## ACSC/MATH 347/447 Conditional Distributions

For jointly distributed random variables, $Y_1$ and $Y_2$, we define the **conditional distribution function of $Y_1$ given $Y_2 = y_2$** to be $F(y_1 \mid y_2) = P(Y_1 \le y_1 \mid Y_2 = y_2)$.

If $Y_1$ and $Y_2$ are discrete, $F(y_1 \mid y_2) = \sum_{t_1 \le y_1} p(t_1 \mid y_2)$.

In the continuous case, we would like to find a density function, $f(y_1 \mid y_2)$, so that

$$F(y_1 \mid y_2) = \int_{-\infty}^{y_1} f(t_1 \mid y_2)\,dt_1 .$$

Since in the discrete case, the conditional probability function is $p(y_1 \mid y_2) = \dfrac{p(y_1, y_2)}{p_2(y_2)}$, it seems natural to try

$$f(y_1 \mid y_2) = \frac{f(y_1, y_2)}{f_2(y_2)},$$

provided $f_2(y_2) > 0$.

Before we can define $f(y_1 \mid y_2)$ in this manner, we need to check two things:

1) $f(y_1 \mid y_2)$ is a density function (nonnegative and integrates to 1); and
2) $f(y_1 \mid y_2)$ is the pdf of $Y_1$ given $Y_2 = y_2$; that is,

$$F(y_1 \mid y_2) = P(Y_1 \le y_1 \mid Y_2 = y_2) = \int_{-\infty}^{y_1} \frac{f(t_1, y_2)}{f_2(y_2)}\,dt_1 .$$

But:

1) $f(y_1 \mid y_2) = \dfrac{f(y_1, y_2)}{f_2(y_2)}$ is a density function since

   a) it is nonnegative as both $f(y_1, y_2)$ and $f_2(y_2)$ are; and

   b) $\displaystyle\int_{-\infty}^{\infty} f(y_1 \mid y_2)\,dy_1 = \int_{-\infty}^{\infty} \frac{f(y_1, y_2)}{f_2(y_2)}\,dy_1$

$$= \frac{1}{f_2(y_2)} \int_{-\infty}^{\infty} f(y_1, y_2)\,dy_1$$

$$= \frac{1}{f_2(y_2)} \cdot f_2(y_2)$$

$$= 1 .$$

2) To show that $f(y_1 \mid y_2)$ is the pdf of $(Y_1 \mid y_2)$, use the continuity of $F(y_1, y_2)$:

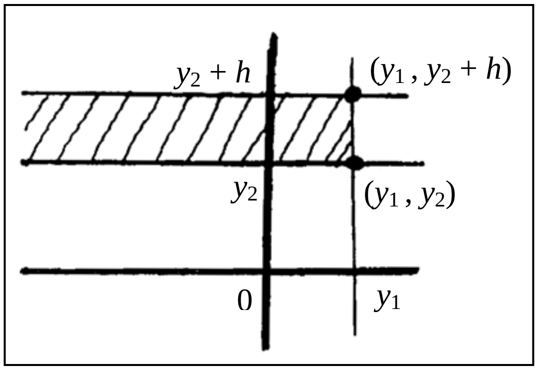
$$F(y_1 \mid y_2) = P(Y_1 \le y_1 \mid Y_2 = y_2)$$

$$= \lim_{h\to 0} P(Y_1 \le y_1 \mid y_2 \le Y_2 \le y_2 + h)$$

$$= \lim_{h\to 0} \frac{P(Y_1 \le y_1,\ y_2 \le Y_2 \le y_2 + h)}{P(y_2 \le Y_2 \le y_2 + h)}$$

$$= \lim_{h\to 0} \frac{\left(\frac{1}{h}\right)}{\left(\frac{1}{h}\right)} \cdot \frac{F(y_1,\ y_2 + h) - F(y_1,\ y_2)}{F(y_2 + h) - F(y_2)}$$

$$= \frac{\lim_{h\to 0} \left(\frac{1}{h}\right) [F(y_1,\ y_2 + h) - F(y_1,\ y_2)]}{\lim_{h\to 0} \left(\frac{1}{h}\right) [F(y_2 + h) - F(y_2)]}$$




Now the denominator is $F'(y_2) = f(y_2)$, and the numerator is

$$\lim_{h\to 0} \left(\frac{1}{h}\right) [F(y_1,\ y_2 + h) - F(y_1,\ y_2)]$$

$$= \lim_{h\to 0} \frac{1}{h} \left[ \int_{-\infty}^{y_1}\int_{-\infty}^{y_2+h} f(t_1, t_2)\, dt_2\, dt_1 - \int_{-\infty}^{y_1}\int_{-\infty}^{y_2} f(t_1, t_2)\, dt_2\, dt_1 \right]$$

$$= \lim_{h\to 0} \frac{1}{h} \int_{-\infty}^{y_1} \left[ \int_{-\infty}^{y_2+h} f(t_1, t_2)\, dt_2\, dt_1 - \int_{-\infty}^{y_2} f(t_1, t_2)\, dt_2\, dt_1 \right]$$

$$= \lim_{h\to 0} \int_{-\infty}^{y_1} \left[ \frac{1}{h} \int_{y_2}^{y_2+h} f(t_1, t_2)\, dt_2 \right] dt_1$$

$$= \int_{-\infty}^{y_1} \lim_{h\to 0} \left[ \frac{1}{h} \int_{y_2}^{y_2+h} f(t_1, t_2)\, dt_1 \right] dt_1$$

$$= \int_{-\infty}^{y_1} f(t_1, y_2)\, dt_1 \quad \text{by the Fundamental Theorem of the Calculus..}$$

Thus

$$F(y_1 \mid y_2) = P(Y_1 \le y_1 \mid Y_2 = y_2) = \frac{\int_{-\infty}^{y_1} f(t_1, y_2)\, dt_1}{f_2(y_2)} = \int_{-\infty}^{y_1} \frac{f(t_1, y_2)}{f_2(y_2)}\, dt_1 = \int_{-\infty}^{y_1} f(t_1 \mid y_2)\, dt_1 .$$

[Note that the above is not completely rigorous, but it is an indication of why the definition $f(y_1 \mid y_2) = \dfrac{f(y_1, y_2)}{f_2(y_2)}$ is reasonable.]