Dr. Z.'s Probability Lecture 14 Handout: Joint Distribution Functions

By Doron Zeilberger

Version of Nov. 26, 2017: Thanks to Bennet Greenberg. (correcting numerous numerical errors in 14.1).

Important Definition (Discrete Case) :

If X and Y are both discrete random variables, the **joint probability mass function** of X and Y is defined by

$$p(x,y) \, = \, P[X=x, Y=y] \quad .$$

The probability mass function of X and Y (called *marginal distributions*) can be gotten from $p(x,y)$ as follows

$$p_X(x) = P[X=x] = \sum_y p(x,y) \quad .$$

(It is adding up all the possibilities for Y)

$$p_Y(y) = P[Y=y] = \sum_x p(x,y) \quad .$$

(It is adding up all the possibilities for X)

Problem 14.1: In a certain development the regulations allow every household to have at most two dogs and at most two cats.

It is found that the probability mass function is

$$p(i,j) = Pr(NumberOfDogs = i, NumberOfCats = j) = \frac{c}{i+j+3} \quad , \quad 0 \leq i \leq 2 \, , \, 0 \leq j \leq 2 \quad ,$$

for some constant c.

(i) Find c. (ii) Find the probability that a family has at least as many cats as dogs.

(iii) Find the expected number of dogs. (iv) Find the expected number of cats.

Sol. to 14.1(i): First let's find c. Adding up all the probabilities, and setting it equal to 1, gives

$$1 = \sum_{i=0}^{2} \sum_{j=0}^{2} \frac{c}{i+j+3}$$

$$c \cdot \left(\frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} \right)$$

$$\frac{401}{210} \cdot c \quad .$$

So

$$c = \frac{210}{401} \quad .$$

Ans. to 14.1(i): $c = \frac{210}{401}$.

Sol. to 14.1(ii):

$$\sum_{0 \leq i \leq j \leq 2} p(i,j) = p(0,0) + p(0,1) + p(0,2) + p(1,1) + p(1,2) + p(2,2)$$

$$= \frac{210}{401} \cdot \left(\frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} \right) =$$

$$\frac{210}{401} \cdot \frac{181}{140} = \frac{543}{802} = 0.677057\ldots \quad .$$

Ans. to 14.1(ii): The probability that the number of cats is $\geq$ than the number of dogs is $\frac{543}{802} = 0.677057\ldots$.

Sol. to 14.1(iii): Let's first find the Dog probability mass function.

$$p_X(i) = \sum_{j=0}^{2} p(i,j) = \frac{210}{401} \sum_{j=0}^{2} \frac{1}{i+j+3} \quad .$$

So

$$p_X(0) = \frac{210}{401} \sum_{j=0}^{2} \frac{1}{0+j+3} == \frac{329}{802} \quad .$$

$$p_X(1) = \frac{210}{401} \sum_{j=0}^{2} \frac{1}{1+j+3} = \frac{210}{401} \left(\frac{1}{4} + \frac{1}{5} + \frac{1}{6} \right) = \frac{259}{802} \quad .$$

$$p_X(2) = \frac{210}{401} \sum_{j=0}^{2} \frac{1}{2+j+3} = \frac{210}{401} \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} \right) = \frac{107}{401} \quad .$$

Hence, $E[X]$, the expected number of dogs is

$$E[X] = \sum_{i=0}^{2} i p_X(i) = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2)$$

$$= 0 \cdot \frac{320}{802} + 1 \cdot \frac{259}{802} + 2 \cdot \frac{214}{802} = \frac{687}{802} = 0.856608\ldots.$$

Ans. to 14.1(iii): The expected number of dogs is $\frac{687}{802} = 0.856608\ldots$.

Sol. to 14.1(iv): You could do it from scratch, but in **this** problem the joint probability mass function is **symmetric** with regards to cats and dogs, so the expected number of cats is also $\frac{687}{802} = 0.856608$.

Important Definition (Continuous Case) :

If X and Y are both continuous random variables, the **joint cumulative probability distribution function** of X and Y is defined by

$$F(x, y) = P[X \leq x \, , \, Y \leq y] \quad .$$

Important Definition/Fact: The **joint probability density function** of X and Y, $f(x, y)$, may be obtained from $F(x, y)$ by taking partial derivatives w.r.t. to x and y

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) \quad .$$

Very Important Fact: For any subset C of the plane, $(-\infty, \infty) \times (-\infty, \infty)$,

$$P[(X, Y) \in C] = \int \int_{(x,y) \in C} f(x, y) \, dx \, dy \quad .$$

Special Case: If C is the *Cartesian product* of A and B (where A and B are subsets of the line), i.e.

$$C = \{(x, y) \,|\, x \in A \, , \, y \in B\},$$

then

$$P[X \in A, Y \in B] = \int_B \int_A f(x, y) \, dx \, dy \quad .$$

Problem 14.2: The joint density function of X and Y is given by

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

Find (i) $P(X < \frac{1}{3} \, , \, Y > \frac{2}{3})$

(ii) $P(0 < X < 1 \, , \, 0 < Y < 1)$

Sol. to 14.2(i): The desired probability is the double-integral

$$\int_0^{\frac{1}{3}} \int_{\frac{2}{3}}^1 4xy \, dy \, dx \quad .$$

The inner integral is

$$\int_{\frac{2}{3}}^1 4xy \, dy \; = \; 4x \int_{\frac{2}{3}}^1 y \, dy = 4x(\frac{y^2}{2})\Big|_{2/3}^1 = 4x \left(\frac{1}{2} - \frac{(2/3)^2}{2} \right) = 4x \frac{5}{18} = \frac{10x}{9} \quad .$$

The outer integral is

$$\int_0^{\frac{1}{3}} \frac{10x}{9}\,dx = \frac{10}{9}(\frac{x^2}{2}\Big|_0^{1/3}) = \frac{10}{9}\frac{1}{18} = \frac{5}{81} \quad .$$

Ans. to 14.2(i): $P(X < \frac{1}{3}, Y > \frac{2}{3}) = \frac{5}{81}$.

Sol. to 14.2(ii): You can do it the same way:

$$\int_0^1 \int_0^1 4xy\,dy\,dx \quad ,$$

but if you **trust** the problem and $f(x,y)$ is a genuine joint probability density function, the answer is obvious! It is 1.

Problem 14.3: A device runs until either of the two components fails, at which point the device stops running. The lifetimes of the two components has a joint probability density function

$$f(x,y) = \frac{4x+2y}{81} \quad , \quad for \quad 0 < x < 3 \quad and \quad 0 < y < 3 \quad ,$$

where x and y are measured in hours. What is the probability that the device fails during the first two hours of operation?

Sol. to 14.3: The probability that both components **survive** beyond 2 hours is

$$\int_2^3 \int_2^3 \frac{4x+2y}{81}\,dx\,dy \quad .$$

We can do iterated integration, but in this problem it is more convenient to break the integral into two parts.

$$\int_2^3 \int_2^3 \frac{4x+2y}{81}\,dx\,dy = \int_2^3 \int_2^3 \frac{4x}{81}\,dx\,dy + \int_2^3 \int_2^3 \frac{2y}{81}\,dx\,dy \quad .$$

The first integral is

$$\int_2^3 \int_2^3 \frac{4x}{81}\,dx\,dy = \left(\int_2^3 \frac{4x}{81}\,dx\right) \cdot \left(\int_2^3 dy\right) = \left(\frac{2x^2}{81}\Big|_2^3\right) \cdot 1 = \frac{2}{81}(3^2 - 2^2) = \frac{10}{81} \quad .$$

The second integral is

$$\int_2^3 \int_2^3 \frac{2y}{81}\,dx\,dy = \left(\int_2^3 dx\right)\left(\int_2^3 \frac{2y}{81}\,dy\right) = 1 \cdot \left(\frac{y^2}{81}\Big|_2^3\right) = \frac{1}{81}(3^2 - 2^2) = \frac{5}{81} \quad .$$

Adding them up gives that the desired probability is $\frac{10}{81} + \frac{5}{81} = \frac{15}{81} = \frac{5}{27}$. Hence the probability that at least one component does **not** make it beyond two hours is the **complimenary** probability is $1 - \frac{5}{27} = \frac{22}{27} = 0.8148148148\ldots$.

Ans. to 14.3: The probability that the device fails during the first two hours of operation is $\frac{22}{27} = 0.8148148148\ldots$.

Important Concept/Formula: If (X, Y) are continuous random variables with joint density function $f(x, y)$, the *marginal distributions* (of the individual random variables X and Y), denoted by $f_X(x)$ and $f_Y(y)$ respectively are

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

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

Note: Often $f(x, y)$ is non-zero only in a finite region, in that case the limits of integration are no longer $-\infty$ and ∞ but some finite numbers, given from the definition of $f(x, y)$.

Problem 14.4: Let X and Y be continuous random variables with joint density function

$$f(x, y) = \begin{cases} \frac{18}{5}(x + y) \quad , \quad for \quad x^4 \leq y \leq x\,; \\ 0 \quad , \quad otherwise. \end{cases}$$

Let $g(y)$ be the marginal density function of Y, and let $h(x)$ be the marginal density function of X. (i) Find $g(y)$ (ii) Find $h(x)$.

The two curves $y = x$ and $y = x^4$ meet when $x - x^4 = 0$, i.e. when $x(x - 1)(x^2 + x + 1) = 0$, i.e. when $x = 0$ and $x = 1$. This region is a subset of $\{(x, y) \,|\, 0 \leq x \leq 1\,,\, 0 \leq y \leq 1\}$.

Sol. of 14.4(i): When you draw the horizontal line $y = y_0$ ($y_0 > 0$) it meets the line $y = x$ at $x = y_0$ and the curve $y = x^4$ at $x = y_0^{1/4}$.

Hence

$$g(y) \;=\; \int_{y}^{y^{1/4}} \frac{18}{5}(x + y)\, dx \;=\; \frac{18}{5}(x^2/2 + xy)\Big|_{x=y}^{x=y^{1/4}} \;=\; \frac{18}{5}(((y^{1/4})^2/2 + y^{1/4}y) - (y^2/2 + y^2))$$

$$= \frac{9}{5}(y^{1/2} + 2y^{5/4} - 3y^2) \quad .$$

Ans. to 14.4(i): The marginal density function of Y is $g(y) = \frac{9}{5}(2y^{1/2} + 2y^{5/4} - 3y^2)$ when $0 < y < 1$, and 0 otherwise. Officially this is written

$$g(y) = \begin{cases} \frac{9}{5}(y^{1/2} + 2y^{5/4} - 3y^2)\,, \quad if \quad 0 < y < 1\,; \\ 0\, otherwise \end{cases}$$

Sol. of 14.4(ii): A vertical line above x meets the region starting at $y = x^4$ and ending at $y = x$, so

$$h(x) = \int_{x^4}^{x} \frac{18}{5}(x + y)\, dy \;=\; \frac{18}{5}(xy + y^2/2)\Big|_{y=x^4}^{y=x} \;=\; \frac{18}{5} \cdot ((x \cdot x + x^2/2) - (x \cdot x^4 + (x^4)^2/2)) \quad .$$

$$= \frac{18}{5} \cdot (x^2 + x^2/2 - x^5 - x^8/2) = \frac{9}{5}(3x^2 - 2x^5 - x^8) \quad .$$

Ans. to 14.4(ii): The marginal density function of X is $h(x) = \frac{9}{5}(3x^2 - 2x^5 - x^8)$.

Officially this is written

$$h(x) = \begin{cases} \frac{9}{5}(3x^2 - 2x^5 - x^8), & if \quad 0 < x < 1\,; \\ 0\, otherwise \end{cases} \quad .$$

Problem 14.5: A company is reviewing hurricane damage claims under a residential insurance policy. Let X be the portion of the claim representing damage to the house and let Y be the portion of the claim representing damage to the front and back yards. The joint density function of X and Y is

$$f(x,y) \;=\; \begin{cases} \frac{3xy}{2} \quad , & for \quad x > 0\,,\, y > 0\,,\, x + y < 2 \\ 0 \quad otherwise. \end{cases}$$

Determine the probability that the portion of a claim representing damage to the house is more than the damage to the rest of the property.

Sol. of 14.5: We need $\int_A f(x,y)\,dx\,dy$ over the part of the region $x > 0, y > 0, x + y < 2$ that satisfies $x \geq y$. The lines $y = x$ and $y = 2 - x$ meet at the point $(1,1)$. The region we need is a triangle with base beteen $(0,0)$ and $(2,0)$ and top vertex $(1,1)$. It is more convenient to do the y integration as the outer-integration and the x-integration as the inner-integration. The projection of our region of interest on the y-axis is $0 \leq y \leq 1$. For every horizontal line corresponding to y the region starts at $x = y$ and ends at $x = 2 - y$. Hence the desired probability is

$$\int_0^1 \int_y^{2-y} \frac{3xy}{2}\,dx\,dy \quad .$$

The inner integral is

$$\int_y^{2-y} \frac{3xy}{2}\,dx \;=\; \frac{3y}{2} \int_y^{2-y} x\,dx$$

$$= \frac{3y}{2} \left(\frac{x^2}{2} \Big|_y^{2-y} \right) = \frac{3y}{4}((2-y)^2 - y^2) = \frac{3y}{4}(4 - 4y) = 3y - 3y^2 \quad .$$

The outer integral is

$$\int_0^1 (3y - 3y^2)\,dy = \left(\frac{3y^2}{2} - y^3 \right) \Big|_0^1 = \frac{3}{2} - 1 = \frac{1}{2} \quad .$$

Ans. to 14.5: the probability that the portion of a claim representing damage to the house is more than the damage to the rest of the property is $\frac{1}{2}$.

Comment: In this *particular* problem (by luck!), there is a short-cut, that does not require any integration. The joint density function $f(x,y) = \frac{3xy}{2}$ **and** the region where it is not zero are **both**

symmetric with respect to x and y. By symmetry, the prob. that the damage to the house exceeds the damage to the yard is the **same** as the prob. that it is the opposite. Since they add-up to 1, they are both $\frac{1}{2}$.

Warning: The above shortcut is **only** applicable if **both** $f(x, y)$ and the region are symmetric! Use with caution.