Joint Distributions (I)

Many experiments/applications require that we use more than one characteristic/variable related to the experimental units.

$\Updownarrow$

① Studying Body mass index (BMI) $\rightarrow$ need height and weight.

② Predicting house prices $\rightarrow$ depends on many factors: location, house area, # bedrooms, bathrooms etc

③ Predicting if the price of a stock $\rightarrow$ depends on many performance metrics

$\Updownarrow$

perhaps how Elon tweets!

In all of these cases $\rightarrow$ need to calculate the probability of two or more events simultaneously.

$\Updownarrow$

Want the "Joint Distribution" of variables of interest.

Joint Distributions for two random Variables

Let X, Y be two r.vs $\rightarrow$ both discrete

or

$\rightarrow$ both continuous.

(I) Two Discrete random Variables

Suppose X, Y are two discrete r.vs $\rightarrow$ $\mathcal{X}$ = values of X, $\mathcal{Y}$ = values of Y

$\Updownarrow$

$$\mathcal{X}\times\mathcal{Y} = \{(x,y)\,;\, x\in\mathcal{X}, y\in\mathcal{Y}\}$$

$\Updownarrow$

"Cartesian Product" of $\mathcal{X}$ and $\mathcal{Y}$.

$\Updownarrow$

the "joint sample space"

The "joint probability mass function", i.e. "joint pmf" of X and Y $\rightarrow$

$$p(x,y) := P(X=x \text{ and } Y=y)$$

$\Updownarrow$

$$P\begin{pmatrix} X \text{ takes value } x \text{ and} \\ Y \text{ takes value } y \end{pmatrix}$$

Note: ① $p(x,y)$ satisfies

(i) $p(x,y) \geq 0 \quad \forall (x,y)\in\mathcal{X}\times\mathcal{Y}$

(ii) $\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y) = 1$.

② If $A \subseteq \mathcal{X}\times\mathcal{Y}$ $\rightarrow$ an event in the joint dist

$\Updownarrow$

$$P(A) = \sum_{(x,y)\in A} p(x,y)$$

③ Marginal and Conditional Distributions

(i) "Marginal probability mass function of X" $\rightarrow$ for fixed $x\in\mathcal{X}$

$$p_X(x) := \sum_{y\in\mathcal{Y}} p(x,y)$$

(ii) "Marginal probability mass function of Y" $\rightarrow$ for fixed $y\in\mathcal{Y}$

$$p_Y(y) := \sum_{x\in\mathcal{X}} p(x,y)$$

(iii) "Conditional Distributions"

$$p_{X|Y}(x|y) := P\begin{pmatrix} X=x \text{ given that} \\ Y=y \end{pmatrix} \Updownarrow := \frac{p(x,y)}{p_Y(y)}$$

$$p_{Y|X}(y|x) := P\begin{pmatrix} Y=y \text{ given} \\ \text{that } X=x \end{pmatrix} \Updownarrow := \frac{p(x,y)}{p_X(x)}$$

Important identity

$$p(x,y) = p_{X|Y}(x|y)\cdot p_Y(y) = p_{Y|X}(y|x)\cdot p_X(x)$$

(II) Two Continuous random Variables

Let X, Y be two continuous r.vs $\rightarrow$ "joint sample space" is given by $\mathbb{R}\times\mathbb{R} = \mathbb{R}^2$.

"The joint probability function" ie "joint pdf" for the r.v's X and Y $\rightarrow$ function $f: \mathbb{R}\times\mathbb{R} \rightarrow \mathbb{R}$ satisfying

$\Updownarrow$

(i) $f(x,y) \geq 0 \quad \forall (x,y)\in\mathbb{R}^2$

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

Note:

① If $A \subseteq \mathbb{R}^2$,

$$P((X,Y)\in A) = \iint_A f(x,y)\,dx\,dy$$

$\Updownarrow$

If $A = \{(x,y) : x\in[a,b], y\in[c,d]\}$ = 2D rectangle $[a,b]\times[c,d]$

$$P((X,Y)\in A) = \int_a^b\int_c^d f(x,y)\,dx\,dy$$

$\|$

$$P(a\leq X\leq b \text{ and } c\leq Y\leq d)$$

② Marginal and Conditional Distributions

(i) "Marginal probability density function for X" $\rightarrow$

$$f_X(x) := \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{for } x\in(-\infty,\infty)$$

(ii) "Marginal probability density function for Y" $\rightarrow$

$$f_Y(y) := \int_{-\infty}^{\infty} f(x,y)\,dx \quad \text{for } y\in(-\infty,\infty)$$

③ Conditional Probability distribution

The "conditional density function of Y given that $X=x$" $\rightarrow$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \quad y\in(-\infty,\infty)$$

Similarly,

The "conditional density function of X given that $Y=y$" $\rightarrow$

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} \quad \text{for } x\in(-\infty,\infty)$$

Important identity:

$$f(x,y) = f_{X|Y}(x|y)\cdot f_Y(y) = f_{Y|X}(y|x)\cdot f_X(x)$$

Independent Random Variables

We say X and Y are independent if

$$p(x,y) = p_X(x)\cdot p_Y(y) \quad \forall (x,y)\in\mathcal{X}\times\mathcal{Y}$$

if X and Y are discrete

or

$$f(x,y) = f_X(x)\cdot f_Y(y) \quad \forall (x,y)\in\mathbb{R}\times\mathbb{R}$$

if X and Y are continuous.

if conditions for independence are not satisfied we say X and Y are dependent.

Note: When X and Y are independent, can calculate the joint pdf by "multiplying" the pdf's of X and Y

$\Updownarrow$

In general, joint pdf/pmf calculation is not so easy!!