Notes on Probability:

Before starting: - Check if they can see/hear

Quick recap:

Random experiment: experiment whose output cannot be surely predicted in advance.

Example: I toss a coin, I cannot predict the output of a single toss, but repeating the experiment several times we can observe some regularities.

4 main ingredients: ① The state space Ω: set of all possible outcomes of a random experiment

Ex: $\Omega = \{\text{head}, \text{tail}\}$

Ex: $\Omega = \mathbb{R}$

② Events: an event is a property which can be observed either to hold or not after the experiment. It is a subset of Ω

Ex: $A = \{\text{head}\}$

$B = 42$

③ The Probability: A function which takes in input an event and outputs its probability, which is a measure on how likely the event is going to be realized a priori, before performing the experiment

Properties: $P(A) \in [0,1] \ \forall A$

$P(\Omega) = 1$

$\forall$ countable sequence of mutually disjoint events $A_1, \dots, A_n, \dots \Rightarrow P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$

④ A random variable: It's a function receiving as an input the outcome of an experiment $\omega \in \Omega$ and giving as output a real number

Ex: $X : \Omega \to \{0,1\}$, $X(\omega) = \begin{cases} 0 & \text{if } \omega = \text{"head"} \\ 1 & \text{if } \omega = \text{"tail"} \end{cases}$

Conditional Probability

Def: Suppose that event B has occurred with $P(B)>0$. We denote by $P(A|B)$ the prob. of the event A, given the fact that event B has occurred, and we define it as: $P(A|B) := \dfrac{P(A,B)}{P(B)}$

Def: A, B independent iff $P(A,B) = P(A)P(B)$

As a consequence, A, B indep $\Rightarrow P(A|B) = P(A)$

Theorem (Useful for HMM): Let $A_1, \dots, A_n$ be events st $P(A_1, \dots, A_n) > 0$. Then

$$P(A_1,\dots,A_n) = P(A_1)\,P(A_2|A_1)\,P(A_3|A_2,A_1)\cdots P(A_n|A_{n-1},\dots,A_1)$$

Proof: (By induction) If $n=2$, $P(A_1,A_2) = P(A_2|A_1)P(A_1)$

Suppose that the theorem holds for $n-1$ events, call $B := A_1 \cap A_2 \cap \dots \cap A_{n-1}$, then by def. of conditional probability $P(B,A_n) = P(A_n|B)\,P(B)$

$$P(A_1,\dots,A_n) = P(A_n|A_{n-1},\dots,A_1)\,P(A_1,\dots,A_{n-1}) =$$

induction hp

$$= P(A_n|A_{n-1},\dots,A_1)\,P(A_{n-1}|A_1,\dots,A_{n-2})\cdots P(A_2|A_1)\,P(A_1)$$

$\square$

Random Variables

$\boxed{\text{Discrete rv}}$ $X: \Omega \to E$, E countable

Example: Bernoulli random variable (can represent a possibly biased coin toss): $X(\omega) = \begin{cases} 1 & \text{if } \omega = \text{"head"} \\ 0 & \text{if } \omega = \text{"tail"} \end{cases}$

$X \sim Be(p)$ $\quad P(X=x) = P_X(x) = p^x (1-p)^{1-x} \mathbb{1}_{\{0,1\}}(x)$, $\quad p \in [0,1]$

Expected value: $\mathbb{E}_{X\sim P_X}[X] = \sum_{k=0}^{1} k P(X=k) = 0P(X=0) + 1P(X=1) = p$

Variance:

$$\text{Var}_{X\sim P_X}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{k=0}^{1} k^2 p_X(k) - p^2 = 0P(X=0) + 1^2 P(X=1) - p^2 =$$

$$= p - p^2 = p(1-p)$$

Example: Binomial random variable: $X \sim Bi(n,p)$ $\quad P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \mathbb{1}_{\{0,1\}}(x)$, $n \in \mathbb{N}_0$, $p \in [0,1]$ $=$

$$= \frac{n!}{x!(n-x)!} p^x (1-p)^{1-x} \mathbb{1}_{\{0,..\}}(x)$$

Expected value: $\mathbb{E}[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} \frac{k \cdot n!}{k!(n-k)!} p^k (1-p)^{n-k} =$

$$= np \sum_{k=0}^{n} \frac{k(n-1)!}{k!(n-k)!} p^{k-1} (1-p)^{(n-1)-(k-1)} = np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} p^{k-1}(1-p)^{(n-1)-(k-1)} =$$

$$= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \overset{l=k-1}{=} np \sum_{l=0}^{n-1} \binom{n-1}{l} p^l (1-p)^{(n-1)-l} \overset{m=n-1}{=}$$

$$= np \sum_{l=0}^{m} \binom{m}{l} p^l (1-p)^{m-l} = np\,(p + (1-p))^m = np$$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Trick: $Y_1, \dots, Y_N \overset{iid}{\sim} Be(p) \Rightarrow X = \sum^{N} Y_n \sim Bi(n,p)$. Then $\mathbb{E}[X] = \mathbb{E}\left[\sum_{n=1}^{N} Y_n\right] = \sum_{n=1}^{N} \mathbb{E}[Y_n] = \sum_{n=1}^{N} \mathbb{E}[Y] = \sum_{n=1}^{N} p = Np$

Example: Poisson r.v. Useful for modelling the number of times an event occurs in an interval of time / space (5)

$X \sim \text{Poisson}(\lambda)$, $X: \Omega \to \mathbb{N}$

$$P(X=x) = \frac{\lambda^x}{x!} e^{-\lambda} \mathbb{1}_{\mathbb{N}}(x), \quad \lambda > 0$$

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k P(X=k) = \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} e^{-\lambda} = \lambda \underbrace{\boxed{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}}}_{= e^{\lambda}} e^{-\lambda} = \lambda e^{\lambda} e^{-\lambda} = \lambda$$

$\boxed{\text{Continuous r.v.}}$ $X: \Omega \to \mathbb{R}$

Gaussian r.v. $X \sim \mathcal{N}(\mu, \sigma^2)$ $\quad p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx = \dots \text{ difficult!}$$

Trick: $X \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow X = \mu + \sigma Z$, $Z \sim \mathcal{N}(0,1)$

$$\mathbb{E}[X] = \mathbb{E}[\mu + \sigma Z] = \mu + \sigma \mathbb{E}[Z] = \mu + \sigma \int_{\mathbb{R}} x \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\} dx = \mu + \frac{\sigma}{\sqrt{2\pi}} \left(-e^{-\frac{x^2}{2}}\right)\Bigg|_{-\infty}^{+\infty} = \mu$$

Rem: $\mathbb{E}[g(x)] = \int_{\mathbb{R}} g(x) p_X(x) dx$

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 =$$

$$\mathbb{E}[\mu^2 + 2\mu\sigma Z + \sigma^2 Z^2] \overset{-\mu^2}{=} \mu^2 + 2\mu\sigma \mathbb{E}[Z] + \sigma^2 \mathbb{E}[Z^2] - \mu^2 = \sigma^2 \mathbb{E}[Z^2] =$$

$$\sigma^2 \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} z^2 \exp\left\{-\frac{z^2}{2}\right\} dz = \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} \overset{g}{z} \cdot \overbrace{z\exp\left\{-\frac{z^2}{2}\right\}}^{f'} dz = \underbrace{\sigma^2 z \cdot \left(-e^{-z^2/2}\right)\Bigg|_{-\infty}^{+\infty}}_{0} - \int_{\mathbb{R}} \frac{\sigma^2}{\sqrt{2\pi}} \left(-\exp\left\{-\frac{z^2}{2}\right\}\right) dz =$$

By parts: $\int f'g = fg - \int g'f$

$$= \sigma^2 \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} dz}_{=1} = \sigma^2$$

Theorem (Strong law of large numbers): Let $(X_n)_{n\geq 1}$ be independent and identically distributed r.v. defined on the same space. Let $\mathbb{E}[X_i] = \mu$, $Var(X_i) = \sigma^2 < +\infty$ and set $S_n = \sum_{j=1}^{n} X_j$.

Then $\lim_{n\to+\infty} \frac{S_n}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} X_j = \mu$ a.s.

Theorem (Central limit theorem): Let $(X_n)_{n\geq 1}$ be a sequence of iid r.v. with $\mathbb{E}[X_i] = \mu$, $Var(X_i) = \sigma^2 < \infty$.

Set $\sum_{j=1}^{n} X_j$ and $Y_n = \frac{S_n - n\mu}{\sqrt{\sigma^2 n}}$. Then $Y_n \xrightarrow{n\to\infty} Y \sim \mathcal{N}(0,1)$

$\boxed{\text{Bayes Theorem}}$

Before measuring the data we have 2 random elements:

A prior distribution $\pi(\theta)$: Our belief that θ represents the true population characteristic

A likelihood $\rho(\underline{X}|\theta)$: Our belief that $\underline{X}$ would be the outcome if θ is the true parameter value (Conditional distribution of data, given parameters)

After measuring data: We update our believes on θ computing the posterior distribution through Bayes theorem:

$$\pi(\theta|\underline{X}) = \frac{\rho(\underline{X}|\theta)\,\pi(\theta)}{\int_{\Theta} \rho(\underline{X}|\theta)\,\pi(\theta)\,d\theta} \longrightarrow P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Example: θ : proportion of infected people in a town : $\theta \in [0,1]$

$X_1, \dots, X_{20}$ $\quad X_i = \begin{cases} 1 & \text{if unit } i \text{ is infected} \\ 0 & \text{otherwise} \end{cases}$

$Y = \sum_{i=1}^{20} X_i$

Useful for modelling the random behaviour of percentages or proportions

Before measuring X: <u>Prior distribution</u> $\theta \sim Beta(a,b)$

$$\pi(\theta) = p_\theta(\theta) = \frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1} \mathbb{1}_{(0,1)}(\theta) \qquad a,b>0$$

$$B(a,b) = \int_0^1 \theta^{a-1}(1-\theta)^{b-1} \, d\theta$$

<u>Likelihood</u>: $X|\theta \sim Be(\theta)$

$Y|\theta \sim Bi(n=20, \theta) \qquad p_Y(Y=y) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$

We compute the posterior using Bayes theorem:

$$\pi(\theta|y) = \frac{p(y|\theta)\,\pi(\theta)}{\int_0^1 p(y|\theta)\,\pi(\theta)\, d\theta} \begin{matrix} \leftarrow (N) \\ \leftarrow (D) \end{matrix} =$$

$$(N) = \frac{\binom{n}{y} \theta^y (1-\theta)^{n-y} \theta^{a-1} (1-\theta)^{b-1}}{B(a,b)} \mathbb{1}_{(0,1)}(\theta) = \frac{\binom{n}{y}}{B(a,b)} \cdot \theta^{y+a-1} (1-\theta)^{n-y+b-1} \mathbb{1}_{(0,1)}(\theta)$$

$$(D) \int_0^1 p(y|\theta)\pi(\theta)\, d\theta = \frac{1}{B(a,b)} \int_0^1 \binom{n}{y} \theta^{\overbrace{y+a-1}^{a^*}} (1-\theta)^{\overbrace{n-y+b-1}^{b^*}} d\theta =$$

$$= \frac{B(a^*,b^*)}{B(a,b)} \binom{n}{y} \underbrace{\int_0^1 \frac{1}{B(a^*,b^*)} \theta^{a^*-1} (1-\theta)^{b^*-1} d\theta}_{=1} = \binom{n}{y} \frac{B(a+y, n-y+b)}{B(a,b)}$$

$$\Rightarrow \pi(\theta|y) = \frac{\theta^{y+a-1}(1-\theta)^{n-y+b-1}}{B(a+y, n-y+b)} = \frac{\theta^{a+\sum x_i - 1}(1-\theta)^{b+n-\sum x_i - 1}}{B(a+\sum x_i, b+n-\sum x_i)} = Beta\left(a+\sum x_i, b+n-\sum x_i\right)$$