Ma 322: Biostatistics
Conditional Probabilities and Continuous Densities

Prof. Wickerhauser

Wednesday, February 6th, 2013

Suppose that X, Y are continuous random variables, each taking values in the real line $\mathbf{R}$, with **joint probability density function** $f(x, y)$. Let us further suppose that this joint pdf f is a nice continuous function of the two variables x, y.

The pdf conditions for f are that $f(x, y) \geq 0$ for all x, y, and

$$\int_{x=-\infty}^{\infty} \int_{x=-\infty}^{\infty} f(x, y)\, dydx = 1.$$

For events $E \subset \mathbf{R}$ and $F \subset \mathbf{R}$, we compute the **probability**

$$\text{Prob}\,(X \in E \text{ and } Y \in F) \stackrel{\text{def}}{=} \int_{x \in E} \int_{y \in F} f(x, y)\, dxdy$$

For example, if

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

then $\text{Prob}\,(0 \leq X \leq \frac{1}{3} \text{ and } \frac{1}{2} \leq Y \leq 1) = 1/6$, the volume under the graph of f over the region $(x, y) \in [0, \frac{1}{3}] \times [\frac{1}{2}, 1] \stackrel{\text{def}}{=} E \times F$. The result would be the same with or without the endpoints of E and F.

Note that $\text{Prob}\,(X \in E \text{ and } Y = y_0) = 0$ for any $E \subset \mathbf{R}$ and any single value y_0, because the length of the one-point set $F = \{y_0\}$ is zero so that the integral over F will be zero. Likewise, $\text{Prob}\,(X = x_0 \text{ and } Y \in F) = 0$ for any $F \subset \mathbf{R}$ and any single $x_0 \in \mathbf{R}$.

The **marginal pdfs** in X and Y are computed from the joint pdf by partial integration:

$$f_X(x) \stackrel{\text{def}}{=} \int_{y=-\infty}^{\infty} f(x, y)\, dy;$$

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

These will be continous if f is continuous (by Fubini's theorem), and they allow us to compute **marginal probabilities**:

$$\text{Prob}_X(x \in E) \stackrel{\text{def}}{=} \int_{x \in E} f_X(x)\, dx = \int_{x \in E} \int_{y=-\infty}^{\infty} f(x, y)\, dydx;$$

$$\text{Prob}_Y(y \in F) \stackrel{\text{def}}{=} \int_{y \in F}^{\infty} f_Y(y)\, dy = \int_{y \in F} \int_{x=-\infty}^{\infty} f(x, y)\, dxdy.$$

By Fubini's theorem, these integrals may be evaluated in either order. For the previous example f, we have

$$f_X(x) = \begin{cases} 1, & \text{if } 0 \le x < 1, \\ 0, & \text{otherwise,} \end{cases}$$

and $f_Y(y)$ is the same. For this simple example, $\text{Prob}_X(x \in E) = |E|$ is just the length of event E, and likewise $\text{Prob}_Y(y \in F) = |F|$.

The two **conditional probabilities**, for X and Y respectively, are defined on pairs of events E, F as follows:

$$\begin{aligned} \text{Prob}\,(X \in E | Y \in F) &\stackrel{\text{def}}{=} \frac{\text{Prob}\,(X \in E \text{ and } Y \in F)}{\text{Prob}_Y(y \in F)}; \\ \text{Prob}\,(Y \in F | X \in E) &\stackrel{\text{def}}{=} \frac{\text{Prob}\,(X \in E \text{ and } Y \in F)}{\text{Prob}_X(x \in E)}. \end{aligned}$$

Notice that if $F = \{y_0\}$ is a single point set, then the numerator and denominator in the definition of $\text{Prob}\,(X \in E|Y \in F) = \text{Prob}\,(X \in E|Y = y_0)$ will both be zero, giving the indeterminate ratio $0/0$. Such expressions may be evaluated as limits, putting $F = [y_0, y_0 + h]$ for some $h > 0$ and letting $h \to 0$. Expanding the numerators and denominators using their definitions as integrals, this gives:

$$\begin{aligned} \text{Prob}\,(X \in E|Y = y_0) &\stackrel{\text{def}}{=} \lim_{h\to 0} \frac{\text{Prob}\,(X \in E \text{ and } Y \in [y_0, y_0 + h])}{\text{Prob}_Y(y \in [y_0, y_0 + h])} \\ &= \lim_{h\to 0} \frac{\int_{x\in E} \int_{y=y_0}^{y_0+h} f(x, y)\, dydx}{\int_{y=y_0}^{y_0+h} f_Y(y)\, dy} \stackrel{\text{def}}{=} \lim_{h\to 0} \frac{p(h)}{q(h)}. \end{aligned}$$

This may be evaluated using l'Hôpital's Theorem: if $p \to 0$ and $q \to 0$, then the limit of p/q is the limit of p'/q'. But we can evaluate $p'(h)$ and $q'(h)$ using the Fundamental Theorem of Calculus:

$$\begin{aligned} p'(h) &= \frac{d}{dh} \int_{x\in E} \int_{y=y_0}^{y_0+h} f(x, y)\, dydx = \int_{x\in E} f(x, y_0)\, dx; \\ q'(h) &= \frac{d}{dh} \int_{y=y_0}^{y_0+h} f_Y(y)\, dy = f_Y(y_0). \end{aligned}$$

Neither $p'(h)$ nor $q'(h)$ depends on h, so we can evaluate

$$\text{Prob}\,(X \in E|Y = y_0) = \lim_{h\to 0} \frac{p'(h)}{q'(h)} = \frac{\int_{x\in E} f(x, y_0)\, dx}{f_Y(y_0)} = \int_{x\in E} f(x|y_0)\, dx,$$

where we have defined the **conditonal pdf** $f(x|y) \stackrel{\text{def}}{=} f(x, y)/f_Y(y)$. Similarly, the other conditional pdf $f(y|x) \stackrel{\text{def}}{=} f(x, y)/f_X(x)$ gives the formula for one-X-point conditional probabilities:

$$\text{Prob}\,(Y \in F|X = x_0) = \frac{\int_{y\in F} f(x_0, y)\, dy}{f_X(x_0)} = \int_{y\in F} f(y|x_0)\, dy.$$

CAUTION: except in one-point cases, the conditional pdf does not, in general, integrate to give the conditional probability function, since the ratio of integrals does not always equal the integral of the ratio.