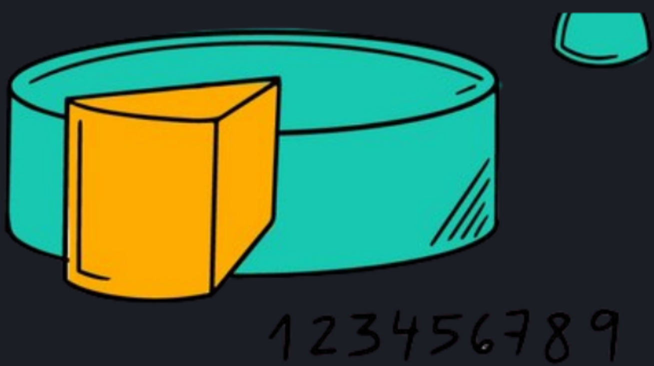
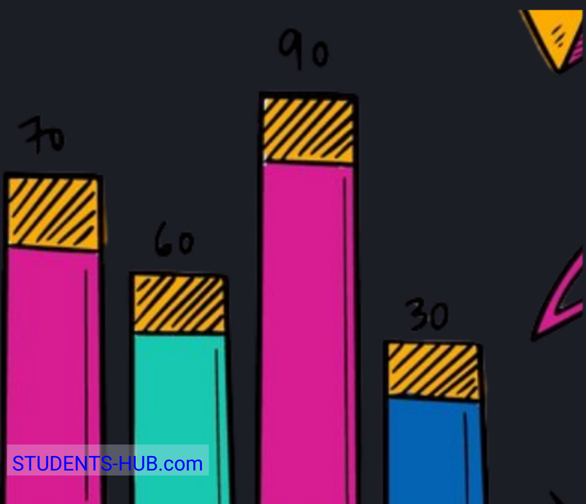
STATISTICS

By Rawan Alfares





1. Intersection $A \cap B$ means And $(\times)$
2. union $A \cup B$ means or $(+)$
3. Mutually Exclusive or disjoint $(A \cap B = \phi)$. $\longrightarrow p(A \cup B) = p(A) + p(B)$.
4. Complement A^c or $\bar{A}$

5. Demorgan's law : $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$
   $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$

6. $p(A) = \dfrac{\text{number of outcomes in event } A}{\text{number of total outcomes in } S}$

7. $1 \geq p(A) \geq 0$

$\Rightarrow$ Conditional Probability and statistical independence.

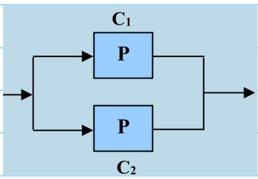
1. $p(A/B) = \dfrac{p(A \cap B)}{p(B)}$

2. $p(A \cap B) = p(A/B) \cdot p(B)$

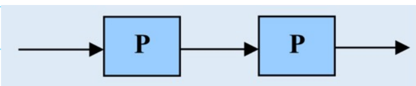
3. $p(A \cap B) = p(A) \cdot p(B) \longrightarrow$ A and B are statistically indep.

* if A and B are S. Indep, then:
1. $p(A/B) = p(A)$
2. $p(B/A) = p(B)$

to prove that A, B and C are S. Indep :-
1. $p(A \cap B) = p(A) \cdot p(B)$
2. $p(A \cap C) = p(A) \cdot p(C)$
3. $p(B \cap C) = p(B) \cdot p(C)$
4. $p(A \cap B \cap C) = p(A) \cdot p(B) \cdot p(C)$.



$p(\text{system works}) = p(C_1 \cup C_2) = p(C_1) + p(C_2) - p(C_1 \cap C_2)$
$= P + P - (p(C_1) \cdot p(C_2))$ Called Realiability of the system.
$= 2P - P^2$



$p(\text{system works}) = p(C_1 \cap C_2) = p(C_1) \cap p(C_2)$
$= P \cap P = P \cdot P = P^2$

Counting techniques :-

1. multiplication Rule :- operation A can be performed in n_1 different ways. and operation B can be performed in n_2 different ways. then the sequence (A, B) Can be performed in $(n_1 \times n_2)$ different ways.

الترتيب مهم / التكرار غير مسموح

2. Permutation :- A- Sampling without replacement :- $N = \dfrac{n!}{(n-R)!}$ احتمالية أمر معين بدون ترجيع

B- Sampling with Replacement :- $N = n^R$ عدد مرات التجربة

عدد الاحتمالات

الترتيب مهم التكرار مسموح

3. Combination :- الترتيب غير مهم، التكرار غير مسموح.

$\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$, n choose k = (k, n)

Random Variable

- Continuoes
  - PDF: $f_X(x)$
  - CDF: $F_X(x)$
- discreate
  - PMF: $P(X=x_i)$ " عند نقطة معينة "
  - CDF: $P(X \leq x) = F_X(x)$ " النسبة التراكمية "

* $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
* $F_X(x) = \int_{-\infty}^{\infty} f_X(x)\,dx$
* $f_X(x) = \frac{dF_X(x)}{dx}$
* $P\{x_1 \leq X \leq x_2\} = \int_{x_1}^{x_2} f_X(x)\,dx$ = Area under the Curve.

$P(a \leq x \leq b) = F_X(b) - F_X(a)$

---

* إذا كان عندي CDF وفيه مجاهيل فبستخدم هدول $\Leftarrow \begin{cases} F_X(-\infty) = 0 \\ F_X(\infty) = 1 \end{cases}$

+ عشان أقدر أحوله لـ pdf وأقدر أوجد احتمالية كل نقطة بستخدم قانون

$$P(X=x) = F(x) - F(x^-) = P(X \leq x) - F(X < x)$$

---

mean, expected value, average: μ_X

discreate:

$$E[g(x)] = \sum_{x=-\infty}^{\infty} g(x)\,P(X=x_i)$$

Continuoes:

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

Variance of x: → يوجد قانونين

1. $Var[X] = \sigma_X^2 = E[(X-\mu_X)^2]$
2. $\sigma_X^2 = E[X^2] - \mu_X^2$

Standard deviation of x :

$\sigma_X = \sqrt{\sigma_X^2}$

The median :

$\int_{-\infty}^{x_0} f_X(x)\,dx = 1/2$ ، النقطة التي يكون الاحتمال قبلها = بعدها

The mod :

$\frac{d f_X(x)}{dx} = 0$ ، قيمة x التي يكون عندها الاقتران أعلى شي

ننتبه إنو قيمة x وليس أعلى قيمة للاقتران

Linear transformation:

$y = aX + b$

$\mu_Y = a\mu_X$ $\qquad$ $\sigma_Y^2 = a^2\sigma_X^2$

ربنا تقبل منا إنك أنت السميع العليم ..

روان الفار

Ch.2

1. Binomial distribution :-

$$p(X=x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & , x = 0, 1, 2, \ldots \\ 0 & , \text{o.w} \end{cases}$$

, n: number of trials.

x: number which success will occure

mean value :- $\mu_x = E(x) = np$

Variance $= \sigma_x^2 = var(x) = np(1-p)$

2. Geometric Distribution :-

$$p(X=x) = \underbrace{(1-p)^{x-1}}_{\text{probability of fail}} \underbrace{p}_{\text{probability of success}} \quad , X = 1, 2, 3, \ldots$$

mean value $= \mu_x = E[x] = \frac{1}{p}$

Variance $= \sigma_x^2 = Var(x) = \frac{1-p}{p^2}$

3. Hype geometric distribution :-

- $p(X=x) = \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$
- mean value $= \dfrac{nK}{N} = np$
- $\sigma_x^2 = np(1-p)\left[\dfrac{N-n}{N-1}\right]$

4. Poisson distribution

$$p(X=x) = e^{-b} \frac{b^x}{x!} \quad , X = 0, 1, 2, \ldots$$

$\mu_x = E[x] = b$, $b = \underbrace{\lambda}_{\text{average of occurences}} \underbrace{T}_{\text{unite}}$

common continuous random variable

1. uniform distribution.

- $f_x(x) = \begin{cases} \frac{1}{b-a} &, a \leq x \leq b \\ 0 &, \text{o.w} \end{cases}$, • $\mu_x = \frac{a+b}{2}$, $\sigma_x^2 = \frac{(b-a)^2}{12}$

2. exponential Distribution :-

$f_x(x) = \begin{cases} \lambda e^{-\lambda x} &, x \geq 0 \\ 0 &, \text{o.w} \end{cases}$, $\mu_x = \frac{1}{\lambda}$, $\sigma_x^2 = \frac{1}{\lambda^2}$

3. Rayleigh Distribution.

$f_x(x) = \frac{2}{b} x e^{\frac{-x^2}{b}}$, $F_x(x) = 1 - e^{\frac{-x^2}{b}}$

$\mu_x = E[x] = \sqrt{\frac{\pi b}{4}}$

$\sigma_x^2 = \frac{b(4-\pi)}{4}$

4. Cauchy Random Variable

$f_x(x) = \frac{\frac{\alpha}{\pi}}{x^2 + \alpha^2}$, $F_x(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{x}{a}\right)$

Gaussian normal distribution

$f_x(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{\frac{-(x-\mu_x)^2}{2\sigma_x^2}}$ $\longrightarrow$ mean = zero , variance = 1

- $P(x \leq 0) = \int_{-\infty}^{0} f_x(x)\,dx = \Phi(0) =$ الجواب من الجدول

- $P(x \leq 1.12) = \int_{-\infty}^{1.12} f_x(x) = \Phi(1.12) =$ من الجدول

- $\Phi(-a) = 1 - \Phi(a)$
- $P(x \geq 3.12) = 1 - P(x \leq 3.12)$

  $= 1 - \Phi(3.12)$
- $P(0.5 \leq x \leq 1.7) = \Phi(1.7) - \Phi(0.5)$

- if $\mu_x \neq 0$ and $\sigma_x \neq 1$ $\rightarrow$ "not standard"

  $\Phi\left(\frac{x-\mu_x}{\sigma_x}\right)$

normal approximation for binomial and poisson distribution

- evalute μ_x and σ_x then use gaussian distribution to find the propability.

$$\Phi\left(\frac{x-\mu_x}{\sigma_x}\right)$$

Transformation of Random Variables

discreate

they are equal

X	P(X=x)	Y=g(x)	P(Y=y)

Continueos :

$$f_Y(y) = \frac{f_X(x)}{\left|\frac{dy}{dx}\right|}$$

* ① $\frac{dy}{dx}$, ② X = in terms of Y , ③ Substitute in $f_Y(y)$ with Cases

---

Note :-

Y is gaussian with mean $\mu_Y = a\mu_X + b$ and variance $\sigma_Y^2 = a^2\sigma_X^2$

ch.3

- $\mu_X = x \sum_{x=0}^{n} p(X = x_i)$
- $\sigma_X^2 = E[x^2] - \mu_X^2$
- Are x and y indep? $p(X=x, Y=y) \overset{??}{=} p(X=x) \cdot p(Y=y)$ ⇝ check at any point you want.

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

- if they are indep. $E[xy] = E[x] \cdot E[y]$

correlation coefficient

- $\rho_{xy} = \dfrac{E[xy] - \mu_x \mu_y}{\sigma_x \sigma_y}$, indep $\xrightarrow{\text{must Be}}$ uncarolated
- Covariance $= E[xy] - \mu_x \mu_y$
- $\rho_{xy} = 0 \rightarrow$ uncarolated , $\rho_{xy} = \mp 1 \rightarrow$ fully Carolated.
- $E[xy] = \sum \sum x\,y\, p(X=x, Y=y)$
- if $Z = X + Y$ then $E[Z] = E[X+Y] = E[X] + E[Y]$
- if $Y = a_1 X_1 + a_2 X_2$

$\quad \hookrightarrow \mu_Y = a_1 \mu_{x_1} + a_2 \mu_{x_2}$

$\quad \hookrightarrow \sigma_Y^2 = a_1^2 \sigma_{x_1}^2 + a_2^2 \sigma_{x_2}^2 + 2a_1 a_2 \sigma_{x_1} \sigma_{x_2} \rho_{x_1 x_2}$ ⇝ equals covariance

$\quad \hookrightarrow \sigma_Y^2 = a_1^2 \sigma_{x_1}^2 + a_2^2 \sigma_{x_2}^2 + 2a_1 a_2 [E[xy] - \mu_{xy}]$.

$\quad \hookrightarrow \sigma_Y^2 = a_1^2 \sigma_{x_1}^2 + a_2^2 \sigma_{x_2}^2$

Covariance $= \sigma_{x_1} \sigma_{x_2} \rho_{x_1 x_2}$

Two Continuous Random Variable

- to find any value in f_{xy} you should do double integration

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{xy}\, dy\, dx \quad \text{or} \quad \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{xy}\, dx\, dy$$

- $p(X \leq x_0)$ → f_{xy}
  → margine function

$$f_x(x) = \int_{y_1}^{y_2} f_{xy}(x,y)\, dy \qquad f_y(y) = \int_{x_1}^{x_2} f_{xy}(x,y)\, dx$$

$* f_{xy} = f_x(x) \cdot f_y(y)$

↳ valid if they are independent

Conditional pdf :-

1. $f_{y/x}(y) = \dfrac{f_{xy}(x,y)}{f_x(x)}$, 2. $f_{x|y}(x) = \dfrac{f_{xy}(x,y)}{f_y(y)}$

* $f_{x/y} = f_x(x)$ and $f_{y/x} = f_y(y)$

then they are statistically indep.

* $0<y<x<2$
  - $\rightarrow 0<x<2$
  - $\rightarrow 0<y<2$
  - $\rightarrow y<x$

Ch. 4

- Sample Mean $\hat{\mu}_x = \frac{1}{n} \sum_{i=1}^{n} x_i$
- Sample Variance $\hat{\sigma}_x^2$
  - when μ_x is known $\hat{\sigma}_x^2 = \frac{1}{n} \sum (x_i - \mu_x)^2$
  - when μ_x is unknown $\hat{\sigma}_x^2 = \frac{1}{n-1} \sum (x_i - \hat{\mu}_x)^2$
  - $\hat{\sigma}_x^2 = \dfrac{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$

- Sample Standard diviation :- $\hat{\sigma}_x = \sqrt{\hat{\sigma}_x^2}$
- Sample Covariance between x and y
  - $C_{xy} = \frac{1}{n-1} \sum (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)$
  - $C_{xy} = \dfrac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n(n-1)}$

- Sample Correlation Coefficient $\hat{\rho}_{xy} = \dfrac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} \rightsquigarrow C_{xy}$

$$y = \alpha x + \beta$$

regression techniques

- $\epsilon = \sum_{i=1}^{n} [y_i - (\alpha x_i + \beta)]^2$ → least square errors.
- $\alpha = \dfrac{C_{xy}}{\hat{\sigma}_{xy}^2}$, $\beta = \hat{\mu}_y - \alpha \hat{\mu}_x$

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

- fitting a polynomial by the method of least squares

$$y = \beta_1 + \beta_2 x + \beta_3 x^2$$

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

Theorem: independent Gaussian Radom variables

- $E[\hat{\mu}_x] = \mu_x$
- $Var[\hat{\mu}_x] = \frac{\sigma_x^2}{n}$

* $y = X_1 + X_2 + \dots + X_n$

$\hookrightarrow E[y] = \mu_{x_1} + \mu_{x_2} + \dots + \mu_{x_n} = n\mu_x$

$\hookrightarrow Var[y] = n\sigma_x^2$

* $\hat{\mu}_x = y = \frac{1}{n} \Sigma X_i$

- $E[\hat{\mu}_x] = \mu_x$
- $Var[\hat{\mu}_x] = \frac{\sigma_x^2}{n}$
- $STD = \frac{\sigma_x}{\sqrt{n}}$

Y gaussian

$\hookrightarrow$ standard $\rightarrow$ var = 1, mean = 0

$\hookrightarrow$ non-standard $\rightarrow \Phi\left(\frac{X - \mu_x}{\sigma_y}\right)$

● استعلاء .. ثمنه التعب

ربنا تقبل منا إنك
أنت السميع العليم ..