Dr. Z.'s Probability Lecture 16 Handout: Conditional Distributions

By Doron Zeilberger

Version of Nov. 17, 2017 (Please discard previous versions). Thanks to Norman Hong.

Important Definition: **Conditional Distribution (Discrete Case)**

If X and Y are discrete random variables, the

conditional probability mass function of X **given that** $Y = y$, denoted by $p_{X|Y}(x|y)$

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

where, as usual, $p_Y(y)$ is the **marginal distribution** of Y

$$p_Y(y) = \sum_x p(x,y) \quad .$$

Problem 16.1: In a certain community, the probability that a family has i boys and j girls is given by

$$p(i,j) = \begin{cases} \frac{c(i+j+1)}{i+j+2} \quad , & if \quad 0 \leq i \leq 2 \quad and \quad 0 \leq j \leq 2; \\ 0 \quad , & otherwise. \end{cases} \quad ,$$

for some positive constant c (that would make it a discrete probability density function). Calculate the conditional probability mass function for the number of boys in families that have exactly 2 girls.

Sol. to 16.1: Let B be the number of boys and G the number of girls. We need $p_{B|G}(i|2)$ for $i = 0, 1, 2$.

$$p_G(2) = p(0,2) + p(1,2) + p(2,2) = c(\frac{3}{4} + \frac{4}{5} + \frac{5}{6}) = \frac{143c}{60} \quad .$$

Hence

$$p_{B|G}(i|2) = \frac{p(i,2)}{\frac{143c}{60}} = \frac{\frac{c(i+3)}{i+4}}{\frac{143c}{60}} = \frac{60(i+3)}{143(i+4)} \quad , \quad 0 \leq i \leq 2 \quad .$$

Spelling it out

$$p_{B|G}(0|2) = \frac{45}{143} \quad , \quad p_{B|G}(1|2) = \frac{48}{143} \quad , \quad p_{B|G}(2|2) = \frac{50}{143} \quad .$$

Note: For this problem (and any conditional distribution problem) c need not be computed, since it cancels out!

Problem 16.2: In a certain village, the probability that a farm has i cows and j sheep is given by

$$p(i,j) = \begin{cases} \frac{i\,j}{3025} \quad & if \quad 0 \leq i \leq 10 \quad and \quad 0 \leq j \leq 10; \\ 0 \quad , & otherwise. \end{cases} \quad ,$$

If it is known that a farm has 6 sheep, what is the probability that it has at most 7 cows?

Sol. to 16.2: Let C be the number of cows and S the number of sheep. First we need $p_{C|S}(i|6)$ for $0 \leq i \leq 10$. We have, by the general formula:

$$p_{C|S}(i|6) = \frac{p(i,6)}{p_S(6)} \quad .$$

The denominator is

$$p_S(6) = \sum_{i=0}^{10} \frac{6i}{3025} = \frac{6}{3025} \sum_{i=0}^{10} i = \frac{6}{3025} \cdot \frac{10 \cdot 11}{2} = \frac{6}{55} \quad .$$

Hence

$$p_{C|S}(i|6) = \frac{6i/3025}{6/55} = \frac{i}{55} \quad .$$

Finally, if it is known that there are 6 sheep, the probability that there are most 7 cows is

$$\sum_{i=0}^{7} p_{C|S}(i|6) = \sum_{i=0}^{7} \frac{i}{55} = \frac{1}{55} \sum_{i=0}^{7} i = \frac{1}{55} \cdot \frac{7 \cdot 8}{2} = \frac{28}{55} \quad .$$

Ans. to 16.2: If it is known that a farm has 6 sheep, the probability that it has at most 7 cows is $\frac{28}{55}$.

Note: In *this* particular problem, it so happens that the probability mass function, $p(i,j)$, is a **product** of a function of i only and a function of j only, and in addition, it it is non-zero in a set of the form $A_1 \leq i \leq A_2, B_1 \leq j \leq B_2$ (i.e. a **Cartesian product**). In that case the number of sheep and the number of cows are **independent** hence the same answer would be no matter how many sheep a farm has. A shortcut (only applicable to cases where the $p(i,j)$ is separable), is to know that the prob. mass function for marginal distribution $p_C(i)$ is ci for some constant c (that happens to be $\frac{1}{55}$, but it is not needed, since we are talking *conditional probability* and the c would cancel). Hence the desired probability is $(\sum_{i=0}^{7} i)/(\sum_{i=0}^{10} i) = \frac{28}{55}$, regardless of the number of sheep.

Important Definition: **Conditional Distribution (Continuous Case)**

If X and Y are continuous random variables with joint density function $f(x,y)$, the

conditional probability density function of X given that $Y = y$, denoted by $f_{X|Y}(x|y)$, is defined by

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

where, as usual, $f_Y(y)$ is the **marginal distribution** of Y

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx \quad .$$

Note: Of course the formula $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$ only makes sense when the denominator, $f_Y(y)$, is larger than 0.

Analogously

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

Problem 16.3 The joint density function of X and Y is given by

$$f(x,y) = \begin{cases} 2x + 4y \quad , \quad if \quad 0 < x \ , \ 0 < y, \ and \ x + y < 1\,; \\ 0 \quad , \quad otherwise. \end{cases} \quad .$$

(i) Compute the conditional density of X given that $Y = y$ and the conditional density of Y given that $X = x$.

(ii) If it is known that $Y = \frac{1}{2}$, what is the probability that $X \leq \frac{1}{4}$?

Sol. of 16.3 (i): First we compute the marginal distributions. Since $f(x,y)$ is non-zero only for $0 < x < 1 - y$, we have

$$f_Y(y) = \int_0^{1-y} (2x + 4y)\, dx \; = \; 1 + 2y - 3y^2 \quad , \quad 0 < y < 1 \quad .$$

(and of course 0 otherwise). (You do it!) Similarly

$$f_X(x) = \int_0^{1-x} (2x + 4y)\, dy \; = \; 2 - 2x \quad , \quad 0 < x < 1 \quad .$$

(and of course 0 otherwise). (You do it!)

Hence for any $0 < y < 1$ we have

$$f_{X|Y}(x|y) = \begin{cases} \frac{2x + 4y}{1 + 2y - 3y^2} \quad , \quad if \quad 0 < x < 1 - y \\ 0 \quad , \quad otherwise. \end{cases} \quad .$$

and for any $0 < x < 1$ we have

$$f_{Y|X}(y|x) = \begin{cases} \frac{2x + 4y}{2 - 2x} \quad , \quad if \quad 0 < y < 1 - x \\ 0 \quad , \quad otherwise. \end{cases} \quad .$$

Sol. of 16.3(ii): At $y = \frac{1}{2}$ we have

$$f_{X|Y}(x|\frac{1}{2}) = \begin{cases} \frac{2x + 4 \cdot \frac{1}{2}}{1 + 2 \cdot \frac{1}{2} - 3(\frac{1}{2})^2} \quad , \quad if \quad 0 < x < 1 - \frac{1}{2} \\ 0 \quad , \quad otherwise. \end{cases} \quad .$$

That simplifies to

$$f_{X|Y}(x|\frac{1}{2}) = \begin{cases} \frac{8}{5}(x+1) \quad , \quad if \quad 0 < x < \frac{1}{2} \\ 0 \quad , \quad otherwise. \end{cases} .$$

Hence

$$P(X \leq \frac{1}{4} \,|\, Y = \frac{1}{2}) = \int_0^{\frac{1}{4}} f_{X|Y}(x|\frac{1}{2})\, dx \;=\; \frac{8}{5} \int_0^{\frac{1}{4}} (x+1)\, dx \;=\; \frac{9}{20} \quad .$$

Ans. to 16.3(ii): If it is known that $Y = \frac{1}{2}$, then the probability that $X \leq \frac{1}{4}$ equals $\frac{9}{20} = 0.45$.

Problem 16.4 The joint density function of X and Y is given by

$$f(x,y) = \begin{cases} \frac{1}{24}\, x(1+y^2) \quad , \quad if \quad 0 < x < 2\,,\, 0 < y < 3; \\ 0 \quad , \quad otherwise. \end{cases} .$$

(i) Compute the conditional density of X given that $Y = y$ and the conditional density of Y given that $X = x$.

(ii) If it is known that $Y = 1$, what is the probability that $X \leq 1$?

Sol. of 16.4 (i): First we compute the marginal distributions.

$$f_Y(y) = \int_0^2 \frac{1}{24} x(1+y^2)\, dx = \frac{1}{24}(1+y^2) \int_0^2 x\, dx = \frac{1}{24}(1+y^2) \left(\frac{x^2}{2} \Big|_0^2 \right)$$

$$= \frac{1}{24}(1+y^2)(2^2 - 0^2)/2 = \frac{1+y^2}{12} \quad .$$

$$f_X(x) = \int_0^3 \frac{1}{24} x(1+y^2)\, dy = \frac{x}{24} \int_0^3 (1+y^2)\, dy = \frac{x}{24} \left(y + \frac{y^3}{3} \Big|_0^3 \right)$$

$$= \frac{x}{2} \quad .$$

Hence

$$f_{X|Y}(x|y) = \begin{cases} \frac{x}{2}\,, \quad if \quad 0 < x < 2; \\ 0\,, \; otherwise \end{cases} ,$$

$$f_{Y|X}(y|x) = \begin{cases} \frac{1+y^2}{12}\,, \quad if \quad 0 < y < 3; \\ 0\,, otherwise \end{cases} .$$

Sol. of 16.4(ii):

$$P(X \leq 1 \,|\, Y = 1) = \int_0^1 f_{X|Y}(x|1)\, dx \;=\; \int_0^1 \frac{x}{2}\, dx \;=\; \frac{1}{2}(\frac{x^2}{2} \Big|_0^1) = \frac{1}{4} \quad .$$

Note: Also here the joint density function was **separable**, and the set where it is non-zero is rectangular. hence X and Y are independent. In such cases $f_{Y|X}(y|x) = f_Y(y)$, regardless of x and $f_{X|Y}(x|y) = f_X(x)$ regardless of y.

Problem 16.5: A company offers a basic life insurance policy to its employees, as well as a supplemental life insurance policy. To purchase the supplemental policy, an employee must first purchase the basic policy.

Let X denote the proportion of employees who purchase the basic policy, and let Y the proportion of the employees who purchase the supplemental policy. Let X and Y have joint density function $f(x,y) = (12x + 18y)/7$ on the region where the density is positive.

Given that 20% of the employees buy the basic policy, determine the probability that more than 7% buy the supplemental policy.

Sol. to 16.5: Since the problem tells you that *always*, $Y \leq X$ (those who purchased the supplemental policy must have also purchased the basic one), the **probability density function** is

$$f(x,y) \;=\; \begin{cases} \frac{12x+18y}{7} \quad if \quad 0 \leq y \leq x \leq 1\,; \\ 0\,,\; otherwise. \end{cases} \quad .$$

The density function of the **marginal distribution** , in general is

$$f_X(x) \;=\; \int_0^x \frac{12x + 18y}{7}\, dy \quad .$$

But we only need its value when $x = 0.2$, so

$$f_X(0.2) \;=\; \int_0^{0.2} \frac{2.4\,+\,18y}{7}\, dy \;=\; \frac{1}{7}(2.4y + 9y^2)\Big|_0^{0.2} = \frac{1}{7}(2.4 \cdot 0,2 + 9 \cdot (0.2)^2) \;=\; 0.12 \quad .$$

Hence

$$f_{Y|X}(y|0.2) \;=\; \frac{(2.4 + 18y)/7}{0.12} \quad ,$$

when $0 \leq y \leq 0.2$ and 0 otherwise. Finally, the desired probability, $P(Y \geq 0.07|X = 0.2)$ is

$$\int_{0.07}^{0.2} f_{Y|X}(y|0.2)\, dy \;=\; \int_{0.07}^{0.2} \frac{(2.4 + 18y)/7}{0.12} = 0.7475000001\ldots \quad .$$

Ans. to 16.5: Given that 20% of the employees buy the basic policy, the probability that more than 7% buy the supplemental policy is %74.75000001.

Problem 16.6: In a certain community of married couples, the maximal income of the wife is 200K and the maximal income of the husband is 100K. Every husband makes **at most** half of his wife's income. Let X denote the the wife's income and let Y denote the husband's income. Let X and Y have joint density function $f(x,y) = 3(x + y)/5$ on the region where the density is positive. The unit of money is 100K.

If it is known that the husband makes 50000 dollars, what is the probability that the wife makes more than 120000?

Sol. to 16.6: The region where $f(x,y)$ is non-zero is $\{(x,y) \,|\, 0 \le x \le 2, 0 \le y \le 1, x \ge 2y\}$.

$$f_Y(0.5) = \int_1^2 f(x, 0.5)\, dx \; = \; \int_1^2 3(x+0.5)/5\, dx = 1.2 \quad .$$

We have:

$$f_{X|Y}(x, 0.5) = \begin{cases} \frac{3(x+0.5)/5}{1.2} = 0.5x + 0.25\,, \quad if \quad 1 \le x \le 2\,; \\ 0\, otherwise. \end{cases}$$

The desired conditional probability is

$$\int_{1.2}^2 f_{X|Y}(x, 0.5)\, dx = \int_{1.2}^2 (0.5x + 0.25)\, dx = 0.8400000000 \quad .$$

Ans. to 16.6: The probability that the wife makes more than 120000 dollars if it is known that her husband makes exactly 50000 dollars is %84.