Lecture Stat 302
Introduction to Probability - Slides 20

AD

April 2010

Conditional Distributions: Discrete Case

- Given a joint p.m.f. for two r.v. X, Y it is possible to compute the conditional p.m.f. X given $Y = y$.
- Assume X, Y are discrete-valued r.v. with a joint p.m.f. $p(x, y)$ then the conditional p.m.f. of X given $Y = y$ is

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

- The conditional expectation of $g(X)$ is given by

$$E(g(X)|Y = y) = \sum_x g(x) \cdot p_{X|Y}(x|y).$$

- $E(g(X)|Y = y)$ is a function of y and $E(g(X)|Y)$ is a r.v..

Example: Warm-up

- Let X, Y be two discrete r.v. of joint p.m.f.

$$p(x, y) = \frac{(x + y)}{21}$$

  for $x = 1, 2, 3$ and $y = 1, 2$.

- (a) Show that X and Y are not independent.
- (b) Compute $p_{X|Y}(x|y)$.
- (c) Compute $E(X|Y = y)$.

Example: Warm-up

- (a) X and Y are independent if and only if $p(x,y) = p_X(x)\, p_Y(y)$. We have

$$\begin{aligned} p_X(x) &= \sum_y p(x,y) = p(x,1) + p(x,2) \\ &= \frac{(x+1)}{21} + \frac{(x+2)}{21} = \frac{2x}{21} + \frac{1}{7} \end{aligned}$$

and

$$\begin{aligned} p_Y(y) &= \sum_x p(x,y) = p(1,y) + p(2,y) + p(3,y) \\ &= \frac{(1+y)}{21} + \frac{(2+y)}{21} + \frac{(3+y)}{21} = \frac{2}{7} + \frac{3y}{21}. \end{aligned}$$

Clearly $p(x,y) \neq p_X(x)\, p_Y(y)$ so X, Y are not independent.

- Remark: A safety check consists of checking that $\sum_x p_X(x) = \sum_y p_Y(y) = 1$.

Example: Warm-up

- (b) The conditional p.m.f. is given by

$$\begin{aligned} p_{X|Y}(x|y) &= \frac{p(x,y)}{p_Y(y)} = \frac{(x+y)/21}{(6+3y)/21} \\ &= \frac{x+y}{6+3y} \end{aligned}$$

- (c) The conditional mean is given by

$$\begin{aligned} E(X|Y=y) &= \sum_x x \cdot p_{X|Y}(x|y) \\ &= \frac{1+y}{6+3y} + 2 \times \frac{2+y}{6+3y} + 3 \times \frac{3+y}{6+3y} \\ &= \frac{14+6y}{6+3y} \end{aligned}$$

Example: Insurance Company

- An insurance company provides insurance to three groups of staff with the following characteristics

Age	# ind.	Proba claim (in a year)	Expect amount (in a year)
<30	20	0.02	\$500
30 to 50	50	0.04	\$1000
>50	30	0.06	\$1500

1. What is the proba that a randomly selected individual is below the age of 30 and will make a claim in a year?
2. What is the proba that a randomly selected individual will not make any claim in a year?
3. Given that the randomly selected individual has made a claim in a year, what are the proba that he/she is below 30, aged 30 to 50 or older than 50?
4. Given that the randomly selected individual has made a claim in a year and is older than 30, what is the expected claim?

Example: Insurance Company

1. $\Pr(\text{"} < 30\text{"} \cap \text{"Claim"}) = \frac{20}{100} \times 0.02 = 0.004.$
2. We have

$$\begin{aligned}
\Pr(\text{No Claim}) &= \Pr(\text{No Claim} \,|\, \text{"} < 30\text{"}) \Pr(\text{"} < 30\text{"}) \\
&+ \Pr(\text{No Claim} \,|\, \text{"}30 - 50\text{"}) \Pr(\text{"}30 - 50\text{"}) \\
&+ \Pr(\text{No Claim} \,|\, \text{"} > 50\text{"}) \Pr(\text{"} > 50\text{"}) \\
&= (1 - 0.02)\,\tfrac{20}{100} + (1 - 0.04)\,\tfrac{50}{100} + (1 - 0.06)\,\tfrac{30}{100} \\
&= 0.958
\end{aligned}$$

3. We want

$$\begin{aligned}
\Pr(\text{"} < 30\text{"} \,|\, \text{Claim}) &= \frac{\Pr(\text{Claim} \,|\, \text{"} < 30\text{"}) \Pr(\text{"} < 30\text{"})}{\Pr(\text{Claim})} \\
&= \frac{0.02 \times (20/100)}{1 - 0.958} = 0.0952.
\end{aligned}$$

Similarly, we obtain $\Pr(\text{"}30 - 50\text{"} \,|\, \text{Claim}) = 0.4762$ and $\Pr(\text{"} > 50\text{"} \,|\, \text{Claim}) = 0.4286$.

Example: Insurance Company

4a. We have

$$\Pr\left(\text{"}30-50\text{"}\,\middle|\,\text{Claim}\cap\{\text{"}30-50\text{"}\cup\text{"}>50\text{"}\}\right)$$
$$=\frac{\Pr\left(\text{Claim}\,|\{\text{"}30-50\text{"}\cup\text{"}>50\text{"}\}\cap\text{"}30-50\text{"}\right)\Pr\left(\text{"}30-50\text{"}\,|\text{"}30-50\text{"}\cup\text{"}>50\text{"}\right)}{\Pr(\text{Claim}\cap\{\text{"}30-50\text{"}\cup\text{"}>50\text{"}\})}$$
$$=\frac{\Pr\left(\text{Claim}\,|\text{"}30-50\text{"}\right)\Pr\left(\text{"}30-50\text{"}\,|\text{"}30-50\text{"}\cup\text{"}>50\text{"}\right)}{\Pr(\text{Claim}\cap\{\text{"}30-50\text{"}\cup\text{"}>50\text{"}\})}$$

where $\Pr\left(\text{Claim}\cap\{\text{"}30-50\text{"}\cup\text{"}>50\text{"}\}\right)=$
$\Pr\left(\text{Claim}|\text{"}30-50\text{"}\right)\Pr\left(\text{"}30-50\text{"}\,|\,\text{"}30-50\text{"}\cup\text{"}>50\text{"}\right)+$
$\Pr\left(\text{Claim}|\text{"}>50\text{"}\right)\Pr\left(\text{"}30-50\text{"}\,|\,\text{"}30-50\text{"}\cup\text{"}>50\text{"}\right)$.

4b. Hence it follows that the expected amount claimed is

$$\begin{aligned}
& E\left[\text{Amount Claim}\,|\,\text{Claim }\cap\{\text{"}30-50\text{"}\cup\text{"}>50\text{"}\}\right] \\
= \;& 1000\times\Pr\left(\text{"}30-50\text{"}\,|\,\text{Claim}\cap\{\text{"}30-50\text{"}\cup\text{"}>50\text{"}\}\right) \\
& +1500\times\Pr\left(\text{"}>50\text{"}\,|\,\text{Claim}\cap\{\text{"}30-50\text{"}\cup\text{"}>50\text{"}\}\right) \\
= \;& \frac{0.4762}{0.4762+0.4286}\times 1000+\frac{0.4286}{0.4762+0.4286}\times 1500 \\
= \;& 1236,8\$
\end{aligned}$$

Conditional Densities: Continuous Case

- Given a joint p.d.f. for two r.v. X, Y it is possible to compute the conditional p.d.f. of X having observed $Y = y$.
- Assume X, Y are continous-valued r.v. with a joint p.d.f. $f(x, y)$ then the conditional pdf of X given $Y = y$ is

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

- This can be heuristically established by noting that

$$\begin{aligned} f_{X|Y}(x|y)\,dx &= \frac{f(x, y)\,dxdy}{f_Y(y)\,dy} \\ &\approx \frac{P(x \leq X \leq x + dx \cap y \leq Y \leq y + dy)}{P(y \leq Y \leq y + dy)} \\ &= P(x \leq X \leq x + dx|\, y \leq Y \leq y + dy). \end{aligned}$$

- In the case where X and Y are independent, we have $f_{X|Y}(x|y) = f_X(x)$ as $f(x, y) = f_X(x)\,f_Y(y)$.

Conditional Densities: Continuous Case

- We have
$$f(x, y) = f_{X|Y}(x|\, y)\, f_Y(y)$$
and similarly
$$f(x, y) = f_{Y|X}(y|\, x)\, f_X(x)$$
- Hence we obtain
$$f_{X|Y}(x|\, y) = \frac{f_{Y|X}(y|\, x)\, f_X(x)}{f_Y(y)}$$
which holds if $f_Y(y) > 0$.

Conditional Expectation and Variance

- We can define the mean, variance of the conditional p.m.f.
- The conditional mean is given by

$$E\left(X|Y=y\right)=\int xf_{X|Y}\left(x|\,y\right)dx$$

- The conditional variance is given by

$$\begin{aligned}Var\left(X|Y=y\right) &= E\left(\left(X-E\left(X|Y=y\right)\right)^2|Y=y\right)\\ &= E\left(X^2|Y=y\right)-\left\{E\left(X|Y=y\right)\right\}^2\end{aligned}$$

  where

$$E\left(X^2|Y=y\right)=\int x^2 f_{X|Y}\left(x|\,y\right)dx$$

- $E\left(X|Y=y\right)$ and $Var\left(X|Y=y\right)$ are functions but $E\left(X|Y\right)$ and $Var\left(X|Y\right)$ are random variables.

Example: Toy example

- Find the conditional pdf of Y given X and $E(Y|X=x)$ when their joint density is

$$f(x,y) = \lambda^2 e^{-\lambda y} \text{ for } 0 \leq x \leq y \leq \infty$$

- We have for $0 \leq x \leq \infty$

$$f_X(x) = \int_x^\infty \lambda^2 e^{-\lambda y} dy = \lambda e^{-\lambda x}$$

so for $0 \leq x \leq y \leq \infty$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \lambda e^{\lambda(x-y)}$$

- Hence, we obtain

$$E(Y|X=x) = \int_x^\infty y f_{Y|X}(y|x)\, dy \text{ (change of var. } u = y-x\text{)}$$
$$= \int_0^\infty (u+x)\lambda e^{-\lambda u} du = \frac{1}{\lambda} + x.$$

Example: Another Toy example

- Find the conditional pdf of Y given X and $E\left(\left. Y \right| X = x \right)$ when their joint density is

$$f\left(x, y\right) = xe^{-x(y+1)} \text{ for } x, y \geq 0$$

- We have for $0 \leq x \leq \infty$

$$f_X\left(x\right) = \int_0^\infty xe^{-x(y+1)} dy = e^{-x}$$

so that for $0 \leq y \leq \infty$

$$f_{Y|X}\left(\left. y \right| x \right) = \frac{f\left(x, y\right)}{f_X\left(x\right)} = xe^{-xy}$$

- It follows that

$$E\left(Y|X = x\right) = \int_0^\infty y f_{Y|X}\left(\left. y \right| x \right) dy = \frac{1}{x}.$$

Optimality of Conditional Expectation

- Let us consider two r.v. X and Y. Assume we observe Y and want to find a way to estimate X based on Y. Then in some sense, $E(X|Y)$ is the best possible estimate of X.
- **Proposition.** Consider an arbitrary function $g(X)$ then, we have

$$E\left[(X - E(X|Y))^2\right] \leq E\left[(X - g(Y))^2\right],$$

  that is the expected square "distance" between $g(Y)$ and X is minimized for $g(Y) = E(X|Y)$.
- This is valid for both discrete and continuous r.v.