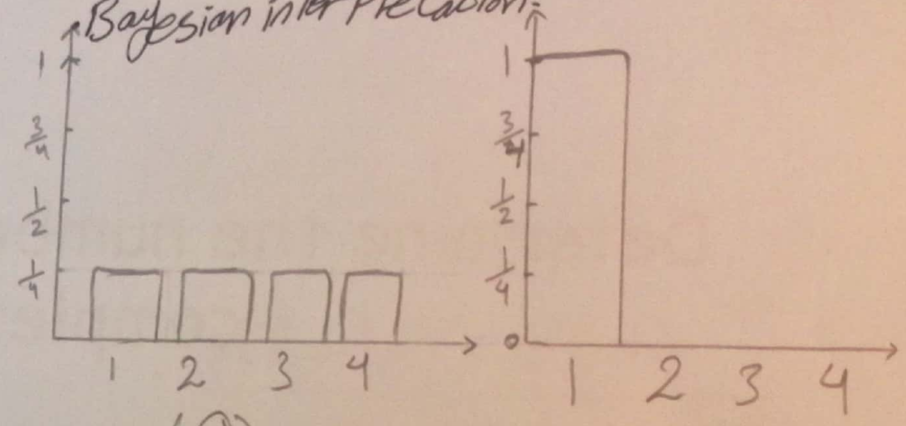
(Chapter 2:- Probability) (Introduction)

1

- frequentist interpretation - (Coin)
- Bayesian interpretation:



(a) (b)

(Uniform distribution) $P(x) = \frac{1}{B}$ (degenerate distribution)

- Discrete random variables: - $P(x) = \mathbb{I}(x=1)$

- $0 \leq P(A) \leq 1$.

$P(\overline{A}) = 1 - P(A)$.

$P() \xrightarrow{\text{يعني}}$ probability mass function (Pmf)

$\mathbb{I}() \xrightarrow{\text{يعني}}$ indicator function.

- fundamental rules:-

1. Probability of a union of two event

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

$= P(A) + P(B)$ في حالة A S B مفيش ما بينهم تقاطع

(if A and B are mutually exclusive)

2. joint probabilities:-

$$P(A, B) = P(A \wedge B) = P(A|B)\,P(B)$$

(Product rule) [illegible]

$$P(A) = \sum_b P(A, B) = \sum_b P(A|B=b)\,P(B=b)$$

(marginal distribution) [illegible]

③ Conditional Probability:

$$P(A|B) = \frac{P(A,B)}{P(B)} \quad \text{if } P(B) > 0$$

$\boxed{2}$

- **Bayes rule:-**

$$P(X=x|Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$$

$$= \frac{P(X=x)\,P(Y=y|X=x)}{\sum_{x'} P(X=x')\,P(Y=y|X=x')}$$

Example:- medical diagnosis

$P(\boxed{x=1}|\boxed{y=1}) = 0.8$ يعني احتمالية أن ال Test يطلع Positive نتيجة $\boxed{0.8}$

$x=1$: حدث أن mammogram إيجابي (نوع من أنواع الأشعة)

$y=1$: حدث أن عندها breast cancer

$P(y=1) = 0.004$ الاحتمالية للإصابة بالسرطان وهي احتمالية ضعيفة جداً.

$P(x=1|y=0) = 0.1$ معناه يطلع ال Test إيجابي غلط يعني احنا مش عيانين

false positive

أو

false alarm

هنا بقى هنطبق القانون اللي فات :-

$$P(X=1|Y=1) = \frac{\overbrace{P(Y=1)}^{0.004}\,\overbrace{P(Y=1|X=1)}^{0.8}}{\underbrace{P(x=1|y=1)}_{0.8}\,\underbrace{P(y=1)}_{0.004} + \underbrace{P(x=1|y=0)}_{0.1}\,\underbrace{P(y=0)}_{=1-P(y=1)=1-0.004=0.996}}$$

$$= \frac{0.004 \times 0.8}{0.8 \times 0.004 + 0.1 \times 0.996}$$

$$= 0.031 \ \#$$

- **Independence and Conditional independence:-**

$$\boxed{X \perp Y} \Longleftrightarrow P(X,Y) = P(X)P(Y)$$

Un Conditionally independent / marginally independent

$$\boxed{P(a < X \le b) = F(b) - F(a)}$$

$P(W) = P(B) - P(A)$ علشان نفتكر بس

$\therefore f(x) = \frac{d}{dx} F(x) \Rightarrow$ Probability density function (pdf)

$\xRightarrow{\text{يعني}} F(x) = \int f(x)\, dx$

$\therefore P(a < X \le b) = \int_a^b f(x)\, dx$

$\xRightarrow[\text{صياغة}]{\text{إعادة}} P(x \le X \le x + dx) \simeq \boxed{P(x)\,dx}$

require $P(x) \geq 0$

Uniform distribution

ex:-

$a = 0$

$b = \frac{1}{2}$

$\text{Unif}(x|a,b) = \frac{1}{b-a} \quad (a \le x \le b)$

$\therefore P(x) = \frac{1}{\frac{1}{2} - 0} = \boxed{2}$ for $x \in [0, \frac{1}{2}]$

- Unconditional independence is rare

نادر الحدوث

$X \perp Y | Z \iff P(X, Y|Z) = P(X|Z)\, P(Y|Z)$ $\boxed{3}$

Conditionally independence

- Continuous random variables:-

$a \le X \le b$

$\therefore A = (X \le a)$, $B = (X \le b)$

$W = (a < X \le b)$

$\therefore B = A \vee W$

$$\boxed{\therefore P(B) = P(A) + P(W)}$$

$$\boxed{P(W) = P(B) - P(A)}$$

$$F(q) \triangleq P(X \le q)$$

Cumulative distribution function (cdf)

- **Mean and Variance :-**
- Mean or expected value μ.
- Variance σ^2.

$$E[X] \triangleq \sum_{x \in X} x\,p(x), \text{ for discrete} \quad \boxed{\text{mean}}$$

$$E[X] \triangleq \int_X x\,p(x)\,dx, \text{ for Continuous.}$$

$$\boxed{\sigma^2} \; Var[X] \triangleq E[(X-\mu)^2] = \int (x-\mu)^2 p(x)\,dx \quad \boxed{\text{Variance}}$$

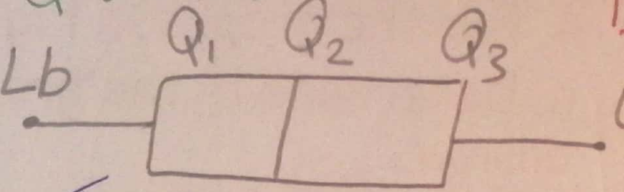
$$= \int x^2 p(x)\,dx + \mu^2 \int p(x)\,dx - 2\mu \int x\,p(x)\,dx$$

$$= E[X^2] - \mu^2$$

$$\therefore E[x^2] = \mu^2 + \sigma^2$$

يعني $\Longrightarrow$ $std[X] \triangleq \sqrt{Var[X]}$ $\boxed{\text{Standard deviation}}$

- **Quantiles:-** $\boxed{4}$



$IQR = Q_3 - Q_1 \rightarrow (1)$

Where:-

Q_1:- is median of 1st 25%

Q_2:- is the median 50%

Q_3:- is median of 2nd 75%

$Lb = Q_1 - (1.5 * IQR)$

$Ub = Q_3 + (1.5 * IQR)$

معظم بيجو بيعبر عنه بطريقة تانية زي
اللي في الكتاب:

$F^{-1}(0.5)$ median Q_2

$F^{-1}(0.25)$ Q_1

$F^{-1}(0.75)$ Q_3

• Some Common discrete distributions

• The binomial and Bernoulli distributions:- $\boxed{5}$

① Binomial:-

هنفترض أننا رمينا Coin

عدد $\boxed{n}$ من المرات و $X \in \{0 \ldots n\}$

عدد الـ heads و عبرنا احتمالية

ظهور الـ heads بـ θ

هنا بنقول أن الـ X دي :: binomial distribution

$$\therefore Bin(k|n,\theta) \triangleq \binom{n}{k}\theta^{k}(1-\theta)^{n-k}$$

Where:- ⇓

$$\binom{n}{k} \triangleq \frac{n!}{(n-k)!\,k!}$$

② Bernoulli:-

هنا بقى هنفترض أننا رمينا Coin مرة واحدة بس

Binary random variable ⇐ $X \in \{0,1\}$

كتابة ملك

((heads) احتمالية الـ) θ وبردو عبرنا

هنا بنقول أن الـ X دي :: Bernoulli distribution

$$\therefore Ber(X|\theta) = \theta^{(x=1)}(1-\theta)^{(x=0)}$$

بطريقة تانية بردو

$$Ber(x|\theta) = \begin{cases} \theta & \text{if } x=1 \\ 1-\theta & \text{if } x=0 \end{cases}$$

ولو لاحظنا أن الـ (Bernoulli) هي حالة خاصة

special case من الـ (Binomial) لما $\boxed{n=1}$

6

- The Multinomial and Multinoulli distributions:-

- Multinomial distributions

$$Mu(X|n,\theta) \triangleq \binom{n}{x_1 \cdots x_k} \prod_{j=1}^{k} \theta_j^{x_j}$$

where:-

$$\binom{n}{x_1 \cdots x_k} \triangleq \frac{n!}{x_1! x_2! \cdots x_k!}$$

is the multinomial coefficient.

$$\boxed{Mu(X|1,\theta) = \prod_{j=1}^{k} \theta_j^{\mathbb{I}(x_j=1)}}$$

Multinoulli distribution

وبرده هنقول هنا أمر إن Multinoulli distribution

دي حالة خاصة من ال Multinomial distribution

وعلشان منفصل عن ال Multinoulli destribution

بنستخدم ب

$$Cat(x|\theta) \triangleq Mu(X|1,\theta)$$

- Poission distribution:-

$$Poi(x|\lambda) = e^{-\lambda} \frac{\lambda^x}{x!}$$

- Empirical distribution:-

$$P_{emp}(A) \triangleq \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}(A)$$

Where:- $\delta_x(A)$ is $\boxed{\text{Dirac measure}}$

$$\delta_x(A) = \begin{cases} 0 & \text{if } x \notin A \\ 1 & \text{if } x \in A \end{cases}$$

$$P(x) = \sum_{i=1}^{N} w_i \delta_{x_i}(x)$$

- where:

$$0 \leq w_i \leq 1$$

$$\sum_{i=1}^{N} w_i = 1$$

- We Can Compute it in term of the error function (erf):-

$$\Phi(x;\mu,\sigma^2) = \frac{1}{2}\left[1+\mathrm{erf}(z/\sqrt{2})\right].$$

Where:-

$$z = (x-\mu)/\sigma \quad \underline{\text{or}} \quad \mathrm{erf}(x) \triangleq \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt$$

- Degenerate Pdf:-

$$\lim_{\sigma^2 \to 0} \mathcal{N}(x|\mu,\sigma^2) = \delta(x-\mu)$$

Where:- δ is Called $\boxed{\text{Dirac delta function}}$

$$\delta(x) = \begin{cases} \infty & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases}$$

$\Downarrow$ is the sifting property

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

- Some Common Continous distributions:- $\boxed{4}$

- Gaussian (normal) distribution:-

$$\mathcal{N}(x|\mu,\sigma^2) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

- If $X \sim \mathcal{N}(0,1)$

X follows a standard normal distribution

- precision of Gaussian mean the inverse variance $\lambda = \frac{1}{\sigma^2}$

- The Cumulative distribution function or cdf of the Gaussian:-

$$\Phi(x;\mu,\sigma^2) = \int_{-\infty}^{x} \mathcal{N}(z|\mu,\sigma^2)\,dz$$

Where $\Gamma(a)$ is the gamma function:

$$\Gamma(x) \triangleq \int_0^\infty u^{x-1} e^{-u} \, du$$

- The special cases of the Gamma function:
  - Exponential distribution:

$$\text{Expon}(x|\lambda) \triangleq Ga(x|1, \lambda)$$

  - Erlang distribution:

$$\text{Erlang}(x|\lambda) = Ga(x|2, \lambda)$$

  - Chi_squared distribution:

$$\chi^2(x|\nu) \triangleq Ga(x|\tfrac{\nu}{2}, \tfrac{1}{2})$$

- Inverse gamma:-

$$IG(x|\text{shape}=a, \text{scale}=b) \triangleq \frac{b^a}{\Gamma(a)} x^{-(a+1)} e^{-b/x}$$

- The Laplace distribution:- $\boxed{8}$

know as the double sided exponential.

$$\text{Lap}(x|\mu, b) \triangleq \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

where:-

μ is a location parameter.

mean $= \mu$, mode $= \mu$

$\underline{\underline{\text{Var} = 2b^2}}$

- The gamma distribution:-

$$Ga(T|\text{shape}=a, \text{rate}=b) \triangleq \frac{b^a}{\Gamma(a)} T^{a-1} e^{-Tb}$$

- Joint Probability distributions:-

  - Covariance and Correlation:-

$$\mathrm{Cov}[X,Y] \triangleq E\left[(X-E[X])(Y-E[Y])\right]$$

$$= E[XY] - E[X]E[Y].$$

$$\therefore \mathrm{Cov}[x] \triangleq E\left[(x-E[x])(x-E[x])^T\right]$$

$$= \begin{pmatrix} \mathrm{Var}[x_1] & \mathrm{Cov}[x_1,x_2] & \cdots \\ \mathrm{Cov}[x_2,x_1] & \mathrm{Var}[x_2] & \cdots \\ \vdots & \vdots & \ddots \\ \mathrm{Cov}[x_d,x_1] & \mathrm{Cov}[x_d,x_2] & \mathrm{Var}[x_d] \end{pmatrix}$$

- $$\mathrm{Corr}[X,Y] \triangleq \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}}$$

$$R = \begin{pmatrix} \mathrm{Corr}[x_1,x_1] & \mathrm{Corr}[x_1,x_2] & \cdots \\ \vdots & \vdots & \ddots \\ \mathrm{Corr}[x_d,x_1] & \mathrm{Corr}[x_d,x_2] & \mathrm{Corr}[x_d,x_d] \end{pmatrix}$$

- The beta distribution:-

$$\mathrm{Beta}(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

Where:-

$$B(a,b) \triangleq \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

- The Pareto distribution:-

$$\mathrm{Pareto}(x|k,m)$$

$$= k m^k x^{-(k+1)} \mathbb{I}(x \geq m).$$

- Multivariate student t distribution:-

$$\mathcal{T}(x|\mu,\Sigma,\nu) = \frac{\Gamma(\nu/2+D/2)}{\Gamma(\nu/2)} \cdot \frac{|\Sigma|^{-\frac{1}{2}}}{\nu^{D/2}\pi^{D/2}}$$

$$\times \left[1+\frac{1}{\nu}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]^{-\left(\frac{\nu+D}{2}\right)}$$

$$= \frac{\Gamma(\nu/2+D/2)}{\Gamma(\nu/2)}|\pi V|^{-\frac{1}{2}}$$

$$\times \left[1+(x-\mu)^T V^{-1}(x-\mu)\right]^{-\left(\frac{\nu+D}{2}\right)}$$

- The multivariate Gaussian:-

$\boxed{10}$

$$\mathcal{N}(x,\mu,\Sigma) \triangleq \frac{1}{(2\pi)^{D/2}\,|\Sigma|^{\frac{1}{2}}} \cdot \exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

- we will work in terms of the **Precision matrix** or **Concentration matrix**

- A **spherial** or **isotropic** Covariance, $\Sigma\ \sigma^2 I_D$ has free Parameter.

- **The Transformations of random Variables:**

  - **Linear Transformations:**

    $f()$ is a linear function.

    $Y = f(x) = Ax + b.$

    $E[Y] = E[Ax+b] = A\mu + b$

    Where: $\mu = E[X]$ is called linearity of expectation.

    $f(x) = a^T X + b.$

    $E[a^T X + b] = a^T \mu + b$

    $Var[y] = Var[a^T X + b]$

    $= a^T \Sigma a.$

  - **Convariance:**

    $Cov[Y] = Cov[Ax+b] = A \Sigma A^T$

    Where: $\Sigma = Cov[X]$

- **Dirichlet distribution:**

  $S_K = \{x : 0 \leq x_k \leq 1, \sum_{k=1}^{K} x_k = 1\}$ Probability simplex

  $Dir(X|\alpha) \triangleq \frac{1}{B(\alpha)} \prod_{k=1}^{K} x_k^{\alpha_k - 1} \mathbb{I}(x \in S_K)$

  Where: $B(\alpha) \triangleq \frac{\prod_{k=1}^{K} \Gamma(\alpha_k)}{\Gamma(\alpha_0)}$

  Where: $\alpha_0 \triangleq \sum_{k=1}^{K} \alpha_k.$

• Central limit theorem:

$$P(S_N = s) = \frac{1}{\sqrt{2\pi N\sigma^2}} \exp\left(\frac{-(s-N\mu)^2}{2N\sigma^2}\right)$$

$$Z_N \triangleq \frac{S_N - N\mu}{\sigma\sqrt{N}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{N}}$$

where:

$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is the sample mean and is called Central limit theorem.

• General transformations:

$\boxed{12}$

• $P_y(y) = \sum_{x: f(x)=y} P_x(x)$

• $P_y(y) \triangleq P(Y \leq y) = P(f(X) \leq y)$

$= P(X \in \{x \mid f(x) \leq y\})$.

• $P_y(y) \triangleq P(f(x) \leq y) = P(X \leq f^{-1}(y))$

$= P_x(f^{-1}(y))$.

• $P_y(y) \triangleq \frac{d}{dy} P_y(y) = \frac{d}{dy} P_x(f^{-1}(y))$

$= \frac{dx}{dy}\frac{d}{dx} P_x(x) = \frac{dx}{dy} P_x(x)$

$\therefore P_y(y) = P_x(x)\left|\frac{dx}{dy}\right|$; is called change of variables