Page: ________
Date: / /

Lec-1

Hypothesis testing

$$\begin{tikzcd}
\to \text{Hypothesis} \arrow[r, shift left, "\text{experiment}"] \arrow[d] & \text{Data} \arrow[l, shift left, "\text{inference}"] \arrow[d] & \\
\text{Translate Hypothesis} \arrow[r, leftrightarrow] \arrow[d] & \text{Plot} \arrow[r] & \text{Inference} \\
\text{Verify} & & (\text{Quant} \to \text{Qualitative})
\end{tikzcd}$$

→ Regression towards the mean

Classification: Validation, Cross validation, K-fold validation, Bootstrapping.

Estimation: Biased or not, confidence interval.

K-fold validation — Train the model K times with different selection of the data and find average of all occurances.

No. of samples required to get a close attraction of parameter

Bias or variance of estimators.

~~→Pro~~ Contents

→ Prob. stat. review
→ Stat. measurements tests
→ Stats using R
→ using Matlab
→ Linear Regression
→ Hypothesis testing
→ Resampling techniques and Bootstrapping.

Grading

Midsem :- 20%

Endsem :- 40%

Programming Assignment :- 40% (No lab exam)

(90% attendance req.)

Book :- Intro. to Mathematical Statistics → V. Hogg, T. Craig ...

Page: ____
Date: / /

Lec-2

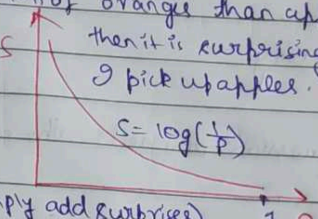
Random variable (X) :- ① Continous R.V ② Discrete R.V

X : $\Omega \rightarrow \mathbb{R}$

Q)- $f_{x_1 x_2} = \begin{cases} 8 x_1 x_2^2, & 0 < x_1 < x_2 < 1 \\ 0, & o/w \end{cases}$

If, there is a group A with lots of orange than apples then it is surprising if I pick up apples.

\# surprise (S) $\propto \frac{1}{\text{Probability (P)}}$



surprise(S) = $\hat{I}(p) = \log_2(1/p) = -\log_2 p$ → (can simply add surprises)

(surprise per item) $H\,[I(x)] = E[I(P)] = \sum p \log(1/p) = -\sum p_i \log(p_i)$

E(surprise)

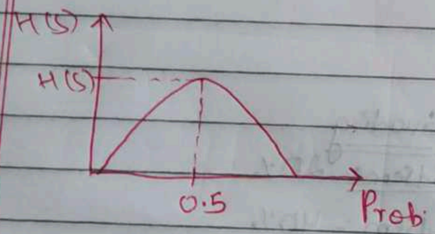
$= -\int p_i(x) \log\,(p_i(x))\,dx$

(Entropy) → used to quantify similarities & dissimilarities.

Ex:- H, T, $P(H) = 0.9$, $P(T) = 0.1$

⇒ $\sum -p_i \log(p)$ → find

Ex:- $H = T = 0.5$



\# fixed length coding, variable length coding.

\# Generally, the no. of occurances of a event A is more than B. then we keep less bit for A than B.

AAAB → 1110
keeping less bit → 10

\# This is generally variable length coding.
better than fixed length coding.

Page:
Date: / /

Prob.

$P(A) = 0.4$, $P(B) = 0.1$, $P(C) = 0.3$, $P(D) = 0.1$, $P(E) = 0.1$

$2^3 = 8$ — — — | $= \Sigma\, \ln(x_k) \cdot P(x_k)$
$= \Sigma\, 3 \cdot P(x_k) = 3[0.4 + 0.1 + \cdots + 0.1] = 3.$

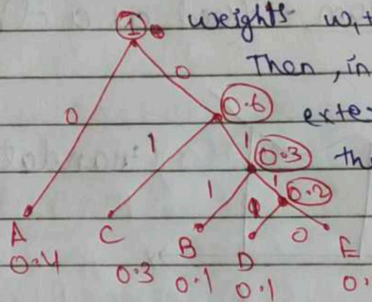
★ Prefix Coding.

Here, weight is prob.

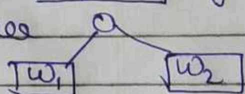
A	B	C	D	E
0.4	0.1	0.3	0.1	0.1

Bottom up approach)

$A = 0$
$B = 011$
$C = 01$
$D = 0111$
$E = 0110$



→ Huffmann Coding Rule (Algorithm) :- Suppose w_1 & w_2 are two minimum weights among the n given weights $w_1, w_2, \cdots, w_n$. Find a tree T' which gives a soln for the (n-1) weights $w_1+w_2, w_3, w_4, \cdots, w_n$. Then, in the tree T', replace the external node $\boxed{w_1+w_2}$ by the subtree

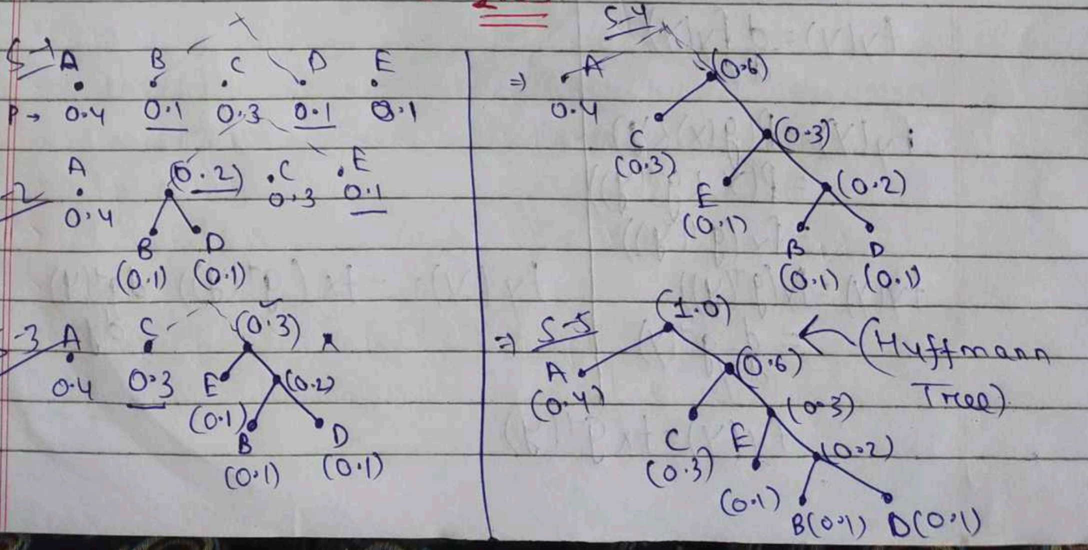


The new 2-tree T is the desired soln.

$\Sigma\, d \times P(i) = 1\times0.4 + 2\times0.3 + 3\times0.1 + 4\times0.1 + 4\times0.1$
(distance from 1) $= \underline{\underline{2.10}}$

$H(C) = -\Sigma\, P_i \log(P_i) = -[0.4\log(0.4) + 0.3\log(0.3) + 3\cdot0.1\log 0.1]$
$= \underline{\underline{2.05}}$

Ex:-



Page: ____ Date: / /

Lec-3

Rand, Randn – distribution.

Random

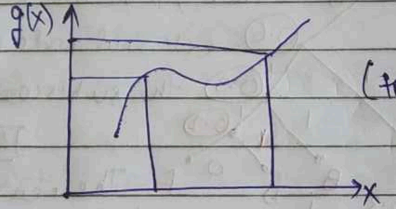
How to generate random no. having specific dist.?

↳ generally Gaussian dist.

Transformation of R.V → (one-one) mapping

$f_Y(Y) = f_X(g^{-1}(Y)) \left| \frac{dg^{-1}(Y)}{dY} \right|$ $Y = g(X)$, $X = g^{-1}(Y)$

Pseudo random → (seed as a clk freq.)



(translated)

One-one mapping

increasing

$f_Y(Y) = \frac{d}{dY} F_Y(Y)$

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

$f_Y(Y) = f_X(g^{-1}(y))$
$= \frac{d}{dy} g^{-1}(y)$

decreasing

$f_Y(Y) = -f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy}$

$f_Y(Y) = f_X\, g^{-1}(y)$

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d\, g^{-1}(y)}{dy} \right|$$

Ex:- $y = x^p$

Special Case:-

$y = g(x) = F_X(x)$ | $f_X(g^{-1}(y)) \cdot \frac{d g^{-1}(y)}{dy}$

$y = g(x) = \int_{-\infty}^{x} f_X(x)\,dx$ | $f_X(x) \cdot \frac{dx}{dy} \Big| x = g^{-1}(y)$

$\boxed{\frac{1}{f_X(x)} \cdot f_X(x) = 1} \rightarrow$

when we take $y = g(x) = F_X(x)$, we get always uniform distribution.

$\boxed{f_X(x) = 1}$

Ex:- $f_X(x) = \begin{cases} \frac{2L}{(x-1)^2}, & 0 \le x \le L+1 \\ 0, & o/w \end{cases}$

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx = 2L \int_{-\infty}^{x} \frac{1}{(x-1)^2}\,dx = \frac{2L}{1-x}$$

when the transformation func. is cdf then we get always $y = F_X(x) \rightarrow$ (uniform distribution)

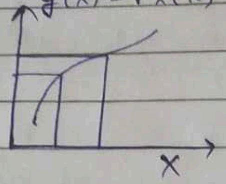
$F_X(x) = \frac{x^2}{(L-1)^2} \Rightarrow, \frac{(L-1)^2}{(L-1)^2} = 1.$

$\frac{\cancel{f_X(x)}}{\sqrt{Y(L-1)^2}}$ $f_X(g^{-1}(y))$ do

Page: ____ Date: / /

To achieve uniform distribution we take $Y = g(x) = F_X(x)$

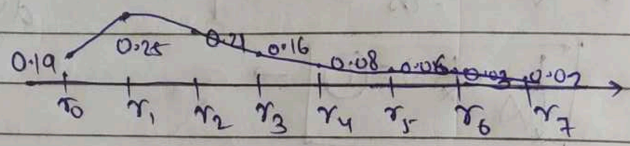
$g(x) = F_X(x)$



To get $f_X(x)$ or any specified distribution
we reverse transform $g(x) = F_X(x)$.

Discrete type (we do approximation)

$r_0 = 0.19$
$r_1 = 0.25$
$r_2 = 0.21$
$r_3 = 0.16$
$r_4 = 0.08$
$r_5 = 0.06$
$r_6 = 0.03$
$r_7 = 0.02$



$S = T(r)$

$S(K) = \sum_{i=0}^{K} P(r_i)$ length $= L = 8$.

$S(K) = (L-1) \sum_{i=0}^{K} P(r_i)$

(to normalise)

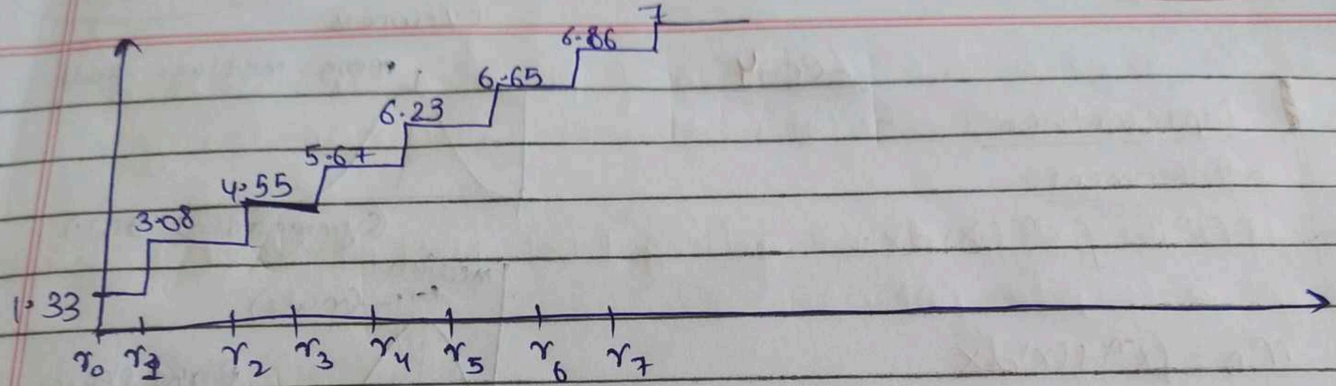
Approx

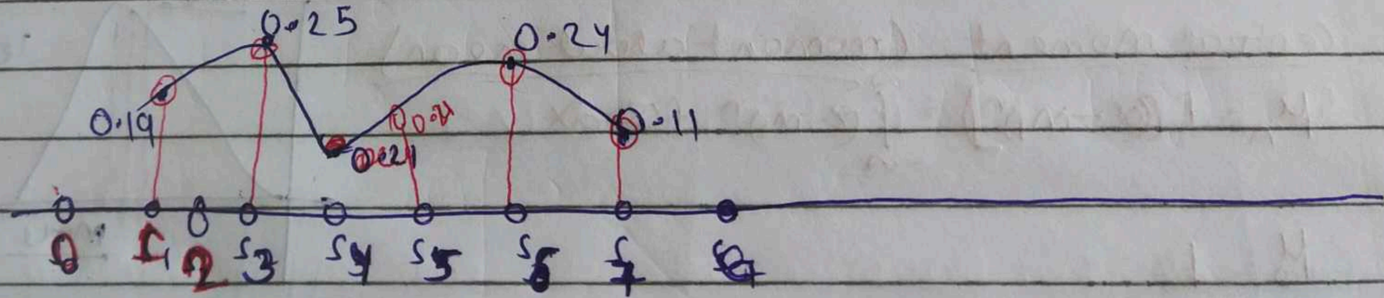
$S(0) = 7 \times P(0) = 7 \times 0.19 = 1.33 \sim 1$
$S(1) = 7 \times \ldots = \quad = 3.08 \sim 3$
$S(2) = \quad = 4.55 \sim 5$
$S(3) = \quad = 5.67 \sim 6$
$S(4) = \quad = 6.23 \sim 6$
$S(5) = \quad = 6.65 \sim 7$
$S(6) = \quad = 6.86 \sim 7$
$S(7) = \quad = 7 \sim 7$

x(x)

ribution

x).





Lec-4

Lec-4

Representation

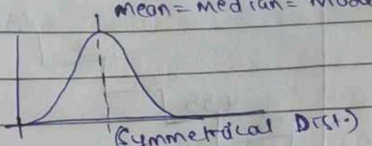
nth moment:-

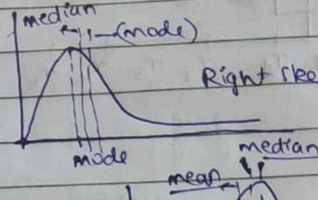
$m_n = E[x^n] = \int x^n f(x)\,dx$.

$m_n = \int x^n f(x)\,dx$

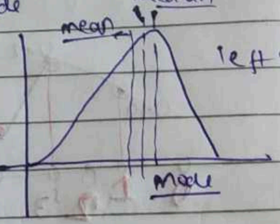
$m_0 = 1,\ m_1 = E[x],\ m_2 = E[x^2]$

Skewness





Right skew.



left skew.

Central moment (moment about mean)

$\mu_n = E[(x-m_1)^n] = \int_{-\infty}^{\infty} (x-m_1)^n f(x)\,dx$

$\mu_0 = 1$

$\mu_1 = 0$

$\mu_2 = \mathrm{var}(x)$

$\mu_3 = E[(x-m_1)]^3 \rightarrow$ skewness.

$\mu_4 = E[(x-m_1)^4] \rightarrow$ Kurtosis. $\mu_4 > 0 \rightarrow$ super Gaussian, $\mu_4 < 0 \rightarrow$ sub Gaussian.

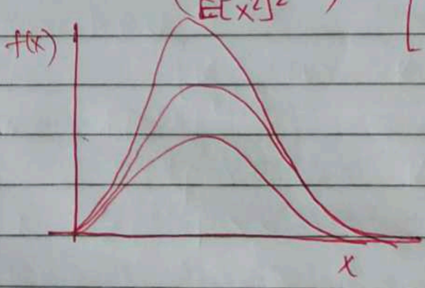
ex:- wealth dist. → (+ve)

Age, wt, Irith. → (0)

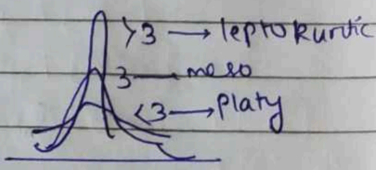
life span of Human ↳ (-ve)

$\mathrm{Kurt}(X) = \left(\dfrac{E[x^4]}{E[x^2]^2} - 3\right)$

[(μ_3, μ_4) — used to determine Non-Gaussianity]



Kurtosis



Characteristics function

$X,\ f_X(x)$

$\phi_X(j\omega) = \int_{-\infty}^{\infty} e^{-j\omega x} f_X(x)\,dx$

$$\boxed{\phi_X(0) \geq \phi_X(j\omega)}$$

Cummulant

(Read Fourier & Laplace Transform)

Page:
Date:

$j\omega = s$

$\phi_x(s) = \int e^{sx} f_x(x)\,dx$ ★ (moment generating func^n)

$\phi_x(s) = E[e^{sx}]$

$\phi_x(s) \xrightarrow{L} f_x(x)$

Moment Theorem

$$\frac{\partial(\phi_x(s))}{\partial s} = \frac{\partial}{\partial s}\int_{-\infty}^{\infty} e^{sx} f_x(x)\,dx = \int x \cdot e^{sx} f_x(x)\,dx$$

$$\phi'_x(s) = \int x \cdot e^{sx} \cdot f_x(x)\,dx$$

$$\phi^n_x(s) = \int x^n \cdot f_x(x) \cdot e^{sx}\,dx$$

$$\phi^n_x(0) = \int x^n \cdot f_x(x)\,dx = E[x^n] = m_n \rightarrow \text{(nth central moment)}$$

So, as it generates moment so, that's why its called ~~MGF~~ MGF

$$\phi_x(s) = \int_{-\infty}^{\infty} e^{sx} \cdot f_x(x)\,dx$$

$$e^z = \sum_{n=0}^{\infty} z^n/n!$$

$$\phi_x(s) = \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} (sx)^n/n! \cdot f_x(x)\,dx$$

$$= \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} x^n f_x(x) \cdot \left(\frac{s^n}{n!}\right) dx$$

$$\phi_x(s) = \sum_{n=0}^{\infty} E[x^n] \cdot \frac{s^n}{n!}$$

$$f_x(x) \longrightarrow \phi_x(s) \longrightarrow E[x^n]$$

ϕ_x

$\Sigma e^{tx} \cdot p^{x}(1-p)^{1-x}$
$\Sigma (pe^{t})^{x}(1-p)^{1-x}$

$e^{a} \times e^{b} = e^{a+b}$

Page: ____
Date: / /

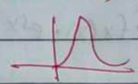
So, to get a complete idea of pdf of a func, we need all $E[X^n] \to E[X], E[X^2] \ldots$ to get an idea of pdf.

Don't say that after getting $E[X]$, $E[X^2]$ or $Var[X]$, we get ~~a~~ the ~~pdf~~ PDF.

It may happen that 2 pdfs having same PDF, $f_X(x)$ & $f_Y(y)$, (let).

So, find all $E[X^n] \to E[X], E[X^2], E[X^3] \ldots$

~~Cummul~~ For Gaussian dist., we can approximately say that by getting $E[X]$ & $E[X^2]$ or $Var(X)$, we get approximately correct distn.



Cummulant :- $\ln[\phi_X(\omega)] = \log_e[\phi_X(\omega)]$

$K_X(t) = \log_e M_X(t)$

Properties

1. Additive property →

$X_1 + X_2 + \ldots + X_n$

$$\boxed{K_{\Sigma X_i}(t) = K_{X_1}(t) + K_{X_2}(t) + \ldots + K_{X_n}(t)}$$

Discrete RV

MGF

$M_X(t) = E[e^{tX}]$

$= \Sigma e^{tx} p(x)$

$= \Sigma e^{tx} p(x)$.

$M'_X(t) = \Sigma e^{tx} \cdot x\, p(x) \Rightarrow M'_X(0) = E(X)$

$\vdots$

$M^n_X(t) = \Sigma e^{tx} \cdot x^n p(x) \Rightarrow M^n_X(0) = E[X^n]$.

$K_{\Sigma X_i}(t) = \log_e M_{\Sigma X_i}(t)$

$= \log_e E[e^{t\Sigma X_i}]$

$= \log_e E[\Pi e^{tX_i}]$

$= \log_e (E[e^{tX_1}] \times E[e^{tX_2}] \times \ldots E[e^{tX_n}])$

$= \Sigma \log E[e^{tX_i}]$

$= \Sigma \log_e M_{X_i}(t)$

$= \Sigma K_{X_i} t$.

2. Effect of change of origin and scale on cgf

$U = \frac{X-a}{h}$, $K_U(t) = \log_e[M_U(t)] = \log_e E e^{tU} = \log_e E[e^{t(X-a/h)}]$

$\log_e E[e^{tX/h} \cdot e^{-ta/h}]$

$= \log_e E[e^{tX/h}] \cdot e^{-ta/h}$

$= \log_e M_X(t/h) \cdot e^{-ta/h} = \log_e e^{-ta/h} + K_X(t/h) = \boxed{K_X(t/h) - ta/h} = K_Y(t)$

Page:
Date: / /

$M_X(t) = \sum_{n=0}^{\infty} e^{tx} p(x)$

$= \sum\left(\sum \frac{(xt)^n}{n!}\right) p(x)$

$= \sum\left(\sum x^n p(x)\right) \cdot \frac{t^n}{n!}$

$M_X(t) = \sum_n \frac{E(X^n) \cdot t^n}{n!}$

Bernoulli

$p(x) = p^x (1-p)^{1-x} \quad x = 0, 1$

(Poisson
Gamma
Beta
Chi-Squared)

$E(X) = p$

$E(X^2) = p$

$E((X-\mu)^2) = p(1-p)$

$M_X(t) = \sum p^x (1-p)^{x} \cdot e^{tx} =$ ~~EP~~

Binomial

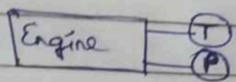
$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$

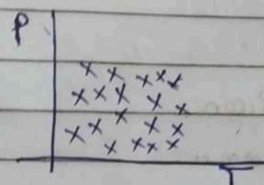
Lec-5

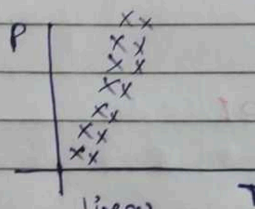
Recap?

(more than 1 variable)

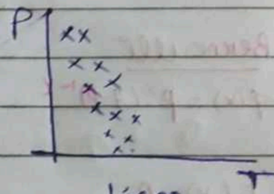
Multivariate Normal Distribution
- Standard Normal Distribution
- Gaussian Normal Distribution



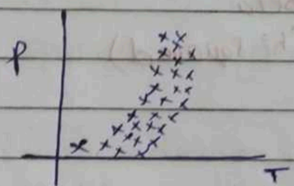




linear



linear



Non-linear

$$\begin{pmatrix} T \rightarrow X_1 \\ P \rightarrow X_2 \end{pmatrix}$$

$f(x_1, x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2$

$P(X_1, X_2)$

$$M_X(t) = E[e^{tX}]$$

$$M_{X_1 X_2}(t_1, t_2) = E[e^{t_1 X_1 + t_2 X_2}]$$

$t' = [t_1, t_2]^T$

$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \Rightarrow M_{X_1, X_2}(t_1, t_2) = E[e^{t'X}]$

Random Vector $\rightarrow X = [X_1, X_2 \cdots X_n]^T$

Page:
Date: / /

Six Sigma (6σ)

↳ ~~[illegible]~~ chances of error is v. v. least.
3 out of 1 million.

So, the company is called 6 sigma.

Ex:- Mumbai Dabbawala

Lec-6

Z

$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \quad (-\infty < z < \infty)$

$\int_{-\infty}^{\infty} f_Z(z)\,dz = 1. \Rightarrow \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) dz = 1$

$F_Z(z) = \int_{-\infty}^{z} f_Z(y)\,dy$, MGF $= M_Z(t) = E(e^{tZ})$

$M_Z(t) = \int_{-\infty}^{\infty} e^{tz} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) dz$

$(\alpha = z - t)$

$= \int_{-\infty}^{\infty} e^{t(\alpha+t)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\alpha+t)^2}{2}\right) d\alpha$

$= \int_{-\infty}^{\infty} e^{t(\alpha+t)} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-\alpha^2}{2}\right) \times \exp\left(\frac{-t^2}{2}\right) \times \exp(-\cancel{\alpha} t)$

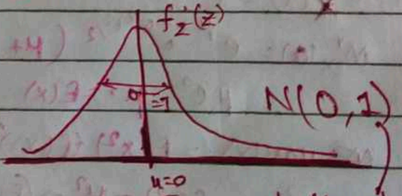
$\int_{-\infty}^{\infty} e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\alpha^2}{2} - \frac{\cancel{t^2}}{2}\right) d\alpha$

$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t^2/2} \cdot \exp\left(-\frac{\alpha^2}{2}\right) d\alpha = \frac{e^{t^2/2}}{\cancel{\sqrt{2\pi}}} \times \cancel{\sqrt{2\pi}} = e^{t^2/2}$

$M_Z(t) = e^{t^2/2}$, $M'_Z(t) = t \cdot e^{t^2/2}$, $M''_Z(t) = t^2 e^{t^2/2} + e^{t^2/2}$

$M_Z(0) = 1$, $\underbrace{M'_Z(0) = 0}_{= E[Z]}$, $\underbrace{M''_Z(0) = 1.}_{= E(Z^2)}$

$Var(Z) = E(Z^2) - (E(Z))^2 = 1.$



★ Standard Normal distribution

$X = bZ + a \Rightarrow Z = (X-a)/b, \left(\frac{dZ}{dX} = \frac{1}{b}\right)$

$y = g(x).$

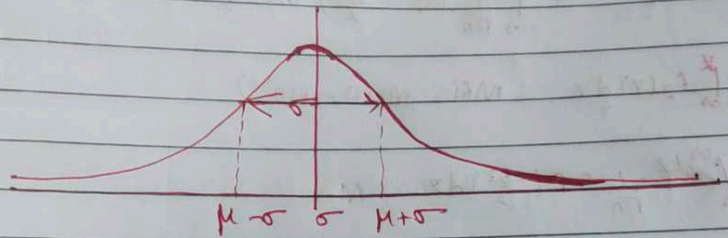
$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d\,g^{-1}(y)}{dy} \right|$

$f_X(x) = \frac{1}{\sqrt{2\pi}\,b} \exp\left(-\frac{1}{2}\right)\left(\frac{x-a}{b}\right)^2$

$E[X] = a = \mu.$

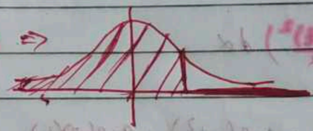
$Var(X) = E[X^2] - (E[X])^2$

$= b^2 = \sigma^2$



$N(\mu, \sigma^2) \longrightarrow$ Normal Distribution.

$P_0(X \le x_0) \Rightarrow$



$X = bZ + a$.

$X = \sigma Z + \mu$.

$M_X(t) = E(e^{tX}) = E[e^{t(\sigma Z + \mu)}] = e^{t\mu} \cdot E[e^{t\sigma Z}]$

$\int_{-\infty}^{\infty} e^{t\sigma z} \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\left(\frac{-\sigma^2 z^2}{2} - \frac{\mu^2}{2} + \sigma\mu z\right) dz$ $\quad \int_{-\infty}^{\infty} e^{t\sigma z} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-\sigma^2 z^2 - \mu^2 + 2\sigma\mu}{2}\right) dz$

$= e^{\sigma^2 t^2/2}$

$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$

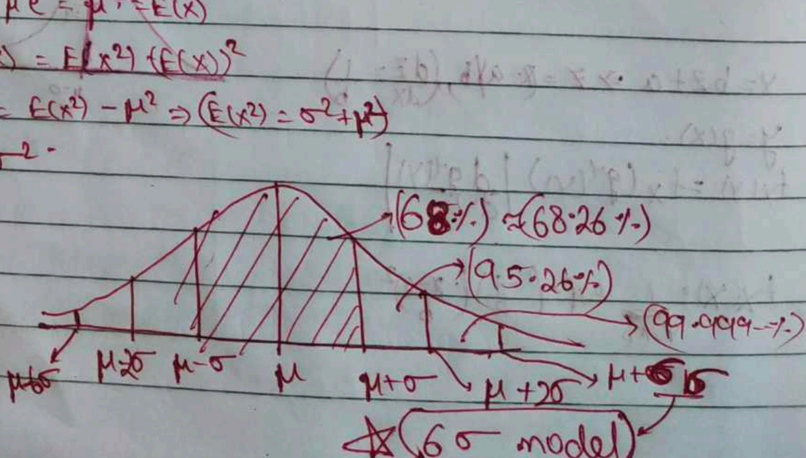
$M'_X(0) = e^{\mu t + \sigma^2 t^2/2} \left(\mu + \frac{\sigma^2 t}{2}\right)$

$\mu e^0 = \mu = E(X)$

$Var(X) = E(X^2) - (E(X))^2$

$\sigma^2 = E(X^2) - \mu^2 \Rightarrow E(X^2) = \sigma^2 + \mu^2$

$Var(X) = \sigma^2$.



Jeffwelth

CMM level-1
level-2
⋮

Page:
Date: / /

ISO.

$Z \sim N(0,1)$

$X = \sigma Z + \mu \sim N(\mu, \sigma^2)$

$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

(# so, going from Z to X and then X to Z.)

$X \sim N(\mu, \sigma^2)$ (Prove it)?

$V = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2_1$

$V = W^2$ where $W = \left(\frac{X-\mu}{\sigma}\right) \Rightarrow (X = \sigma W + \mu)$ $W \sim N(0,1)$

Chi-Squared Distribution

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(r/2)\, 2^{r/2}} \cdot x^{r/2-1} \cdot e^{-x/2} & 0 \le x < \infty \\ 0 & , \ o/w. \end{cases}$$

~~$f_W(w) = F$~~

CDF of V $= G(v) = P_r(V \le v) = P_r[W^2 \le v]$ ★(Solve it & prove it) where $r=1$.

$X_1, X_2, X_3 - - - X_n$, $X_i \sim N(\mu_i, \sigma_i^2)$

$S = \sum_{i=1}^{N} K_i X_i$, $M_S(t) = E[e^{tS}] = E[e^{t\sum_{i=1}^{N} K_i X_i}]$

$= \exp\left(t\sum_{i=1}^{n} K_i \mu_i + \frac{t^2}{2}\sum_{i=1}^{n} \sigma_i^2 K^2\right)$

★ Book, Ch-3

Lec-7

Multivariate Normal Distribution

$X = bZ + a$

$Z \sim N(0,1)$

$X \sim N(\mu, \sigma^2)$ $\boxed{X = \sigma Z + \mu}$

$Z \rightarrow X$	$X_1, X_2, \ldots X_n \rightarrow iid$
$Z \leftarrow \left(\frac{X-\mu}{\sigma^2}\right)$	$M_{\underline{X}}(t) = \Sigma M_{X_i}(t)$

$Z = (Z_1, Z_2, \ldots Z_n)$

Pdf = ___

mgf = ___

$E(XY) = \frac{\partial^2 m_{(\underline{x})}(0,0)}{\partial t_1 \partial t_2} = \rho\sigma_1\sigma_2 + \mu_1\mu_2$

$Cov(X,Y) = E(XY) - E(X) \cdot E(Y)$

$= \rho\sigma_1\sigma_2 + \mu_1\mu_2 - \mu_1\mu_2$

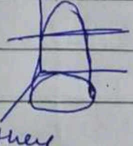
$= \rho\sigma_1\sigma_2$

If $\rho = 0 \Rightarrow cov = 0 \Rightarrow$ uncorrelated.

$\boxed{M_{\underline{X}}(t_1, t_2) = M_{X_1}(t_1) * M_{X_2}(t_2)}$

Contour $(f(x,y) = c)$.

↳ when we chop the top we get a contour, which is a ellipsoid if they are related else circle if they are independent and pdf at every points of contour is same.



$Z_1, \dots Z_n$.
$Z_i \sim N(0,1)$
$Z_1, \dots Z_n$ are independent.

$$f(\underline{Z}) = \prod_{i=1}^{n} f(Z_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{Z_i^2}{2}\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(\frac{-\sum Z_i^2}{2}\right)$$

$$= \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{Z^T Z}{2}\right)$$

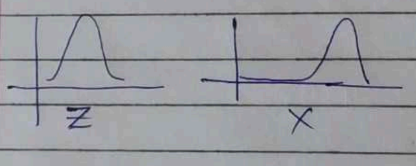
inner product → single value
outer " → vector of value

$$E(Z) = \begin{bmatrix} E(Z_1) \\ E(Z_2) \\ \vdots \\ E(Z_n) \end{bmatrix} \qquad \operatorname{cov} Z = E\left[(Z-\mu)(Z-\mu)^T\right]$$

$$\mu = E(Z)$$

$$E(Z Z^T) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} = I_n.$$

see visually what will happen, if we transform linearly.



$$X \leftarrow \Phi Z$$
$(R^{n \times n})$

$\Sigma \in \mathbb{R}^{n \times n}$
└→ (positive definite matrix)

Matrix decomposition

$\Sigma = S^T \Lambda S$ $\quad (SS^T = I)$

$$S = \begin{bmatrix} | & | & & | \\ s_1 & s_2 & \cdots & s_n \\ | & | & & | \end{bmatrix}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix} \quad \boxed{\lambda_1 > \lambda_2 > \lambda_3 \cdots > \lambda_n}$$

Page:
Date: / /

$\Lambda^{1/2} = \text{diag}(\sqrt{\Lambda_1}, \sqrt{\Lambda_2} \ldots \sqrt{\Lambda_n})$

$\Sigma = (S^T \Lambda^{1/2} S) \cdot (S^T \Lambda^{1/2} S)$.

$= \boxed{S^T \Lambda S}$

$\boxed{\Sigma^{1/2} = S^T \Lambda^{1/2} S}$

$X = bZ + \mu$

$\boxed{X = \Sigma^{1/2} Z + \mu}$

$E(X) = \Sigma^{1/2} E(Z) + \mu$.

$= \mu \quad$ as $E(Z) = 0$.

$\text{cov}(X) = E[(X-\mu)(X-\mu)^T]$

$= \Sigma$.

$\cancel{E(X) = \sum x f(x)}$

$\cancel{= \sum (\Sigma^{1/2} Z + \mu)}$

$(X-\mu) = \Sigma^{1/2} Z$

$(X-\mu)^T = Z^T \Sigma^{1/2}$

$= \Sigma$.

missed

if it is a diagonal matrix then, what will happen?.

$N(\mu, \Sigma)$

$M_X(t) = \exp\{t^T\mu + \frac{1}{2} t^T \Sigma t\}$.

Transformation

$Y = g(X)$

$f_Y(Y) = f_X(g^{-1}(Y)) \left| \frac{d g^{-1}(Y)}{dY} \right|$

Jacobian matrix.

$f_Z = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{Z^T Z}{2}\right)$

$X = \Sigma^{1/2} Z + \mu$

$\boxed{(X-\mu)\Sigma^{-1/2} = Z}$

$f_X(X) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\left((X-\mu)^T \Sigma^{-1/2}\right)\left(\Sigma^{-1/2}(X-\mu)\right)\right\}$

Page: ______
Date: / /

Page: ______
Date: / /

For contour,

$(X-\mu)^T \Sigma^{-1} (X-\mu) = c^2$, $c>0$.

$(\Sigma^{1/2} Z)^T \Sigma^{-1} (\Sigma^{1/2} Z)$

$Z^T \Sigma^{+1/2} \Sigma^{-1} \Sigma^{+1/2} Z = c^2$

$Z^T \Sigma^{0} Z = c^2 \Rightarrow \boxed{c^2 = Z^T Z}$ $\boxed{c^2 = \sum_{i=1}^{n} Z_i^2}$ $\sim \chi^2_n$.

$Z_i \sim N(0,1)$

$\left(\frac{Z_i - 0}{1}\right)^2 \sim \chi^2_1$.

$\boxed{\Sigma(Z_i^2) \sim \chi^2_n}$

$P\left[(X-\mu)^T \Sigma^{-1} (X-\mu) \leq c^2\right]$

$P[Y \leq c^2] = F_{\chi^2_n}(c^2)$