Math 464 Fall 2012
Notes on Marginal and Conditional Densities
klin@math.arizona.edu
October 18, 2012

Marginal densities. Suppose you have 3 continuous random variables X, Y, and Z, with joint density f(x,y,z). The marginal density for X is the PDF of X. In class we gave a formula for this. It is just

$$f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y,z)\ \ dy\ \ dz \tag{1}$$

Similarly, the PDF for Y and Z are

$$f_Y(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y,z)\ \ dx\ \ dz$$
$$f_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y,z)\ \ dx\ \ dy$$

Notice that to get the density for one of the variables, what we do is integrate over all the other variables. This generalizes naturally to n continuous random variables $X_1, X_2, \cdots, X_n$ with joint density $f(x_1, x_2, \cdots, x_n)$: we define the ith marginal density of f to be the PDF of X_i . The formula is

$$f_{X_i}(x) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\cdots,x_{i-1},x,x_{i+1},\cdots,x_n)\ \ dx_1\cdots dx_{i-1}\ \ dx_{i+1}\cdots dx_n \tag{2}$$

Notice that we integrate exactly n-1 times.

We gave a derivation of this formula in class. Here it is again, for n=3. To get the marginal density of X, the basic idea is to first find the CDF F_X, then differentiate it. Here goes: let X, Y, Z, and f be as above, and let x be any real number. We start with the observation that

$$\begin{aligned} F_X(x) &= P(X \le x) \\ &= P\Big((-\infty, x] \times \mathbb{R} \times \mathbb{R}) \end{aligned}$$

The 2nd line just says that we want the probability of the event that $X \le x$ and that we don't care about the values of Y and Z. From this and the definition of the joint density, we get

$$F_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{x} f(x',y,z)\ \ dx'\ \ dy\ \ dz$$

Differentiating, we get[1]

$$\begin{aligned}\frac{d}{dx}F_X(x) &= \frac{d}{dx}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{x} f(x',y,z)\ \ dx'\ \ dy\ \ dz\right)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{d}{dx}\left(\int_{-\infty}^{x} f(x',y,z)\ \ dx'\right)\ \ dy\ \ dz\end{aligned}$$

By the fundamental theorem of calculus, we have

$$\frac{d}{dx}\left(\int_{-\infty}^{x} f(x',y,z)\ \ dx'\right) = f(x,y,z)$$

So

$$f_X(x) = \frac{d}{dx}F_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y,z)\ \ dy\ \ dz$$

as desired.

Example. In the darts example with radius 1, the PDF is

$$f(x,y) = \begin{cases} 1/\pi, & x^2+y^2 \le 1 \\ 0, & \texttt{otherwise}\end{cases}$$

If we want the PDF for the X variable alone, this is given by

$$\begin{aligned}f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\ \ dy\\ &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 1/\pi\ \ dy\\ &= \frac{2}{\pi}\sqrt{1-x^2} \qquad \texttt{if } -1 \le x \le 1;\end{aligned}$$

$f_X(x) = 0$ otherwise. The first line is just the formula for the marginal of X. To go from the first to the second line, it is important to realize that f(x,y) is nonzero if and only if (x,y) lies on or inside the unit circle.

Conditional density, book's version. Grinstead and Snell gave the following definition of a conditional density: Let X be a random variable with PDF f, and let E be an event with $P(E) > 0$. Then we define the <u>conditional density of X given E has occurred</u> to be

$$f(x|E) = f(x)/P(E)$$

[1]This derivation is not meant to be a proof. To <u>prove</u> the formula along these lines, we would have to justify exchanging the order of differentiation and integration. That belongs in a course on real analysis.

This extends naturally to 2 or more random variables. For exmaple, if f(x,y) is the joint density for X and Y, and E is an event, then the conditional density of (X,Y) given E is

$$f(x,y|E) = f(x,y)/P(E)$$

The key thing about these formulas is that they make sense only when $P(E) > 0$. If $P(E) = 0$, then we would be dividing by 0.

As an example, in the darts example, if we let E be the event ''the dart lands in the first quadrant,'' then we have

$$\begin{aligned} f(x,y|E) &= f(x,y)/P(E) \\ &= \begin{cases} 4/\pi, & x \geq 0 \text{ and } y \geq 0 \text{ and } x^2 + y^2 \leq 1 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

Important! A key property of $f(x|E)$ is that it is still a probability density. So if we want the conditional probability $P(A|E)$ for some event A, we would evaluate $\int_A f(x|E)\ dx$. A consequence is that $\int_{-\infty}^{\infty} f(x|E)\ dx = 1$.

Conditional density of Y given X=x. A closely related concept is the conditional density of one random variable Y given that a second random variable X has a certain value x. This is different, but related to, the concept of conditional density given in the book. Both notions are useful, but in different ways.

If X and Y are two continuous random variables and x is a real number, the conditional density of Y given X=x is defined to be

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \quad . \tag{3}$$

The key property of this is that it is a PDF: if we want the probability of an event A given X=x, then

$$P(A|X = x) = \int_A f_{Y|X}(y|x)\ \ dy$$

(pay attention to what variable is being integrated, and what's being left alone). In particular, we have

$$\int_{-\infty}^{\infty} f_{Y|X}(y|x)\ \ dy = 1. \tag{4}$$

The reason we need a separate definition when conditioning on X=x is that the event (X=x) has probability 0. So there is no way we can use the book's definition (or something like it) to define the conditional density.

In class I gave a motivation for Eq. (3). Here it is, in a more general form: let $\delta > 0$ be any number, and $[a,b]$ be any interval. Then

$$P(\underbrace{a \le Y \le b}_{\text{call this A}} \mid \underbrace{x-\delta \le X \le x+\delta}_{\text{call this B}}) = \frac{P(A \cap B)}{P(B)}$$

We have

$$P(A \cap B) = \int_a^b \left[\int_{x-\delta}^{x+\delta} f(x', y) \ \ dx' \right] \ \ dy$$

We can approximate the term in the square brackets by the rectangle rule, to get

$$\int_{x-\delta}^{x+\delta} f(x', y) \ \ dx' \approx 2\delta \cdot f(x, y)$$

so that

$$P(A \cap B) \approx 2\delta \int_a^b f(x, y) \ \ dy$$

We can treat P(B) similarly:

$$\begin{aligned} P(B) &= \int_{x-\delta}^{x+\delta} \Big[\int_{-\infty}^{\infty} f(x', y) \ \ dy \Big] dx' \\ &= \int_{x-\delta}^{x+\delta} f_X(x') \ \ dx' \\ &\approx 2\delta \cdot f_X(x) \end{aligned}$$

So we get

$$\begin{aligned} P(a \le Y \le b | x-\delta \le X \le x+\delta) &= \frac{P(A \cap B)}{P(B)} \\ &\approx \frac{2\delta \int_a^b f(x,y) \ \ dy}{2\delta \cdot f_X(x)} \\ &\approx \frac{\int_a^b f(x,y) \ \ dy}{f_X(x)} \end{aligned}$$

One can show that in the limit as $\delta \to 0$, the ''$\approx$'' becomes exact ''='', so we get

$$P(a \leq Y \leq b | X = x) = \frac{\int_a^b f(x,y) \ \ dy}{f_X(x)}$$

In particular, for any real number y, we have

$$P(Y \leq y | X = x) = \frac{\int_{-\infty}^y f(x,y') \ \ dy'}{f_X(x)}$$

Differentiating gives us Eq. (3). (Notice that the variable x should be treated as a constant in the above.)

Finally, there is a second way to see why the definition in Eq. (3) is reasonable: suppose we want to define $f_{Y|X}(y|x)$. What properties should it have? At the minimum, it should be a PDF, i.e., Eq. (4) should hold. A second property we might insist on is that for each given value of x, $f_{Y|X}(y|x)$ should be proportional to f(x,y) as a function of y, so that events have the right relative frequency.[2] With these conditions, we have

$$f_{Y|X}(y|x) = C_x f(x,y)$$

where C_x is a ''constant'' (it really depends on x) to be chosen so that Eq. (4) holds. It is easy to check (you should do it!) that $C_x = f_X(x)$.

[2] This statement will take some digesting...