Section 2.3. Conditional Distributions and Expectations

Note. We now consider conditional distributions in which a random variable is assumed to assume a specific value (in the support of that random variable) and then the distribution of a second random variable is defined by taking into consideration of the value of the first ransom variable. This generalizes the conditional concept introduced in 1.4. Conditional Probability and Independence.

Note 2.3.A/Definition. Let X_1 and X_2 be discrete random variables with joint probability mass function $p_{X_1,X_2}(x_1, x_2)$ and support $\mathcal{S}$. Let $p_{X_1}(x_1)$ and $p_{X_2}(x_2)$ be the marginal probability mass functions of X_1 and X_2 as defined in Section 2.1. For x_1 in the support $\mathcal{S}_{X_1}$ of p_{X_1}, we have by the definition of conditional probability (Definition 1.4.1) that

$$P(X_1 = x_2 \mid X_1 = x_1) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = x_1)} = \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_1}(x_1)}$$

for all x_2 in the support $\mathcal{S}_{X_2}$ of X_2. Denote this function as

$$p_{X_1|X_2}(x_2 \mid x_1) = \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_1}(x_1)} \text{ for } x_2 \in \mathcal{S}_{X_2}.$$

Notice that $p_{X_2|X_1}(x_2 \mid x_1) \geq 0$ for all $x_2 \in X_2$ and

$$\sum_{x_2 \in X_2} p_{X_2|X_1}(x_2 \mid x_1) = \sum_{x_2 \in X_2} \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_1}(x_1)} = \frac{1}{p_{X_1}(x_1)} \sum_{x_2 \in X_2} p_{X_1,X_2}(x_1, x_2)$$

$$= \frac{1}{p_{X_1}(x_1)} p_{X_1}(x_2) = 1,$$

so $p_{X_2|X_1}(x_2 \mid x_1)$ is in fact a probability mass function called the *conditional probability mass function* of X_2 given $X_1 = x_1$. We similarly define for $x_2 \in \mathcal{S}_{X_2}$

$$p_{X_1|X_2}(x_1 \mid x_2) = \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_2}(x_2)} \text{ for all } x_1 \in \mathcal{S}_{X_1}.$$

We may denote these functions as

$$p_{X_1|X_2}(x_1 \mid x_2) = p_{1|2}(x_1 \mid x_2) \text{ and } p_{X_2|X_1}(x_2 \mid x_1) = p_{2|1}(x_2, x_1).$$

Note/Definition. Let X_1 and X_2 be continuous random variables with joint probability density function $f_{X_1,X_2}(x_1, x_2)$ and support $\mathcal{S}$. Let $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$ be the marginal probability density functions as defined in Section 2.1. Motivated by the discrete case above, for $x_{\in} X_1$ with $x_{X_1}(x_1) > 0$ define

$$f_{X_2|X_1}(x_2 \mid x_1) = f_{2|1}(x_2 \mid x_1) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)}.$$

We have $f_{X_2|X_1}(x_2 \mid x_1)$ nonnegative and

$$\int_{-\infty}^{\infty} f_{X_2|X_1}(x_2 \mid x_1) = \int_{-\infty}^{\infty} \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)}\, dx_2 = \frac{1}{f_{X_1}(x_1)} \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\, dx_2$$
$$= \frac{1}{f_{X_1}(x_1)} f_{X_1}(x_1) = 1,$$

so $f_{X_2|X_1}(x_2 \mid x_2)$ is in fact a probability density function called the *conditional probability density function* of X_2 given $X_1 = x_1$. Similarly, for $x_2 \in X_2$ with $f(x_2) > 0$ we have

$$f_{X_1|X_2}(x_1 \mid x_2) = f_{1|2}(x_1 \mid x_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}.$$

Note. The conditional probability that $a < X_2 < b$ given $X_1 = x_1$ is computed as

$$P(a < X_2 < b \mid X_1 = x_1) = P(a < X_2 < b \mid x_1) = \int_a^b f_{X_2|X_1}(x_2 \mid x_1)\, dx_2.$$

Definition. If $u(X_2)$ is a function of random variable X_2 then the *conditional expectation* of $u(X_2)$ given $X_1 = x_1$ (if it exists) is

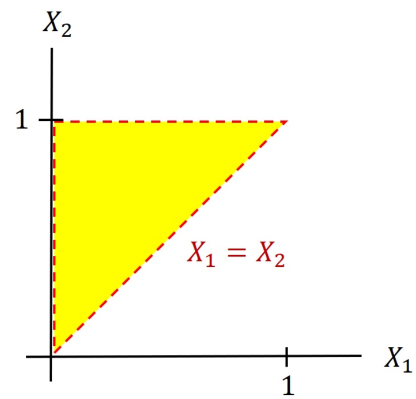
$$E[u(X_2) \mid X_1 = x_1] = E[u(X_2) \mid x_1] = \int_{-\infty}^{\infty} u(x_2) f_{X_2|X_1}(x_2 \mid x_1)\, dx_2.$$

If it exists then $E[X_2 \mid x_1]$ is the *conditional mean* of the conditional distribution of X_2 given $X_1 = x_1$. It it exists then $E[(x_2 - E[X_2 \mid x_1])^2 \mid x_1] = \mathrm{Var}(X_2 \mid x_1)$ is the *conditional variance* of the conditional distribution of X_2 given $X_1 = x_1$. We can similarly define these quantities for X_1 given $X_2 = x_2$.

Example 2.3.1. Let X_1 and X_2 be continuous random variables with joint probability density function

$$f_(x_1, x_2) = \begin{cases} 2 & \text{for } 0 < x_1 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

The support of $f(x_1, x_2)$ is:



The marginal probability density functions are

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2 = \int_{x_2=x_1}^{x_2=1} d\, dx_2 = 2(1 - x_1) \text{ for } 0 < x_1 < 1,$$

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1 = \int_{x_1=0}^{x_1=x_2} 2\, dx_1 = 2x_2 \text{ for } 0 < x_2 < 1$$

and each is 0 elsewhere. The conditional probability density function of X_1 given $X_2 = x_2$ for $0 < x_2 < 1$ is

$$f_{X_1|X_2}(x_1 \mid x_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)} = \frac{2}{2x_2} = \frac{1}{x_2}.$$

The conditional mean of X_1 given $X_2 = x_2$ is

$$E[X_1 \mid x_2] = \int_{-\infty}^{\infty} x_1 f_{X_1|X_2}(x_1 \mid x_2)\, dx_1 = \int_0^{x_2} x_1 \frac{1}{x_2}\, dx_1$$
$$= \frac{1}{x_2}\left(\frac{1}{2}(x_2)^2 - \frac{1}{2}(0)^2\right) = \frac{x_2}{2} \text{ for } 0 < x_2 < 1,$$

so the conditional variance of X_1 given X_2 is

$$\mathrm{Var}(X_1 \mid x_2) = E[(X_1 - E[X_1 \mid x_1])^2] = \int_{-\infty}^{\infty} \left(x_1 - \frac{x_2}{2}\right)^2 f_{X_1,X_2}(x_1 \mid x_2)\, dx_1$$
$$= \int_0^{x_1} \left(x_1 - \frac{x_2}{2}\right)^2 \left(\frac{1}{x_2}\right) dx_1 = \int_0^{x_2} \frac{x_1^2 - x_1 x_2 + x_2^2/4}{x_2}\, dx_1$$
$$= \frac{1}{x_2}\left(\frac{1}{3}x_2^3 - \frac{1}{2}x_2^3 + \frac{1}{4}x_2^3\right) = \frac{x_2^2}{12} \text{ for } 0 < x_2 < 1.$$

Also,

$$P\left(0 < X_1 < \frac{1}{2} \,\middle|\, X_2 = \frac{3}{4}\right) = \int_0^{1/2} f_{X_1,X_2}\left(x_1 \,\middle|\, \frac{3}{4}\right) dx_1$$
$$= \int_0^{1/2} \frac{1}{3/4}\, dx_1 = \frac{4}{3}\left(\frac{1}{2} - 0\right) = \frac{2}{3}.$$

(We can use the marginal distribution $f_{X_1}(x_1)$ to find that $P(0 < X_1 < 1/2) = 3/4$, so the value of X_2 affects probabilities of the value of X_1.)

Note. In the previous example, we have $P(X_2 = 3/4) = 0$ since X_2 is a continuous, so in computing $P(0 < X_1 < 1/2 \mid X_2 = 3/4)$ we have **conditioned on a probability** 0 **event**, and yet the conditional probability is still defined!

Theorem 2.3.1. Let (X_1, X_2) be a random vector such that the variance of X_2 is finite. Then

(a) $E[E[X_2 \mid X_1]] = E[X_2]$, and

(b) $\text{Var}([E[X_2 \mid X_1]) \leq \text{Var}(X_2)$.

Note. The book comments (see page 115) that, since the mean of both X_2 and $E[X_2 \mid X_1]$ is μ_2, if we wish to approximate μ_2 then because $\text{Var}(X_2) \geq \text{Var}(E[X_2 \mid X_1])$ then we can put more "reliance" on estimates of $E[X_1 \mid X_2]$ to guess the unknown μ_2.

Revised: 3/27/2019