A crash course on Probability: (Not a replacement for a proper text book)

A discrete probality space (Ω, P) is given by

- A finite set or a countable set Ω
  e.g. $\{a, b, c, d\}$ $\quad\quad$ $\{1, 2, 3, 4, \cdots\}$
- A probality mass function $P: \Omega \to [0,1]$ s.t.
  ① $\forall w \in \Omega$, $P(w) \geq 0$
  ② $\sum_{w \in \Omega} P(w) = 1$

For $w \in \Omega$, $P(w)$: the probality the case w happen

An <u>event</u> A is a subset $A \subseteq \Omega$.

$$P(A) = \sum_{w \in A} P(w)$$

$P: \Omega \to [0,1]$ induce a $\boxed{\text{probalility distribution}}$ $P: 2^{\Omega} \to [0,1]$, $2^{\Omega} = \{A \mid A \subseteq \Omega\}$

also denoted by P.

① if $A \cap B = \emptyset$ $\quad P(A \cup B) = P(A) + P(B)$

② $\quad P(\Omega) = 1$

p.m.f $\longleftrightarrow$ p.d.

$P(w)$ $\quad\quad$ $P(\{w\})$

Example 1: Rolling a fair die 

$\Omega = \{1, 2, 3, 4, 5, 6\}$ $\quad\quad$ $P(i)$ = Probality of face i happens

$P(1) = P(2) = \cdots = P(6) = \frac{1}{6}$

Event $A = \{$ out come is even $\}$

$= \{2, 4, 6\}$

$P(A) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2}$

Let $A, B \subseteq \Omega$.

The condition probability $P(A|B) = \dfrac{P(A\cap B)}{P(B)}$

Example: A fair die $\Omega = \{1,2,3,4,5,6\}$

$A = \{\omega \text{ is even}\}$ $B = \{w \geq 4\}$ $P(A) = \frac{1}{2} = P(B)$

$P(A\cap B) = P(\{4,6\}) = \frac{1}{3}$

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Bayes' Rule $P(B|A) = \dfrac{P(A|B)\cdot P(B)}{P(A)}$

Pf: $P(B|A)\,P(A) = P(A\cap B) = P(A|B)\cdot P(B)$

Two events A and B are independent if

$P(A\cap B) = P(A)\cdot P(B)$ $(\Longleftrightarrow P(A|B) = P(A))$

Example. Flip a fair coin twice.

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

$A = \{\text{first outcome is } H\}$ $P(A) = \frac{1}{2}$ $P(B) = \frac{1}{2}$

$B = \{\text{second — is } T\}$ $P(A\cap B) = P(\{HT\}) = \frac{1}{4}$

A <u>Random variable</u> is a function $X: \Omega \rightarrow \mathcal{X}$ from a prob. space $(\Omega, P) \longrightarrow$ a target space $\mathcal{X}$

X is discrete if $\mathcal{X}$ is discrete. (We always in this case as $X(\Omega)$ is discrete)

$\mathcal{X} = \{x_1, x_2, \cdots\}$ is called the alphabet of X

X induce a distribution on $\mathcal{X}$

$\forall x \in \mathcal{X}, P_X(x) = P(\{\omega \mid X(\omega) = x\})$

In many cases, $(\mathcal{X}, P_X)$ capture all the information we need from RV X.

(or law)

$X \sim P_X$ means X has distribution P_X on $\mathcal{X}$.

Example, X be the rank of a poker card randomly picked from a 52-card deck

$\Omega = \{$ all cards in a 52-card deck $\}$ $\quad P(\omega) = \frac{1}{52} \ \forall \omega \in \Omega$

$X: \Omega \longrightarrow \mathcal{X} = \{2, 3, \cdots\cdots 10, J, Q, K, A\}$

$P_X(A) = P_X(2) \cdots \cdots = \frac{1}{13}$

$X: \Omega \rightarrow \mathcal{X}, \ Y: \Omega \rightarrow \mathcal{Y}$ two random variables

Joint distribution on $\mathcal{X} \times \mathcal{Y}$: $\quad P_{XY}(X=x, Y=y) = P(\{X(\omega)=x, Y(\omega)=y\})$

$A \subseteq \mathcal{X}, B \subseteq \mathcal{Y} \qquad P_{XY}(X \in A, Y \in B) = P(\{X(\omega) \in A, Y(\omega) \in B\})$

P_{XY} is a distribution on the produce space $(\mathcal{X} \times \mathcal{Y}, P_{XY})$

Example. A fair die $\quad \Omega = \{1, 2, 3, 4, 5, 6\}$

$X(\omega) = \begin{cases} \text{Large if } \omega \geq 4 \\ \text{Small if } \omega \leq 3 \end{cases} \qquad (\omega \geq 4\}$

$Y(\omega) = \begin{cases} \text{Even} & \text{if } \omega \text{ even} \\ \text{odd} & \text{if } \omega \text{ odd} \end{cases}$

$P_{XY}(\text{Large \& Even}) = P(\{w \text{ even}, w \geq 4\}) = P(\{4,6\}) = \frac{1}{3}$

$P_{XY}(\text{B \& Odd}) = P(\{5\}) = \frac{1}{6}$

$P_{XY}(\text{Small \& E}) = P(\{2\}) = \frac{1}{6}$

$P_{XY}(\text{S \& O}) = P(\{1,3\}) = \frac{1}{3}$

Example. Flip a fair coin twice. X : outcome of first flip

Y : - - - second - -

$\mathcal{X} = \{H, T\}$ $\quad \mathcal{Y} = \{H, T\}$ $\quad \mathcal{X} \times \mathcal{Y} = \{HH, TH, HT, TT\}$

$P_X(H) = P_X(T) = \frac{1}{2} = P_Y(H) = P_Y(T)$

$P_{XY}(HH) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} = P_{XY}(HT)$ - - -

Two R.V. X and Y are independent if $P_{XY}(X=x, Y=y) = P_X(x) P_Y(y)$

$\Longleftrightarrow P_{XY}(X \in A, Y \in B) = P_X(A) P_Y(B)$

Product prob. (Ω_1, P_1) (Ω_2, P_2). two prob. spaces.

$P_1 \times P_2 (A \times B) = P_1(A) P_2(B)$ $\quad A \subseteq \Omega_1, B \subseteq \Omega_2$

is the product probability on $\Omega_1 \times \Omega_2$

Prop. X, Y independent $\Longleftrightarrow$ $P_{XY} = P_X \times P_Y$ product prob.

Example. Randomly pick one from 52-card deck

X : the rank, $\quad Y$ the type

$X : \Omega \to \mathcal{X} = \{2, 3, \cdots 10, J, Q, K, A\}$

$$Y : \Omega \to \mathcal{Y} = \{ \overset{S}{\text{spade}}, \overset{H}{\text{heart}}, \overset{C}{\text{club}}, \overset{D}{\text{diamond}} \}$$

$$P_{XY}(\diamond 10) = \frac{1}{52} = P_X(X=10)\, P_Y(Y=\diamond) = \frac{1}{13} \times \frac{1}{4}$$

Real Random variables.

A Real R.V. is a function $X : \Omega \to \mathbb{R}$.

e.g. The height of a random person. , Value of a die

$X :=$ the rank of poker card is real R.V.

if we identify $A=1$ $J=11$. $Q=12$. $K=13$

$Y :=$ type of poker card is not

In the discrete case : if $X : \Omega \to \mathcal{X}$ is a R.V.

the prob. distribution $P_X : \mathcal{X} \to [0,1]$ is a Real R.V.

For two R.V. $X : \Omega \to \mathcal{X}$

$Y : \Omega \to \mathcal{Y}$

one can define $P_{X|Y} : \mathcal{X} \times \mathcal{Y} \to [0,1]$ a Real R.V.

$P_{X|Y}(x|y) = P(X=x \mid Y=y)$

What is special of Real random variables?

Given $X, Y : \Omega \to \mathbb{R}$

We can define : $X+Y$, $X \cdot Y$, $f(X)$ as real R.V.

where $f : \mathbb{R} \to \mathbb{R}$ is a real function.

Expection and Variance.

Let $X : \Omega \to \mathcal{X} \subseteq \mathbb{R}$ be a discrete real R.V.

① Expectation (or mean)

$$\mathbb{E}X = \sum_{x \in \mathcal{X}} x P_X(x) = \sum_{w \in \Omega} X(w) P(w)$$

② Variance

$$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}|X - \mathbb{E}X|^2$$

Example: $X :=$ Value of a fair die $\quad X : \Omega \to \{1,2,3,4,5,6\}$

$$\mathbb{E}X = \sum_{j=1}^{6} \frac{1}{6} j = \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

$$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$$
$$= \sum_{j=1}^{6} \frac{1}{6} j^2 - (3.5)^2 = \frac{91}{6} - (3.5)^2 = \frac{35}{12}$$

Prop. Let $X, Y : \Omega \to \mathbb{R}$.

① $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$, $\quad \mathbb{E}(cX) = c\,\mathbb{E}X \quad c \in \mathbb{R}$

$Var(cX) = c^2 Var(X)$

If X, Y are independent

② $\mathbb{E}(X \cdot Y) = (\mathbb{E}X)(\mathbb{E}Y)$

③ $Var(X+Y) = Var(X) + Var(Y)$

Pf: $\mathbb{E}(X+Y) = \sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) P(\omega) = \sum_{\omega \in \Omega} X(\omega) P(\omega) + \sum_{x \in \Omega} Y(\omega) P(\omega)$

$= \mathbb{E}X + \mathbb{E}Y$

② $\mathbb{E}(X \cdot Y) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} xy\, P_{XY}(X=x, Y=y) = \sum xy\, P_X(X=x) \cdot P_Y(Y=y)$

independence

$= \left(\sum_x x P_X(X=x)\right)\left(\sum_y y\, P_Y(Y=y)\right)$

$= \mathbb{E}X \;\mathbb{E}Y$

③ $Var(X+Y) = \mathbb{E}(X+Y)^2 - (\mathbb{E}(X+Y))^2$

$= \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}X + \mathbb{E}Y)^2$

$= \mathbb{E}X^2 + \mathbb{E}Y^2 + \underbrace{2\mathbb{E}XY} - [(\mathbb{E}X)^2 + \underbrace{2\mathbb{E}X \cdot \mathbb{E}Y} + (\mathbb{E}Y)^2]$

$= Var(X) + Var(Y)$

Law of large number

A sequence of R.V. $X_1, X_2, \cdots X_n$ i.i.d $\sim P_X$ if

① $\forall i \quad X_i \sim P_X$

② $X_1 \cdots X_n$ mutually independent

e.g. $(X_1, X_3, X_6) : \Omega \to \mathbb{R}^3$

is independent to (X_2, X_4, X_7)

Thm (Weak L.L.N)

Let $X_i, i \in \mathbb{N}$ be an infinite i.i.d sequence subject to P_X.

Denote $\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$. Suppose $Var(X)$ and $\mathbb{E}X < +\infty$

For $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(|\overline{X}_n - \mathbb{E}X| < \varepsilon) = 1$$

Chebyshev's Inequality

$$P(|X - \mathbb{E}X| > \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

Pf: $$Var(X) = \mathbb{E}|X - \underset{\underset{\mathbb{E}X}{\parallel}}{u}|^2 \geq \sum_{|x-u|>\varepsilon} |x-u|^2 P(x) + \sum_{|x-u|<\varepsilon} |x-u| P(x)$$

$$\geq \sum_{|x-u|>\varepsilon} \varepsilon^2 P(x) \geq \varepsilon^2 P(|X-u|>\varepsilon)$$

Pf of Weak L.L.N: $$\mathbb{E}\overline{X}_n = \mathbb{E}\frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}\mathbb{E}X_1 + \cdots \mathbb{E}X_n$$
$$= \frac{1}{n} \cdot n\mathbb{E}X = \mathbb{E}X$$

$$Var(\overline{X}_n) = Var(\frac{1}{n}(X_1 + \cdots X_n))$$
$$= \frac{1}{n^2} Var(X_1 + \cdots + X_n)$$

$$= \frac{1}{n^2} Var(X_1) + \cdots Var(X_n)$$

$$= \frac{1}{n^2} \cdot n \, Var(X) = \frac{1}{n} Var(X)$$

Then $P(|\overline{X_n} - \mathbb{E}X| \geq \varepsilon) = P(|\overline{X_n} - \mathbb{E}\overline{X_n}| \geq \varepsilon) \leq \frac{Var(\overline{X_n})}{\varepsilon^2}$

$$= \frac{Var(X)}{n\varepsilon^2} \to 0$$

So $P(|\overline{X_n} - \mathbb{E}X| < \varepsilon) = 1 - P(|\overline{X_n} - \mathbb{E}X| \geq \varepsilon) \to 1$ $\square$

<u>Example</u>: A Bernoulli R.V. has distribution ($X \in \{0,1\}$)

$$P_X(X=1) = p \qquad P_X(X=0) = 1-p$$

$\mathbb{E}X = p \qquad Var(X) = p(1-p) \qquad \lim_{n\to\infty} \frac{1}{n}(X_1 + \cdots X_n) = p$ almose. surely.

Vector valued R.V.: $\mathbb{R}^n = \{(a_1, a_2 \cdots a_n) \mid a_i \in \mathbb{R}\}$

$$X = (X_1, X_2, \cdots, X_n) : \Omega \to \mathbb{R}^n$$

has $\mathbb{E}X$. $Var(X)$. L.L.N as above