Discrete Multivariate Distributions

REVIEW

Multivariate Distributions

* often, there is need to define multiple RVs in experiment

For RVs X, Y, the joint/bivariate distribution is the collection of probabilities

$$P((X,Y) \in B),\ \forall B \subseteq \mathbb{R}^2$$

(note that B represents an arbitrary subset of $\mathbb{R}^2$)

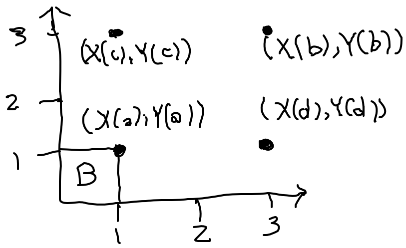
Example

$S = \{a, b, c, d\}$

$$X(s) = \begin{cases} 1: s=a, s=c \\ 3: s=b, s=d \\ 0: \text{o/w} \end{cases}$$

$$Y(s) = \begin{cases} 1: s=a, s=d \\ 3: s=b, s=c \\ 0: \text{o/w} \end{cases}$$

If $B = [0,1] \times [0,1]$ (unit square)



Note that only $s=a$ resuts in $(X(s), Y(s)) \in B$

Then $P((X,Y) \in B) = P(\{a\}) = \dfrac{|\{a\}|}{|S|} = \dfrac{1}{4}$

Joint PMF

For discrete RVs X, Y, the joint/bivariate PMF is defined as

$$p_{X,Y}(x,y) = P(X=x, Y=y) = P(\{X=x\} \cap \{Y=y\})$$

Marginal PMF

Sometimes, you may want the distribution of single variable from joint distribution (called marginal distribution)

$$p_X(x) = P(X=x) = \sum_y P(X=x, Y=y) = \sum_y p_{X,Y}(x,y)$$

Example

$S = \{a, b, c, d\}$

$$X(s) = \begin{cases} 1 : s=a, s=c \\ 3 : s=b, s=d \\ 0 : o/w \end{cases}$$

$$Y(s) = \begin{cases} 1 : s=a, s=d \\ 3 : s=b, s=c \\ 0 : o/w \end{cases}$$

Joint PMF

$$p_{X,Y}(x,y) = P(X=x, Y=y)$$

$$= \begin{cases} 1/4 : X=1, Y=1 \\ 1/4 : X=1, Y=3 \\ 1/4 : X=3, Y=1 \\ 1/4 : X=3, Y=3 \\ 0 : o/w \end{cases}$$

Marginal PMFs

$$p_X(1) = P(X=1) = \sum_y P(X=1, Y=y)$$
$$= P(X=1, Y=1) + P(X=1, Y=3)$$
$$= 1/4 + 1/4$$
$$= 1/2$$

$$p_X(3) = P(X=3) = \sum_y P(X=3, Y=y)$$
$$= P(X=3, Y=1) + P(X=3, Y=3)$$
$$= 1/2$$

Thus,

$$p_X(x) = \begin{cases} 1/2 : x=1 \\ 1/2 : x=3 \\ 0 : o/w \end{cases}$$

Similar for $p_Y(y)$

Conditional PMF

For RVs X, Y, the conditional PMF of X given $Y=y$ is

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

Remember: $p_{X|Y}(x|y)$ is just a function of x – don't be thrown off by notation

This means $p_{X|Y}(x|y)$ must be proper PMF

- $p_{X|Y}(x|y) \geq 0$
- $\sum_x p_{X|Y}(x|y) = 1$, for some arbitrary y

Example

$S = \{a, b, c, d\}$

$$X(s) = \begin{cases} 1 : s=a, s=c \\ 3 : s=b, s=d \\ 0 : o/w \end{cases}$$

$$Y(s) = \begin{cases} 1 : s=a, s=d \\ 3 : s=b, s=c \\ 0 : o/w \end{cases}$$

$$p_{X,Y}(x,y) = \begin{cases} 1/4 : x=1, y=1 \\ 1/4 : x=3, y=1 \\ 1/4 : x=1, y=3 \\ 1/4 : x=3, y=3 \\ 0 : o/w \end{cases}$$

$$p_Y(y) = \begin{cases} 1/2 : y=1 \\ 1/2 : y=3 \\ 0 : o/w \end{cases}$$

For $y=1$ (similar for $y=3$)

$$p_{X|Y}(x|1) = \frac{p_{X|Y}(x,1)}{p_Y(1)}$$

$$= \frac{1}{\frac{1}{2}} \cdot p_{X,Y}(x,1)$$

$$= \begin{cases} 2/4 : x=1 \\ 2/4 : x=3 \\ 0 : o/w \end{cases}$$

Independent RVs

For discrete RVs X, Y, X and Y are independent when
$p_{X,Y}(x,y) = P(X=x, Y=y) = P(X=x)\cdot P(Y=y) = p_X(x)p_Y(y) \quad \forall x,y$

Conditional Variables from Independent Variables

If X, Y are indepent

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

Example

$S = \{a, b, c, d\}$

$$X(s) = \begin{cases} 1: s=a, s=c \\ 3: s=b, s=d \\ 0: o/w \end{cases}$$

$$Y(s) = \begin{cases} 1: s=a, s=d \\ 3: s=b, s=c \\ 0: o/w \end{cases}$$

Turns out $X \perp Y$ (i.e. $p_{X,Y}(x,y) = p_X(x)p_Y(\cancel{x}y)$)

We have

$$p_X(x) = \begin{cases} 1/2: x=1 \\ 1/2: x=3 \\ 0: o/w \end{cases}$$

$$p_Y(y) = \begin{cases} 1/2: y=1 \\ 1/2: y=3 \\ 0: o/w \end{cases}$$

$$p_{X,Y}(x,y) = \begin{cases} 1/4: x=1, y=1 \\ 1/4: x=1, y=3 \\ 1/4: x=3, y=1 \\ 1/4: x=3, y=3 \\ 0: o/w \end{cases}$$

Easy to see $p_{X,Y}(x,y) = p_X(x)p_Y(y)$

This also means $p_{X|Y}(x|y) = p_X(x)$
$p_{Y|X}(y|x) = p_Y(y)$

Q&A

7.a)

$P_Q = \frac{4}{52} = \frac{1}{13}$ (probability queen is drawn on single trial)

$P_K = \frac{1}{13}$ (probability king is drawn on single trial)

$P_O = 1 - \underset{(P_K)}{\frac{1}{13}} - \underset{(P_Q)}{\frac{1}{13}} = \frac{11}{13}$ (probability neither is drawn on single trial)

From Q6 (MULTINOMIAL DISTRIBUTION)

Say you three categories (5 independent trials) $[n=5]$

Also note that we can convert this into a two variable problem $\left[\begin{matrix} X+Y+Z=5 \\ Z=5-X-Y \end{matrix}\right]$

$$P(X=x, Y=y) = \binom{n}{x \;\; y \;\; n-x-y} (p_a)^x (p_k)^y (p_c)^{5-x-y}$$

*Use when considering independent trials with more then two outcomes

7.b)

From 5 (MULTIVARIATE HYPERGEOMETRIC DISTRIBUTION)

$n=5$ w/o replacement

$m=52$

$a=4$ (Queens) $\quad b=4$ (Kings) $\quad c=44$ (other) $\left(\begin{matrix} a+b+c=52 \\ c=5\cancel{4}2-4-4 \end{matrix}\right)$

$$P(X=x, Y=y) = \frac{\binom{a}{x}\binom{b}{y}\binom{c}{5-x-y}}{\binom{m}{n}}$$

$$P(X=x, Y=y) = \frac{\binom{4}{x}\binom{4}{y}\binom{44}{5-x-y}}{\binom{52}{5}}$$

*Use when considering trials w/o replacement and multiple outcomes

3. $$P(X=x, Y=y) = \begin{cases} \frac{1}{mn}, & (x,y) \in \{1,\dots,n\} \times \{1,\dots,m\} \\ 0, & o/w \end{cases}$$

$$P(X \le x, Y \le y) = \begin{cases} 0, & x < 1 \text{ or } y < 1 \\ \frac{xy}{mn}, & 1 \le x < n,\ 1 \le y \le m \\ \frac{x}{n}, & 1 \le x < n,\ y \ge m \\ \frac{y}{m}, & x \ge n,\ 1 \le y < m \\ 1, & x \ge n,\ y \ge m \end{cases}$$