Conditional distribution

Math 477

December 2, 2014

1 Conditional distribution

1.1 Definition

We have discussed conditional probability $P(A|B)$, which is the probability that A happened given the knowledge that B has happened. In a similar way, for 2 RVs X, Y, we can talk about the probability that X takes some value x given that we know Y has taken some y. If X and Y are correlated in some way, the fact that we have seen Y taking some value should change the probability that X taking value x. Formally, we define, for 2 discrete RVs X, Y

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

1.2 Examples

Example 1.1. *We have discussed that the sum of two independent Poissons λ_1, λ_2 is a Poisson$(\lambda_1 + \lambda_2)$. We can ask a converse question: conditioning on their sum being a particular value, what can we say about the distribution of the individual X, Y? More specifically:*

Let X, Y be independent Poissons (λ_1, λ_2). Find the conditional distributio of X given that $X + Y = n$.

Ans: For $k \leq n$

$$\begin{aligned} P(X = k|X + Y = n) &= \frac{P(X = k, Y = n - k}{X + Y = n} = \frac{e^{-\lambda_1} \frac{(\lambda_1)^k}{k!} e^{-\lambda_2} \frac{(\lambda_2)^{n-k}}{(n-k)!}}{e^{-\lambda_1 + \lambda_2} \frac{((\lambda_1 + \lambda_2))^n}{n!}} \\ &= \binom{n}{k} \Big(\frac{\lambda_1}{\lambda_1 + \lambda_2} \Big)^k \Big(\frac{\lambda_2}{\lambda_1 + \lambda_2} \Big)^{n-k}. \end{aligned}$$

That is conditional on $X+Y=n$, X has distribution $\text{Bin}(n, \frac{\lambda_1}{\lambda_1+\lambda_2})$.

Example 1.2. *We will now do a computation that confirms our intuition that in n independent trials with success probability p for each trial, the all possible orderings of successes and failures are equally likely. More specifically:*

Consider n independent trials, each with success probability p. Given that there are k sucesses, show that all possible orderings of k successes and $n-k$ failures are equally likely.

Ans: Let E be the event of a particular ordering of k successes and $n-k$ failures (for example (1,0,1, 0,0 , 1, 1) if we $n=7$ and $k=4$). Then

$$P(E|X=k) = \frac{P(E, X=k)}{P(X=k)} = \frac{P(E)}{P(X=k)} = \frac{p^k(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{1}{\binom{n}{k}}.$$

Example 1.3. *Joint multinomial distribution*

Definition 1.4. *$X_1, X_2, \cdots, X_k$ are said to have a joint multinomial distributions $(n, p_1, p_2, \cdots, p_k)$, $\sum_i p_i = 1$, if for all tuples $(n_1, n_2, \cdots, n_k)$ such that $\sum_i n_i = n$ we have*

$$P(X_i = n_i, i = 1, \cdots, k) = \binom{n}{n_1, n_2, \cdots, n_k} \prod_{i=1}^{k} p_i^{n_i}.$$

Remark: When $n=2$ this is just a Binomial distribution where X_1 denotes the number of successes and X_2 the number of failures. In general, we think of n trials with k possible outcomes, each outcome with probability p_i.

For simplicity, suppose $k=3$. What is the conditional distribution of X_1, X_2 on $X_3 = n_3$?

Ans: It is reasonable to suspect that they still remain joint multinomial. Specifically:

$$P(X_i = n_i, i = 1, 2 | X_3 = n_3) = \frac{P(X_i = n_i, i = 1, 2, 3)}{P(X_3 = n_3)}.$$

Now you can verify that

$$P(X_3 = n_3) = \binom{n}{n_3} p_3^{n_3}(1-p_3)^{n-n_3}.$$

Therefore,

$$\begin{aligned} P(X_i = n_i, i = 1,2|X_3 = n_3) &= \frac{\binom{n}{n_1,n_2,n_3} p_1^{n_1} p_2^{n_2} p_3^{n_3}}{\binom{n}{n_3} p_3^{n_3} (1-p_3)^{n-n_3}} \\ &= \binom{n-n_3}{n_1} \left(\frac{p_1}{1-p_3}\right)^{n_1} \left(\frac{p_2}{1-p_3}\right)^{n_2}, \end{aligned}$$

where of course $n_1 + n_2 = n - n_3$ and $p_1 + p_2 = 1 - p_3$ therefore $0 \le \frac{p_1}{1-p_3}, \frac{p_2}{1-p_3} \le 1$.

That is, conditioned on $X_3 = n_3$, X_1, X_2 has joint multinomial distribution $n - n_3, \frac{p_1}{1-p_3}, \frac{p_2}{1-p_3}$.

You can see that generally, conditioned on $X_i = n_i, i = r, r+1, \cdots, n$, the remaining $X_1, X_2, \cdots, X_{r-1}$ will have joint multinomial distribution $n - \sum_{i=r}^{n} n_i$ and success probabilities $\frac{p_1}{1-\sum_{i=r}^{n} p_i}, \frac{p_2}{1-\sum_{i=r}^{n} p_i}, \cdots, \frac{p_{r-1}}{1-\sum_{i=r}^{n} p_i}$.

Example 1.5. *An urn has 5 white balls and 8 blue balls. Suppose each white balls are numbered $(W_1, W_2, \cdots, W_5)$. A sample of 3 balls are selected from the urn* **with replacement.** *Let Y_i be 1 if the ith white ball is selected, 0 otherwise. Find the conditional distribution of Y_1 given Y_2.*

Ans:

We want to find $P(Y_1 = 0|Y_2 = 0)$ and $P(Y_1 = 0|Y_2 = 1)$ (the other two can be deduced from these answers). There are several ways to approach these questions. We'll discuss all possibilities. Let's first compute $P(Y_1 = 0|Y_2 = 0)$.

The first is to just interpret the conditional statement. When we condition on $|Y_2 = 0$, it means *we never selected the 2nd white ball from the urn.* Thus the situation is the same as if we select from an urn with 12 balls, where the 2nd white ball is not present.

Thus

$$P(Y_1 = 0|Y_2 = 0) = \left(\frac{11}{12}\right)^3.$$

The second is to just go with the definition. That is

$$P(Y_1 = 0|Y_2 = 0) = \frac{P(Y_1 = 0, Y_2 = 0)}{P(Y_2 = 0)}.$$

Now $P(Y_2 = 0) = \left(\frac{12}{13}\right)^3$. $P(Y_1 = 0, Y_2 = 0) = \left(\frac{11}{13}\right)^3$. Thus you see that the answer is the same as the one we got above.

Now let's compute $P(Y_1 = 0|Y_2 = 1)$. Interpreting the conditional statement here is difficult, *because we do not know how many times we have picked the 2nd white* *ball in our sample.* For example, an incorrect way to interpret this is to think the situation is the same as if our sample size is reduced to 2 balls (but *the urn remains the same as before, with 13 balls,* as we can still pick the 2nd white ball among the other 2 balls in our reduced sample). Thus

$$P(Y_1 = 0|Y_2 = 1) = \left(\frac{12}{13}\right)^2.$$

(It is incorrect because the answer is different from the one we get below, which is the correct one).

A correct way to count our conditional sample space is to note that the sample size is $13^3 - 12^3$ (the total number of ways we can select 3 balls without replacement, minus the number of ways we can select 3 balls but NOT the 2nd white ball). Then the size of our conditional event is

$$\binom{3}{1}11^2 + \binom{3}{2}11^2 + 1,$$

which is just listing out the scenarioes where we picked the 2nd white balls once, twice and 3 times.

Another way to count the size of our conditonal event is to note that since $Y_1 = 0$, we only have 12 choices for each our 3 picks: 12^3. The negation of requiring the 3 picks to have at least 1 of them as the 2nd white ball is to require all 3 picks to not have the 2nd white ball (and also not the 1st white ball by our previous condition): 11^3. Thus the size is $12^3 - 11^3$. You can check that these two give the same answers.

If we use definition,

$$P(Y_1 = 0|Y_2 = 1) = \frac{P(Y_1 = 0, Y_2 = 1)}{P(Y_2 = 1)}.$$

The easiest way to compute $P(Y_2 = 1)$ is to use $1 - P(Y_2 = 0)$. Note that it would be wrong to compute $P(Y_2 = 1) = \binom{3}{1}\frac{1}{13}$, with the reasoning that we choose one of the 3 balls to be the 2nd white ball and let the other 2 be whatever choice there is. This is wrong because it *overcounts the possibilities* (for example it counts the possibility we get the 2nd white balls on all three draws three times). Also the easiest way to compute $P(Y_1 = 0, Y_2 = 1)$ is to use $P(Y_1 = 0) - P(Y_1 = 0, Y_2 = 0) = \left(\frac{12}{13}\right)^3 - \left(\frac{11}{13}\right)^3$.

Note that you can also argue

$$P(Y_1 = 0, Y_2 = 1) = P(Y_2 = 1|Y_1 = 0)P(Y_1 = 0) = (1 - \left(\frac{11}{12}\right)^3)\left(\frac{12}{13}\right)^3.$$

You can check that this is the same as the answer above.

Thus we get

$$P(Y_1 = 0|Y_2 = 1) = \frac{\left(\frac{12}{13}\right)^3 - \left(\frac{11}{13}\right)^3}{1 - \left(\frac{12}{13}\right)^3}.$$

2 Conditional density

2.1 Definition

For continuous RVs, we cannot talk about the probablity that X takes some value, given that we have observed Y taking some value. The reason is the probability that Y taking some value is 0, since it is a continuous RV. This poses a slight problem, since in reality, we always observe Y taking some particular value, even if it is a continuous RV (think about the amount of time you wait for the bus to arrive, for example. You always have to wait a particular amount of time until the bus arrives, even if the probability that the continuous random variable representing the time you wait taking that particular value is 0). So for continuous RVs, we talk about the conditional density instead. Formally, we define, for 2 continuous RVs X, Y

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}.$$

Remark: In the conditional distribution or conditional density formula above, we think of y as *fixed*, and x as taking any possible values in the range of X. Thus the conditional distribution, or conditional density, is a function of x, given a fixed value y. Moreover, for a fixed y, the conditional distribution (or probability density), is a probability distribution (or density). That is

$$\sum_x P(X = x|Y = y) = 1;$$

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx = 1$$

.

Exercise: Prove these two equalites.

Note also that $f_{X|Y}(x|y) = f_X(x)$ if and only if X and Y are independent.

2.2 Conditional probability

We define

$$P(X \leq a|Y = y) = \int_{-\infty}^{x} f_{X|Y}(x|y)dx.$$

Note that if X, Y are independent, then $P(X \leq a|Y = y) = P(X \leq a)$. Moreover, we also have

$$f_{X|Y}(x|y) = \frac{\partial}{\partial x} P(X \leq x|Y = y).$$

2.3 Examples

Example 2.1. *The time it takes for John Solve-alot to finish his first midterm is a Uniform[0, Y], where Y is a Uniform[3/4,5/4] random variable. What is the probability that Mr. Solve-alot will take less than 1 hour to finish his exam?*

Ans: Let X have distribution Uniform(Y). Then

$$f_{X,Y}(x, y) = f_{X|Y}(x|y)f_Y(y) = \frac{2}{y}, 0 \leq x \leq y, 3/4 \leq y < 5/4.$$

It is better to integrate in the order of $dxdy$ here, to get rid of the $1/y$ term. Therefore (draw a plot of the region if it's not clear to you)

$$\begin{aligned} P(X \leq 1) &= \int_{3/4}^{1} \int_{0}^{y} \frac{2}{y} dxdy + \int_{1}^{5/4} \int_{0}^{1} \frac{2}{y} dxdy \\ &= \int_{3/4}^{1} 2dy + \int_{1}^{5/4} \frac{2}{y} dy \\ &= 1/2 + 2\log(5/4). \end{aligned}$$

Example 2.2. *The t-distribution*

Let Z be a standard Normal and Y a Chi-square with n degrees of freedom, Z, Y are independent. Let

$$T = \frac{Z}{\sqrt{Y/n}}.$$

We say T has a t-distribution with n degrees of freedom. What is the conditional distribution of $T|Y$?

Ans: Using the result above for conditional probability, we have

$$\begin{aligned} P(T \le t|Y=y) &= P(\frac{Z}{\sqrt{Y/n}} \le t|Y=y) = P(\frac{Z}{\sqrt{y/n}} \le t|Y=y) \\ &= P(\frac{Z}{\sqrt{y/n}} \le t), \end{aligned}$$

by independence between Z and Y. That is conditioned on $Y = y$, $T|Y = y$ has $\sqrt{n/y}Z$ distribution. We can easily write down its density. Now using the fact that

$$f_{T|Y}(t|y) = \frac{f_{TY}(t,y)}{f_Y(y)},$$

and we know the density of the Chi-square distribution, we can write down the joint density of T and Y. Finally integrating over y for f_{TY} we will obtain the distribution of t. The detailed calculation can be found in Ross.

Example 2.3. *The bivariate Normal distribution*

Let

$$\underline{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \underline{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \boldsymbol{\sigma} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}, -1 < \rho < 1.$$

X, Y is said to have a bivariate Normal distribution with mean $\underline{\mu}$ and Covariance matrix $\boldsymbol{\sigma}$ if

$$f_{XY}(x,y) = \frac{1}{\sqrt{2\pi Det(\boldsymbol{\sigma})}} e^{-\frac{(\underline{x}-\underline{\mu})^T \boldsymbol{\sigma}^{-1} (\underline{x}-\underline{\mu})}{2}}.$$

Remark: If $\underline{y} = \boldsymbol{A}\underline{x} + \underline{b}$ for a martix $\boldsymbol{A}$ and vector $\underline{b}$ then $\underline{y}$ also have Normal distribution with mean $\boldsymbol{A}\underline{\mu} + \underline{b}$ and Covariance matrix $\boldsymbol{A}\boldsymbol{\sigma}\boldsymbol{A}^T$. The result can be proved by characteristic function, which we'll cover later.

In particular, by choosing A to be the basis vectors of $\mathbb{R}^2$ and $\underline{b} = 0$, we conclude that X, Y have marginal distributions Normal(μ_x, σ_x^2) Normal(μ_y, σ_y^2), respectively.

We're interested in computing the conditional density $f_{X|Y}(x|y)$ if x, y haver bivariate Normal distribution.

Writing out explicitly, we have

$$f_{XY}(x,y) = \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_x^2\sigma_y^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]},$$

and

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}.$$

Conditioning on y means we treat y as a CONSTANT. Recall that $f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$. Thus we only need to find the form of $f_{X|Y}(x|y)$, as a function OF x, and neglecting all the constants (involving those that include y). Thus

$$\begin{aligned} f_{X|Y}(x|y) &= Ce^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{x(y-\mu_y)}{\sigma_x\sigma_y}\right]} \\ &= Ce^{-\frac{1}{2(1-\rho^2)\sigma_x^2}\left[x-\left(\mu_x+\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right)\right]^2}. \end{aligned}$$

That is, conditioning on $Y = y$, X has a Normal$\Big(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y), (1-\rho^2)\sigma_x^2\Big)$ distribution. Similarly, conditioning on $X = x$, Y has a Normal$\Big(\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x), (1-\rho^2)\sigma_y^2\Big)$.

3 Conditional expectation

3.0.1 Discrete

Let X, Y be discrete RVs. We define the conditional expectation of X, given $Y = y$ as

$$E(X|Y = y) = \sum_x xP(X = x|Y = y).$$

3.0.2 Continuous

Let X, Y be continuous RVs. We define the conditional expectation of X, given $Y = y$ as

$$E(X|Y = y) = \int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx.$$

Interpretation: Besides the fact that conditional expectation is the average (or mean) value of X given $Y = y$, it is also the *best guess* of X given $Y = y$, in some precise sense that we will discuss below.