \text{ row of } P \text{ matrix}} \begin{bmatrix} F_{11} \\ F_{21} \\ F_{31} \\ F_{41} \end{bmatrix} \begin{array}{l} \leftarrow 1 \\ \leftarrow F_{21} \\ \leftarrow 0 \\ \leftarrow 0 \end{array}$$

$F_{21} = 0.1 + 0.3 F_{21}$

$F_{21} = \frac{0.1}{0.7} = 1/7$

F_{11}, F_{31}, F_{41} in this problem are written very simply from "boundary" conditions.

Notes:

For this example, the boundary conditions are written by simply noting the class of the recurrent states.

$$\text{Classes} = \underbrace{\{2\}}_{\text{transient}}, \underbrace{\{1\}}_{\text{recurrent}}, \underbrace{\{3,4\}}_{\text{recurrent}}$$

1. $F_{ij} = 1$, if states i and j belong to the same recurrent class.

2. $F_{ij} = 0$, if states i and j does not belong to the same recurrant class.

F_{ij} is not trivially written for <u>transient</u> states, we need to some calculations to find

$P\{\text{ever reaching a state} \mid X_0 \text{ is a transient state}\}$

a recurrent state
or
possibly another transient
state

For example, $F_{21} = 1/7$ is the transient to a recurrent class ever reaching probability

22/12/2014

$\underline{F}$ Matrix calculation (Transient $\rightarrow$ Recurrent states)

$$F_{ij}^{(n)} = P\{ X_n = j,\ X_{n-1} \neq j,\ X_{n-2} \neq j,\ --,\ X_1 \neq j \mid X_0 = i \}$$

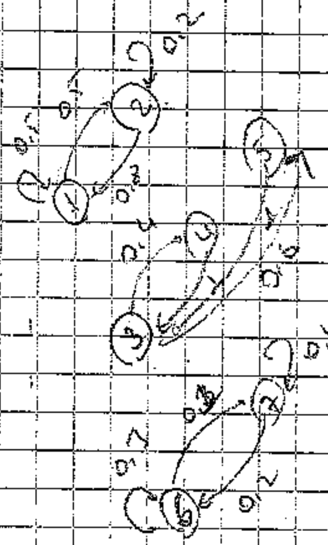
$$= P\{ \text{reaching state } j \text{ first time at } n \mid X_0 = i \}$$

$$F_{ij} = P\{ \text{ever reaching state } j \mid X_0 = i \} = \sum_{n=1}^{\infty} F_{ij}^{(n)}$$

This lecture, we focus on F_{ij} calculation from $i \in$ Transient states to $j \in$ Recurrent states

Ex: (Gialar p.150)

$$\underline{P} = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \left[\begin{array}{cc|ccc|cc} 0.5 & 0.5 & & 0 & & & 0 \\ 0.8 & 0.2 & & & & & \\ \hline & & 0 & 0.4 & 0.6 & & \\ & 0 & 1 & 0 & 0 & & 0 \\ & & 1 & 0 & 0 & & \\ \hline 0.1 & 0 & 0.2 & 0.2 & 0.1 & 0.3 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0 & 0.1 & 0.2 & 0.4 \end{array} \right]$$



$P_{ij} = \text{Prob}\{ X_n = j \mid X_{n-1} = i \}$

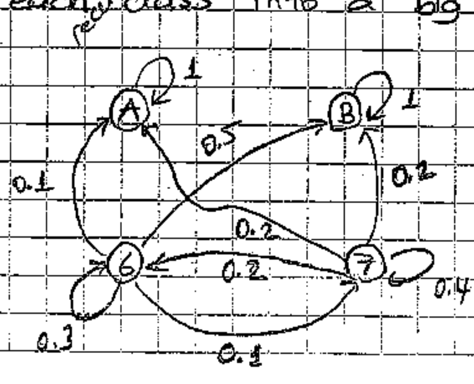
$$\text{Classes} = \{ \underbrace{\{1,2\}}_{\text{recurrent}}, \underbrace{\{3,4,5\}}_{\text{recurrent}}, \underbrace{\{6,7\}}_{\text{transient}} \}$$

Goal: $F_{61} = ?$, $F_{73} = ?$

Approach: lump states of each recurrent class into a "big" state

Ⓐ $= \{1, 2\}$

Ⓑ $= \{3, 4, 5\}$



Write the new state transition matrix $\hat{P}$ by defining transient states as the last states

$$\underline{\underline{\hat{P}}} = \left[\begin{array}{c|c} I & 0 \\ \hline \underline{\underline{B}} & \underline{\underline{Q}} \end{array}\right]$$

$$\underline{\underline{\hat{P}}} = \begin{array}{c} A \\ B \\ 6 \\ 7 \end{array} \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0.1 & 0.5 & 0.3 & 0.1 \\ 0.2 & 0.2 & 0.2 & 0.4 \end{array}\right]$$
$\qquad\qquad\qquad B \qquad\qquad Q$

$$\underbrace{\left(\underline{\underline{\hat{P}}}\right)^n}_{\text{n-step transition matrix}} = \left[\begin{array}{c|c} I & 0 \\ \hline \underline{\underline{B_n}} & \underline{\underline{Q}}^n \end{array}\right] \qquad \underline{\underline{B_n}} = (I + \underline{\underline{Q}} + \underline{\underline{Q}}^2 + --- + \underline{\underline{Q}}^{n-1})\,\underline{\underline{B}}$$

$F_{6A}^{(n)} = P\{\text{reaching } A \text{ for the first time at time } n \mid X_0 = 6\}$

transient states $= \{6, 7\}$

$= P\{X_n = A, X_{n-1} \in T, X_{n-2} \in T, ---, X_1 \in T \mid X_0 = 6\}$

$F_{6A}^{(1)} = 0.1 \leftarrow B_{11}$

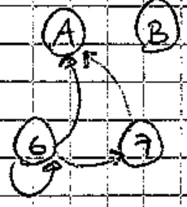
$$F_{6A}^{(2)} = \underbrace{[1\ \ 0]\,\underline{\underline{Q}}}_{[0.3\ \ 0.1]} \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} \qquad = (0.3)(0.1) + (0.1)(0.2) = 0.05$$

B_{11}: $0.1 \leftarrow P_{6A}$

B_{21}: $0.2 \leftarrow P_{7A}$

0.3: $P\{X_1 = 6 \mid X_0 = 6\}$ — transient states

0.1: $P\{X_1 = 7 \mid X_0 = 6\}$ — transient state



$$F_{6A}^{(n)} = \begin{bmatrix} 1 & 0 \end{bmatrix} \underline{\underline{Q}}^{\,n-1} \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0.3 & 0.1 \\ 0.2 & 0.4 \end{bmatrix}^{n-1} \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$$

$$F_{6B}^{(n)} = \begin{bmatrix} 1 & 0 \end{bmatrix} Q^{n-1} \begin{bmatrix} B_{12} \\ B_{22} \end{bmatrix} \begin{array}{l} \leftarrow P_{6B} \\ \leftarrow P_{7B} \end{array}$$

$$F_{7A}^{(n)} = \overbrace{\begin{bmatrix} 0 & 1 \end{bmatrix} Q^{n-1}}^{\begin{array}{l} P\{X_{n-1}=6 \mid X_0=7\} \\ P\{X_{n-1}=7 \mid X_0=7\} \end{array}} \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}$$

$$F_{7B}^{(n)} = \begin{bmatrix} 0 & 1 \end{bmatrix} Q^{n-1} \begin{bmatrix} B_{12} \\ B_{22} \end{bmatrix}$$

then,

$$\underline{\underline{F}}^{(1)} = \begin{bmatrix} F_{6A}^{(n)} & F_{6B}^{(n)} \\ F_{7A}^{(n)} & F_{7B}^{(n)} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \underline{\underline{Q}}^{\,n-1} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

$$= \underline{\underline{Q}}^{\,n-1} \underline{\underline{B}}$$

$$F_{6A} = \sum_{n=1}^{\infty} F_{6A}^{(n)}$$

$$\begin{bmatrix} F_{6A} & F_{6B} \\ F_{7A} & F_{7B} \end{bmatrix} = \underline{\underline{F}}^{(1)} + \underline{\underline{F}}^{(2)} + \dots + \underline{\underline{F}}^{(n)} + \dots$$

$$= \left(\underline{\underline{I}} + \underline{\underline{Q}} + \underline{\underline{Q}}^2 + \underline{\underline{Q}}^3 + \dots + \underline{\underline{Q}}^n + \dots \right) \underline{\underline{B}}$$

24/12/2014

$$\underline{\underline{F}}^{(n)} = \begin{bmatrix} F_{6A}^{(n)} & F_{6B}^{(n)} \\ F_{7A}^{(n)} & F_{7B}^{(n)} \end{bmatrix} = \underline{\underline{Q}}^{n-1} \cdot \underline{\underline{B}}$$

$$F_{6A} = \sum_{n=1}^{\infty} F_{6A}^{(n)}$$

$$\underline{\underline{F}} = \begin{bmatrix} F_{6A} & F_{6B} \\ F_{7A} & F_{7B} \end{bmatrix} = \begin{bmatrix} \sum F_{6A}^{(n)} & \sum F_{6B}^{(n)} \\ \sum F_{7A}^{(n)} & \sum F_{7B}^{(n)} \end{bmatrix} = \sum_{n=1}^{\infty} \underline{\underline{F}}^{(n)} = \sum_{n=1}^{\infty} \left(\underline{\underline{Q}}^{n-1} \cdot \underline{\underline{B}}\right)$$

$$= \left(\sum_{n=1}^{\infty} \underline{\underline{Q}}^{n-1}\right) \cdot \underline{\underline{B}}$$

$$= \left(\sum_{n=0}^{\infty} \underline{\underline{Q}}^{n}\right) \cdot \underline{\underline{B}}$$

$$= \left(\underline{\underline{I}} - \underline{\underline{Q}}\right)^{-1} \cdot \underline{\underline{B}}$$

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}, \quad |r| < 1$$

$$\sum_{n=0}^{\infty} Q^n = S$$

$$I + Q + Q^2 + \dots = S$$

$$I + Q\underbrace{[I + Q + \dots]}_{S} = S \qquad I + QS = S$$

$$I = (\underline{I} - \underline{Q})\underline{S}$$

$$\underline{\underline{S}} = (I - Q)^{-1}$$

Then for example given,

$$\underline{\underline{F}} = (\underline{I} - \underline{Q})^{-1} B = \begin{bmatrix} 1.5 & 0.25 \\ 0.5 & 0.75 \end{bmatrix} \begin{bmatrix} 0.1 & 0.5 \\ 0.2 & 0.2 \end{bmatrix} = \begin{bmatrix} \overset{F_{6A}}{0.2} & \overset{F_{6B}}{0.8} \\ \underset{F_{7A}}{0.4} & \underset{F_{7B}}{0.6} \end{bmatrix}$$

Then

$$F = \begin{array}{c c} A\left\{\begin{array}{c} 1 \\ 2 \end{array}\right. & \\ B\left\{\begin{array}{c} 3 \\ 4 \\ 5 \end{array}\right. & \\ 6 & \\ 7 & \end{array} \left[\begin{array}{cc|ccc|cc} 1 & 1 & & \underline{0} & & & \underline{0} \\ 1 & 1 & & & & & \\ \hline & & 1 & 1 & 1 & & \\ & \underline{0} & 1 & 1 & 1 & & \underline{0} \\ & & 1 & 1 & 1 & & \\ \hline 0.2 & 0.2 & 0.8 & 0.8 & 0.8 & 1/3 & 1/7 \\ 0.4 & 0.4 & 0.6 & 0.6 & 0.6 & 1/3 & 3/7 \end{array}\right]$$

How to find the ever reach prob. from transient state to transient state.

Expected Number of Steps Until Absorption

$$\hat{P} = \begin{array}{c} A \\ B \\ 6 \\ 7 \end{array} \left[\begin{array}{cc|cc} 1 & 0 & & 0 \\ 0 & 1 & & \\ \hline 0.1 & 0.5 & 0.3 & 0.1 \\ 0.2 & 0.2 & 0.2 & 0.4 \end{array}\right]$$

Q: What is the expected number of steps from $t=0$ until absorption into either A or B given that X_0 is a transient state?

A: $$R_6^{abs} = E\{\underbrace{\text{steps until absorption after } t=0}_{\text{Event} = Ev} \mid X_0 = 6\}$$

After first step transition ($X^{(1)}$)

$$= \underbrace{E\{Ev \mid X_1 = A, X_0 = 6\}}_{1.} P\{X_1 = A \mid X_0 = 6\} +$$

$$\underbrace{E\{Ev \mid X_1 = B, X_0 = 6\}}_{1.} P\{X_1 = B \mid X_0 = 6\} +$$

$$\underbrace{E\{Ev \mid X_1 = 6, X_0 = 6\}}_{1 + R_6^{abs}} P\{X_1 = 6 \mid X_0 = 6\} +$$

$$\underbrace{E\{Ev \mid X_1 = 7, X_0 = 6\}}_{1 + R_7^{abs}} \underbrace{P\{X_1 = 7 \mid X_0 = 6\}}_{\text{We know the transition probabilities}}$$

$$R_6^{abs} = 1 + R_6^{abs}(0.3) + R_7^{abs}(0.1)$$

$$R_7^{abs} = 1 + R_6^{abs}(0.2) + R_7^{abs}(0.4)$$

$$\begin{bmatrix} R_6^{abs} \\ R_7^{abs} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \underbrace{\begin{bmatrix} 0.3 & 0.1 \\ 0.2 & 0.4 \end{bmatrix}}_{\underline{\underline{Q}}} \begin{bmatrix} R_6^{abs} \\ R_7^{abs} \end{bmatrix}$$

$$(\underline{\underline{I}} - \underline{\underline{Q}}) \begin{bmatrix} R_6^{abs} \\ R_7^{abs} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} R_6^{abs} \\ R_7^{abs} \end{bmatrix} = (I - Q)^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.75 \\ 2.25 \end{bmatrix} //$$

Steady-State Probabilities of Markov Chains

Assume that we have a Markov chain with M states.

Let

$$\underline{p_0}^T = [\, P\{X_0 = 1\} \;\; P\{X_0 = 2\} \;\; \cdots \;\; P\{X_0 = M\}\,]^T$$

So

$\underline{p_0}^T$: $1 \times M$ vector (row vector) indicating prob. of each state at time $n=0$

then

$$\underline{p_1}^T = \underline{p_0}^T \underset{\uparrow \text{ 1-step prob. transition matrix}}{\underline{\underline{P}}} \;\rightarrow\; \underline{p_1}^T = [\, P\{X_1 = 1\} \;\cdots\; P\{X_1 = M\}\,]^T$$

$$\underline{p_2}^T = \underline{p_1}^T \cdot \underline{\underline{P}} = \underline{p_0}^T \cdot \underline{\underline{P}}^2$$

then

$$\underline{p_N}^T = \underline{p_0}^T \cdot \underset{\uparrow \text{ N-step prob. transition matrix}}{\underline{\underline{P}}^N}$$

Q: If I wait long enough, does $\underline{p_N}^T$ converges to something?

A: Assume it converges,

$$\underline{\Pi}^T = \lim_{N\to\infty} \underline{p_N}^T = \lim_{N\to\infty} \underline{p_0}^T \underline{\underline{P}}^N = \underline{p_0}^T \left(\lim_{N\to\infty} \underline{\underline{P}}^N\right)$$

exists due to the assumption

then $\underline{\Pi}^T \cdot \underline{\underline{P}} = \underline{\Pi}^T$

so, Π is an "invariant" distribution for 1-step transition matrix $\underline{\underline{P}}$

$\underline{\pi}$ is <u>left</u> eigenvector of $\underline{P}$ with eigenvalue of 1.

steady-state dist.

<u>Notes:</u> (1) $\underline{\underline{A}} \cdot \underline{e_k} = \lambda_k \cdot \underline{e_k}$ → $\underline{e_k}$ is a right eigenvector of A with eigenvalue λ_k.

Transpose

$\underline{e_k}^T \underline{\underline{A}}^T = \lambda_k \cdot \underline{e_k}^T$ → then if e_k is a right eigenvector of A then $\underline{e_k}^T$ is a left eigenvector of A^T with the same eigenvalue.

(2) Since $\det(\lambda \underline{\underline{I}} - \underline{\underline{A}}) = \det\left((\lambda \underline{\underline{I}} - \underline{\underline{A}})^T\right) = \det(\lambda I - A^T)$

the eigenvalues of $\underline{\underline{A}}$ and $\underline{\underline{A}}^T$ are the same.

<u>Q1:</u> Do I have always a solution for

$$\underline{\pi}^T \underline{\underline{P}} = \underline{\pi}^T \ ?$$

<u>A1:</u> $\underline{\underline{P}} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = 1. \underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{\underline{1}}$ → then $\underline{1}$ is a right eigenvector with eigenvalue 1.

1-step transition matrix (row sums = 1)

So, there's a left eigenvector with eigenvalue of 1 from the notes (1) and (2)

that means, we always have a solution for

$$\underline{\pi}^T \underline{P} = \underline{\pi}^T$$

Ex:
$$\begin{array}{c} \\ \underline{P} = \end{array}\begin{array}{c} 1 \\ 2 \\ 3 \end{array}\begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.6 & 0 & 0.4 \\ 0 & 0.4 & 0.6 \end{bmatrix}$$

Classes $= \underbrace{\{1,2,3\}}_{\text{A recurrent class}}$, No transient class

what happens as $N \to \infty$

$$\underline{p}_N^T = \underline{p}_0^T \underline{P}^N$$

$$\underline{\pi}^T = \underline{\pi}^T \underline{P} \longrightarrow [\pi_1 \ \pi_2 \ \pi_3] = [\pi_1 \ \pi_2 \ \pi_3]\begin{bmatrix} \\ \underline{P} \\ \\ \end{bmatrix}$$

(1) $\pi_1 = 0.3\pi_1 + 0.6\pi_2$

(2) $\pi_2 = 0.5\pi_1 + 0.4\pi_3$

(3) $\pi_3 = 0.2\pi_1 + 0.4\pi_2 + 0.6\pi_3$

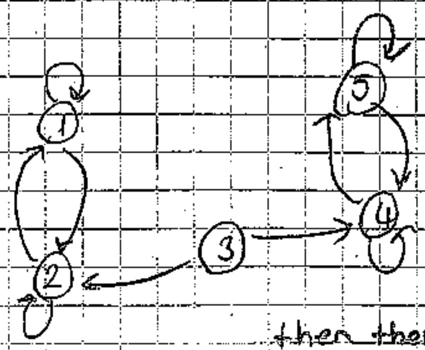
Let's set $\pi_1 = 60 \xrightarrow[(1)]{} \pi_2 = 70 \xrightarrow[(2)]{} \pi_3 = 100$

$$\underline{\pi}^T[\underline{I} - \underline{P}] = \underline{0}^T$$

$$\underline{\pi} \propto \begin{bmatrix} 60 \\ 70 \\ 100 \end{bmatrix} \rightarrow \underline{\pi} = \begin{bmatrix} 60/230 \\ 70/230 \\ 100/230 \end{bmatrix} = \underbrace{\begin{bmatrix} 6/23 \\ 7/23 \\ 10/23 \end{bmatrix}}_{\text{prob. dist. for } X_N \text{ for large } N.}$$

Q2: Is the steady-state prob. dist. unique? (In other words, can there be a multiplicity of eigenvalues of 1?)

A2:



The steady state prob. dist is unique for each recurrent class.

If there are multiple recurrent classes then there are multiple steady-state prob dist. for each class.

Q3: since $\underline{\Pi}$ is independent from initial probability assignment at $n=0$, and

$$\underline{\Pi} = \underline{p_0}^T \underbrace{\lim_{N\to\infty} \underline{\underline{P}}^N}_{\underline{\underline{P}}^\infty},$$

then for any p_0,

$$\underline{\Pi}^T = \underline{p_0}^T \underline{\underline{P}}^\infty$$

and then

$$\underline{\underline{P}}^\infty = \begin{bmatrix} \Pi^T \\ \Pi^T \\ \vdots \\ \Pi^T \end{bmatrix}$$ rows of $\underline{\underline{P}}^\infty$ are all $\underline{\Pi}^T$ vector!

Under what conditions, this result is correct?

A3: this result is correct for a finite-state Markov-chain with a single recurrent class and no transient classes.

$$\underline{\underline{P}} = E \, \Lambda \, E^{-1} \qquad \underline{\underline{E}} = [\underline{e_1} \ \underline{e_2} \ -- \ \underline{e_N}]$$

Λ: diag $(\lambda_1, \lambda_2, --, \lambda_N)$

e_k's are right eigenvectors with eigenvalue λ_k

eigendecomposition of $\underline{\underline{P}}$ matrix

$$\underline{\underline{P}}^N = \underline{\underline{E}} \, \Lambda^N \, \underline{\underline{E}}^{-1}$$

Fact: $$\underline{P} = [\underline{e_1} \ \underline{e_2} \ -- \ \underline{e_m}] \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_m \end{bmatrix} \begin{bmatrix} \Pi_1^T \\ \Pi_2^T \\ \vdots \\ \Pi_m^T \end{bmatrix}$$

$\underline{e_k}$ and $\underline{\Pi_k}^T$ are right and left eigenvectors for the eigenvalue λ_k

(since left and right eigenvectors are related with $\underline{P}$ and $\underline{P}^T$ matrices)

In homework #4, it has been noted that eigenvalues of $\underline{P}$ matrix is less than or equal to 1 in magnitude.

$$|\lambda_k| \leq 1 \quad \forall k.$$

Then let $\underline{e_1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ and Π_1 be the associated left eigenvector corresponding to the steady-state prob. distribution.

If there is only one recurrent class, then there is only one eigenvalue with the value 1. i.e all others are less than 1 in magnitude.

$$\underline{P}^N = [e_1 \; \cdots \; e_M] \begin{bmatrix} \lambda_1^N & & 0 \\ & \ddots & \\ 0 & & \lambda_M^N \end{bmatrix} \begin{bmatrix} \Pi_1^T \\ \Pi_2^T \\ \vdots \\ \Pi_M^T \end{bmatrix}$$

$$= \sum_{k=1}^{M} \lambda_k^N e_k \Pi_k^T = \overset{1}{\lambda_1} e_1 \Pi_1^T + \sum_{k=2}^{M} \lambda_k^N e_k \Pi_k^T$$

$|\lambda_k| < 1 \quad k = \{2, \ldots, M\}$

then as $N \to \infty$

$$\underline{P}^\infty = \underline{e_1} \Pi_1^T + 0 = \underline{1} \cdot \Pi_1^T = \begin{bmatrix} \Pi_1^T \\ \vdots \\ \Pi_1^T \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} [\Pi_1^T]$$