STAT 345/445 Lecture 14

Multiple Random Variables

Conditional Distributions and Independence – Section 4.2

Discrete conditional distributions

- Recall the definition of conditional probability:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- Let (X, Y) be a discrete random vector with joint pmf $f(x, y)$ and marginal pmf's $f_X(x)$ and $f_Y(y)$. $(\ast)$
- For any x such that $f(x) > 0$ we define the **conditional pmf of Y given $X = x$** as

$$f(y \mid x) = \frac{f(x,y)}{f_X(x)} \quad \left(= \frac{P(Y=y, X=x)}{P(X=x)} \right)$$

- Note that $f(y \mid x)$ is a pmf:

(*) Then :

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

$f(x,y)$

$f_X(x)$

$f(y|x)$: x is fixed, y is the variable.

Shorthand for $f(y|X=x)$

pmf? — as a function of y

$f(y|x) \geq 0 \quad \forall y$, true since $f(x,y) \geq 0$
and $f_X(x) > 0$ (condition)

and $\sum_{\text{all } y} f(y|x) = \sum_{\text{all } y} \frac{f(x,y)}{f_X(x)} = \frac{1}{f_X(x)} \sum_y f(x,y)$

$= \frac{1}{f_X(x)} \cdot f_X(x) = 1$

Continuous conditional distributions

- Let (X, Y) be a continuous random vector with joint pdf $f(x, y)$ and marginal pdf's $f_X(x)$ and $f_Y(y)$.

- For any x such that $f(x) > 0$ we define the **conditional pdf of Y given $X = x$** as

$$f(y \mid x) = \frac{f(x, y)}{f_X(x)}$$

- Note that $f(y \mid x)$ is a pdf:

$$\int f(y \mid x)\, dy = \frac{1}{f_X(x)} \int f(y, x)\, dy = \frac{1}{f_X(x)} \cdot f_X(x) = 1$$

For $n > 2$

$$f(x_1, \ldots, x_q \mid x_{q+1}, \ldots, x_n) = \frac{f(x_1, \ldots, x_n)}{f(x_{q+1}, \ldots x_n)}$$

Conditional expectation

- Discrete case

$$E(g(Y) \mid X = x) = \sum_y g(y) f(y \mid x)$$

- Continuous case

$$E(g(Y) \mid X = x) = \int_{-\infty}^{\infty} g(y) f(y \mid x) dy$$

= Expectation w.r.t. a conditional distr.

Example : 3 coins

A fair coin is tossed three times. Let

- X = number of heads on the first toss
- Y = total number of heads

The joint pmf $f(x, y)$ can be given in a table:

x	y 0	1	2	3	$f_X(x)$
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0	$\frac{1}{2}$
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{2}$
$f_Y(y) =$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	

Find the

- conditional pmf of Y given $X = 0$
- conditional pmf of X given $Y = 1$

Given that we did not get a head in the first toss,

$Bin(n=2, p=\frac{1}{2})$

Y is just counting the number of heads in the last 2 tosses

3 coins

Y can take values $\{0,1,2,3\}$

$f(y|0) = \frac{f(0,y)}{f_X(0)}$ for $y \in \{0,1,2,3\}$

i.e. $f(0|0) = \frac{f(0,0)}{f_X(x)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$

$f(1|0) = \frac{f(0,1)}{f_X(0)} = \frac{\frac{2}{8}}{\frac{1}{2}} = \frac{1}{2}$

$f(2|0) = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$

$f(3|0) = \frac{0}{\frac{1}{2}} = 0$

$X \mid Y=1, \quad x \in \{0,1\}$

$$f(x|1) = \frac{f(x,1)}{f_Y(1)}$$

$$f(0|1) = \frac{f(0,1)}{f_Y(1)} = \frac{\frac{2}{8}}{\frac{3}{8}} = \frac{2}{3}$$

$$f(1|1) = \frac{f(1,1)}{f_Y(1)} = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{3}$$

Bernoulli $(p = \frac{1}{3})$

Given $y=1$, i.e. we got exactly 1 head in the 3 tosses, (don't know in what order) the prob of a head in the first toss goes down to $\frac{1}{3}$.

Example: Multinomial distribution

$$f(\mathbf{x} \mid m, \mathbf{p}) = \frac{m!}{x_1!x_2!\cdots x_n!} p_1^{x_1} p_2^{x_2} \cdots p_n^{x_n} \qquad \text{for } \mathbf{x} \in \mathbb{N}^n$$

where $x_1 + x_2 + \cdots + x_n = m$ $\quad p_1 + p_2 \cdots + p_n = 1$

- What is the conditional distribution of $(X_1, X_2, \ldots, X_{n-1})$ given $X_n = x_n$?

Multinomial distribution

Recall: The marginal of X_n is Binomial (m, p_n), So the pdf

$$f(x_1, x_2, \cdots x_{n-1} \mid x_n) = \frac{f(x_1, x_2, \cdots x_{n-1}, x_n)}{f_n(x_n)} = \frac{\frac{\cancel{m!}}{x_1! \cdots \cancel{x_n!}} p_1^{x_1} \cdots p_n^{x_n}}{\frac{\cancel{m!}}{\cancel{x_n!}(m-x_n)!} p_n^{x_n} (1-p_n)^{m-x_n}} \quad \cdots (*)$$

$\binom{m}{x_n} \longrightarrow$

Note: $m - x_n = x_1 + x_2, \cdots x_{n-1}$

$$(1-p_n)^{m-x_n} = (1-p_n)^{x_1 + \cdots + x_{n-1}}$$

$$= (1-p_n)^{x_1} \cdots (1-p_n)^{x_{n-1}}$$

$$(*) = \frac{(m-x_n)!}{x_1! \cdots x_{n-1}!} \frac{p_1^{x_1} \cdots p_{n-1}^{x_{n-1}}}{(1-p_1)^{x_1} \cdots (1-p_n)^{x_{n-1}}}$$

$\leftarrow$ pmf for Multinomial$\left(m-x_n, \left(\frac{p_1}{1-p_n}, \dots, \frac{p_{n-1}}{1-p_n}\right)\right)$

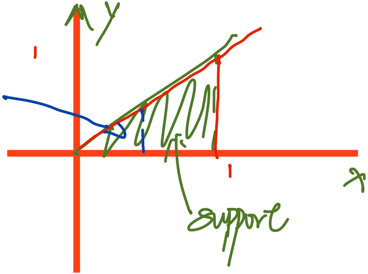
Continuous example 2

$$f(x,y) = \begin{cases} 8xy & \text{if } 0 < y < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

- Find the conditional pdf of Y given $X = x$
- Find the conditional mean and variance of Y given $X = x$ and compare to marginal mean and variance of Y.

Always sketch the support!

Possible range of Y when $X = x$



Continuous pdf example 2 Already found: $f_X(x) = 4x^3$, $0 < x < 1$, $f_Y(y) = 4y - 4y^3$, $0 < y < 1$

For any $x \in (0,1)$

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{8xy}{4x^3} = \frac{2y}{x^2} \quad \text{for } 0 < y < x < 1$$

e.g. if $x = \frac{1}{2}$ we get $f(y|\frac{1}{2}) = \frac{2y}{(\frac{1}{2})^2} = 8y$, for $0 < y < \frac{1}{2}$

Conditional mean:

$$E(Y|X=x) = \int_{-\infty}^{\infty} y\, f(y|x)\, dy = \int_0^x \frac{2y^2}{x^2}\, dy = \frac{2y^3}{3x^2}\Big|_0^x = \frac{2}{3}x$$

$\Longrightarrow$ Depends on the value of x.

Marginal mean: $E(Y) = \int_0^1 y(4y - 4y^3)\,dy = \frac{4}{3}y^3 - \frac{4y^5}{5}\Big|_0^1$

$= \frac{8}{15} \approx 0.533$

Conditional Variance:

$$E(Y^2 \mid X=x) = \int_0^x \frac{2y^3}{x^2}\,dy = \frac{2y^4}{4x^2}\Big|_0^x = \frac{1}{2}x^2 \Longrightarrow$$

$$Var(Y \mid X=x) = E(Y^2 \mid X=x) - (E(Y \mid X=x))^2 = \frac{1}{2}x^2 - \left(\frac{2}{3}x\right)^2$$

$$= \frac{1}{18}x^2$$

Marginal Variance:

$$E(Y^2) = \int_0^1 4y^3 - 4y^5\,dy = \frac{4y^4}{4} - \frac{4}{6}y^6\Big|_0^1 = 1 - \frac{4}{6} = \frac{1}{3}$$

$$\mathrm{Var}(y) = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{225} \approx 0.043$$

E.g. if $x = \frac{1}{2}$, $V(y \mid x = \frac{1}{2}) = \frac{1}{18} \cdot \frac{1}{4} = 0.0139$

Usually $\mathrm{Var}(y \mid X = x) < \mathrm{Var}(y) \Rightarrow$ knowing the value of X lowers uncertainty about y.

Can we get $\mathrm{Var}(y \mid X = x) > \mathrm{Var}(y)$ for some x?

$$\frac{x^2}{18} > \frac{11}{225} \Rightarrow x = 0.938$$

Yes, for large values of x.

Independent random variables

Definition

Let (X, Y) be a random vector with joint pdf/pmf $f(x, y)$ and marginal pdfs/pmfs $f_X(x)$ and $f_Y(y)$. If

$$f(x, y) = f_X(x)f_Y(y) \qquad \text{for all } x \in \mathbb{R} \text{ and } y \in \mathbb{R}$$

then X and Y are called **independent random variables**.

- If X and Y are independent then

$$f(y \mid x) = f_Y(y) = \frac{f(x,y)}{f_X(x)} = \frac{f_X(x)f_Y(y)}{f_X(x)}$$

$$= f_Y(y)$$

Example: 3 coins

The joint pmf $f(x, y)$ and marginal pmfs:

x	y = 0	1	2	3	$f_X(x)$
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0	$\frac{1}{2}$
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{2}$
$f(y) =$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	

Note: $f(x,y) = f_X(x) \cdot f_Y(y)$, $(*)$ has to hold for all x's and all y's for X and y to be independent.

- Are X and Y independent?

$\Rightarrow$ To show that X and Y are not independent, i.e. is enough to find one x and one y where $*$ does not hold.

E.g. $x = 1$, $y = 0$

$f(1, 0) = 0$

$f_X(1) f_Y(0) = \frac{1}{2} \cdot \frac{1}{8} \neq 0$

or take $x = 1$ and $y = 2$, $f(1, 2) = \frac{2}{8} \neq \frac{1}{2} \cdot \frac{3}{8}$ $\Rightarrow$ X independent.

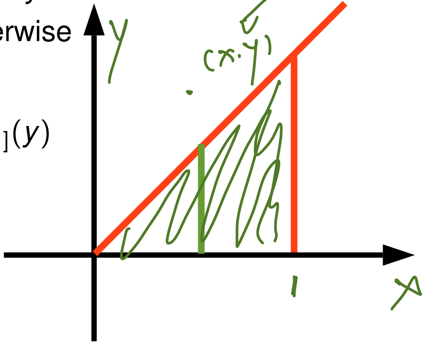
Continuous example 2

Usually, when support is not rectangular, X and Y can't be independent.

$$f(x,y) = \begin{cases} 8xy & \text{if } 0 < y < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_X(x) = 4x^3 \; I_{[0,1]}(x)$$

$$f_Y(y) = 4(y - y^3) \; I_{[0,1]}(y)$$



- Are X and Y independent?

E.g. $x = \frac{1}{2}$, $y = \frac{3}{4}$

$f(\frac{1}{2}, \frac{3}{4}) = 0$

$f_X(\frac{1}{2}) \cdot f_Y(\frac{3}{4}) = 0.72 \neq 0 \Rightarrow$ No, not independent.

Continuous example 1

$$f(x,y) = \begin{cases} 4xy & \text{if } 0 < y < 1 \text{ and } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_X(x) = 2x\ I_{[0,1]}(x)$$

$$f_Y(y) = 2y\ I_{[0,1]}(y)$$

Yes, x and y are independent.

- Are X and Y independent?

$$f_X(x) \cdot f_Y(y) = 2x\ I_{[0,1]}(x)\ 2y\ I_{[0,1]}(y)$$

$$= 4xy\ I_{[0,1]}(x,y) = f(x,y),\ \forall x, y$$

Independent random variables

Lemma

Let (X, Y) be a random vector with join pdf/pmf $f(x, y)$.
X and Y are independent if and only if there exist functions $g(x)$ and $h(y)$ such that

Look out for non-rectangular support.

$$f(x, y) = g(x)h(y) \qquad \text{for all } x \in \mathbb{R} \text{ and } y \in \mathbb{R}$$

Proof... Lemma: $\underbrace{f(x,y) = g(x)h(y) \ \forall x, y}_{(*)} \iff x, y$ independent

" $\Leftarrow$ " Suppose X and Y are independent, then by def.

$f(x,y) = f_X(x) \cdot f_Y(y) \quad \forall x, y$

$g(x) = f_X(x)$ and $h(y) = f_Y(y)$ and we have $(*)$

"$\Rightarrow$" Suppose $f(x,y) = g(x)h(y)$ $\forall x,y$ then

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{-\infty}^{\infty} g(x)h(y)\,dy = g(x)\overbrace{\int_{-\infty}^{\infty} h(y)\,dy}^{=c}$$

$$= c\,g(x)$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{-\infty}^{\infty} g(x)h(y)\,dx = h(y)\underbrace{\int_{-\infty}^{\infty} g(x)\,dx}_{=d}$$

$$= d\,h(y)$$

Also, $cd = 1$ since

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

$$= \int_{\infty}^{-\infty} h(y)\underbrace{\left(\int_{\infty}^{-\infty} g(x)\,dx\right)}_{d} dy = d\int_{-\infty}^{\infty} h(y)\,dy = cd$$

$\Rightarrow f_X(x) f_Y(y) = c\,g(x)\,d\,h(y) = g(x)h(y) = f(x,y) \Rightarrow x, y$ are indep.

Continuous examples 1 and 2

Example 1: $$f(x,y) = \begin{cases} 4xy & \text{if } 0 < y < 1 \text{ and } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Example 2: $$f(x,y) = \begin{cases} 8xy & \text{if } 0 < y < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

- Determine independence using the lemma

① : Set $g(x) = 4x\, I_{[0,1]}(x)$ and $h(y) = y\, I_{[0,1]}(y)$

$\Rightarrow$ indep. by Lemma.

② : $f(x,y) = 8xy\, \underbrace{I_{[0,x]}(y)}_{}\, I_{[0,1]}(x)$

can't write as a function of x only or y only

More on independent random variables

$E(XY) = E(X)E(Y)$ if X, Y independent

Theorem

Let X and Y be *independent* random variables. Then

(a) For any $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$ we have Think $P(A \cap B) = P(A)P(B)$

special case of (b) since $E[I_A(X)] = P(X \in A)$

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

i.e. $\{X \in A\}$ and $\{Y \in B\}$ are independent events

(b) Let $g(x)$ be a function of x only and let $h(y)$ be a function of y only. Then

$$E_{X,Y}(g(X)\, h(Y)) = E_X(g(X))\, E_Y(h(Y))$$

Proof...

- Note that $E(I_A(X)) = P(X \in A)$

(a) is special case of (b) since $E[I_A(x)] = P(X \in A)$, recall: $I_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases}$

$$E(I_A(x)) = \int_{-\infty}^{\infty} I_A(x) f(x)\,dx = \int_A f(x)\,dx = P(X \in A)$$

Or think $Y = I_A(x) \sim \text{Bernoulli}(p)$

$$p = P(Y=1) = P(X \in A)$$

Know: $E(Y) = p \Rightarrow E(I_A(x)) = P(X \in A)$

Proof of (b): Just write out the integrals.

$$E(g(X)h(Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)\,h(y)\,f_X(x)\,f_Y(y)\,dx\,dy$$

since $f(x,y) = f_X(x)\,f_Y(y)$

$$= \int h(y)\,f_Y(y) \underbrace{\int_{-\infty}^{\infty} g(x)\,f_X(x)\,dx}_{=E(g(X))}\,dy$$

$$= E(g(X)) \int_{-\infty}^{\infty} h(y)\,f_Y(y)\,dy = E(g(X))\,E(h(Y))$$

More on independent random variables

Theorem

Let X and Y be independent random variables with mgf's $M_X(t)$ and $M_Y(t)$. Then the mgf of $Z = X + Y$ is

X only   Y only

$$M_Z(t) = E(e^{tZ}) = E(e^{t(X+Y)}) = E(e^{tX} \cdot e^{tY})$$

$$= E(e^{tX}) \cdot E(e^{tY})$$ since X, Y indep

$$= M_X(t) M_Y(t)$$

Example: Sum of independent r.v.

- Let $X \sim \mathrm{Poisson}(\lambda_1)$, $Y \sim \mathrm{Poisson}(\lambda_2)$ be independent r.v. What is the distribution of $X + Y$?

Recall: $M_X(t) = \exp(\lambda_1 - \lambda_1 e^t)$, $M_Y(t) = \exp(\lambda_2 - \lambda_2 e^t)$

$$\Rightarrow M_{X+Y}(t) = M_X(t)\, M_Y(t)$$

$$= \exp(\lambda_1(1-e^t)) \cdot \exp(\lambda_2(1-e^t))$$

$$= \exp((\lambda_1+\lambda_2)(1-e^t))$$

$$= \text{mgf of Poisson}(\lambda_1+\lambda_2)$$

→ Notice how handy it can be that a mgf uniquely identifies a distribution.

Example: Sum of independent r.v.

- Let $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$ be independent r.v. What is the distribution of $X + Y$?

$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = \exp(\mu_1 t + \sigma_1^2 t^2/2) \cdot \exp(\mu_2 t + \sigma_2^2 t^2/2)$

mgf of normal:

$\exp(\text{mean}\, t + \text{Var}\, t^2/2)$

$= \exp((\mu_1+\mu_2)t + (\sigma_1^2+\sigma_2^2)t^2/2)$

$=$ mgf of $N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$

$X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$