Lecture 17

Conditional distributions (discrete case)

The basic idea behind conditional distributions is simple: Suppose (X, Y) is a jointly-distributed random vector with a discrete joint distribution. Then we can think of $P\{X = x \mid Y = y\}$, for every fixed y, as a distribution of probabilities indexed by the variable x.

Example 1. Recall the following joint distribution from your text:

		possible value for X		
		1	2	3
possible	3	1/6	1/6	0
values	2	1/6	0	1/6
for Y	1	0	1/6	1/6

Now suppose we know that $Y = 3$. Then, the conditional distribution of X, given that we know $Y = 3$, is given by the following:

$$P\{X = 1 \mid Y = 3\} = \frac{P\{X = 1\,, Y = 3\}}{P\{Y = 3\}} = \frac{1/6}{1/3} = \frac{1}{2},$$

$$P\{X = 2 \mid Y = 3\} = \frac{1}{2},$$

similarly, and $P\{X = x \mid Y = 3\} = 0$ for all other values of x. Note that, in this example, $\sum_x P\{X = x \mid Y = 3\} = 1$. Therefore, the "conditional distribution" of X given that $Y = 3$ is indeed a total collection of probabilities. As such, it has an expectation, variance, etc., as well. For instance,

$$E(X \mid Y = 3) = \sum_x P\{X = x \mid Y = 3\} = \left(1 \times \frac{1}{2}\right) + \left(2 \times \frac{1}{2}\right) = \frac{3}{2}.$$

That is, if $Y = 3$, then our best predictor of X is 3/2. Whereas, $EX = 2$, which means that our best predictor of X, in light of no additional information, is 2. You should check that also

$$E(X \mid Y = 1) = \frac{5}{2}, \quad E(X \mid Y = 2) = 2.$$

Similarly,

$$E(X^2 \mid Y = 3) = \left(1^2 \times \frac{1}{2}\right) + \left(2^2 \times \frac{1}{2}\right) = 3,$$

whence

$$\mathrm{Var}(X \mid Y = 3) = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}.$$

You should compare this with the unconditional variance $\mathrm{Var}X = 8/3$ (check the details!).

Fix a number y. In general, the conditional distribution of X given that $Y = y$ is given by the table [function of x]:

$$P\{X = x \mid Y = y\} = \frac{P\{X = x\,, Y = y\}}{P\{Y = y\}} = \frac{P\{X = x\,, Y = y\}}{\sum_a P\{X = a\,, Y = y\}}.$$

It follows easily from this that: (i) $P\{X = x \mid Y = y\} \geq 0$; and (ii) $\sum_x P\{X = x \mid Y = y\} = 1$. This shows that we really are studying a [here conditional] probability distribution. As such,

$$E\left[g(X) \,\middle|\, Y = y\right] = \sum_x g(x) P\{X = x \mid Y = y\},$$

as long as the sum converges absolutely [or when $g(x) \geq 0$ for all x].

Example 2. Choose and fix $0 < p < 1$, and two integers $n, m \geq 1$. Let X and Y be independent random variables; we suppose that X has a binomial distribution with parameters n and p; and Y has a binomial distribution with parameters m and p. Because $X + Y$ can be thought of as the total number of successes in $n + m$ tosses of independent p-coins, it follows that $X + Y$ has a binomial distribution with parameters $n + m$ and p. Our present goal is to find the conditional distribution of X, given that $X+Y = s$, for a fixed integer $0 \leq s \leq n + m$.

$$\begin{aligned} P\{X = x \mid X + Y = s\} &= \frac{P\{X = x\,, X + Y = s\}}{P\{X + Y = s\}} \\ &= \frac{P\{X = x\} \cdot P\{Y = s - x\}}{P\{X + Y = s\}}. \end{aligned}$$

The numerator is zero unless $x = 0\,, \ldots, s$. But if $0 \leq x \leq s$, then

$$\begin{aligned} P\{X = x \mid X + Y = s\} &= \frac{\binom{n}{x} p^x q^{n-x} \cdot \binom{m}{s-x} p^{s-x} q^{m-s+x}}{\binom{n+m}{s} p^s q^{n+m-s}} \\ &= \frac{\binom{n}{x} \cdot \binom{m}{s-x}}{\binom{n+m}{s}}. \end{aligned}$$

That is, given that $X + Y = s$, then the conditional distribution of X is a hypergeometric distribution! We can now read off the mean and the variance from facts we know about hypergeometrics. Namely, according to Example 1 on page 61 of these notes,

$$E(X \mid X + Y = s) = \frac{n}{n+m} \times s,$$

and Example 2 on page 61 tells us that

$$\operatorname{Var}(X \mid X + Y = s) = s \frac{n}{n+m} \frac{m}{n+m} \frac{n+m-s}{s-1}.$$

(Check the details! Some care is needed; on page 61, the notation was slightly different than it is here: The variable B there is now n; N there is now $n+m$, etc.)

Example 3. Suppose X_1 and X_2 are independent, distributed respectively as Poisson(λ_1) and Poisson(λ_2), where $\lambda_1, \lambda_2 > 0$ are fixed constants. What is the distribution of X_1, given that $X_1 + X_2 = s$ for a positive [fixed] integer s?

We will need the distribution of $X_1 + X_2$. Therefore, let us begin with that: The possible values of $X_1 + X_2$ are $0, 1, \ldots$; therefore the "convolution" formula for discrete distributions implies that for all $a \geq 0$,

$$\begin{aligned}
P\{X_1 + X_2 = a\} &= \sum_x P\{X_1 = x\} \cdot P\{X_2 = a - x\} \\
&= \sum_{x=0}^{\infty} \frac{\lambda_1^x e^{-\lambda_1}}{x!} \cdot P\{X_2 = a - x\} \\
&= \sum_{x=0}^{a} \frac{\lambda_1^x e^{-\lambda_1}}{x!} \cdot \frac{\lambda_2^{a-x} e^{-\lambda_2}}{(a-x)!} = \frac{e^{-(\lambda_1+\lambda_2)}}{a!} \cdot \sum_{x=0}^{a} \binom{a}{x} \lambda_1^x \lambda_2^{a-x} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{a!} \cdot (\lambda_1 + \lambda_2)^a \qquad \text{[binomial theorem]};
\end{aligned}$$

and $P\{X_1 + X_2 = a\} = 0$ for other values of a. In other words, $X_1 + X_2$ is distributed as Poisson($\lambda_1 + \lambda_2$).[1]

[1] **Aside:** This and induction together prove that if $X_1, \ldots, X_n$ are independent and all distributed respectively as Poisson(λ_1), ..., Poisson(λ_n), then $X_1 + \cdots + X_n$ is distributed according to a Poisson distribution with parameter $\lambda_1 + \cdots + \lambda_n$.

Now,

$$P\{X_1 = x \mid X_1 + X_2 = s\} = \frac{P\{X_1 = x\} \cdot P\{X_2 = s - x\}}{P\{X_1 + X_2 = s\}}$$
$$= \frac{\dfrac{\lambda_1^x e^{-\lambda_1}}{x!} \cdot \dfrac{\lambda_2^{s-x} e^{-\lambda_2}}{(s-x)!}}{\dfrac{(\lambda_1 + \lambda_2)^s e^{-(\lambda_1+\lambda_2)}}{s!}}$$
$$= \binom{s}{x} \mathcal{P}^x \mathcal{Q}^{s-x},$$

where

$$\mathcal{P} := \frac{\lambda_1}{\lambda_1 + \lambda_2}, \quad \mathcal{Q} := 1 - \mathcal{P} = \frac{\lambda_2}{\lambda_1 + \lambda_2}.$$

That is, the conditional distribution of X_1, given that $X_1 + X_2 = s$, is Binomial$(s\,,\mathcal{P})$.

Continuous conditional distributions

Once we understand conditional distributions in the discrete setting, we could predict how the theory should work in the continuous setting [although it is quite difficult to justify that what is about to be discussed is legitimate].

Suppose $(X\,,Y)$ is jointly distributed with joint density $f(x\,,y)$. Then we *define* the conditional density of X given $Y = y$ [assuming that $f_Y(y) \neq 0$] as

$$f_{X|Y}(x\,|\,y) := \frac{f(x\,,y)}{f_Y(y)} \qquad \text{for all } -\infty < x < \infty.$$

This leads to conditional expectations:

$$E\left[g(X)\mid Y = y\right] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x\,|\,y)\,dx,$$

etc. As such we have now $E(X \mid Y = y)$, $\mathrm{Var}(X \mid Y = y)$, etc. in the usual way using [now conditional] densities.

Example 4 (Uniform on a triangle, p. 414 of your text)**.** Suppose $(X\,,Y)$ is chosen uniformly at random from the triangle $\{(x\,,y) :\ x \geq 0,\, y \geq 0,\, x + y \leq 2\}$. Find the conditional density of Y, given that $X = x$ [for a fixed $0 \leq x \leq 2$].

We know that $f(x\,,y) = \frac{1}{2}$ if $(x\,,y)$ is in the triangle and $f(x\,,y) = 0$ otherwise. Therefore, for all fixed $0 \leq x \leq 2$,

$$f_X(x) = \int_0^{2-x} f(x\,,y)\,dy = \frac{2-x}{2},$$

and $f_X(x) = 0$ for other values of x. This yields the following: For all $0 \le y \le 2 - x$,

$$f_{Y|X}(y\,|\,x) = \frac{f(x\,,y)}{f_X(x)} = \frac{1/2}{(2-x)/2} = \frac{1}{2-x},$$

and $f_{Y|X}(y\,|\,x) = 0$ for all other values of y. In other words, given that $X = x$, the conditional distribution of Y is Uniform$(0\,,2-x)$. Thus, for instance, $P\{Y > 1\,|\,X = x\} = 0$ if $2 - x < 1$ [i.e., $x > 1$] and $P\{Y > 1\,|\,X = x\} = (1-x)/(2-x)$ for $0 \le x \le 1$. Alternatively, we can work things out the longer way by hand:

$$P\{Y > 1\,|\,X = x\} = \int_1^\infty f_{Y|X}(y\,|\,x)\,dy.$$

Now the integrand is zero if $y > 2 - x$. Therefore, unless $0 \le x \le 1$, the preceding probability is zero. When $0 \le x \le 1$, then we have

$$P\{Y > 1\,|\,X = x\} = \int_1^{2-x} \frac{1}{2-x}\,dy = \frac{1-x}{2-x}.$$

As another example within this one, let us compute $E(Y\,|\,X = x)$:

$$\begin{aligned} E(Y\,|\,X = x) &= \int_{-\infty}^{\infty} y f_{Y|X}(y\,|\,x)\,dy \\ &= \int_0^{2-x} \frac{y}{2-x}\,dy = \frac{2-x}{2}. \end{aligned}$$

[Alternatively, we can read this off from facts that we know about uniforms; for instance, we should be able to tell—without computation—that $\mathrm{Var}(Y\,|\,X = x) = (2-x)^2/12$. Check this by direct computation also!]