Probability & statistics (II)

Dr Mervat Mikhail

Lecture Notes

Fall 2022

Lec (1)

Revision

<u>Sample space</u> : the set of all Possible outcomes of a random experiment

Event : is a subset of sample space S
A

$$P(A) = \frac{N(A)}{N(S)}$$

A, B are indep when $P(A \cap B) = P(A)P(B)$

A, B are disjoint when $P(A \cap B) = \phi$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

if A, B disjoint $P(A \cup B) = P(A) + P(B)$

<u>Conditional Probability</u>

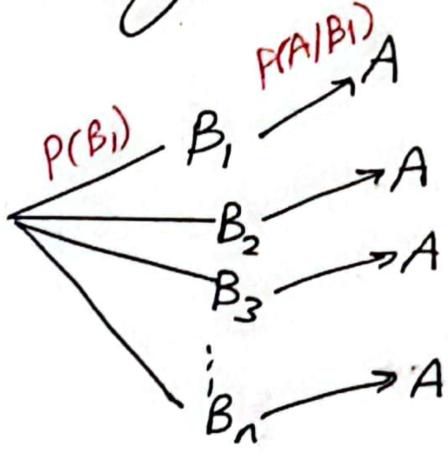
$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \quad P(B/A) = \frac{P(A \cap B)}{P(A)}$$

$$\therefore P(A \cap B) = P(A/B)\,P(B) = P(B/A)\,P(A)$$

if A, B indep $\begin{cases} P(A/B) = P(A) \\ P(A \cap B) = P(A)P(B) \end{cases}$

Total Probability Rule (Bayes' Theorem)

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + \cdots + P(B_n)P(A/B_n)$$



Random Variable

Is a f^n that assocaate a real number with each element in Sample space

Discrete

$X = x_1, x_2, \ldots$

e.g. a coin is tossed twice

X : No of appearing heads

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

$x = \{0, 1, 2\}$

PMF
Probability mass function

$P(x_i) = P(X = x_i)$

$$PMF = \begin{cases} P(0) = 1/4 \\ P(1) = 2/4 \\ P(2) = 1/4 \end{cases}$$

$\sum P(x_i) = 1$

Continuous

$a \leq X \leq b$

e.g. length of a product between 2cm to 3cm

PDF
Probability density function

$f(x) \geq 0$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$F(b_1) - F(a_1) = P(a_1 \leq X \leq b_1) = \int_{a_1}^{b_1} f(x)\,dx$$

$P(X = a_1) = 0$

CDF Cumulative distr. fn

discrete

$F(x) = P(X \leq x)$

$= \sum_{x_i \leq x} P(x_i)$

Cont.

$F(x) = \int_{-\infty}^{x} f(x)\,dx$

Mean (Expectation) average $E(x) = \mu$

discr

$E(x) = \sum x_i f(x_i)$

Cont

$E(x) = \int_{-\infty}^{\infty} x\, f(x)$

if c is a Constant

$E(c) = c$

$E(c\, g(x)) = c\, E(g(x))$

$E(ax+b) = a\, E(x) + b$

$E(g_1(x) + g_2(x)) = E(g_1(x)) + E(g_2(x))$

Variance σ^2 Standard deviation $= \sigma$

$V(x) = E(x-\mu)^2 = E(x^2) - \mu^2$

Mean square deviation

Deviation $\longrightarrow$ $(x_1-\mu)\,(x_2-\mu)\,(x_3-\mu)\ldots(x_n-\mu)$

if c is a constant

$V(c) = 0$, $V(c\, g(x)) = c^2 V(g(x))$

$V(ax+b) = a^2\, V(x)$ e.g $V(2X-5) = 4\, V(X)$

MGF Moment generating Functions

$$M_x(t) = E(e^{xt})$$ f^n of t

disr

$$M_x(t) = \Sigma\, e^{xt} f(x)$$

Cont

$$M_x(t) = \int_{-\infty}^{\infty} e^{xt} f(x)\, dx$$

why we use MGF ?

$$\mu_r = \left. \frac{d^r M_x(t)}{dt^r} \right|_{t=0}$$

$\therefore\; M'_x(0) = E(x) = \mu_1$

$M''_x(0) = E(x^2) = \mu_2$

$\vdots$

$$\boxed{\mu_r = E(x^r)}$$

If pdf is not easy to get
You can use MGF to get the mean
and variance

$$V(x) = E(x^2) - \mu^2 = \mu_2 - \mu_1^2$$

Special discrete Probability distributions

1 Bernoulli Trial

random experiment with two Possible outcomes success or fail

Prob of success $= P$

" " failure $= 1 - P = q$

$$f(x) = \begin{cases} P & x = 1 \\ 1-P & x = 0 \end{cases}$$

PMF

$E(x) = \mu = P$

$V(x) = Pq$

2 Binomial experiment

Consists of n trial such that

1- trials are indep

2- each trial result in only two possible outcomes success or fail (Bernoulli trial)

3- Prob of success in each trial $= p$

$X \sim Bin(n, p)$ X: No of successes in n trials

$f(x) = {}^{n}C_{x} \, p^{x} \, q^{n-x}$ $x = 0, 1, 2, \dots n$

Mean $\mu = np$

Variance $\sigma^2 = npq$

$M_x(t) = (pe^{t} + q)^{n}$ Check $\Rightarrow M'(0) = np$

where $q = 1 - p$ Prob of failure

Check

$M'(t) = n(pe^{t} + q)^{n-1} pe^{t}$

put $t = 0$ $M'(0) = n(p+q)^{n-1} p$

$\because p + q = 1 \Rightarrow M'(0) = np = E(X)$

this is why $M_x(t)$ is very important to get $\mu_1, \mu_2, \mu_3, \dots,$

3 Poisson distribution

when $n \to \infty$, $p \to 0$, $np \to \lambda$

$X \sim poiss(\lambda)$

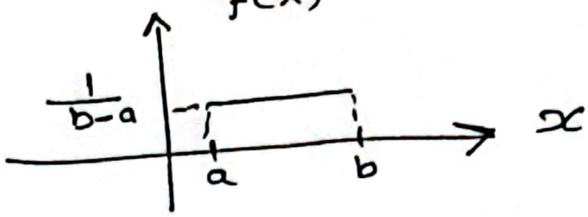
$$f(x) = \frac{e^{-\lambda} \lambda^{x}}{x!} \qquad x = 0, 1, 2, \dots$$

$E(x) = \lambda$

$V(x) = \lambda$

Special Cont. distribution

1 Uniform distr.



$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & o.w \end{cases}$$

$$E(x) = \frac{a+b}{2}$$

$$V(x) = \frac{(a-b)^2}{12}$$

2 Exponential distr.

If $X \sim Pois(\lambda)$

X: No of events per unit of time

then $T \sim exp(\theta)$

T: time to first Poisson event

or Interarrival time between successive Poisson events

$$\boxed{\theta = \frac{1}{\lambda}}$$

θ: Mean time

$$f(t) = \frac{1}{\theta} e^{-\frac{t}{\theta}} \qquad t > 0$$

$$Pr(T \leq t) = F(t) = 1 - e^{-\frac{t}{\theta}}$$

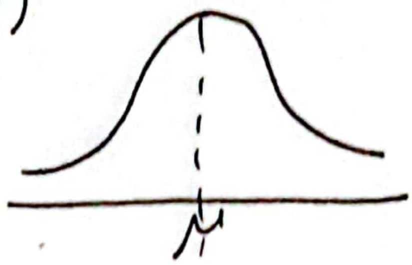
$$Pr(T \geq t) = 1 - F(t) = e^{-\frac{t}{\theta}}$$

$$Pr(t_1 \geq T \geq t_2) = F(t_2) - F(t_1)$$
$$= e^{-\frac{t_1}{\theta}} - e^{-\frac{t_2}{\theta}}$$

$$\mu = \frac{1}{\lambda}, \quad \sigma^2 = \frac{1}{\lambda^2}$$

3 Normal distribution

if $X \sim \text{Normal}(\mu, \sigma)$



$Z \sim \text{Normal}(0,1)$
standard normal

$P(z_1 < Z < z_2) = \Phi(z_2) - \Phi(z_1)$

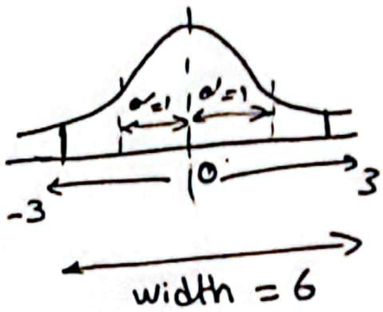
CDF of standard normal
(found in tables for +ve values only)

$P(Z < z_1) = \Phi(z_1)$

$P(Z > z_1) = 1 - \Phi(z_1)$

$Z = \frac{X-\mu}{\sigma}$ to standarize any normal R.V.

Properties of standard normal Curve



① $P(-3 < Z < 3) \simeq 1$

② $\Phi(0) = \frac{1}{2}$

③ $\Phi(\infty) = 1$, $\Phi(-\infty) = 0$
$\Phi(3) = 1$ $\quad \Phi(-3) = 0$

④ $\Phi(-z) = 1 - \Phi(z)$
$\Phi(-2) = 1 - \Phi(2)$ No -ve values of CDF in table

⑤ $P(a < Z < b) = \Phi(b) - \Phi(a)$

If More than one R.v

$\overbrace{\hspace{10em}}$

Discrete joint pdf

$f_{XY}(x,y) = P(X=x, Y=y)$
Table

Marginal PMF

$f_X(X) = \sum_Y f_{XY}(x,y)$

$f_Y(x) = \sum_X f_{XY}(x,y)$

Cont. joint pdf

$f_{XY}(x,y)$ is region

$f_X(x) = \int_y f_{XY}(x,y)\,dy$
vertical strip

$f_Y(y) = \int_x f_{XY}(x,y)\,dx$
horizontal strip

If X, Y are indep

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$$

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

Conditional = marginal

When X, Y are two R.V.

Covariance

$$Cov(X,Y) = E\big((X-\mu_x)(Y-\mu_y)\big)$$

$$= E(XY) - \mu_x \mu_y$$

$$= \sigma_{xy}$$

Cov: $+$: $Y > \mu_y$, $X > \mu_x$, or $X < \mu_x$, $Y < \mu_y$

$-$: $X > \mu_x$, $Y < \mu_y$ or $X < \mu_x$, $Y > \mu_y$

if x increas $\Rightarrow$ y increases $\Rightarrow \sigma_{xy}$ +ve

if x " $\rightarrow$ y decrease $\Rightarrow \sigma_{xy}$ -ve

Correlation

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)\,V(Y)}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \rho$$

$-1 \leq \rho \leq 1$

if $Cov(X,Y) = 0 \Rightarrow \rho(X,Y) = 0$

X, Y are said to be uncorrelated

* X, Y are said to be independent

iff $\underset{\text{joint pdf}}{f(x,y)} = f_x(x)\, f_y(y)$

Marginal pdf

$\rho = 1$ perfect +ve correlated

$\rho = -1$ perfectly -ve correlat

$\rho = 0$ No correlat

Regards

☺