Lecture 7:

Properties of the probability density function (pdf):

1. Since $F_X(x)$ is non-decreasing, $f_X(x) \geq 0$.

2. $$F_X(x) = \int_{-\infty}^{x} f_X(\alpha)\, d\alpha .$$

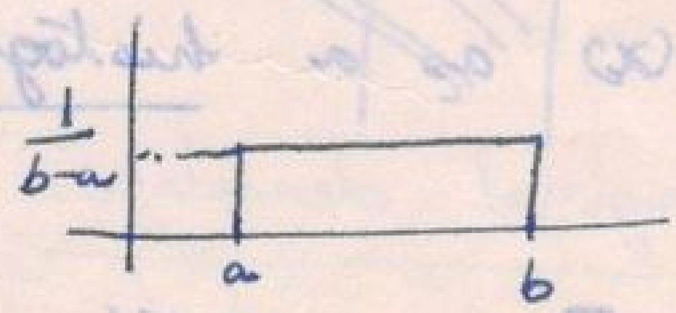
3. $$\int_{-\infty}^{\infty} f_X(x)\, dx = 1 .$$

4. $$F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\, dx .$$

Examples:

① Uniform random variable in $[a, b]$ $a, b \in R$ & $a < b$.

$$f_X(x) = \frac{1}{b-a} \quad \forall \; x \in [a, b]$$



② Gaussian (normal) random variable

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad x \in (-\infty, \infty)$$

(Widely used model for noise description in communication theory)

σ^2 is any positive real number.

m is any real number.

Mixture pdf's :

Consider 2 random variables X_1 and X_2 defined on S with pdf's $f_{X_1}(x)$ and $f_{X_2}(x)$ respectively. Observe that for $\alpha, \beta > 0$ and $\alpha + \beta = 1$, the function

$$f_X(x) = \alpha f_{X_1}(x) + \beta f_{X_2}(x) \text{ is also a}$$

density function.

Physical interpretation and determination of a PDF:

$$f_X(x) \approx \frac{P[X \leq x + dx] - P[X \leq x]}{dx}$$

$$f_X(x)\, dx \approx P[x \leq X \leq x + dx].$$

We can calculate $f_X(x)$ as a histogram.

A random experiment E is repeated n times. Let $N(x)$ denote the number of trials such that

$$x \leq X(s) \leq x + \Delta x .$$

It follows that

$$f_X(x)\, \Delta x \approx \frac{N(x)}{n}.$$

Two random variables:

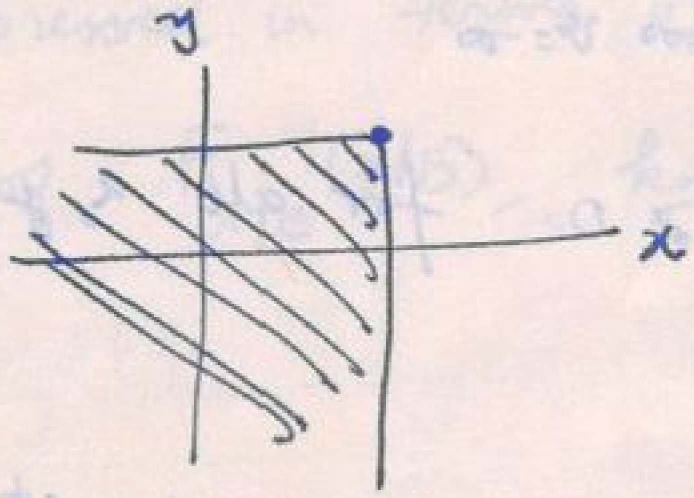
Suppose that we have 2 random variables X and Y, and we wish to determine the probability (region can be arbitrary subject to the condition that it can expressed as a countable union or intersection of rectangles). The distribution functions $F_X(x)$ and $F_Y(y)$ are not sufficient in general. Therefore, we need to define the following joint distribution.

The joint distribution $F_{X,Y}(x,y)$ of two random variables X and Y is the probability of the event $\{X \le x, Y \le y\}$

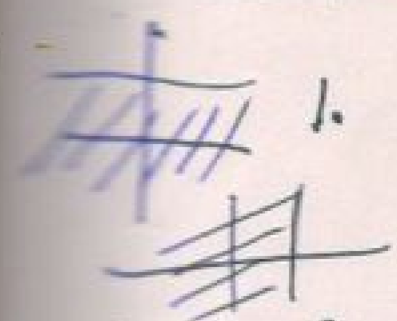
$$F_{X,Y}(x,y) = P[X \le x, Y \le y]$$

comma indicates "AND" here.

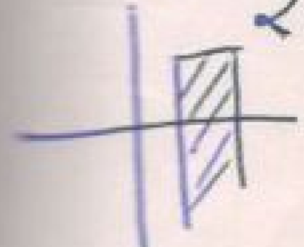
x, y are dummy variables. We could have used a, b instead.



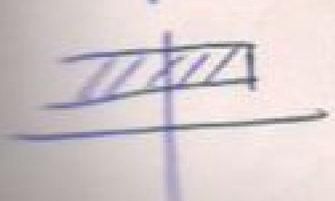
Properties:



1. $F_{XY}(-\infty, y) = 0, \quad F_{XY}(x, -\infty) = 0, \quad F_{XY}(\infty, \infty) = 1.$

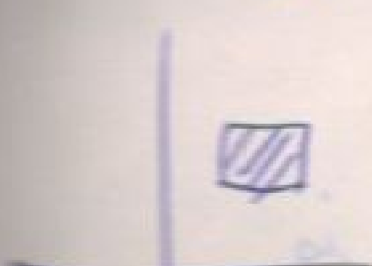


2. $P[x_1 < X \le x_2, Y \le y] = F_{X,Y}(x_2, y) - F_{X,Y}(x_1, y)$



$P[X \le x, y_1 < Y \le y_2] = F_{X,Y}(x, y_2) - F_{X,Y}(x, y_1)$

3. $P[x_1 < X \le x_2, y_1 < Y \le y_2]$



$= F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1)$

4. $0 \leq F_{X,Y}(x,y) \leq 1$ for all x, y

5. Non-decreasing in both x & y.

6. $F_X(x) = F_{X,Y}(x, \infty)$

$F_Y(y) = F_{X,Y}(\infty, y)$

Joint density:

Let $F_{X,Y}(x,y)$ be continuous and differentiable for all x & y. Then, the joint density $f_{X,Y}(x,y) \triangleq \dfrac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$.

$$F_{X,Y}(x,y) = \int_{u=-\infty}^{x} \int_{v=-\infty}^{y} f_{X,Y}(u,v)\, dv\, du$$

$$f_{X,Y}(x,y) \geq 0 \quad \text{for all } x, y.$$

Marginal distributions & marginal densities:

$$F_X(x) = F_{X,Y}(x, \infty)$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy \quad \leftarrow$$

$$F_Y(y) = F_{X,Y}(\infty, y)$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx.$$

Lecture 8:

Conditional distributions and densities:

Probability of an event A given event B is

The conditional distribution $F_{X|B}(x|B)$ of an RV X given B is defined as

$$F_{X|B}(x|B) = P(X \le x | B) = \frac{P[X \le x, B]}{P[B]}.$$

$$f_{X|B}(x|B) = \frac{d}{dx} F_{X|B}(x|B).$$

($F_{X|B}(x|B)$ & $f_{X|B}(x|B)$ satisfy the same properties as $F_X(x)$ & $f_X(x)$).

To find $F_{X|B}(x|B)$, we must, in general, know the underlying experiment. However, if B is an event that can be expressed in terms of X, then, for the determination of $F_{X|B}(x|B)$, knowledge of $F_X(x)$ is sufficient.

Case 1 : $F_X(x | X \le a)$ $\quad B = \{X \le a\}$.

$$F_X(x|X \le a) = P[X \le x | X \le a]$$

$$= \frac{P[X \le x, X \le a]}{P[X \le a]}$$

If $x \geq a$, then

$$F_X(x|X \le a) = \frac{P[X \le a]}{P[X \le a]} = 1.$$

$$F_X(x/X\le a] = \frac{P[X\le x]}{P[X\le a]} = \frac{F_X(x)}{F_X(a)}$$

$$f_X(x/X\le a) = \begin{cases} \dfrac{f_X(x)}{F_X(a)} = \dfrac{f_X(x)}{\int_{-\infty}^{a} f_X(x)dx} & \text{for } x<a \\ 0 & \text{for } x\geqslant a. \end{cases}$$

Case 2. Suppose $B = \{b < X \le a\}$.

$$F_X(x|b<X\le a) = \frac{P[X\le x,\ b<X\le a]}{P[b<X\le a]}$$

If $x \geqslant a$, then

$$F_X(x|b<X\le a) = \frac{P[b<X\le a]}{P[b<X\le a]} = 1.$$

If $b \le x < a$, then

$$F_X(x|b<X\le a) = \frac{P[X\le x,\ b<X\le a]}{P[b<X\le a]}$$

$$= \frac{P[b<X\le x]}{P[b<X\le a]} = \frac{F_X(x)-F_X(b)}{F_X(a)-F_X(b)}$$

If $x < b$, then

$$F_X(x \mid b < X \le a) = \frac{P[X \le x, b < X \le a]}{P[b < X \le a]}$$

$$= \cancel{F_X(\le x)} \quad 0 .$$

Therefore,

$$f_X(x \mid b < X \le a) = \begin{cases} 0 & x < b \\[2ex] \dfrac{f_X(x)}{F_X(a) - F_X(b)} & b \le x < a \\[2ex] 1 & x \ge a . \end{cases}$$

Total probability and Bayes' Theorem:

From (13); If $A_1, A_2, \dots, A_n$ is a partition of S and B is an arbitrary event, then

$$P(B) = P(B/A_1)P(A_1) + \cdots + P(B/A_n)P(A_n)$$

$$P(A/B) = \frac{P(B/A)\,P(A)}{P(B)}$$

1. $$P(A \mid X \le x) = \frac{P(X \le x \mid A)\,P(A)}{P(X \le x)} = \frac{F_X(x \mid A)}{F_X(x)}\,P(A)$$

2. $$P(A \mid x_1 < X \le x_2) = \frac{P(x_1 < X \le x_2 \mid A)}{P(x_1 < X \le x_2)}\,P(A)$$

3. $P(A|X=x)$ cannot be defined as in $P(A/B) = \frac{P(AB)}{P(B)}$

because, if X is a continuous r.v, $P(X=x)=0$.

Therefore, we define the following limit

$$P(A|X=x) = \lim_{\Delta x \to 0} P\left(A / x < X \leq x+\Delta x\right)$$

$$= \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x|A) - F_X(x/A)}{F_X(x+\Delta x) - F_X(x)} P(A)$$

$$= \lim_{\Delta x \to 0} \frac{\dfrac{F_X(x+\Delta x/A) - F_X(x/A)}{\Delta x}}{\dfrac{F_X(x+\Delta x) - F_X(x)}{\Delta x}} P(A)$$

$$\Rightarrow \boxed{P(A|X=x) = \frac{f_X(x/A)}{f_X(x)} P(A)} \quad \text{— I}$$

$$\text{I} \Rightarrow \int_{-\infty}^{\infty} P(A|X=x) f_X(x)\, dx = \int_{-\infty}^{\infty} f_X(x/A)\, P(A)\, dx$$

$$= P(A)\left[\int_{-\infty}^{\infty} f_X(x/A)\, dx\right]$$

$$= P(A)\,[1]$$

$$= P(A)$$

$$\Rightarrow \boxed{P(A) = \int_{-\infty}^{\infty} P(A|X=x) f_X(x)\, dx}$$

Total Probability Thm.

From (14): $\boxed{P(A_i|B) = \dfrac{P(B/A_i)\,P(A_i)}{\sum_{i=1}^{n} P(B/A_i)\,P(A_i)}}$

Similarly, here we have

$$\frac{P(A/X=x)}{P(A)}\, f_X(x) = f_X(x/A)$$

$$\Rightarrow \quad f_X(x/A) = \frac{f_X(x)\cdot\left[P(A/X=x)\right]}{\int_{-\infty}^{\infty} P(A|X=x)\, f_X(x)\,dx}.$$

Bayes' Theorem: $\boxed{f_X(x/A) = \dfrac{P(A|X=x)\, f_X(x)}{\int_{-\infty}^{\infty} P(A|X=x)\, f_X(x)\,dx}}$

From page (31): $F_X(x/B) = P(X \le x/B) = \dfrac{P[X\le x, B]}{P(B)}$

Case 1: $B = \{Y=y\}$ where Y is a ~~contin~~ discrete r.v.

$$F_{X/Y}(x/Y=y) = \frac{P(X\le x,\ Y=y)}{P(Y=y)}.\ (\text{assuming } P(Y=y)\neq 0).$$

$$\Rightarrow \quad P[X\le x,\ Y=y] = F_{X/Y}(x/Y=y)\ P[Y=y]$$

$$\sum_y P[X\le x,\ Y=y] = \sum_y F_{X/Y}(x/Y=y)\ P[Y=y]$$

$$\Rightarrow \quad \boxed{P[X\le x] = \sum_y F_{X/Y}(x/Y=y)\ P[Y=y] = F_X(x)}$$

$$\Rightarrow \boxed{f_X(x) = \sum_y f_{X/Y}(x/Y=y)\, P[Y=y].}$$

Case 2: $B = \{Y = y\}$ where Y is a continuous r.v.

$$F_{X/Y}(x/\, y<Y\le y+\Delta y) = \frac{P[X\le x,\ y<Y\le y+\Delta y]}{P[y<Y\le y+\Delta y]}$$

$$= \frac{F_{X,Y}(x, y+\Delta y) - F_{X,Y}(x,y)}{F_Y(y+\Delta y) - F_Y(y)}$$

$$= \frac{\dfrac{F_{X,Y}(x,y+\Delta y) - F_{X,Y}(x,y)}{\Delta y}}{\dfrac{F_Y(y+\Delta y) - F_Y(y)}{\Delta y}}$$

As $\Delta y \to 0$, we get

$$F_{X/Y}(x/\,Y=y) = \frac{\frac{\partial}{\partial y} F_{X,Y}(x,y)}{f_Y(y)} = \int_{-\infty}^{x} f_{X,Y}(a,y)\,da$$

$$\left[\text{Since } F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(a,b)\,db\,da\right]$$

$$\boxed{f_{X/Y}(x/\,Y=y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}} = \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X/Y}(x/Y=y)\, f_Y(y)\,dy$$

Slice of $f_{X,Y}(x,y)$ at $Y=y$

Lecture 9:

Independence

Let A be any event generated by the random variable X only, and B be any event generated by the random variable Y only. The two random variables are independent if and only if

$$P(AB) = P(A)P(B)$$

for all such events A, B.

(Equivalent statement): The random variables X and Y are independent if and only if

$$F_{X,Y}(x,y) = F_X(x)\,F_Y(y) \quad \text{for all } (x,y) \in \mathbb{R}^2.$$

<u>If</u>: Show $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ if X, Y are independent.

$$A = \{X \le x\}, \quad B = \{Y \le y\}$$

$$P(AB) = P(X \le x, Y \le y) = F_{X,Y}(x,y)$$

$$P(A)P(B) = F_X(x)F_Y(y)$$

If X, Y are independent, A and B are independent events $\Rightarrow F_{X,Y}(x,y) = F_X(x)\,F_Y(y)$.

<u>Only if</u>: Show that if $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, then X and Y are independent.

Rigorous proof is very difficult.

We will show this for a specific type of sets A, B.

$$P[a < X \le b,\ c < Y \le d] = F_{X,Y}(b,d) - F_{X,Y}(b,c) - F_{X,Y}(a,d) + F_{X,Y}(a,c)$$

$$= F_X(b)F_Y(d) - F_X(b)F_Y(c) - F_X(a)F_Y(d) + F_X(a)F_Y(c)$$

$$= (F_X(b) - F_X(a))(F_Y(d) - F_Y(c))$$

Differentiate $F_{X,Y}(x,y) = F_X(x) F_Y(y)$

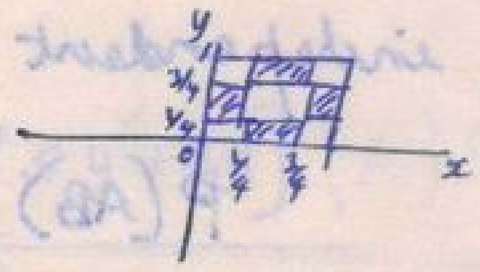
The random variables X and Y are independent if and only if $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ for all $(x,y) \in \mathbb{R}^2$

Also, $f_{X/Y}(x/y) = f_X(x)$

$f_{Y/X}(y/x) = f_Y(y)$.

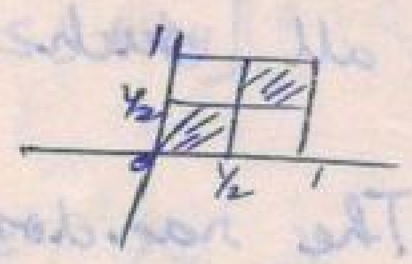
Examples : $\longrightarrow$ More examples: (Uniform marginals but different joint PDF)

(2a)



Joint PDF uniform in shaded region
$X \sim$ uniform$[0,1]$
$Y \sim$ uniform$[0,1]$
Dependent.

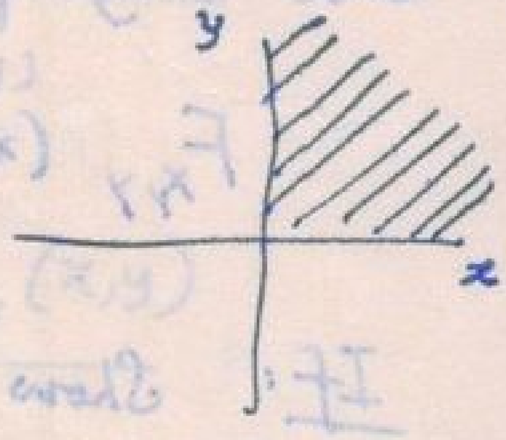
(2b)



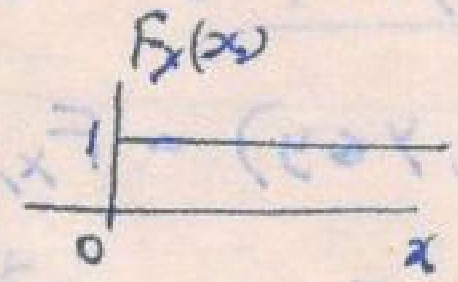
Joint PDF uniform in shaded region
X, Y dependent
$X \sim$ uniform $[0,1]$
$Y \sim$ uniform $[0,1]$.

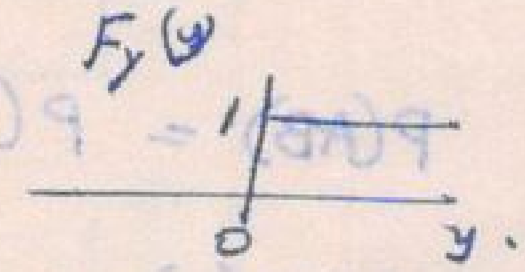
①. Two independent random variables.

$$F_{X,Y}(x,y) = \begin{cases} 1 & x, y \geq 0 \\ 0 & \text{else.} \end{cases}$$



$P[X=0, Y=0] = 1 \qquad P[X=0] = 1. \qquad P[Y=0] = 1.$





(2c) Two dependent random variables. (can be skipped)

$$f_{X,Y}(x,y) = \begin{cases} g(x,y) > 0 & 3 \leq x^2 + y^2 \leq 4 \\ 0 & \text{else} \end{cases}$$

Exact form of $g(\,)$ does not matter.

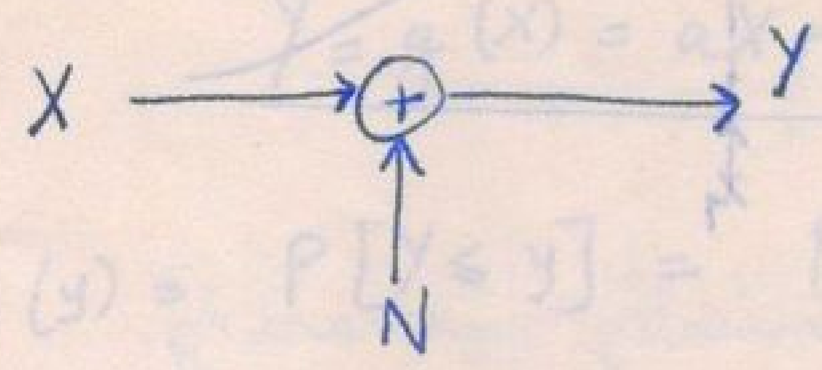
$$f_X(0) = \int_{-\infty}^{\infty} f_{X,Y}(0,y)\,dy > 0$$

$$f_Y(0) = \int_{-\infty}^{\infty} f_{X,Y}(x,0)\,dx > 0$$

{ Need to find only one pair (x,y) such that $f_{X,Y}(x,y) \neq f_X(x) f_Y(y)$ }

$$f_{X,Y}(0,0) = 0 \neq f_X(0) f_Y(0)$$

Example : (Do this slowly!)



$$Y = X + N$$

X is a discrete r.v $\quad X = \begin{cases} x_0 & P(x_0) \\ x_1 & P(x_1) \end{cases}$

N & X are independent.

Need to find a decision rule that maps received Y to x_0 or x_1. Need to maximize P(correct decision).

$$P(\text{correct decision}) = \int_{-\infty}^{\infty} P(\text{correct decision} \,/\, Y = y)\, f_Y(y)\, dy$$

For every received value y, choose $\hat{x}_i$ to maximize $P(\text{correct decision} / Y = y)$.

Choose $\hat{x}_i = x_0$ if

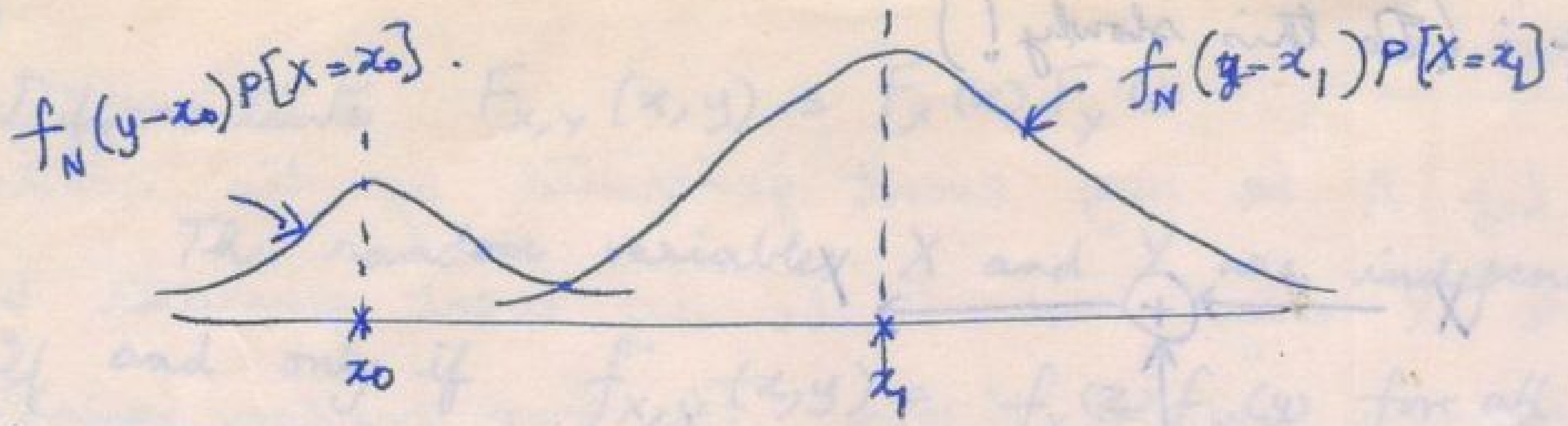
$$P(x_0 / Y = y) > P(x_1 / Y = y)$$

[Using (I) from page (34)]
$$\frac{f_Y(y/X = x_0)\, P[X = x_0]}{f_Y(y)} > \frac{f_Y(y/X = x_1)\, P[X = x_1]}{f_Y(y)}$$

$$f_Y(y/X = x_0)\, P[X = x_0] > f_Y(y/X = x_1)\, P[X = x_1]$$

$$f_Y(y/X = x_0) = f_N(y - x_0 / X = x_0) = f_N(y - x_0).$$

$$f_Y(y/X = x_1) = f_N(y - x_1)$$



Lecture 10:

Functions (Transformations) of one random variable:

Define a new random variable Y as a function of r.v X

$$Y = g(X).$$

For a given $s \in S$, $X(s)$ is the number mapped to by the r.v. X, $g(X(s)) = Y(s)$ is the number mapped to by the r.v. Y. (memoryless transformation).

The c.d.f. of Y, $F_Y(y)$, is

$$F_Y(y) = P[Y \leq y] = P[g(X) \leq y]$$

For a specific y, let R_y denote the set of values of x such that $g(x) \leq y$.

$$F_Y(y) = P[X \in R_y].$$

For $g(X)$ to be a r.v, the function $g(x)$ must have the following properties:

(i) Its domain must include the range of X.

(ii) For every y, the set R_y, such that $g(x) \leq y$, must consist of the union & intersection of a countable number of intervals. This will ensure that $Y \leq y$ is an event.

(iii) The events $g(X) = \pm\infty$ must have zero probability.

1. Linear transformation: (Assume that X is a continuous r.v.)

$$Y = g(X) = aX + b.$$

$$F_Y(y) = P[Y \le y] = P[aX + b \le y]$$

$$= P[aX \le y - b].$$

If $a > 0$, then

$$F_Y(y) = P\left[X \le \frac{y-b}{a}\right] = F_X\left(\frac{y-b}{a}\right)$$

If $a < 0$, then

$$F_Y(y) = P\left[X \ge \frac{y-b}{a}\right] = 1 - F_X\left[\frac{y-b}{a}\right]$$

$$f_Y(y) = \begin{cases} \frac{1}{a} f_X\left(\frac{y-b}{a}\right) & \text{if } a > 0 \\ \frac{-1}{a} f_X\left(\frac{y-b}{a}\right) & \text{if } a < 0. \end{cases}$$

2. $Y = g(X) = X^2$

$$F_Y(y) = P[Y \le y] = P[X^2 \le y]$$

for $y < 0$, $F_Y(y) = P[X^2 \le y] = 0$.

for $y \ge 0$, $F_Y(y) = P[-\sqrt{y} \le X \le \sqrt{y}]$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) & y \ge 0 \end{cases}$$

3. $g(x)$ is an invertible function.

$$F_Y(y) = P[g(X) \le y]$$

If $g()$ is monotonically increasing

$$F_Y(y) = P[X \le g^{-1}(y)] = F_X(g^{-1}(y))$$

If $g()$ is monotonically decreasing

$$F_Y(y) = P[X \ge g^{-1}(y)] = 1 - F_X(g^{-1}(y))$$

$$f_Y(y) = \begin{cases} \dfrac{d}{dy} F_X(g^{-1}(y)) = \left[\dfrac{d}{dy}(g^{-1}(y))\right] f_X(g^{-1}(y)) & \text{if } g() \text{ is monotonically increasing} \\[2ex] -\dfrac{d}{dy} F_X(g^{-1}(y)) = -f_X(g^{-1}(y)) \dfrac{d}{dy}(g^{-1}(y)) & \text{if } g() \text{ is monotonically decreasing} \end{cases}$$

(Relate dx and dy using slope)

$$f_Y(y) = \left|\frac{dh}{dy}\right| f_X(g^{-1}(y)). \qquad (h = g^{-1})$$

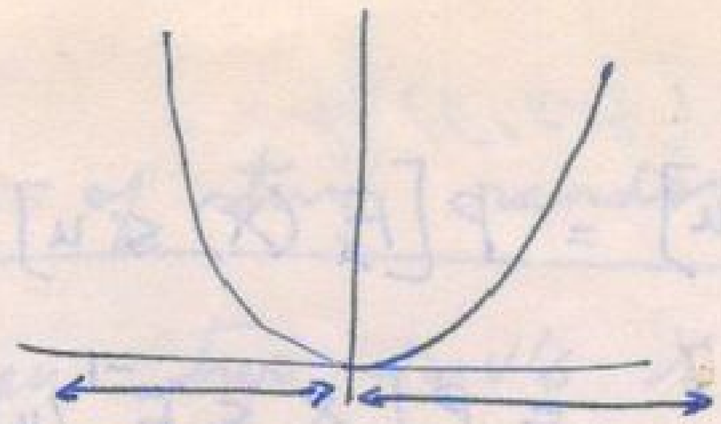
4. $g(x)$ is such that the x-axis can be divided into regions where $g(x)$ is invertible.

Let $g(x) = g_i(x)$ in the i^{th} such interval.

Let $h_i(y)$ denote the inverse of $g_i(x)$.

$$f_Y(y) = \sum_{i=1}^{n} \left|\frac{dh_i(y)}{dy}\right| f_X(h_i(y))$$

Ex: $Y = X^2$

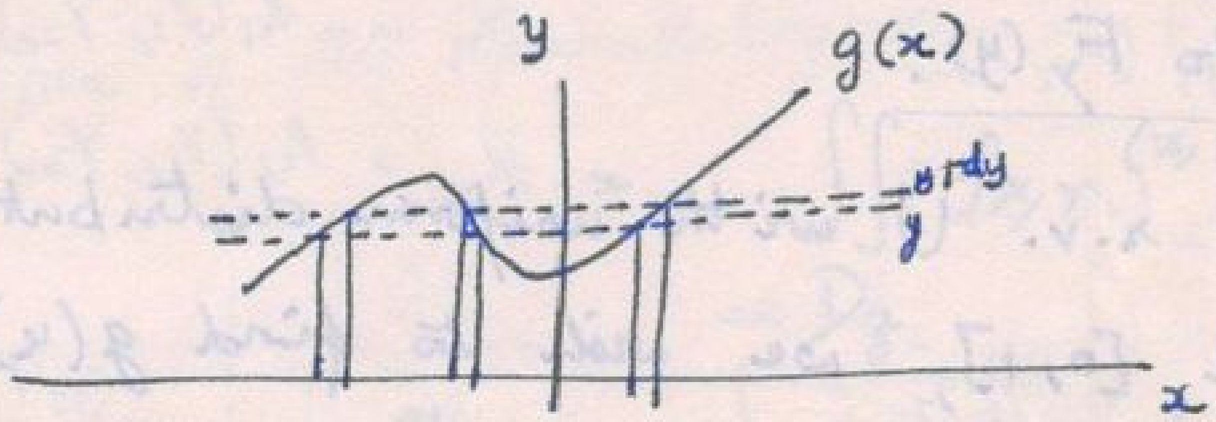


$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) \quad \text{for } y>0.$$

5. Another approach to write $f_Y(y)$ in terms of $f_X(x)$ (Papoulis).

Solve the equation $y=g(x)$. Let the real roots be denoted by x_n.

$$y = g(x_1) = g(x_2) = \cdots = g(x_n) = \cdots$$



$$\frac{dy}{|g'(x_1)|} = |dx_1|$$

$$f_Y(y)\,dy = \frac{f_X(x_1)\,dy}{|g'(x_1)|} + \cdots + \frac{f_X(x_n)\,dy}{|g'(x_n)|} + \cdots$$

$$f_Y(y) = \frac{f_X(x_1)}{|g'(x_1)|} + \cdots + \frac{f_X(x_n)}{|g'(x_n)|} + \cdots$$

6. <u>From $F_X(x)$ to uniform distribution:</u>

Given a r.v X with cdf $F_X(x)$, we wish to find a function $g(x)$ such that $U=g(X)$ is uniformly distributed in the interval $[0,1]$.

<u>Solution</u>: $g(x) = F_X(x)$.

i.e. if $U = F_X(X)$ then $F_U(u) = u$ for $0<u\le 1$.

Proof:

$$F_U(u) = P[U \le u] = P[F_X(X) \le u]$$

$$= P[X \le F_X^{-1}(u)]$$

$$= P[X \le F_X^{-1}(F_X(x))]$$

$$= P[X \le x]$$

$$= F_X(x)$$

$$= u.$$

$$X \longrightarrow \boxed{F_X(\,)} \longrightarrow U.$$

(Discuss about random number generation programs)

7. From uniform to $F_Y(y)$.

Given a r.v. U with uniform distribution in the interval $[0,1]$, we wish to find $g(u)$ such that r.v. $Y = g(U)$ has a cdf $F_Y(y)$.

Solution: $Y = F_Y^{-1}(U)$.

Proof: $P[Y \le y] = P[F_Y^{-1}(U) \le y]$

$$= P[U \le F_Y(y)]$$

$$= F_Y(y).$$

8. Under a transformation, a continuous pdf can become a discrete pdf or partly discrete pdf.

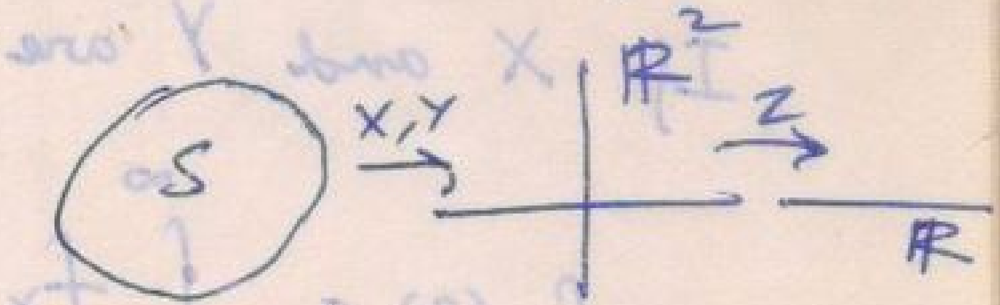
Eg: $Y = \mathrm{sgn}(X)$.

Lecture 11:

One function of two random variables:

Given two r.v's X and Y and a function $g(x,y)$, we form the r.v

$$Z = g(X, Y).$$

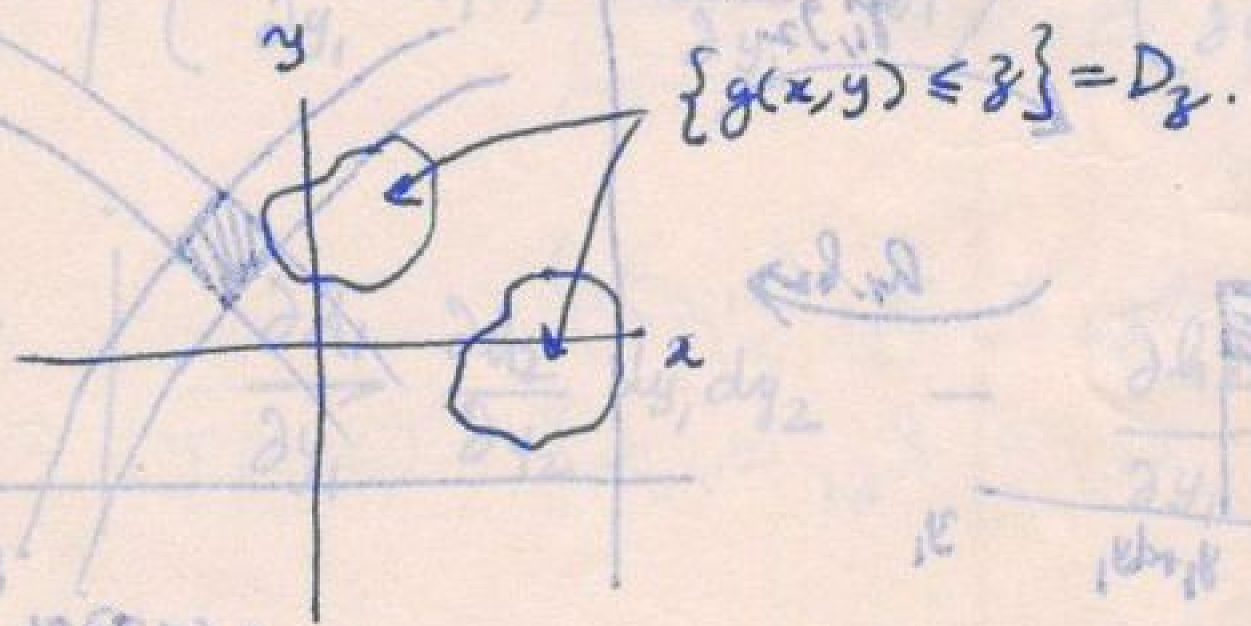


For a given z, we denote by D_z the region on the xy-plane such that $g(x,y) \le z$.

$$F_Z(z) = P[Z \le z] = P[g(X,Y) \le z]$$

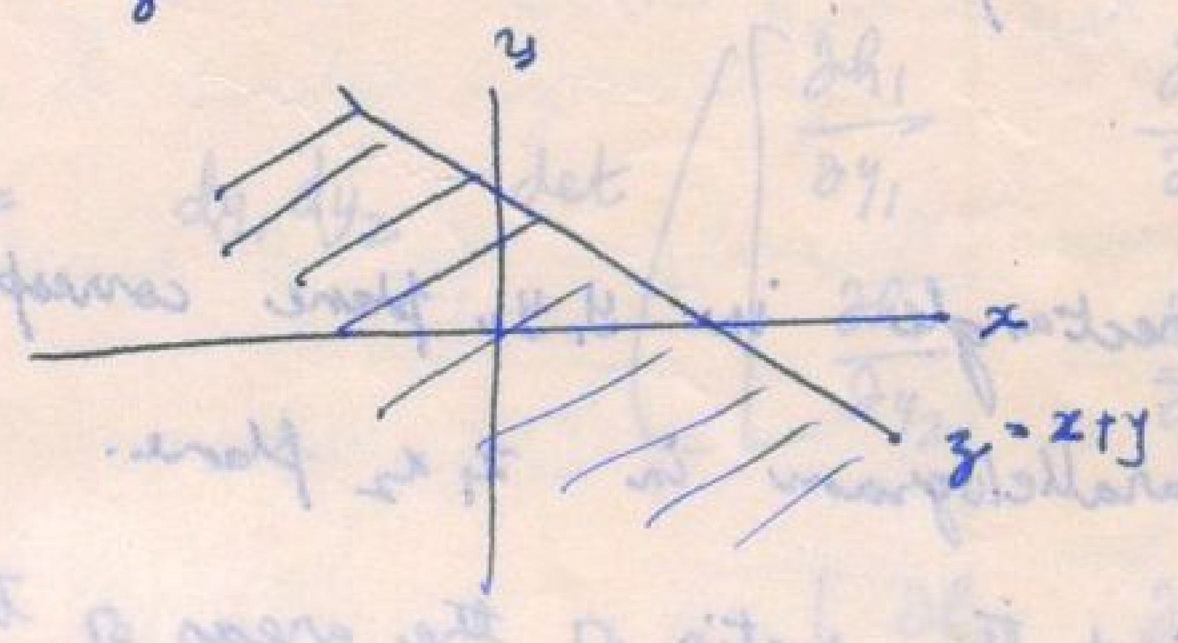
$$= P[(X,Y) \in D_z]$$

$$= \iint_{D_z} f_{X,Y}(x,y)\,dx\,dy$$



Eg: $Z = X + Y$

$$D_z = \{(x,y) : x + y \le z\}$$



$$F_Z(z) = \iint_{D_z} f_{X,Y}(x,y)\,dx\,dy$$

If X and Y are independent r.v's, then

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(z-y, y)\, dy.$$

If X and Y are independent r.v's, then

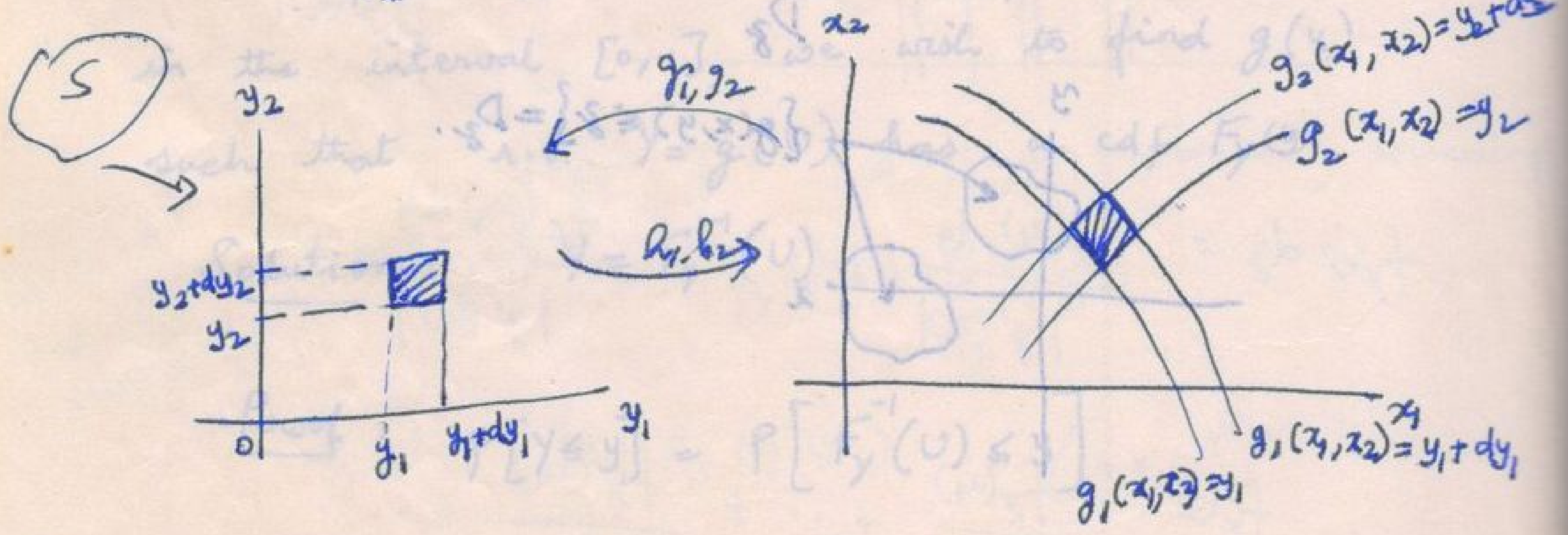
$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy. \quad \left(\begin{array}{c}\text{Example in} \\ \text{assignment}\end{array}\right)$$

$\Rightarrow$ The PDF of Z is the convolution of the PDF's of X and Y.

Two functions of two random variables:

$Y_1 = g_1(X_1, X_2)$ $\qquad X_1 = h_1(Y_1, Y_2)$

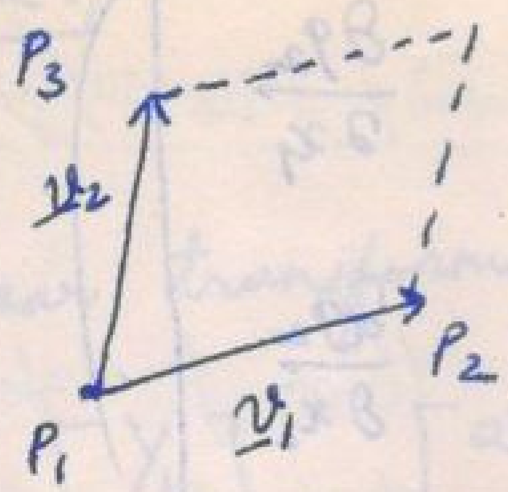
$Y_2 = g_2(X_1, X_2)$ $\qquad X_2 = h_2(Y_1, Y_2)$.



Assume that the functions are one-to-one and differentiable.

Infinitesimal rectangle in $y_1 y_2$ plane corresponds to infinitesimal parallelogram in $x_1 x_2$ plane.

We need to find the ratio of the areas of these two regions.



$P_1 : \quad (x_1, x_2)$

$P_2 : \quad \left(x_1 + \frac{\partial h_1}{\partial y_1} dy_1 , \quad x_2 + \frac{\partial h_2}{\partial y_1} dy_1 \right)$

$P_3 : \quad \left(x_1 + \frac{\partial h_1}{\partial y_2} dy_2 , \quad x_2 + \frac{\partial h_2}{\partial y_2} dy_2 \right)$

Area of the parallelogram $= |\underline{v}_1 \times \underline{v}_2|$

$$= \left| \left(\frac{\partial h_1}{\partial y_1} dy_1 , \frac{\partial h_2}{\partial y_1} dy_1 \right) \times \left(\frac{\partial h_1}{\partial y_2} dy_2 , \frac{\partial h_2}{\partial y_2} dy_2 \right) \right|$$

$$= \left| \frac{\partial h_1}{\partial y_1} \frac{\partial h_2}{\partial y_2} dy_1 dy_2 - \frac{\partial h_2}{\partial y_1} \frac{\partial h_1}{\partial y_2} dy_1 dy_2 \right|$$

$$= dy_1 dy_2 \left| \frac{\partial h_1}{\partial y_1} \frac{\partial h_2}{\partial y_2} - \frac{\partial h_2}{\partial y_1} \frac{\partial h_1}{\partial y_2} \right|$$

$$= dy_1 dy_2 \det \left(\begin{bmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_2}{\partial y_1} \\ \frac{\partial h_1}{\partial y_2} & \frac{\partial h_2}{\partial y_2} \end{bmatrix} \right)$$

Ratio of the areas $= \det \begin{bmatrix} \frac{\partial h_1}{\partial y_2} & \frac{\partial h_2}{\partial y_1} \\ \frac{\partial h_1}{\partial y_2} & \frac{\partial h_2}{\partial y_2} \end{bmatrix} = J_h$

$$\frac{1}{\text{Ratio}} = \det\left(\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_2}{\partial x_1} \\ \frac{\partial g_1}{\partial x_2} & \frac{\partial g_2}{\partial x_2} \end{bmatrix}\right)$$

$$\Rightarrow f_{X_1,X_2}(x_1, x_2) = \frac{f_{Y_1,Y_2}(y_1,y_2)}{|J_h(y_1,y_2)|} = |J_g| f_{Y_1,Y_2}(y_1,y_2)$$

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{f_{X_1,X_2}(x_1,x_2)}{|J_g(x_1,x_2)|}.$$

Suppose there is more than one solution to the equations

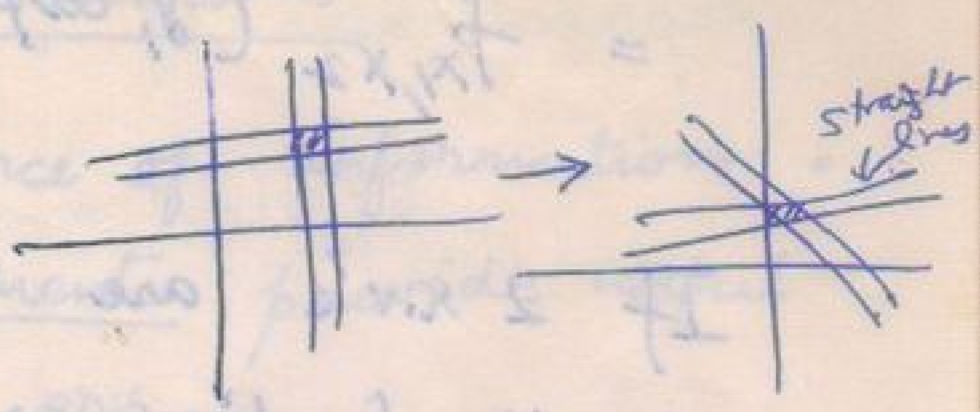
$$x_{1,1} = h_{1,1}(y_1, y_2)$$
$$x_{2,1} = h_{2,1}(y_1, y_2)$$
$$\vdots$$
$$x_{1,N} = h_{1,N}(y_1, y_2)$$
$$x_{2,N} = h_{2,N}(y_1, y_2)$$

$$f_{Y_1,Y_2}(y_1,y_2) = \sum_{i=1}^{n} \frac{f_{X_1,X_2}(x_{1_i}, x_{2_i})}{|J_g(x_{1_i}, x_{2_i})|}.$$

Lecture 12:

Examples:

① Linear transformation:



$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

$$J(y_1, y_2) = \det \begin{bmatrix} A & C \\ B & D \end{bmatrix} = AD - BC.$$

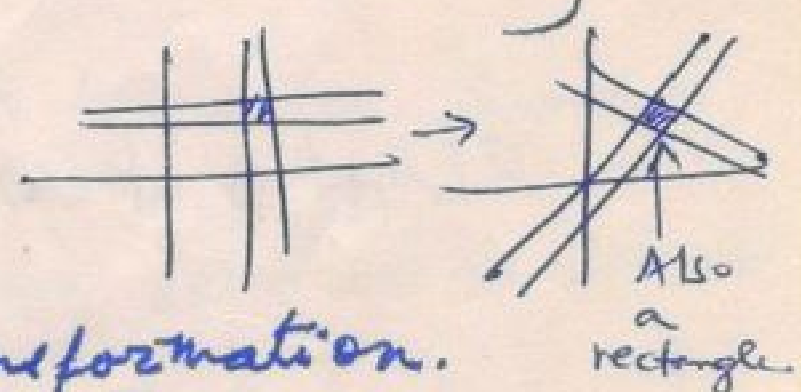
$$f_{Y_1,Y_2}(y_1, y_2) = |AD - BC|\, f_{X_1,X_2}(Ay_1 + By_2,\ Cy_1 + Dy_2).$$

$$= \frac{1}{|ad - bc|} f_{X_1,X_2}(Ay_1 + By_2,\ Cy_1 + Dy_2).$$

$$\left\{ J(y_1, y_2) = \frac{1}{J(x_1, x_2)}, \quad J(x_1, x_2) = ad - bc. \right.$$

$$A = \frac{d}{ad-bc},\ B = \frac{-c}{ad-bc},\ C = \frac{-b}{ad-bc},\ D = \frac{a}{ad-bc} \}$$

② Rotation & circular symmetry:



Special case of a linear transformation.

$a = \cos\phi$

$b = \sin\phi$

$c = -\sin\phi$

$d = \cos\phi$

$$\begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}$$

$f\ (\ldots) = f\ (Ay_1 + By_2,\ Cy_1 + Dy_2)$

$$= f_{X_1,X_2}\left(y_1\cos\phi - y_2\sin\phi,\ y_1\sin\phi + y_2\cos\phi\right)$$

If 2 r.v's are rotated by ϕ, their probability density functions are rotated by ϕ in the opposite direction.

If X_1, X_2 and Y_1, Y_2 have the same statistics for all rotations ϕ, then their joint density is circularly symmetric.

(3) Rectangular to Polar coordinates:

$$R = \sqrt{X^2+Y^2}$$

$$\Theta = \tan^{-1}\left(\frac{Y}{X}\right).$$

$$X = R\cos\Theta, \quad Y = R\sin\Theta.$$

$$J(r,\theta) = \det\begin{bmatrix} \cos\theta & -R\sin\theta \\ \sin\theta & R\cos\theta \end{bmatrix}$$

$$= R\cos^2\theta + R\sin^2\theta$$

$$= R.$$

$$f_{R,\Theta}(r,\theta) = r\, f_{X,Y}(r\cos\theta, r\sin\theta)$$

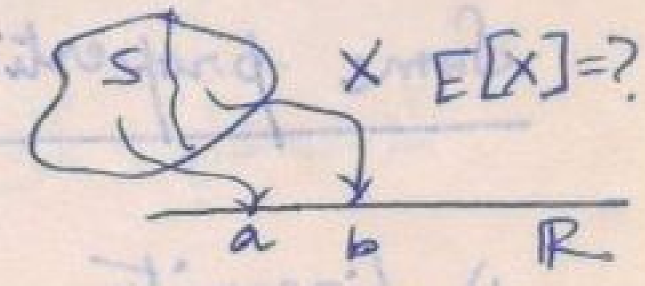
(4) X, Y

$U = \max(X,Y)$, $V = \min(X,Y)$?

(in the assignment)

Partial information about random variables:

PDF is the complete source of information about a random variable. Moments provide useful partial information about the PDF.



Expectation:

The expected value of a random variable X is defined as

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx \quad \text{(for continuous r.v's)}$$

Also called mean.

(for discrete r.v's) $f_X(x) = \sum_i p_i \,\delta(x - x_i)$ where X takes on values x_i & $P[X = x_i] = p_i$:

$$E[X] = \int_{-\infty}^{\infty} x \sum_i p_i\, \delta(x - x_i)\,dx = \sum_i p_i x_i$$

Variance:

$$E\left[(X - E[X])^2\right] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)\,dx = \sigma^2$$

where σ is called standard deviation.

N^{th} moment:

$$m_n = E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\,dx$$

n^{th} central moment:

$$\mu_n = E\left[(X - E[X])^n\right] = \int_{-\infty}^{\infty} (x - E[X])^n f_X(x)\,dx$$

n^{th} absolute moment:

$$E[|X|^n],\ E\left[|X - E[X]|^n\right]$$

Generalized moments: $E[(X-a)^n]$, $E[|X-a|^n]$

Expected value of a function of a random variable:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx .$$

Some properties of the expectation operator:

1) Linearity

$$E[a_1 g_1(X) + \cdots + a_n g_n(X)] = a_1 E[g_1(X)] + \cdots + a_n E[g_n(X)]$$

$$E[aX+b] = aE[X] + b .$$

2) If $X > 0$, $E[X] > 0$.

3) $E[c] = c$ where c is any constant.

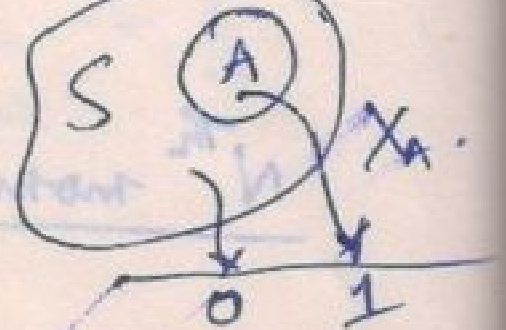
4) If $Y = X - E[X]$, $E[Y] = E[X - E[X]] = E[X] - E[E[X]]$
$= E[X] - E[X] = 0 .$

5) $E[X^2] = E[(X - E[X])^2] + (E[X])^2$.

$\boxed{\text{Lecture 13:}}$

Probability of any event A as the expected value of a random variable (Indicator fn)

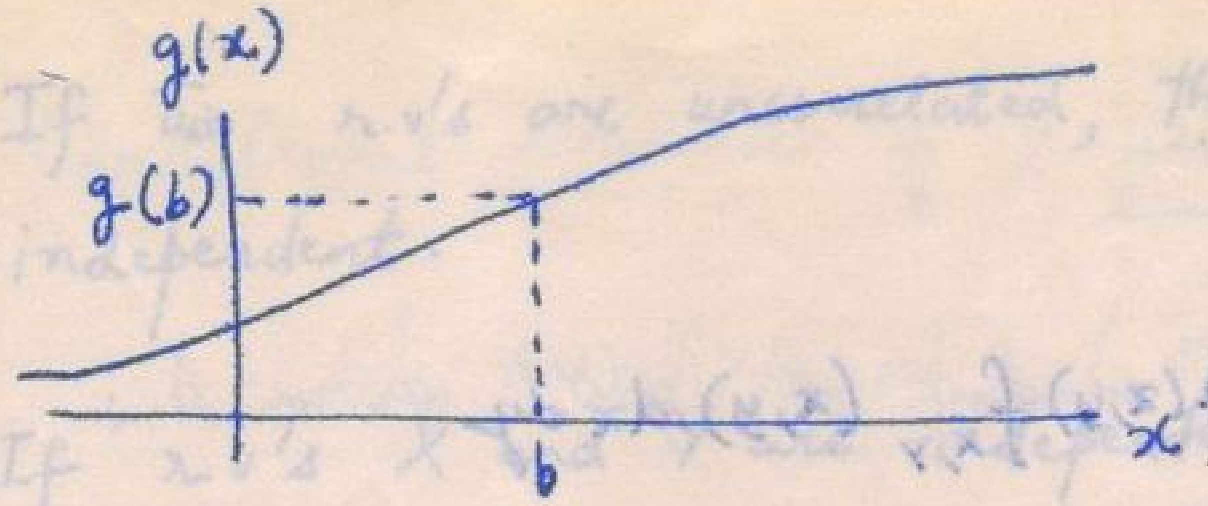
Define X_A such that $X_A(s) = \begin{cases} 1 & s \in A \\ 0 & s \notin A \end{cases}$

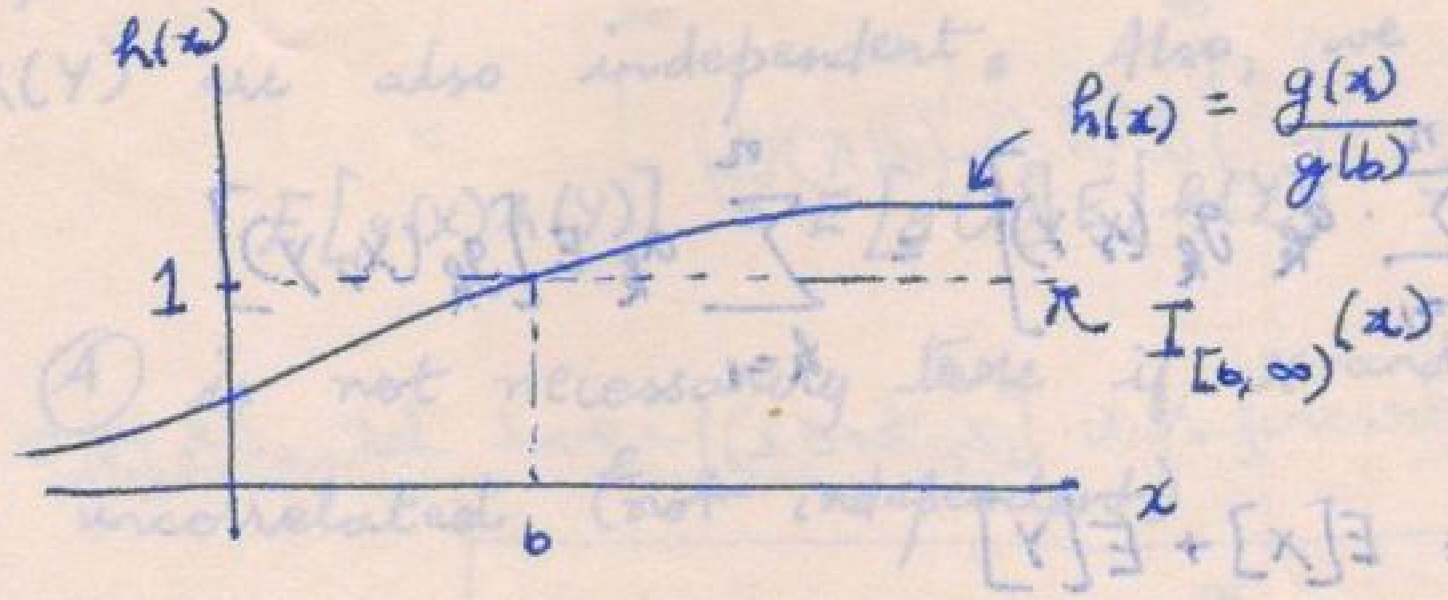


$$E[X_A] = P(A)\cdot 1 + P(\bar{A})\cdot 0 = P(A).$$

Inequalities:

Suppose $g()$ is any non-negative non-decreasing function and suppose $b \in \mathbb{R}$ with $g(b) > 0$.





$$h(x) \geq I_{[b,\infty)}(x)$$

$$E\left[h(X)\right] \geq E\left[I_{[b,\infty)}(X)\right]$$

$$E\left[h(X)\right] \geq P\left[X \geq b\right]$$

$$\Rightarrow \boxed{\frac{E\left[g(X)\right]}{g(b)} \geq P\left[X \geq b\right]}$$

(i) Let $X > 0$ and $g(x) = x$:

$$P\left[X \geq b\right] \leq \frac{E\left[X\right]}{b}$$ Markov inequality

(ii) Let $Y = X^2$.

$$P\left[X^2 \geq b\right] \leq \frac{E\left[X^2\right]}{b}$$ (As long as we choose $b > 0$) (also, $g()$ is not non-decreasing)

$$P\left[|X| \geq a\right] \leq \frac{E\left[X^2\right]}{a^2}$$ where $a^2 = b$.

$$P\left[|X - m| \geq \delta\right] \leq \frac{E\left[|X - m|^2\right]}{\delta^2}$$ Chebyshev inequality.

(iii) (i) $\Rightarrow P\left[|X - a|^n \geq \epsilon^n\right] \leq \frac{E\left[|X - a|^n\right]}{}$

Joint moments:

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y)\, dx\, dy$$

Linearity: $E\left[\sum_{k=1}^{n} a_k\, g_k(X,Y)\right] = \sum_{k=1}^{n} a_k\, E[g_k(X,Y)]$

e.g.: $E[X+Y] = E[X] + E[Y]$

Correlation of two r.v's X & Y.

$$R_{XY} = E[XY] \neq E[X]E[Y] \text{ in general.}$$

Covariance of two r.v's X & Y

$$C_{XY} = E\left[(X-E[X])(Y-E[Y])\right] = E[XY] - E[X]E[Y].$$

Uncorrelatedness: $C_{XY} = 0 \Rightarrow E[XY] = E[X]E[Y]$.

Orthogonality: $E[XY] = 0$.

Theorem: If two r.v's X, Y are independent, then they are uncorrelated.

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{X,Y}(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_X(x)\, f_Y(y)\, dx\, dy \quad \{\text{since } X, Y \text{ are independent}\}$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x f_X(x)\, dx\right] y f_Y(y)\, dy$$

$$= E[X]E[Y] \Rightarrow X \;\&\; Y \text{ are uncorrelated}$$

* If two r.v's are uncorrelated, they are not necessarily independent.

* If r.v's X and Y are independent, then r.v's $g(X)$ and $h(Y)$ are also independent. Also, we have

$$E[g(X)h(Y)] = E[g(X)]\,E[h(Y)]. \text{ — } (A)$$

(A) is not necessarily true if X and Y are merely uncorrelated (not independent)

Cauchy-Schwarz Inequality:

$$(E[XY])^2 \leq E[X^2]\,E[Y^2]$$

Proof: $\quad E[(aX-Y)^2] \geq 0 \quad$ (since $(aX-Y)^2 \geq 0$).

$$a^2 E[X^2] - 2a E[XY] + E[Y^2] \geq 0.$$

This is a quadratic in 'a' and is non-negative. So, its discriminant has to be ≤ 0 (since it can have only complex roots or equal roots). Therefore,

$$4(E[XY])^2 - 4E[X^2]E[Y^2] \leq 0$$

$$\Rightarrow (E[XY])^2 \leq E[X^2]\,E[Y^2].$$

Correlation matrix for two r.v's X & Y: ← (Do this when random vectors are discussed)

$$R = E\left[\begin{bmatrix} X \\ Y \end{bmatrix}[X \;\; Y]\right] = \begin{bmatrix} E[X^2] & E[XY] \\ E[XY] & E[Y^2] \end{bmatrix}$$

Similarly, covariance matrix $C = \begin{bmatrix} E[(X-E[X])^2] & E[(X-E[X])(Y-E[Y])] \\ E[(X-E[X])(Y-E[Y])] & E[(Y-E[Y])^2] \end{bmatrix}$

Conditional Expectation:

The conditional mean of $g(Y)$ given A:

$$E[g(Y)/A] = \int_{-\infty}^{\infty} g(y)\, f_Y(y/A)\, dy$$

Choose A to be $\{X \in [x, x+\Delta x]\}$ and let $\Delta x \to 0$.

$$\boxed{\begin{aligned} E[g(Y)/X=x] &= \int_{-\infty}^{\infty} g(y)\, f_Y(y/X=x)\, dy \\ E[Y/X=x] &= \int_{-\infty}^{\infty} y\, f_Y(y/X=x)\, dy \end{aligned}}$$

The conditional mean $E[Y/X=x]$ is a function of x, say $\varphi(x)$.

Now, consider the r.v $\varphi(X)$.

$$E[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x)\, f_X(x)\, dx$$

$$= \int_{-\infty}^{\infty} f_X(x) \int_{-\infty}^{\infty} y\, f_Y(y/X=x)\, dy\, dx$$

Since $f_{X,Y}(x,y) = f_X(x)\, f_Y(y/X=x)$, we have

$$E\left[E[Y/X=x]\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{X,Y}(x,y)\, dy\, dx = E[Y].$$

i.e. $\boxed{E[Y] = E\left[E[Y/X=x]\right]}$.

Conditional variance of Y given X is

$$E\left[\left(Y - E[Y/X=x]\right)^2 / X=x\right]$$

$$= \int_{-\infty}^{\infty} \left(y - E[Y/X=x]\right)^2 f_Y(y/X=x)\,dx.$$

Application area: Mean-square estimation.

$$E\left[g(X,Y)/A\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y/A)\,dx\,dy$$

$$E\left[g(X,Y)/X=x\right] = \int_{-\infty}^{\infty} g(x,y)\, f_Y(y/X=x)\,dy$$

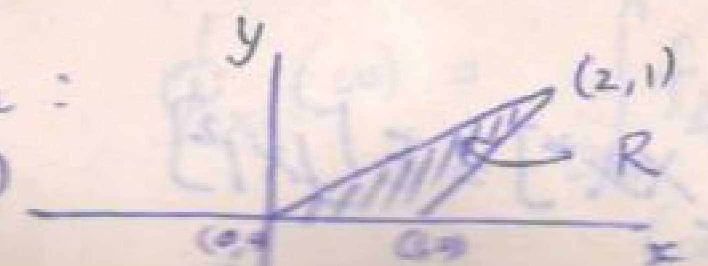
Also note that

$$E\left[g(x,Y)/X=x\right] = \int_{-\infty}^{\infty} g(x,y)\, f_Y(y/X=x)\,dy$$

i.e. $\quad E\left[g(X,Y)/X=x\right] = E\left[g(x,Y)/X=x\right]$

The functions $g(X,Y)$ and $g(x,Y)$ have the same expected value given $X=x$, but they are not equal. $g(X,Y)$ is a function of r.v's X & Y and for a specific outcome s, it takes the value $g(X(s),Y(s))$. $g(x,Y)$ is a function of the real variable x and r.v Y and for a specific s, it takes the value $g(x,Y(s))$.

Example: [1 lecture almost]



$$f_{X,Y}(x,y) = \begin{cases} c & (x,y)\in R. \\ 0 & \text{else.} \end{cases}$$

Complex random variables: (* cover later)

* A complex random variable X is defined as a random variable of the form

$$X = X_r + j X_i \qquad (j = \sqrt{-1}).$$

where X_r and X_i are real random variables.

* Expectation of a random variable is naturally generalized to the complex case as

$$E[X] = E[X_r] + j E[X_i].$$

* The statistical properties of $X = X_r + jX_i$ are determined by the joint pdf $f_{X_r X_i}(x_r, x_i)$ of X_r and X_i.

* $E[g(X)] = E[Re\{g(X)\}] + j E[Im\{g(X)\}]$

* Covariance matrix & Pseudo-covariance matrix.

For two complex random vectors $\underline{X}$ and $\underline{Y}$

covariance matrix $= E[(\underline{X} - E[\underline{X}])(\underline{Y} - E[\underline{Y}])^H]$

& pseudo-covariance matrix $= E[(\underline{X} - E[\underline{X}])(\underline{Y} - E[\underline{Y}])^T]$

* For X, we can define $E[XX^*] = E[|X|^2]$
(zero-mean X)
& $E[XX] = E[X^2]$

$$E\left[|X|^2\right] = E\left[X_r^2 + X_i^2\right] = E\left[X_r^2\right] + E\left[X_i^2\right]$$

$$E\left[X^2\right] = E\left[(X_r + jX_i)(X_r + jX_i)\right]$$

$$= E\left[X_r^2 - X_i^2 + 2j X_i X_r\right]$$

$$= E\left[X_r^2\right] - E\left[X_i^2\right] + 2j\, E\left[X_r X_i\right].$$

* A complex random variable X is proper if

(for zero mean X) $E\left[X^2\right] = 0$, i.e., $E\left[X_r^2\right] = E\left[X_i^2\right]$ and $E\left[X_r X_i\right] = 0$.

(for general X) $E\left[(X - E[X])^2\right] = 0$, i.e., $E\left[(X_r - E[X_r])^2\right] = E\left[(X_i - E[X_i])^2\right]$, $E\left[(X_r - E[X_r])(X_i - E[X_i])\right] = 0$

* A complex random vector $\underline{X}$ is proper if

$$E\left[\underline{X}\,\underline{X}^T\right] = \underline{0}. \quad (\text{for zero mean } \underline{X})$$

$$E\left[(\underline{X} - E[\underline{X}])(\underline{X} - E[\underline{X}])^T\right] = \underline{0}.$$

* Two complex random vectors $\underline{X}$ and $\underline{Y}$ are uncorrelated

(zero mean case) if and only if $E\left[\underline{X}\,\underline{Y}^H\right] = \underline{0}$ & $E\left[\underline{X}\,\underline{Y}^T\right] = \underline{0}$.

(covariance) (pseudo-covariance).

* Cauchy-Schwarz inequality.

$$\left|E\left[XY^*\right]\right|^2 \leq E\left[|X|^2\right] E\left[|Y|^2\right].$$

$\boxed{\text{Lecture 15:}}$

Characteristic function:

The characteristic function of a r.v. X is by definition

$$\phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x)\, e^{j\omega x}\, dx = E\left[e^{j\omega X}\right]$$

(continuous X)

or

$\phi_X(\omega)$ is the Fourier transform of $f_X(x)$.

Since $f_X(x) \geq 0$, $|\phi_X(\omega)| \leq \phi_X(0) = 1$.

$$\phi_X(s) = \int_{-\infty}^{\infty} f_X(x) e^{sx} dx = E\left[e^{sX}\right]$$

$\phi_X(s)$ is called the moment generating function.

used to calculate
- calculate moments
- Chernoff bound
- PDF of the sum of independent r.v's.

Moment theorem:

$$\phi_X^{(n)}(s) = E\left[X^n e^{sX}\right]$$

$$\Rightarrow \phi_X^{(n)}(s)\Big|_{s=0} = E\left[X^n\right].$$

Expand $\phi_X(s)$ into a series near the origin,

$$\phi_X(s) = \sum_{n=0}^{\infty} \frac{E[X^n]}{n!} s^n .$$

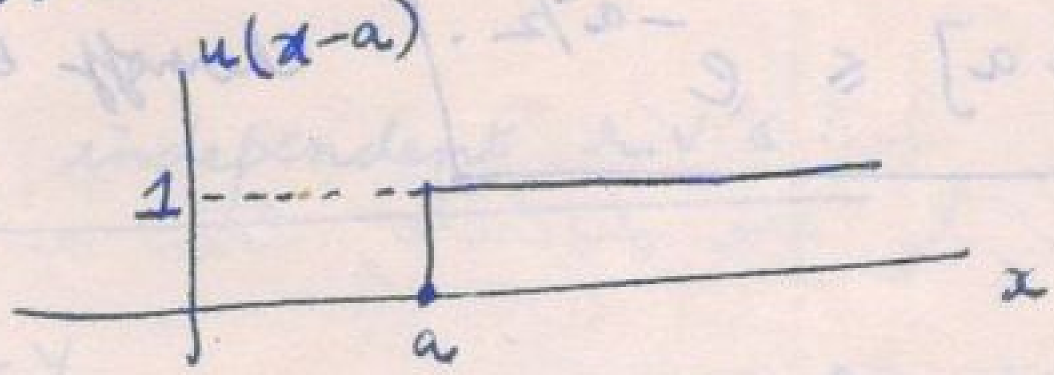
This series expansion is valid only if all the moments are finite and the series converges absolutely near $s=0$.

Since $f_X(x)$ can be determined from $\phi_X(s)$, under the conditions stated above, the pdf of a r.v is uniquely determined if all its moments are known.

Chernoff bound:

* Upper bound on the tail probability $P[X \geq a]$.

* First note that $u(x-a) \leq e^{s(x-a)}$ for any $s>0$.



$$e^{s(x-a)} = \begin{cases} 1 & x = a \\ >1 & x > a \\ <1 & x < a. \end{cases}$$

$$P[X \geq a] = \int_a^\infty f_X(x)\,dx = \int_{-\infty}^{\infty} f_X(x)\, u(x-a)\,dx$$

$$\leq \int_{-\infty}^{\infty} f_X(x)\, e^{s(x-a)}\,dx \quad \text{for any } s>0.$$

$$\Rightarrow \quad P[X \geq a] \leq e^{-as}\,\phi_X(s) \quad \text{for any } s>0.$$

The bound can be minimized w.r.t. s.

$$P[X \geq a] \leq \min_{s>0} e^{-as}\,\phi_X(s).$$

Example: Consider a Gaussian r.v. X with $E[X]=0$ and $Var(X) = 1$. $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

$s^2/2$

$$P[X \geq a] \leq e^{-as} e^{\frac{s^2}{2}} = e^{-s\left(a - \frac{s}{2}\right)}$$

Differentiating w.r.t. s, we see that the minimum is achieved when $s = a$. Thus

$$\boxed{P[X \geq a] \leq e^{-a^2/2}.} \quad \text{Chernoff bound.}$$

Joint characteristic functions:

The joint characteristic function of the pair of r.v's (X, Y) is defined as

$$\phi_{X,Y}(u, v) = E\left[e^{juX + jvY}\right]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{jux + jvy} f_{X,Y}(x, y)\, dx\, dy$$

The n^{th} order moment of X and Y is the expectation of the form

$$\mu_{k,l} = E\left[X^k Y^l\right] \text{ where}$$

$$k + l = n \text{ and } k, l \in [0, 1, 2 \ldots, n]$$

Similarly, n^{th} central moment is

$$m_{k,l} = E\left[(X - E[X])^k (Y - E[Y])^l\right]$$

$$\text{where } k + l = n \text{ \& } k, l = [0, 1, \ldots, n]$$

$$\phi_{X,Y}(s_1, s_2) = E\left[e^{s_1 X + s_2 Y}\right]$$

$$E\left[X^k Y^l\right] = \left.\frac{\partial^n}{\partial s_1^k \partial s_2^l} \phi_{X,Y}(s_1, s_2)\right|_{\substack{s_1 = 0 \\ s_2 = 0}}$$

Sum of two independent r.v's:

$$Z = X + Y$$

$$\phi_Z(\omega) = E\left[e^{j\omega Z}\right] = E\left[e^{j\omega(X+Y)}\right]$$

$$= E\left[e^{j\omega X} e^{j\omega Y}\right]$$

If X & Y are independent,

$$\phi_Z(\omega) = E\left[e^{j\omega X}\right] E\left[e^{j\omega Y}\right]$$

$$= \phi_X(\omega)\, \phi_Y(\omega)$$

$$\Rightarrow f_Z(z) = f_X(z) \circledast f_Y(z)$$

(Convolution)

$$= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$$

$\boxed{\text{Lecture 16:}}$

Random vectors:

Random vectors can be formed by considering a number n of random variables related to a random experiment. Borel sets in n-dimensions are used to define the probability measure.

$\underline{X} = [X_1 \; X_2 \; \ldots \; X_n]^T$

Joint CDF:

$$F_{\underline{X}}(\underline{x}) = P[X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n].$$

If $X_1, X_2, \ldots, X_n$ are discrete, then a joint pmf is defined as

$$P_{\underline{X}}(\underline{x}) = P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n].$$

Joint PDF :

If the random vector $\underline{X}$ is jointly continuous, we can define a n-dimensional joint density such that

$$F_{\underline{X}}(\underline{x}) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} f_{\underline{X}}(\underline{x})\, dx_n\, dx_{n-1} \cdots dx_1 .$$

Marginal PDF: $f_{X_j}(x_j) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{\underline{X}}(\underline{x})\, dx_1 \quad dx_{j-1}\, dx_{j+1} \cdots dx_n.$

Independence:

The elements of the random vector $\underline{X}$ are independent if

events $\{s \in S : X_1(s) \in B_1\}, \ldots \{s \in S : X_n(s) \in B_n\}$

are independent events for all $B_1, B_2, \ldots, B_n \in \mathcal{B}$

← Borel field.

$$F_{\underline{X}}(\underline{x}) = F_{X_1}(x_1)\, F_{X_2}(x_2) \cdots F_{X_n}(x_n).$$

Expectation :

$$E[g(\underline{X})] = \int_{R^n} g(\underline{x}) f_{\underline{X}}(\underline{x})\, d\underline{x} . \qquad E[\underline{X}] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix};$$

Conditional distributions & densities : Chain rule.

$$f_{\underline{X}}(\underline{x}) = f_{X_1}(x_1)\, f_{X_2}(x_2 / X_1 = x_1) \cdots f_{X_n}(x_n / X_1 = x_1, \ \ X_{n-1} = x_{n-1})$$

$$\underset{n\times 1}{\underline{X}} = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix} \begin{matrix} \leftarrow k\times 1 \\ \leftarrow (n-k)\times 1 \end{matrix} \qquad f_{\underline{X}_1 | \underline{X}_2 = \underline{x}_2}(\underline{x}_1) = \frac{f_{\underline{X}}\left(\begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \end{bmatrix}\right)}{f_{\underline{X}_2}(\underline{x}_2)} .$$

Characteristic function & Moment generating function :

$$\phi_{\underline{X}}(\underline{s}) = E\left[e^{\underline{s}^T \underline{X}} \right]$$

Correlation matrix :

$$E[\underline{X}\,\underline{X}^T].$$

This is a non-negative definite matrix, i.e.,

$$\underline{a}^T R\, \underline{a} \geq 0 \text{ for any } \underline{a} \neq \underline{0} .$$

For a complex random vector

$$E[\underline{X}\,\underline{X}^H].$$

Covariance matrix :

$$E\left[(\underline{X} - E[\underline{X}])(\underline{X} - E[\underline{X}])^H \right].$$

Proper complex random vectors : (* cover later)

A complex random vector is proper if

$$E\left[(\underline{X} - E(\underline{X}))(\underline{X} - E(\underline{X}))^T \right] = 0$$

$$\Rightarrow E\left[\ldots \right.$$

$$E\left[(\underline{X}_r - E[\underline{X}_r])(\underline{X}_r - E[\underline{X}_r])^T\right]$$

$$= E\left[(\underline{X}_i - E[\underline{X}_i])(\underline{X}_i - E[\underline{X}_i])^T\right].$$

$$E\left[(\underline{X}_r - E[\underline{X}_r])(\underline{X}_i - E[\underline{X}_i])^T\right]$$

$$= -E\left[(\underline{X}_i - E[\underline{X}_i])(\underline{X}_r - E[\underline{X}_r])^T\right].$$

* Two complex random vectors are uncorrelated only if covariance & pseudo-covariance are both vanishing. (0)

Conditional PDF : $f_{X_j}(x_j \mid X_1 = a_1, \ldots X_{j-1} = a_{j-1}, X_{j+1} = a_{j+1}, \cdots X_n = a_n)$

$$= \frac{f_{\underline{X}}(a_1, a_2, \ldots, x_j, \cdots, a_n)}{\int_{-\infty}^{\infty} f_{\underline{X}}(a_1, a_2, \ldots, x_j, \cdots, a_n)\, dx_j}$$

Gaussian random variables :

A Gaussian random variable X has the pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Start with → PDF → mean m

→ CDF?

↓ $\phi_X(s)$

↓ moments for $N(0, \sigma^2)$

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

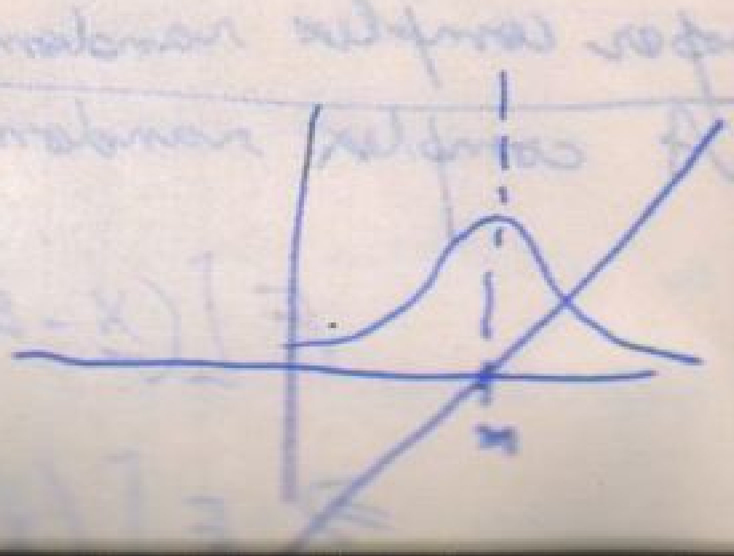
$$= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

$$t = (x-m)^2 / 2\sigma^2$$

$$dt = \frac{2(x-m)\,dx}{2\sigma^2}$$

$$= \int_{-\infty}^{\infty} \frac{(x-m)}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx + m$$

$$= \int_{-\infty}^{\infty} \cancel{\frac{\sigma^2 t}{\sqrt{2\pi}\,\sigma} e^{-t} dt} + m$$



CDF: $F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$

$y = \frac{x-m}{\sigma}$

$dy = \frac{dx}{\sigma}$

$$F_X(x) = \int_{-\infty}^{\frac{x-m}{\sigma}} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-y^2/2} \sigma\, dy$$

$$= \int_{-\infty}^{\frac{x-m}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

$$F_X(x) = 1 - Q\left(\frac{x-m}{\sigma}\right)$$

where $Q(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\alpha^2/2} d\alpha$.

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\alpha^2} d\alpha$$

$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x).$$

$$Q(x) = \frac{1}{2} \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right).$$

$\phi_X(\omega)$ & $\phi_X(s)$:

$$\phi_X(\omega) = E\left[e^{j\omega X}\right]$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} e^{j\omega x} dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{\left(j\omega x - \frac{(x-m)^2}{2\sigma^2}\right)} dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\left(\frac{x^2-2mx+m^2-2j\omega x\sigma^2}{2\sigma^2}\right)}\, dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-(m+j\omega\sigma^2))^2}{2\sigma^2}}\, e^{-\frac{\omega^2\sigma^2}{2}+j\omega m}\, dx$$

Using

$$x^2-2mx+m^2-2j\omega x\sigma^2 = x^2-2(m+j\omega\sigma^2)x+m^2$$

$$= x^2-2(m+j\omega\sigma^2)x+(m+j\omega\sigma^2)^2 - (m+j\omega\sigma^2)^2+m^2$$

$$= (x-(m+j\omega\sigma^2))^2 + \cancel{m^2} - \cancel{m^2} + \cancel{j}\,\omega^2\sigma^4 - 2j\omega m\sigma^2$$

$$= e^{j\omega m-\frac{\omega^2\sigma^2}{2}}$$

$$\phi_X(s) = e^{ms+\frac{s^2\sigma^2}{2}}$$

$$\frac{d}{ds}\phi_X(s) = (m+s\sigma^2)\, e^{ms+\frac{s^2\sigma^2}{2}}$$

$$E[X] = m.$$

$$\frac{d^2}{ds^2}\phi_X(s) = (m+s\sigma^2)^2\, e^{ms+\frac{s^2\sigma^2}{2}} + e^{ms+\frac{s^2\sigma^2}{2}}(\sigma^2)$$

$$E[X^2] = m^2+\sigma^2$$

Suppose $m = 0$.

$$\phi_X(s) = e^{\frac{s^2\sigma^2}{2}}$$

$$= 1 + \frac{s^2\sigma^2}{2} + \frac{1}{2!}\left(\frac{s^2\sigma^2}{2}\right)^2 + \cdots$$

n^{th} derivative when n is odd (at $s = 0$) $= 0$.

n^{th} derivative when n is even (at $s=0$) $= \sigma^{2m}(1.3.\cdots(2m-1))$
$(n = 2m)$

$\boxed{\text{Lecture 17:}}$

N-dimensional Gaussian random vector:

$\underline{X}$ is a Gaussian random vector whose pdf is defined as follows:

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2}|C|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x}-\underline{m})^T C^{-1}(\underline{x}-\underline{m})\right\}$$

with parameters $\underline{m}$ and C (a $n \times n$ symmetric non-negative definite matrix).

$$N(\underline{m}, C).$$

$$\phi_{\underline{X}}(\underline{s}) = \exp\left\{\underline{s}^T\underline{m} + \frac{1}{2}\underline{s}^T C \underline{s}\right\}.$$

$$\int_{\underline{x}} f_{\underline{X}}(\underline{x})\,d\underline{x} = \phi_{\underline{X}}(\underline{0}) = 1.$$

Properties:

① $\underline{m} = E[\underline{X}]$

② $C = E\left[(\underline{X} - E[\underline{X}])(\underline{X} - E[\underline{X}])^T\right]$

③ Marginal density of any l-dimensional sub-vector $\underline{Y}$ $(l < n)$ is also Gaussian.

④ Consider $\underline{Y} = A\underline{X} + \underline{b}_{l \times 1}$ where $\underline{X} \sim N(\underline{m}, C)$.

$$\underline{Y} \sim N\left(A\underline{m} + \underline{b},\ ACA^T\right).$$

(5) The conditional pdf of any subvector of $\underline{X}$ given any other subvector is also jointly Gaussian.

(6) If C is a diagonal matrix, the r.v's in $\underline{X}$ are uncorrelated. For jointly Gaussian $\underline{X}$, this means that X_i are independent.

Proof of Properties (1) to (6):

(1) $E[\underline{X}] = \underline{m}$.

$$\underline{s}^T\underline{m} = \sum_{i=1}^{n} s_i m_i$$

$$\underline{s}^T C\,\underline{s} = \sum_{i=1}^{n} s_i\left(\sum_{j=1}^{n} C_{ij} s_j\right)$$

$$\frac{\partial}{\partial s_k}\left[\phi_{\underline{X}}(\underline{s})\right] = \phi_{\underline{X}}(\underline{s})\left[\frac{1}{2}\left(\sum_{\substack{j=1\\ j\neq k}}^{n} C_{kj} s_j\right) + \frac{1}{2}\left(\sum_{\substack{i=1\\ i\neq k}}^{n} C_{ik} s_i\right) + C_{kk} s_k + m_k\right]$$

$$= \phi_{\underline{X}}(\underline{s})\left[C_{kk} s_k + \frac{1}{2}\sum_{\substack{j=1\\ j\neq k}}^{n} (C_{kj} + C_{jk}) s_j + m_k\right]$$

$$E[X_k] = \frac{\partial}{\partial s_k}\left[\phi_{\underline{X}}(\underline{s})\right]\Bigg|_{\underline{s}=0} = \phi_{\underline{X}}(\underline{0})\left[m_k\right] = m_k$$

$$\Rightarrow E[\underline{X}] = \underline{m}.$$

(2) $C = E\left[(\underline{X}-E[\underline{X}])(\underline{X}-E(\underline{X}))^T\right]$

$$E\left[(\underline{X}-E[\underline{X}])(\underline{X}-E[\underline{X}])^T\right] = E\left[\underline{X}\,\underline{X}^T\right] - E\left[\underline{X}\,E[\underline{X}]^T\right] - E\left[E[\underline{X}]\,\underline{X}^T\right] + E\left[E[\underline{X}]E[\underline{X}]^T\right]$$

$$= E\left[\underline{X}\,\underline{X}^T\right] - E[\underline{X}]E[\underline{X}]^T$$

$$C = R - \underline{m}\,\underline{m}^T.$$

Need to show $R = C + \underline{m}\,\underline{m}^T$.

$$R_{kl} = \frac{\partial}{\partial s_l}\frac{\partial}{\partial s_k}\phi_{\underline{X}}(\underline{s})\Big|_{\underline{s}=0}$$

$$= \frac{\partial}{\partial s_l}\left[\phi_{\underline{X}}(\underline{s})\left[m_k + C_{kk}s_k + \frac{1}{2}\sum_{\substack{j=1\\ j\neq k}}^{n}(C_{kj}+C_{jk})s_j\right]\right]\Bigg|_{\underline{s}=0}.$$

$$R_{kk} = \left\{\phi_{\underline{X}}(\underline{s})\left[C_{kk}\right] + \left[m_k + C_{kk}s_k + \frac{1}{2}\sum_{\substack{j=1\\ j\neq k}}^{n}(C_{kj}+C_{jk})s_j\right]^2\phi_{\underline{X}}(\underline{s})\right\}\Bigg|_{\underline{s}=0}$$

$$= C_{kk} + m_k^2.$$

$$R_{kl} = \left\{\phi_{\underline{X}}(\underline{s})\left[\frac{1}{2}(C_{kl}+C_{lk})\right] + \phi_{\underline{X}}(\underline{s})\left[m_k + C_{kk}s_k + \frac{1}{2}\sum_{\substack{j=1\\ j\neq k}}^{n}(C_{kj}+C_{jk})s_j\right]\left[m_l + C_{ll}s_l + \frac{1}{2}\sum_{\substack{j=1\\ j\neq l}}^{n}(C_{lj}+C_{jl})s_j\right]\right\}\Bigg|_{\underline{s}=0}$$

$$= \frac{1}{2}(C_{kl}+C_{lk}) + m_k m_l = C_{kl} + m_k m_l$$

3. $$\phi_{\underline{X}}(\underline{s}) = \exp\left\{ \underline{s}^T \underline{m} + \frac{1}{2} \underline{s}^T C \underline{s} \right\}.$$

$\underline{Y}$ is a $l \times 1$ sub-vector of $\underline{X}$ $(l < n)$.

$$\phi_{\underline{Y}}(\underline{s}^{(l)}) = \phi_{\underline{X}}(\underline{s}) \Big|_{s_i = 0 \text{ for all } i:\ X_i \text{ is not a component of } \underline{Y}.}$$

$$\phi_{\underline{Y}}(\underline{s}^{(l)}) = \exp\left\{ \underline{s}^{(l)T} \underline{m}_Y + \frac{1}{2} \underline{s}^{(l)T} C_Y \underline{s}^{(l)} \right\}$$

where C_Y is obtained by retaining the columns & rows of C for all i such that X_i is a component of $\underline{Y}$ and $\underline{m}_Y$ is obtained by retaining the components of $\underline{m}$ for all i such that X_i is a component of $\underline{Y}$.

Since $\phi_{\underline{Y}}(\underline{s}^{(l)})$ also has the form of a Gaussian r.v's characteristic fn, $\underline{Y}$ is also jointly Gaussian.

4. $$\underset{l\times 1}{\underline{Y}} = \underset{l\times n}{A}\, \underset{n\times 1}{\underline{X}} + \underset{l\times 1}{\underline{b}}$$

Point out this result:
For any $\underline{X}$, (not necessarily Gaussian)
$E[A\underline{X} + \underline{b}] = A\underline{m} + \underline{b}$
& $C_{A\underline{X}+\underline{b}} = ACA^T$

$$E\left[e^{\underline{s}^T \underline{Y}} \right] = E\left[e^{\underline{s}^T (A\underline{X} + \underline{b})} \right]$$

$$= e^{\underline{s}^T \underline{b}} \cdot E\left[e^{\underline{s}^T A \underline{X}} \right]$$

$$= e^{\underline{s}^T \underline{b}} \cdot E\left[e^{(A^T \underline{s})^T \underline{X}} \right]$$

$$= e^{\underline{s}^T \underline{b}} \exp\left\{ (A^T\underline{s})^T \underline{m} + \frac{1}{2} (A^T \underline{s})^T C (A^T \underline{s}) \right\}$$

$$= e^{\underline{s}^T \underline{b}} \exp\left\{ \underline{s}^T A \underline{m} + \frac{1}{2} \underline{s}^T A C A^T \underline{s} \right\}$$

$$= \exp\left\{ \underline{s}^T (A\underline{m} + \underline{b}) + \frac{1}{2} \underline{s}^T (ACA^T) \underline{s} \right\}.$$

$$\underline{Y} \sim N\left((A\underline{m} + \underline{b}), ACA^T \right).$$

(5) Suppose $\underline{X} = \begin{bmatrix} \underline{X}^{(1)} \\ \hline \underline{X}^{(2)} \end{bmatrix} \begin{array}{l} k \\ n-k \end{array}$

$$\underline{m} = \begin{bmatrix} \underline{m}^{(1)} \\ \underline{m}^{(2)} \end{bmatrix} \qquad C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}.$$

Since C is symmetric, $C_{21} = C_{12}^T$, $C_{11} = C_{11}^T$, $C_{22} = C_{22}^T$.

$$f_{\underline{X}^{(2)}/\underline{X}^{(1)}} \left(\underline{x}^{(2)} / \underline{X}^{(1)} = \underline{x}^{(1)} \right) = \frac{f_{\underline{X}}(\underline{x})}{f_{\underline{X}^{(1)}}(\underline{x}^{(1)})}$$

$$= N(\underline{m}', C')$$

where $\underline{m}' = \underline{m}^{(2)} + C_{21} C_{11}^{-1} (\underline{x}^{(1)} - \underline{m}^{(1)})$

$$C' = C_{22} - C_{21} C_{11}^{-1} C_{12}$$

→ {Prove (5) for the 2-D Gaussian) (Derive 2-D Gaussian PDF in terms of ρ)

(6) If C is diagonal, i.e. $C = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \sigma_n^2 \end{bmatrix}$

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2} \prod_{i=1}^{n} \sigma_i} \exp\left\{ - \sum_{i=1}^{n} \frac{(x_i - m_i)^2}{2\sigma_i^2} \right\}$$

$$= f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

$\Rightarrow X_1, X_2, \ldots, X_n$ are independent.

$\boxed{\text{Lecture 18:}}$

Given $\underline{X} \sim N(\underline{m}, C)$, we can find A such that $AA^T = C$

2-D Gaussian random vector

(Do this with property (5) in previous page)

$$f_{\underline{x}}(\underline{x}) = \frac{1}{(2\pi)\,|C|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x}-\underline{m})^T C^{-1} (\underline{x}-\underline{m})\right\}$$

$$C = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_1, X_2) & Var(X_2) \end{bmatrix} \qquad \underline{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

Define $\rho = \dfrac{Cov(X_1, X_2)}{\sqrt{Var(X_1)\,Var(X_2)}}$. (correlation coefficient)

(Measure of correlation of 2 random variables: If ρ is high, then if X_1 is large/small X_2 is also likely to be large/small).

Since $|Cov(X_1, X_2)| \leq \sqrt{Var(X_1)\,Var(X_2)}$

(Cauchy-Schwartz inequality)

$|\rho| \leq 1$.

$$C = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$C^{-1} = \frac{1}{\sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

$$= \frac{1}{(1-\rho^2)\,\sigma_1^2\sigma_2^2} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} = \frac{1}{1-\rho^2} \begin{bmatrix} 1/\sigma_1^2 & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & 1/\sigma_2^2 \end{bmatrix}$$

$$|C|^{1/2} = \sigma_1\sigma_2(1-\rho^2)^{1/2}$$

$$(\underline{x}-\underline{m})^T C^{-1}(\underline{x}-\underline{m}) = \frac{(x_1-m_1)}{(1-\rho^2)}\left(\frac{(x_1-m_1)}{\sigma_1^2} - \frac{\rho(x_2-m_2)}{\sigma_1\sigma_2}\right)$$

$$+ \frac{(x_2-m_2)}{(1-\rho^2)}\left(\frac{(x_2-m_2)}{\sigma_2^2} - \frac{\rho}{\sigma_1\sigma_2}(x_1-m_1)\right)$$

$$= \frac{1}{(1-\rho^2)}\left[\frac{(x_1-m_1)^2}{\sigma_1^2} + \frac{(x_2-m_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2}\right]$$

2-D Gaussian PDF (general case)

$$\boxed{f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi(1-\rho^2)^{1/2}\sigma_1\sigma_2}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-m_1)^2}{\sigma_1^2} + \frac{(x_2-m_2)^2}{\sigma^2} - \frac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2}\right]\right\}}$$

$\boxed{\text{Consider } \underline{m} = \underline{0} \text{ \& } \sigma_1, \sigma_2 = 1.}$ → ($\underline{m} \neq \underline{0}$ case is just a shifted PDF. So, we will always use $\underline{m} = \underline{0}$)

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi(1-\rho^2)^{1/2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[x_1^2 + x_2^2 - 2\rho x_1 x_2\right]\right\}$$

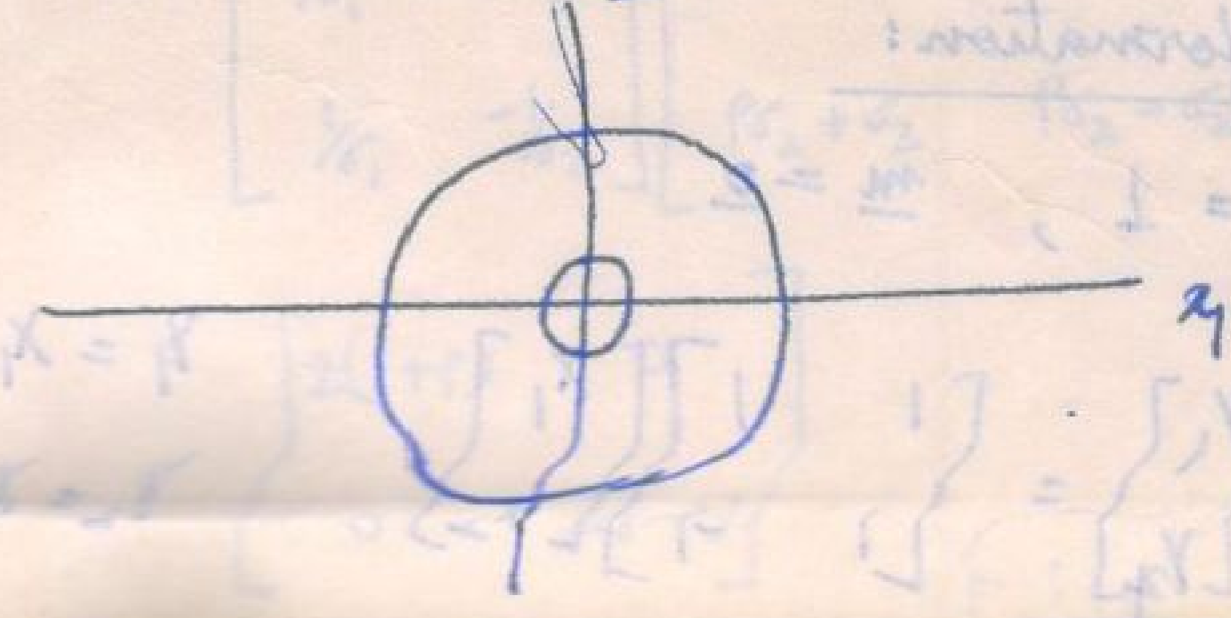
Contours of this pdf: (Do contours after whitening transformation is done)

$$x_1^2 + x_2^2 - 2\rho x_1 x_2 = \text{constant}.$$

If $\rho = 0$, $x_1^2 + x_2^2 = \text{constant}$. (circle).

$\boxed{\rho = 0,\ \sigma_1 \neq \sigma_2}$

(1)

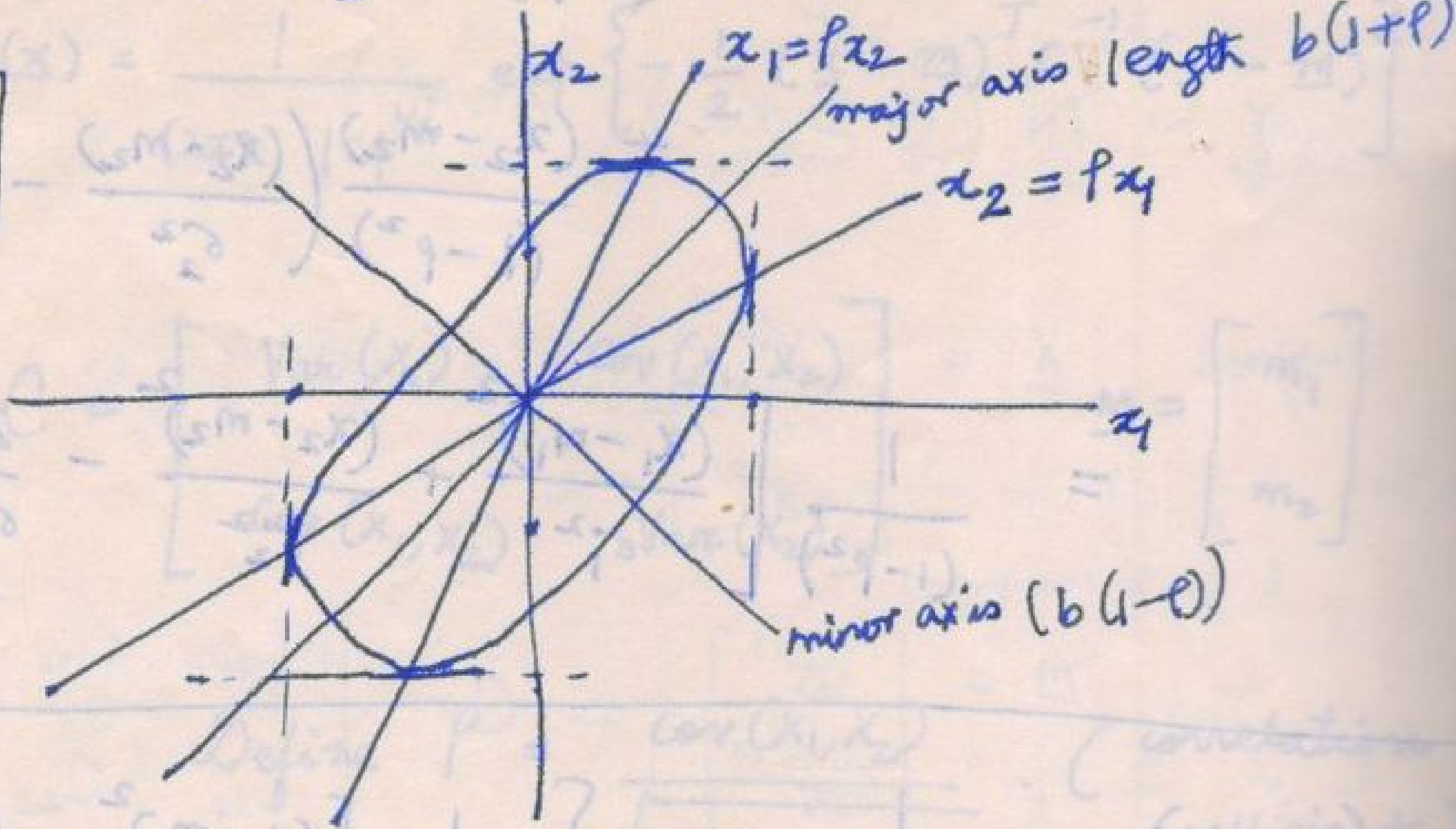


(2)

(3)(a) $\boxed{\rho \neq 0 \cdot \ \rho > 0 \quad \sigma_1, \sigma_2 = 1}$

$$x_1^2 + x_2^2 - 2\rho x_1 x_2 = \text{constant} \, . = b^2$$

$$C = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

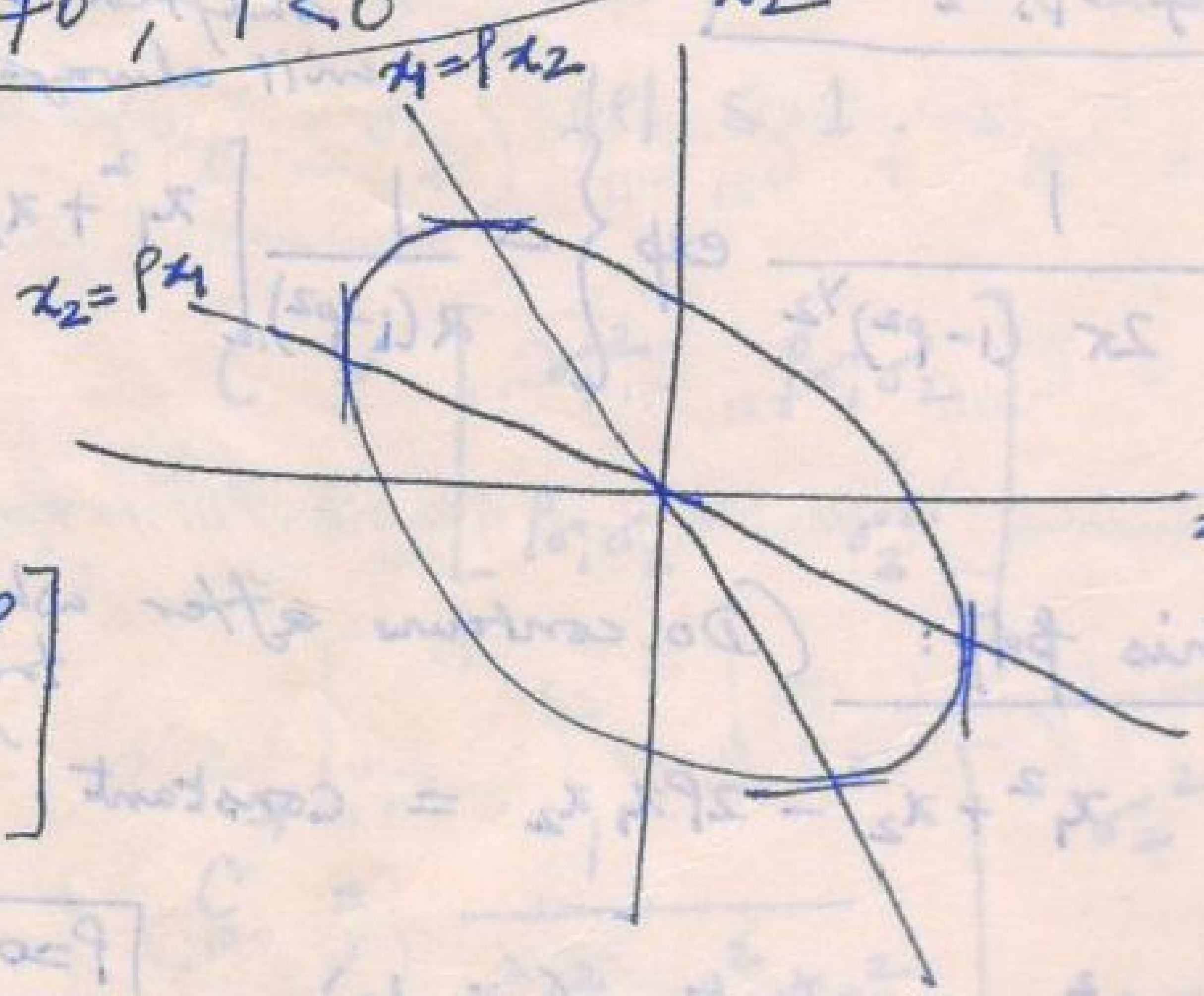


$$\frac{\partial}{\partial x_1}\left(x_1^2 + x_2^2 - 2\rho x_1 x_2 - \text{constant}\right) = 2x_1 - 2\rho x_2$$

$$\frac{\partial}{\partial x_2}\left(x_1^2 + x_2^2 - 2\rho x_1 x_2 - \text{constant}\right) = 2x_2 - 2\rho x_1$$

(b) $\boxed{\rho \neq 0, \ \rho < 0 \quad \sigma_1, \sigma_2 = 1}$

$$C = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$



<u>Whitening transformation:</u>

<u>I:</u> $\sigma_1, \sigma_2 = 1 \, , \quad \underline{m} = \underline{0}$

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \qquad \begin{aligned} Y_1 &= X_1 + X_2 \\ Y_2 &= X_1 - X_2 \end{aligned}$$

$$C_y = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1+\rho & 1-\rho \\ 1+\rho & \rho-1 \end{bmatrix}$$

$$= \begin{bmatrix} 2(1+\rho) & 0 \\ 0 & 2(1-\rho) \end{bmatrix}$$

Y_1 & Y_2 are uncorrelated & independent.

II: $\underline{m} = \underline{0} \qquad \sigma_1, \sigma_2$

$$C = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Another transformation that will also result in $N(\underline{0}, I)$ is $\begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & (1-\rho^2)\sigma_2^2 \end{bmatrix}^{-1}$. This will be related to A by a rotation matrix.

Similarly, there are infinitely many transformations

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 1/\sigma_1 & 1/\sigma_2 \\ 1/\sigma_1 & -1/\sigma_2 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \qquad Y_1 = \frac{X_1}{\sigma_1} + \frac{X_2}{\sigma_2} \qquad Y_2 = \frac{X_1}{\sigma_1} - \frac{X_2}{\sigma_2}$$

$$C_y = ACA^T = \begin{bmatrix} 1/\sigma_1 & 1/\sigma_2 \\ 1/\sigma_1 & -1/\sigma_2 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \begin{bmatrix} 1/\sigma_1 & 1/\sigma_1 \\ 1/\sigma_2 & -1/\sigma_2 \end{bmatrix}$$

$$= \begin{bmatrix} 1/\sigma_1 & 1/\sigma_2 \\ 1/\sigma_1 & -1/\sigma_2 \end{bmatrix} \begin{bmatrix} \sigma_1 + \rho\sigma_1 & \sigma_1 - \rho\sigma_1 \\ \rho\sigma_2 + \sigma_2 & \rho\sigma_2 - \sigma_2 \end{bmatrix}$$

$$= \begin{bmatrix} 2(1+\rho) & 0 \\ 0 & 2(1-\rho) \end{bmatrix}$$

$\checkmark A$

Example of 2 r.v's X & Y being individually Gaussian but not jointly Gaussian:

$$f_{X,Y}(x,y) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(x^2+y^2)\right\}\left(1 + xy\exp\left[-\frac{1}{2}(x^2+y^2-2)\right]\right)$$

(Try a mesh plot of this)

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \quad \text{Gaussian}$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\} \quad \text{Gaussian}$$

X, Y are not jointly Gaussian.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

$$= \frac{1}{2\pi}\exp\left\{-\frac{x^2}{2}\right\}\int_{-\infty}^{\infty} e^{-y^2/2}\,dy$$

$$+ \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}\int_{-\infty}^{\infty} xy\, e^{-y^2} \cdot e^{-x^2/2} \cdot e\, dy$$

$$= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\} + \frac{1}{\sqrt{2\pi}} e^{-x^2} x.e \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} y e^{-y^2} dy$$

$$= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\} + 0$$

$$= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}.$$

Conditional pdf of X_1 given X_2 when X_1 & X_2 are jointly Gaussian (Proof of property 5 in page 73 for $n=2$)

$$f_{X_1|X_2}(x_1 \mid X_2 = x_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}$$

$$= \frac{\dfrac{1}{2\pi(1-\rho^2)^{1/2}\sigma_1\sigma_2} \exp\left\{-\dfrac{1}{2(1-\rho^2)}\left[\dfrac{(x_1-m_1)^2}{\sigma_1^2} + \dfrac{(x_2-m_2)^2}{\sigma_2^2} - \dfrac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2}\right]\right\}}{\dfrac{1}{\sqrt{2\pi}\,\sigma_2} \exp\left\{-\dfrac{1}{2}\dfrac{(x_2-m_2)^2}{\sigma_2^2}\right\}}$$

$$= \frac{1}{\sqrt{2\pi}\,(1-\rho^2)^{1/2}\sigma_1} \exp\left\{-\frac{1}{2}\,\frac{\dfrac{(x_1-m_1)^2}{\sigma_1^2} + \dfrac{\rho^2(x_2-m_2)^2}{\sigma_2^2} - \dfrac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2}}{(1-\rho^2)}\right\}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_1(1-\rho^2)^{1/2}} \exp\left\{-\frac{1}{2}\,\frac{\left(\dfrac{x_1-m_1}{\sigma_1} - \dfrac{\rho(x_2-m_2)}{\sigma_2}\right)^2}{(1-\rho^2)}\right\}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\,\frac{\left(x_1 - m_1 - \dfrac{\rho\sigma_1}{\sigma_2}(x_2-m_2)\right)^2}{\sigma_1^2(1-\rho^2)}\right\}$$

$$X_1 / X_2 = x_2 \sim N\left(m_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2 - m_2),\ \sigma_1^2(1-\rho^2)\right).$$

$\downarrow$ if $\rho = \pm 1$, $\sigma_1^2(1-\rho^2) = 0$.

Back to the contours of a 2-D Gaussian pdf:

I: (4)(a) $\boxed{\rho \neq 0, \quad \sigma_1 = \sigma_2, \quad m_1 = m_2 = 0}$

Eqn of the contour

$$\frac{x_1^2}{\sigma^2} + \frac{x_2^2}{\sigma^2} - \frac{2\rho x_1 x_2}{\sigma^2} = b^2 = \text{constant}$$

rotation by 45°

Define $\begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = -\frac{1}{\sqrt{2}} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

$$\Rightarrow \frac{(u+v)^2}{2\sigma^2} + \frac{(u-v)^2}{2\sigma^2} - \frac{2\rho(u^2 - v^2)}{2\sigma^2} = b^2$$

$$u^2 + v^2 - \rho(u^2 - v^2) = b^2\sigma^2$$

$$u^2(1-\rho) + v^2(1+\rho) = b^2\sigma^2$$

$\frac{u^2}{1+\rho} + \frac{v^2}{1-\rho} = \frac{b^2\sigma^2}{1-\rho^2}$ (Ellipse with major & minor axes along u, v ~~directions~~

Assuming $\rho > 0$,

Major axis length $= \frac{b\sigma}{\sqrt{1-\rho^2}}(1+\rho)$

Minor axis length $= \frac{b\sigma}{\sqrt{1-\rho^2}}(1-\rho)$

II 4(b) $\boxed{\rho \neq 0, \quad \sigma_1 \neq \sigma_2, \quad m_1 = m_2 = 0.}$

Eqn. of contour.

$$\frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2} - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} = d^2 = \text{constant}.$$

Rotated ellipse.

Find a rotation such that

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

& the eqn. of the contour becomes

$$\frac{u^2}{a^2} + \frac{v^2}{b^2} = c^2 .$$

Leave this part as an assignment.

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} .$$

$x_1 = u\cos\theta - v\sin\theta$

$x_2 = u\sin\theta + v\cos\theta .$

$$\frac{(u\cos\theta - v\sin\theta)^2}{\sigma_1^2} + \frac{(u\sin\theta + v\cos\theta)^2}{\sigma_2^2} - \frac{2\rho(u\cos\theta - v\sin\theta)(u\sin\theta + v\cos\theta)}{\sigma_1\sigma_2} = d^2$$

$$\frac{u^2\cos^2\theta + v^2\sin^2\theta - 2uv\sin\theta\cos\theta}{\sigma_1^2} + \frac{u^2\sin^2\theta + v^2\cos^2\theta + 2uv\sin\theta\cos\theta}{\sigma_2^2}$$

$$- \frac{2\rho\left(u^2\sin\theta\cos\theta + uv\cos^2\theta - uv\sin^2\theta - v^2\sin\theta\cos\theta\right)}{\sigma_1\sigma_2} = d^2$$

uv term should have coefficient 0. $\Rightarrow$

$$\frac{-\sin 2\theta}{\sigma_1^2} + \frac{\sin 2\theta}{\sigma_2^2} - \frac{2\rho}{\sigma_1\sigma_2}(\cos^2\theta - \sin^2\theta) = 0$$

$$\Rightarrow \quad \sin 2\theta\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right) = \frac{2\rho}{\sigma_1\sigma_2}\left(\frac{1+\cos 2\theta}{2} - \frac{1-\cos 2\theta}{2}\right)$$

$$\Rightarrow \quad \sin 2\theta\left(\frac{\sigma_1^2 - \sigma_2^2}{\ldots}\right) = \frac{2\rho}{\ldots}\cos 2\theta$$

$$\Rightarrow \tan 2\theta = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}.$$

if $\sigma_1 = \sigma_2$, rotation by $45°$

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}\right).$$

$\boxed{\text{Lecture 19:}}$

N-dimensional proper complex Gaussian random vector: (cover all of complex r.v's here)

Let $\underline{X}$ be a proper complex N-dimensional Gaussian random vector with mean $\underline{m}$ and nonsingular covariance matrix $C = E[(\underline{X}-\underline{m})(\underline{X}-\underline{m})^H]$.

Then, the pdf of $\underline{X}$ is given by

$$f_{\underline{X}}(\underline{x}) = \frac{1}{\pi^N |C|} \exp\left\{-(\underline{x}-\underline{m})^H C^{-1} (\underline{x}-\underline{m})\right\}$$

$E[(\underline{X}-\underline{m})(\underline{X}-\underline{m})^T] = 0$ (since $\underline{X}$ is proper)

Note that the pdf is completely specified by the vector of means & the conventional covariance matrix.

Sequences of random variables:

The limit of a sequence of numbers $x_1, x_2, x_3, \ldots$ is x if for any $\epsilon > 0$ there exists an N such that

$$|x_n - x| \leq \epsilon \text{ for all } n \geq N.$$

$$\left(\lim_{n\to\infty} x_n = x.\right)$$

Talk of a [illegible] of [illegible] [illegible] convergence [illegible] convergence [illegible]

Now, consider a sequence of random variables.

$X_1(\zeta), X_2(\zeta), \ldots, X_n(\zeta), \ldots$

The sequence converges to a random variable $X(\zeta)$ if for any $\epsilon > 0$, there exists an N such that

$$|X_n(s) - X(s)| \leq \epsilon \text{ for all } n \geq N$$

$$\& \text{ for all } s \in S.$$

(Convergence everywhere) or Pointwise convergence.

Other weaker types of convergence:

① Almost sure convergence:

$\{X_n\}_{n=1}^{\infty}$ converges almost surely to the r.v X if the set of $s \in S$ such that $\lim_{n\to\infty} X_n(s) = X(s)$ has probability 1.

$$P(\Omega_0) = 1 \quad \text{where } \Omega_0 = \{s \in S \mid \lim_{n\to\infty} X_n(s) = X(s)\}.$$

② Convergence in the mean-square sense:

$\{X_n\}_{n=1}^{\infty}$ converges in the mean square sense, to X if

$$\lim_{n\to\infty} E[(X_n - X)^2] = 0.$$

③ Convergence in probability:

$\{X_n\}_{n=1}^{\infty}$ converges to X in probability if for any $\epsilon > 0$

$$P(|X_n - X| > \epsilon) \longrightarrow 0 \quad \text{as } n \to \infty$$

i.e. $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0.$

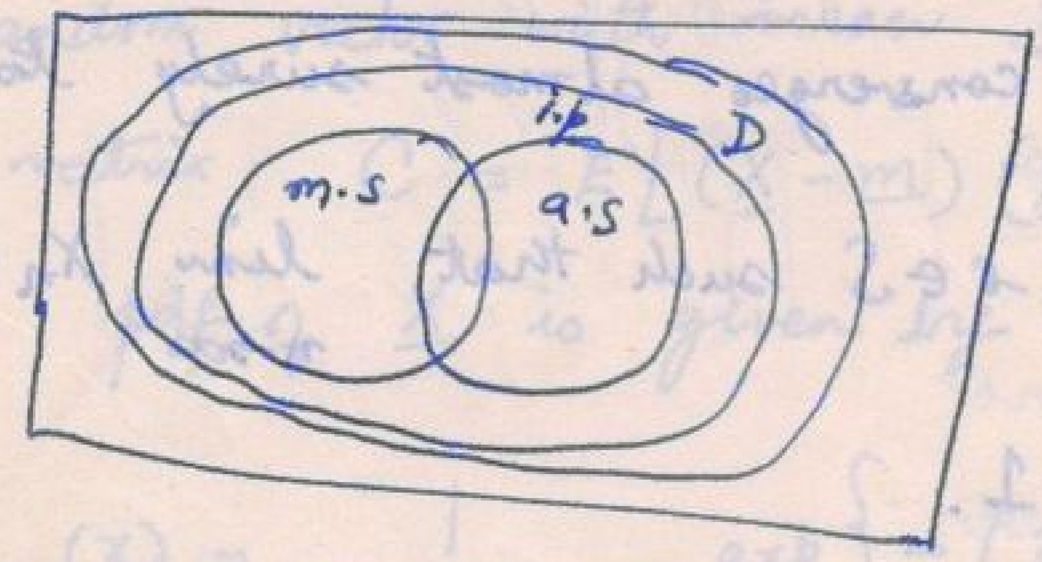
④ Convergence in Distribution:

$\{X_n\}_{n=1}^{\infty}$ converges in distribution to X if

$$\lim_{n\to\infty} F_{X_n}(b) = F_X(b) \quad \text{for all continuity points of } F_X(x).$$

almost sure convergence $\Rightarrow$ convergence in probability $\Rightarrow$ convergence in distribution

mean-square convergence $\Rightarrow$ convergence in probability



Sum of random variables: example

Suppose $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of r.v's such that $E[X_i] = \mu$ for all i and $Var(X_i) < M$ for all i and $Cov(X_i, X_j) = 0$ for all $i \neq j$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{m.s} \mu.$$

$$E\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2\right] = \frac{1}{n^2}\sum_{i=1}^{n} E\left[(X_i-\mu)^2\right]$$

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

$$< \frac{Mn}{n^2} = \frac{M}{n}.$$

$$\lim_{n\to\infty} E\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2\right] = 0.$$

Central limit theorem (CLT):

If $\{X_n\}_{n=1}^{\infty}$ is a sequence of i.i.d r.v's with mean μ and variance $\sigma^2 < +\infty$, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow{D} X$$

where $X \sim N(0, \sigma^2)$.

Proof:

Lemma 1: Let $\{F_n\}_{n=1}^{\infty}$ be a sequence of cdf's and let $\{\phi_n\}_{n=1}^{\infty}$ be the corresponding sequence of their characteristic functions. Then, if there is a characteristic fn $\phi(\omega)$ of a cdf $F(\alpha)$ such that

$$\lim_{n \to \infty} \phi_n(\omega) = \phi(\omega) \quad \forall \omega$$

then $\lim_{n \to \infty} F_n(\alpha) = F(\alpha) \quad \forall \alpha$ pointwise convergence.

Lemma 2:

Suppose X is a r.v. with $E[X^2] < \infty$, then

$$\phi_X(\omega) = E\left[e^{j\omega X}\right]$$

$$= 1 + j\omega E[X] + \frac{j^2 \omega^2 E[X^2]}{2!} + \cdots$$

$$= 1 + j\omega E[X] - \frac{\omega^2}{2}\left[E[X^2] + d(\omega)\right]$$

where $d(\omega) \to 0$ as $\omega \to 0$ (if the moments are finite).

$\omega^2 d(\omega) \to 0$ as long as $E[X]$ and $E[X^2]$ are finite.

Now, $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ (Assume $E[X_i] = 0$).

$$\phi_{S_n}(\omega) = E\left[e^{j\omega \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k}\right]$$

$$= E\left[\prod_{k=1}^{n} e^{j\frac{\omega}{\sqrt{n}} X_k}\right]$$

$$\text{(i.i.d)} \quad = \prod_{k=1}^{n} E\left[e^{j\frac{\omega}{\sqrt{n}} X_k}\right] = \prod_{k=1}^{n} \phi_{X_k}\left(\frac{\omega}{\sqrt{n}}\right)$$

$$\text{(i.i.d)} \quad = \left[\phi_{X_k}\left(\frac{\omega}{\sqrt{n}}\right)\right]^n$$

$$= \left[1 + j\frac{\omega}{\sqrt{n}} E[X] - \frac{\omega^2}{2n}\left(E[X_k^2] + d\left(\frac{\omega}{\sqrt{n}}\right)\right)\right]^n$$

$$= \left[1 - \frac{\omega^2}{2n}\left(E[X_k^2] + d\left(\frac{\omega}{\sqrt{n}}\right)\right)\right]^n$$

(From next J

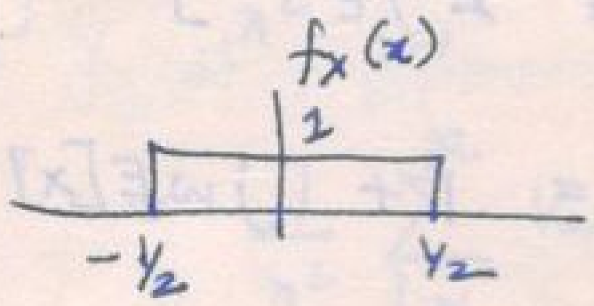
$$\left(\text{using } \lim_{n\to\infty} \left(1 + \frac{y}{n}\right)^n = e^y\right)$$

$$\lim_{n\to\infty} \phi_{S_n}(\omega) = e^{-\frac{\omega^2}{2}\sigma^2} = \phi_X(\omega)$$

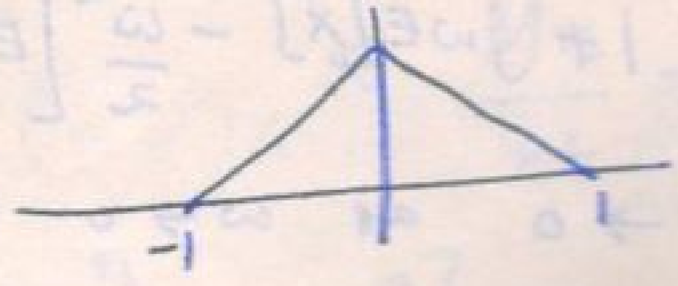
where $X \sim N(0, \sigma^2)$

CLT Eg:

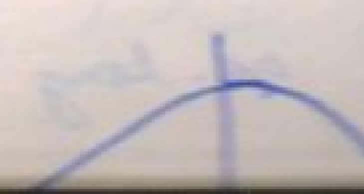
$X_i \sim$



$X_1 + X_2$



$X_1 + X_2 + X_3 + X_4$



CLT powerful near the mean, not so near the tail.

$$\text{cdf of } \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(X_i-\mu) \to \text{cdf of } N(0,\sigma^2)$$

as $N\to\infty$

If X_i's are continuous r.v's, $f_{S_N}(\alpha)$ approaches a normal pdf. If X_i's are discrete, cdf $F_{S_N}(\alpha)$ is a staircase function approaching the normal cdf. The pmf is not necessarily related to the normal density except for lattice-type r.v's that take equidistant values.

When is the CLT approx. good? Depends on the dist of X_i.

(From Wozencraft & Jacobs)

Eg:

$$P\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i \geqslant a\right] \approx Q\left(\frac{a}{\sigma}\right)$$

If N is "sufficiently large" and 'a' is "sufficiently small" compared to σ.

If $a=\epsilon\sqrt{N}$ (increasing w.r.t. N), then it is a bad approximation.

Suppose $X_i \sim$ Bernoulli $(\frac{1}{2},\frac{1}{2})$.

$$P\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}X_i \geqslant \underbrace{2\sqrt{N}}\right] = P\left[\frac{1}{N}\sum_{i=1}^{N}X_i \geqslant 2\right] = 0.$$

This is too far in the tail of the PDF!

But CLT-based

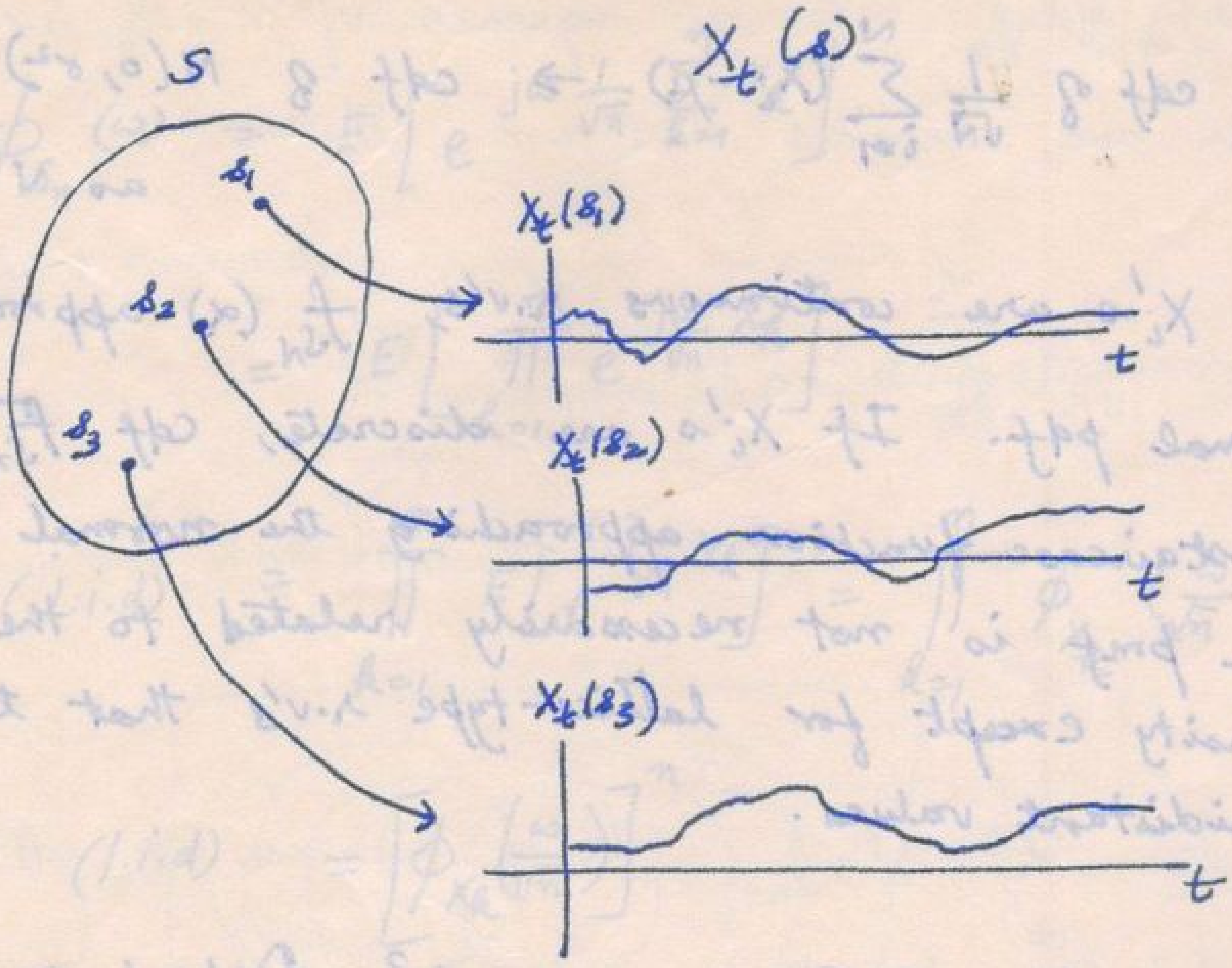
$$Q\left(\frac{\epsilon\sqrt{N}}{\sigma}\right) > 0.$$

Example: $N=100$.

$$P\left[\frac{1}{10}\sum^{100}X_i \geqslant 20\right] = 0. \quad Q\left(\frac{20}{1/2}\right) = Q(40).$$

Lecture 20:

RANDOM PROCESSES:



First interpretation:

A random process is an indexed collection of random variables $\{X_t(s) : t \in T\}$ where T is the indexing set.

If $T = \mathbb{R}$ (a continuous set), X_t is a continuous-time random process.

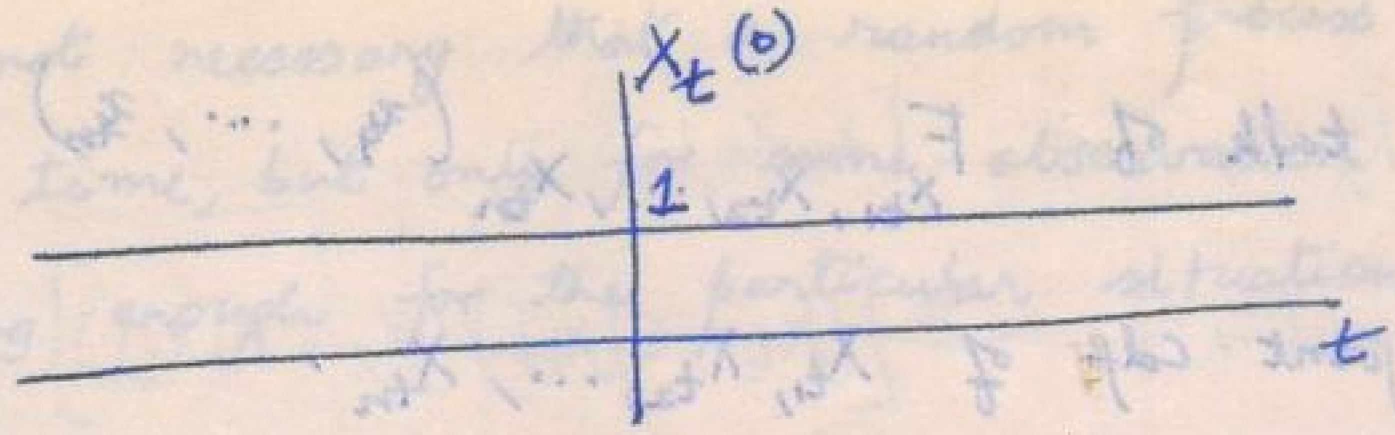
$T = \mathbb{Z}$ or N (discrete), X_t is a discrete-time random process.

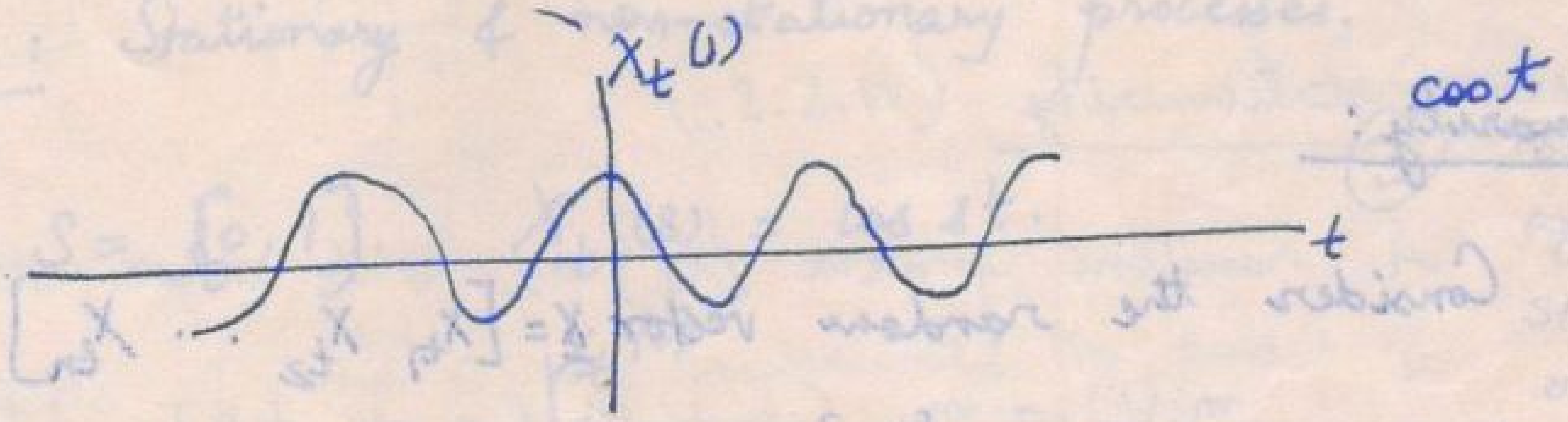
Second interpretation:

A random process is an ensemble (collection) of sample functions that occur randomly, i.e, for all $s \in S$, we have $X_t(s)$. $X_t(s)$ is a function of t and is referred to as a sample function for each $s \in S$.

Example: $S = \{0, 1\}$

For each s, $X_t(s) = \cos st$.





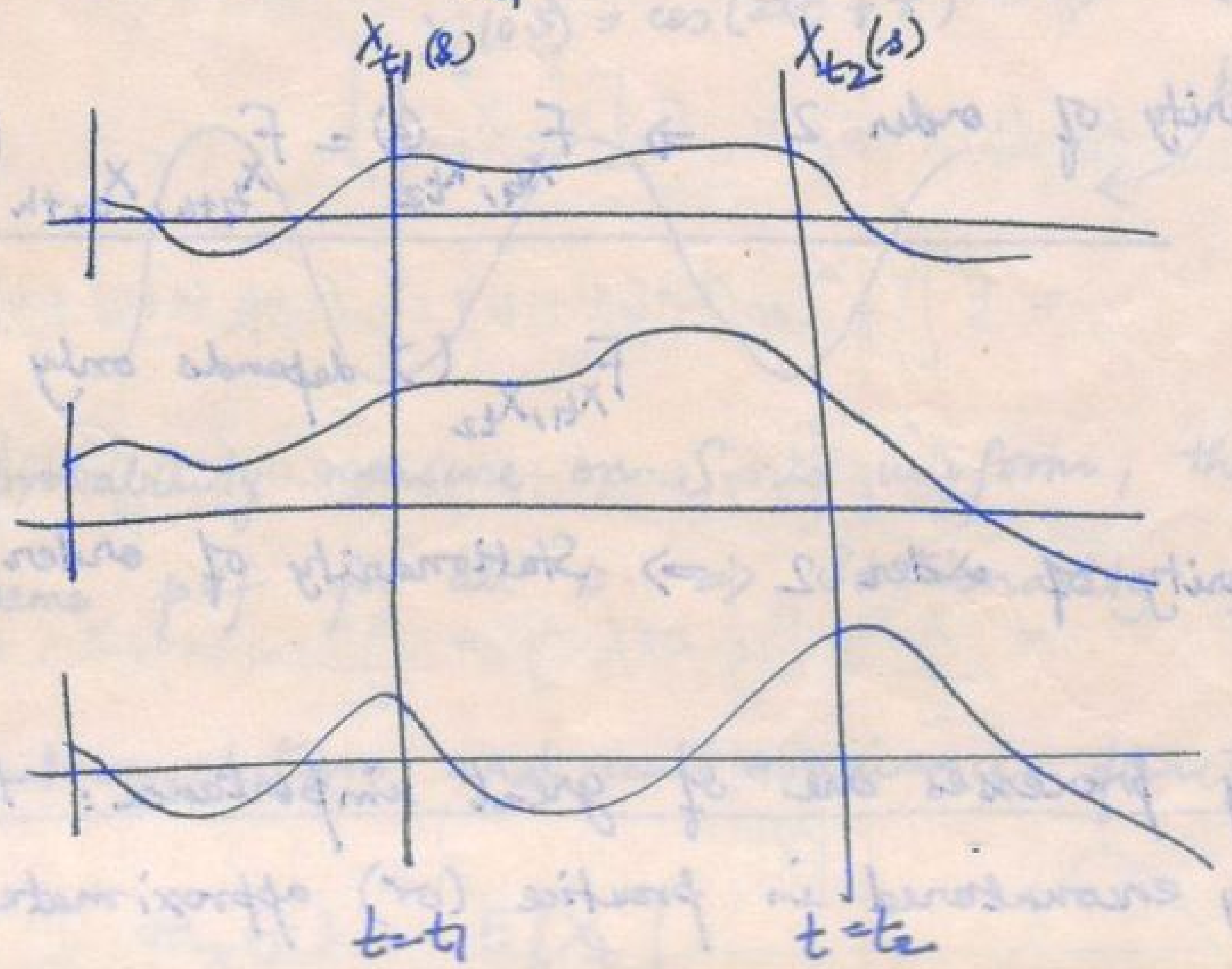
At any given time t_0, $X_{t_0}(s)$ is a binary random variable.

$$X_{t_0}(s) = \begin{cases} 1 & P[s=0] \\ \cos t_0 & P[s=1] = 1 - P[s=0]. \end{cases}$$

Random Vectors obtained from random processes:

As discussed above, X_t is an infinite collection of random variables. We can describe the statistical properties of the random process by describing the joint statistics of every finite subcollection of random variables obtained by sampling $X_t(s)$.

We can talk about $F_{X_{t_1}}(x_{t_1})$: cdf of X_{t_1} (sample of X_t at $t=t_1$)



Similarly, we can talk of $F_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(x_1, \ldots, x_n)$

joint cdf of $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$.

Stationarity:

Consider the random vector $\underline{X} = [X_{t_1} \; X_{t_2} \; \cdots \; X_{t_n}]^T$ obtained by sampling $\{X_t\}$. The random process $\{X_t\}_{t \in T}$ is said to be strictly stationary if

$$F_{\underline{X}}(\underline{x}) = F_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$$

$$= F_{X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h}}(x_{t_1}, x_{t_2}, \ldots, x_{t_n})$$

for all $h \in T$ and $(x_{t_1}, x_{t_2}, \ldots, x_{t_n}) \in S^n$ and all n

If this is true only for $k \le n$, then the random process is stationary to order n.

Stationarity of order 1 $\Rightarrow F_{X_{t_1}}(a) = F_{X_{t_1+h}}(a)$ for all $h \in T$, ~~t_1~~

$\Rightarrow F_{X_t}(a) = F_X(a)$ independent of t.

Stationarity of order 2 $\Rightarrow F_{X_{t_1}, X_{t_2}}(a) = F_{X_{t_1+h}, X_{t_2+h}}(a) \;\; \forall h \in T$, ~~$t_1$~~

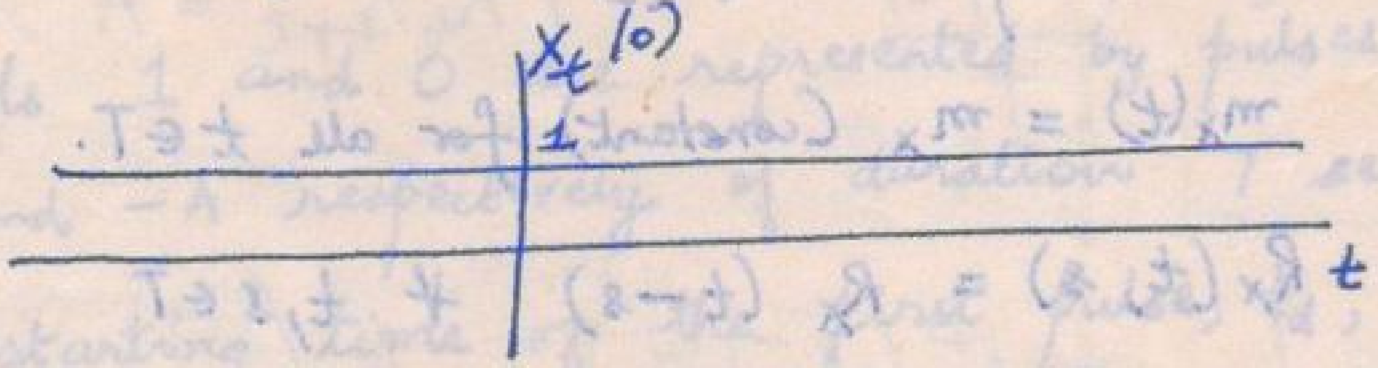
$F_{X_{t_1}, X_{t_2}}(\,)$ depends only on $t_2 - t_1$.

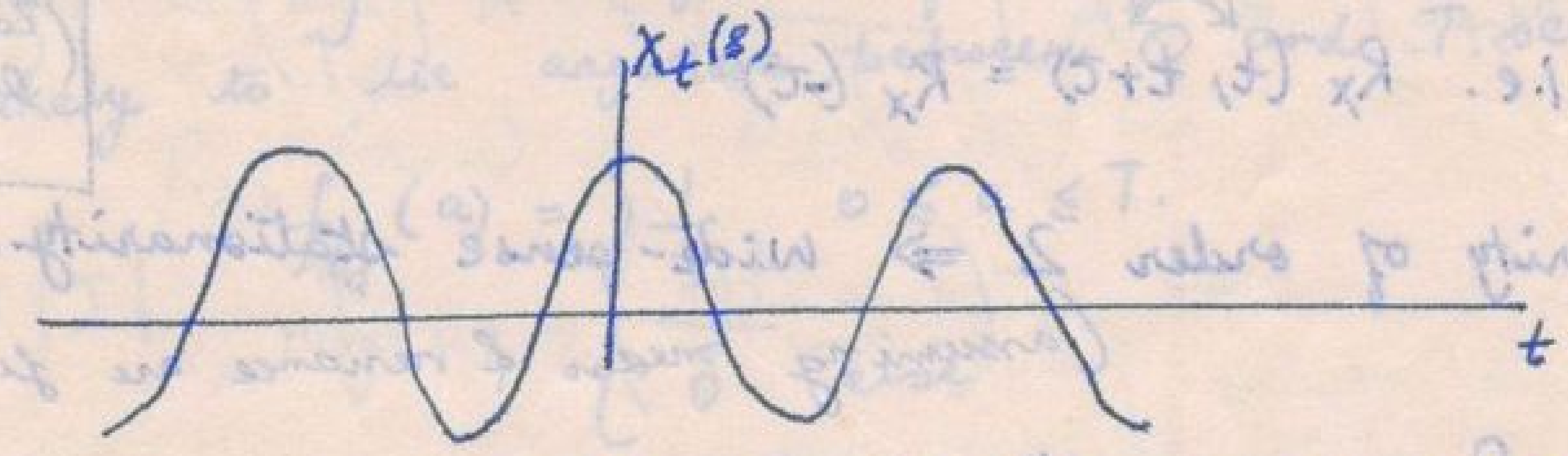
Stationarity of order 2 $\Rightarrow$ Stationarity of order 1.

Stationary processes are of great importance: they are frequently encountered in practice (or) approximated to a high degree of accuracy. From a practical point of view it is not necessary that a random process be stationary at all time, but only for some observation interval that is long enough for the particular situation. (Haykin).

Examples : Stationary & non-stationary processes.

① $S = \{0, 1\}$ $X_t(s) = \cos st$.





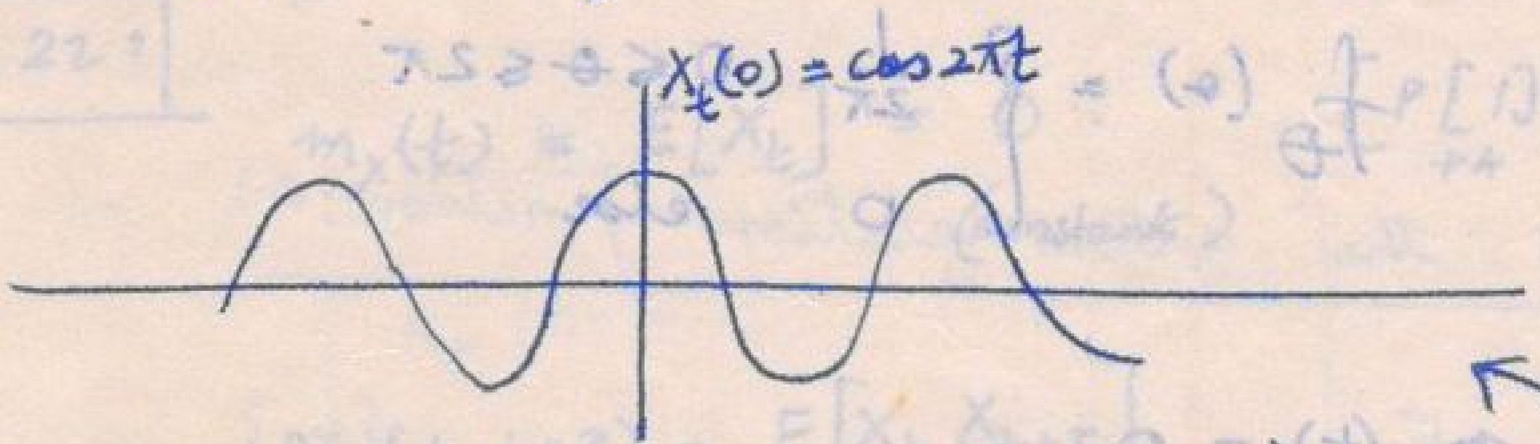
$X_0(s) = 1$ for all $s \in S$. $X_\pi(s) = \cos \pi s = \begin{cases} 1 & s = 0 \\ -1 & s = -1 \end{cases}$

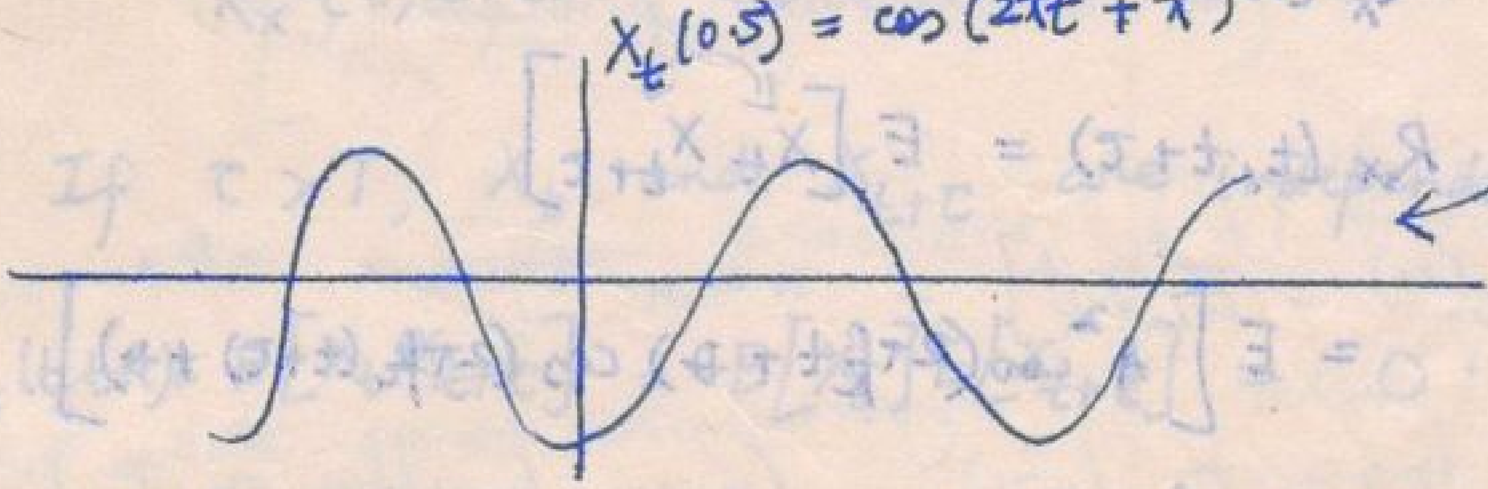
Not stationary. ~~in general, except for trivial cases where $P[s=1]$~~

③ Toss a coin again & again. Stationarity of order 1 $\Rightarrow$ P(head) is the same for each toss.

$\boxed{\text{Lecture 21.}}$

② $S = [0,1]$. $X_t(s) = \cos(2\pi t + 2\pi s)$.





Same frequency but different phase

If probability measure on S is uniform, then X_t has the same pdf for all t. $\Rightarrow$ Stationarity of order 1.

Mean, autocorrelation, and autocovariance functions:

$$m_X(t) = E[X_t]$$

$$R_X(t,s) = E[X_t X_s]$$

$$C_X(t,s) = E\left[(X_t - E[X_t])(X_s - E[X_s])\right]$$

Wide-sense stationarity (W.S.S.)

A random process $\{X_t\}$ is wide-sense stationary if

$$m_X(t) = m_X \text{ (constant) for all } t \in T.$$

$$R_X(t,s) = R_X(t-s) \quad \forall\, t,s \in T$$

define this as $E[X_{t+\tau} X_t]$ ← i.e. $R_X(t, t+\tau) = R_X(-\tau)$

$R_X(t,s)$ is a 2-D fn. For W.S.S. process it has some structure (constant along lines $t-s=$ ...

Stationarity of order 2 ⇒ Wide-sense stationarity. (assuming mean & variance are finite) Therefore ...

Example: Sine wave with random phase.

$$X_t = A\cos(2\pi f_c t + \Theta)$$

where A, f_c are constants and Θ is a random variable that is uniformly distributed over a range of 0 to 2π.

$$f_\Theta(\theta) = \begin{cases} \frac{1}{2\pi} & 0 \le \theta \le 2\pi \\ 0 & \text{else} \end{cases}$$

Same frequency but different phase.

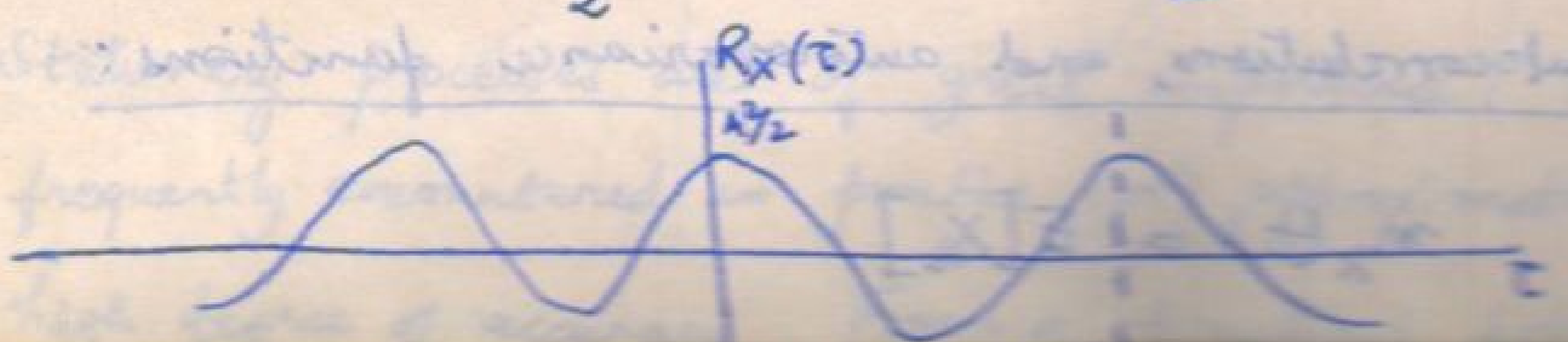
$$m_X(t) = 0$$

$$R_X(t, t+\tau) = E[X_t X_{t+\tau}]$$

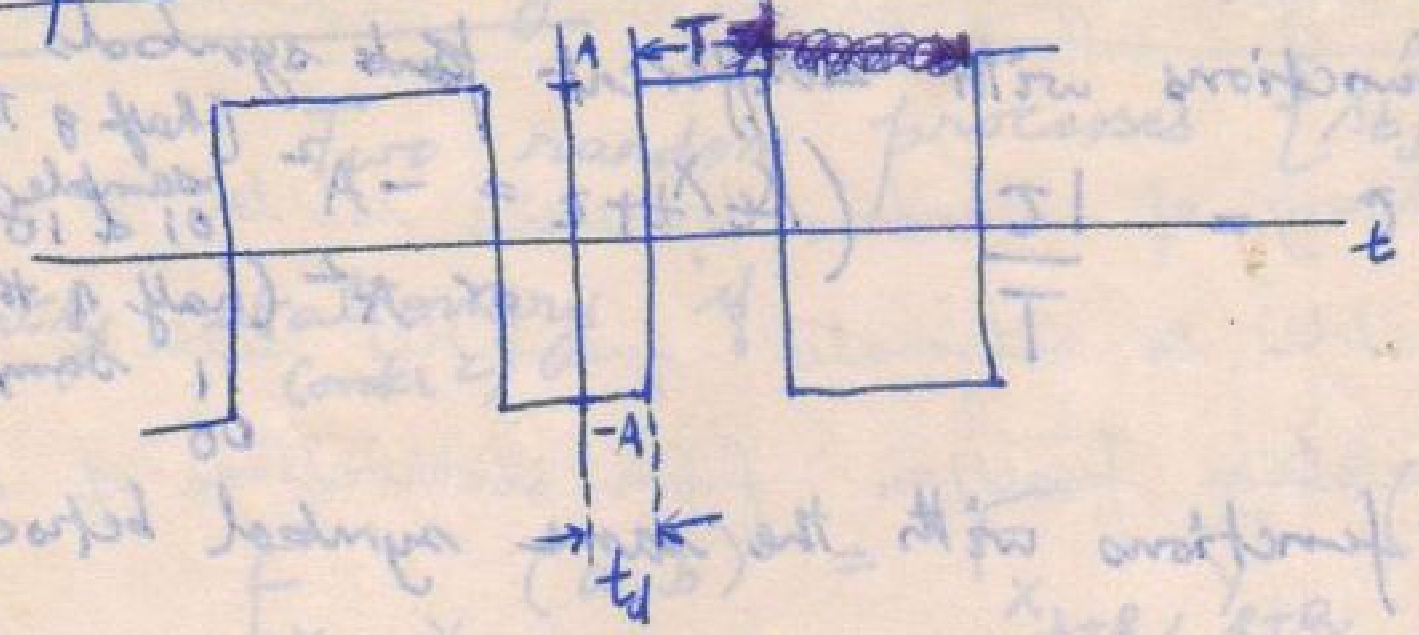
$$= E\left[A^2 \cos(2\pi f_c t + \theta)\cos(2\pi f_c (t+\tau) + \theta)\right]$$

$$= \frac{A^2}{2} E\left[\cos(4\pi f_c t + 2\pi f_c \tau + 2\theta)\right] + \frac{A^2}{2} E[\cos 2\pi f_c \tau]$$

$$= \frac{A^2}{2} E[\cos 2\pi f_c \tau] = \frac{A^2}{2}\cos 2\pi f_c \tau. = R_X(\tau)$$



Example: Random binary wave.



$$X_t = \sum_{n=-\infty}^{\infty} A_n \, rect(t-nT-t_d)$$

1. Symbols 1 and 0 are represented by pulses of amplitude $+A$ and $-A$ respectively of duration T seconds.
2. The starting time of the first pulse, t_d, is equally likely to lie anywhere between 0 and T seconds.

$$f_{t_d}(a) = \begin{cases} \frac{1}{T} & 0 \le a \le T. \\ 0 & \text{else} \end{cases}$$

3. During any time interval $(n-1)T \le t - t_d < nT$, where n is an integer, $P[1] = P[0]$ and the value in each interval is independent of the value in all other intervals.

Lecture 22:

$$m_x(t) = E[X_t] = 0 . \quad \left[P[\underset{+A}{1}] = P[\underset{-A}{0}]\right].$$
(constant)

$$R_X(t, t+\tau) = E[X_t X_{t+\tau}] .$$

* If $\tau > T$, X_t and $X_{t+\tau}$ are independent. Therefore,

$$R_x(t,t+\tau) = E[X_t X_{t+\tau}] = E[X_t]E[X_{t+\tau}] = 0 .$$

* If $\tau < T$, $X_t X_{t+\tau}$ takes on values A^2 or $-A^2$.

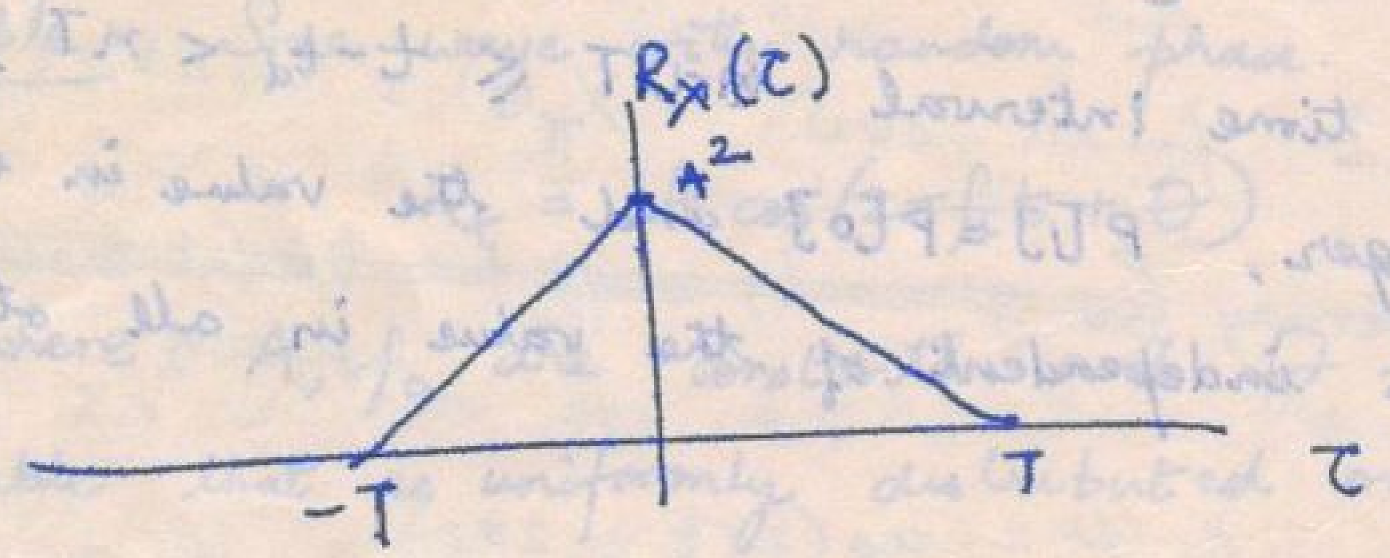
We need to find $E[X_t X_{t+\tau}]$.

Fraction of sample functions with 2 different ~~fent~~ symbols between t & $t+\tau$ $= \dfrac{|\tau|}{T}$ $\Big(X_t X_{t+\tau} = -A^2$ (half of these sample functions) 01 & 10 · or A^2 (half of these sample functions) 11 00

Fraction of sample functions with the same symbol between t & $t+\tau$ $= \dfrac{T-|\tau|}{T}$ $\quad \left(X_t X_{t+\tau} = A^2 \right)$

$$R_X(t,t+\tau) = E\left[X_t X_{t+\tau}\right] = A^2\left(\frac{T-|\tau|}{T}\right) + A^2\left(\frac{|\tau|}{2T}\right) - A^2\left(\frac{|\tau|}{2T}\right)$$

$$= A^2\left(1 - \frac{|\tau|}{T}\right) \quad \text{for } |\tau| \leq T.$$



Properties of the autocorrelation function:

For a stationary process $\{X_t\}$

1. $R_X(0) = E[X_t^2]$ = mean-square value

2. $R_X(\tau) = R_X(-\tau)$ (even function of τ)

3. $|R_X(\tau)| \leq R_X(0)$ (Cauchy-Schwartz inequality).

↳ Do this for complex W.S.S. processes too.

Joint stationarity:

Two random processes $\{X_t\}$ & $\{Y_t\}$ are jointly stationary (order 2) if

$$F_{X_t, Y_s}(a,b) = F_{X_{t+h}, Y_{s+h}}(a,b) \quad \forall (a,b),\ \forall h,\ \forall t, s.$$

↳ Explain by fixing $t = t_0, s = s_0$. $t = t_1, s = s_1$, and so on.

Cross-correlation functions:

$$R_{XY}(t,u) = E[X_t Y_u]$$

$$R_{YX}(t,u) = E[Y_t X_u]$$

If the correlation matrix

$$\begin{bmatrix} R_X(t,u) & R_{XY}(t,u) \\ R_{YX}(t,u) & R_Y(t,u) \end{bmatrix}$$

can be written as

$$\begin{bmatrix} R_X(t-u) & R_{XY}(t-u) \\ R_{YX}(t-u) & R_Y(t-u) \end{bmatrix}$$

then the random processes $\{X_t\}$ & $\{Y_t\}$ are each wide-sense stationary & also jointly wide-sense stationary. (Also, $m_X(t) = m_X$ & $m_Y(t) = m_Y$).

$$R_{XY}(\tau) = R_{YX}(-\tau)$$

— Examples to be added

Time averages and ergodicity:

If the theory of random processes is to be useful, we have to be able to estimate quantities such as the mean & autocorrelation function from observations of a random process.

In many practical situations, the only thing that might be available is the recording of one (or a small number) of sample functions of the random process.

Lecture 23:

For a stationary process,

$$m_X = E[X_t] = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

$$R_X(\tau) = E[X_t X_{t+\tau}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{X_t, X_{t+\tau}}(x,y)\,dx\,dy$$

Ensemble averaging is needed to determine m_X & $R_X(\tau)$.

What can we do with time averages of individual sample functions of a random process?

Time-averaged mean value: of the sample function of X_t is

$$\langle X_t(s)\rangle = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} X_t(s)\,dt$$

Time-averaged autocorrelation function: of the sample function of X_t is

$$\langle X_t(s) X_{t+\tau}(s) \rangle = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X_t(s)\, X_{t+\tau}(s)\, dt$$

$\langle X_t(s) \rangle$ & $\langle X_t(s) X_{t+\tau}(s) \rangle$ are random variables.

In general, ensemble averages and time averages are not equal.

A random process X_t is said to be ergodic (in the most general form) if all of its statistical properties can be determined (with probability one) from any one sample function.

It is necessary for a random process to be stationary in the strict sense for it to be ergodic. However, not all stationary processes are ergodic.

Ergodicity of the mean: (ergodicity in a limited sense).

A random process X_t is ergodic in the mean if

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X_t(s)\, dt = m_X$$

with probability one, i.e., its time-averaged and ensemble-averaged mean values must be equal with probability one.

↳ (Almost-sure convergence?)

The necessary and sufficient condition for the ergodicity of the mean is that the variance of $\frac{1}{2T}\int_{-T}^{T} X_t(s)\,dt$ approaches 0 as $T \to \infty$.

{For a stationary process,

$$E\left[\frac{1}{2T}\int_{-T}^{T} X_t\,dt\right] = \frac{1}{2T}\int_{-T}^{T} m_x\,dt = m_x.$$

I.e, the time average provides an unbiased estimate of m_x. }

{If we have a r.v. X such that $E[X^2] = 0$, then $P[X \neq 0] = 0$.

Proof: Suppose that this is not true. We can find an $\epsilon > 0$ such that

$$P(|X| > \epsilon) = \int_{|x|>\epsilon} f_X(x)\,dx \neq 0.$$

$$\Rightarrow E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx \geq \int_{|x|>\epsilon} x^2 f_X(x)\,dx \geq \epsilon^2 \underbrace{\int_{|x|>\epsilon} f_X(x)\,dx}_{>0}$$

This a contradiction.

Therefore, if $E[X^2] = 0$, then $P[X \neq 0] = 0$. }

Ergodicity of the autocorrelation function:

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} X_t X_{t+\tau}\, dt = R_X(\tau) \text{ with probability } 1.$$

For a stationary process,

$$E\left[\frac{1}{2T}\int_{-T}^{T} X_t X_{t+\tau}\, dt\right] = R_X(\tau).$$

The necessary and sufficient condition for the ergodicity of the autocorrelation function is that the variance of $\frac{1}{2T}\int_{-T}^{T} X_t X_{t+\tau}\,dt \to 0$ as $T\to\infty$.

In order to test ergodicity of the mean, it suffices to know the mean m_X and the autocorrelation function $R_X(\tau)$ of the process. However, in order to test the ergodicity of the autocorrelation function, we have to know fourth-order moments of the process. Therefore, it is usually difficult to establish if a random process meets the conditions for the ergodicity of the mean & autocorrelation function.

In practice, we are usually forced to consider the physical origin of the random process, and thereby make a somewhat intuitive judgement as to whether it is reasonable to interchange time & ensemble averages.

Example:

$$X_t = A\cos(2\pi f_c t + \theta)$$

A, f_c are constants

$\theta \sim$ uniform $[0, 2\pi]$.

$$m_X(t) = \int_{-\infty}^{\infty} A\cos(2\pi f_c t + \theta)\, f_\theta(\theta)\, d\theta$$

$$= 0.$$

$$R_X(\tau) = \frac{A^2}{2}\cos 2\pi f_c \tau.$$

$$\langle X_t(s) \rangle = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} A\cos(2\pi f_c t + \theta(s))\, dt$$

$$= 0$$

$$\langle X_t(s)\, X_{t+\tau}(s) \rangle = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} A^2 \cos(2\pi f_c t + \theta)\cos(2\pi f_c(t+\tau)+\theta)\, dt$$

$$= \lim_{T\to\infty} \frac{A^2}{4T}\int_{-T}^{T} [\cos 2\pi f_c \tau + \cos(4\pi f_c t + 2\pi f_c \tau + 2\theta)]\, dt$$

$$= \frac{A^2}{2}\cos 2\pi f_c \tau$$

$\Rightarrow$ This random process is ergodic in both the mean and autocorrelation function.