Continuous conditional distributions

Márton Balázs* and Bálint Tóth*

October 7, 2022

This little write-up is part of important foundations of probability that were left out of the unit Probability 1 due to lack of time and prerequisites. Here we define the conditional distributions and conditional expectations in the jointly continuous case. This requires some knowledge of two dimensional calculus, and we also assume a general knowledge of joint continuous distributions. The discrete counterpart of all this has been covered in Probability 1.

1 Conditional distributions

We would like to talk about distributions under a condition of the form $\{Y = y\}$. This could easily be done in the case of discrete random variables. However, the event to condition on has zero probability when Y has a continuous distribution. This poses technical difficulties which we get through below. We will assume that the marginal density f_Y of the random variable Y is continuous at y, and also that $f_Y(y) > 0$. Then for every $\varepsilon > 0$ the event $\{Y \in (y,\, y+\varepsilon]\}$ is of positive probability, and we can condition on it without problems.

$$\begin{aligned}\mathbf{P}\{X \le x \,|\, Y \in (y,\, y+\varepsilon]\} = \frac{\mathbf{P}\{X \le x,\ Y \in (y,\, y+\varepsilon]\}}{\mathbf{P}\{Y \in (y,\, y+\varepsilon]\}} &= \frac{F(x,\, y+\varepsilon) - F(x,\, y)}{F_Y(y+\varepsilon) - F_Y(y)} = \\ &= \frac{[F(x,\, y+\varepsilon) - F(x,\, y)]/\varepsilon}{[F_Y(y+\varepsilon) - F_Y(y)]/\varepsilon} \underset{\varepsilon\searrow 0}{\longrightarrow} \frac{\partial F(x,\, y)/\partial y}{f_Y(y)},\end{aligned}$$

and a similar argument holds for $\varepsilon \nearrow 0$. This motivates a meaningful way of defining conditional probabilities given the zero-probability event $\{Y = y\}$:

Definition 1 *Let X and Y be jointly continuous random variables, f_Y continuous at y, and $f_Y(y) > 0$. Then the* conditional distribution function of X, given the condition $\{Y = y\}$ *is defined by*

$$F_{X\,|\,Y}(x \,|\, y) := \lim_{\varepsilon\searrow 0} \mathbf{P}\{X \le x \,|\, Y \in (y,\, y+\varepsilon]\} = \frac{\partial F(x,\, y)/\partial y}{f_Y(y)}.$$

Differentiating this the conditional density function of X, given the condition $\{Y = y\}$ *is*

$$f_{X\,|\,Y}(x \,|\, y) := \frac{\partial^2 F(x,\, y)/[\partial x\, \partial y]}{f_Y(y)} = \frac{f(x,\, y)}{f_Y(y)}.$$

Naturally, fixing y, $F_{X\,|\,Y}(\cdot \,|\, y)$ and $f_{X\,|\,Y}(\cdot \,|\, y)$ are proper distribution and density functions, respectively. It is also clear that X and Y are independent if and only if $F_{X\,|\,Y}(x \,|\, y) = F_X(x)$ and $f_{X\,|\,Y}(x \,|\, y) = f_X(x)$ for all x and y for which these quantities are defined.

Example 2 *Let the joint density of X and Y be given by*

$$f(x,\, y) = \begin{cases} \mathrm{e}^{-x/y} \cdot \mathrm{e}^{-y}/y, & \textit{if } x,\, y > 0, \\ 0, & \textit{otherwise.} \end{cases}$$

Find the conditional density of X under the condition $Y = y$ ($y > 0$).

We start with the marginal density of Y. Let $y > 0$, otherwise the marginal density is zero.

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,\, y)\, \mathrm{d}x = \int_0^{\infty} \mathrm{e}^{-x/y} \cdot \mathrm{e}^{-y}/y\, \mathrm{d}x = \mathrm{e}^{-y}.$$

*University of Bristol / Budapest University of Technology and Economics

Thus $Y \sim \mathrm{Exp}(1)$, and its marginal distribution function is $F_Y(y) = 1 - \mathrm{e}^{-y}$ for positive y, and zero for negative y. Notice that finding the marginal of X would be rather troublesome.

Next the conditional density. According to the above, for $x > 0$ we have

$$f_{X \mid Y}(x \mid y) = \frac{f(x,\, y)}{f_Y(y)} = \frac{\mathrm{e}^{-x/y} \cdot \mathrm{e}^{-y}/y}{\mathrm{e}^{-y}} = \frac{1}{y} \cdot \mathrm{e}^{-x/y},$$

and zero for $x \le 0$. We recognise the $\mathrm{Exp}(\frac{1}{y})$ distribution here, therefore $(X \mid Y = y) \sim \mathrm{Exp}(\frac{1}{y})$, or in short $(X \mid Y) \sim \mathrm{Exp}(\frac{1}{Y})$.

Example 3 *Let $X \sim U(0,\, 1)$ uniformly distributed, and $(Y \mid X = x) \sim U(0,\, x)$, when $0 < x < 1$. Determine the marginal density of Y, and the conditional density of $(X \mid Y = y)$.*

Notice that the information given is sufficient to calculate the joint density of X and Y:

$$f(x,\, y) = f_{Y \mid X}(y \mid x) \cdot f_X(x) = \frac{1}{x} \cdot 1, \qquad \text{if } 0 < y < x < 1$$

(and zero otherwise). This is not constant, hence the joint distribution is *not* uniform on the triangle. Now, the marginal density of Y: let $0 < y < 1$, then

$$f_Y(y) = \int\limits_{-\infty}^{\infty} f(x,\, y)\, \mathrm{d}x = \int\limits_{y}^{1} \frac{1}{x}\, \mathrm{d}x = -\ln y, \qquad \text{and 0 otherwise.}$$

This blows up near 0, as Y prefers to stay close to the origin. For the conditional density, let $0 < y < x < 1$, then

$$f_{X \mid Y}(x \mid y) = \frac{f(x,\, y)}{f_Y(y)} = \frac{1/x}{-\ln y}.$$

The conditional density is not defined for $y \notin (0,\, 1)$, and equals zero if $y \in (0,\, 1)$ but $x \notin (y,\, 1)$. We see that the conditional distribution of X given Y is not uniform either.

Next we show an interesting mixed case. Let N be an integer-valued, and X a continuous random variable. Fix x where the marginal density f_X is continuous and positive, and n such that $p_N(n) > 0$. We can define the following quantities:

- The conditional mass function of N, given $X = x$:

$$p_{N \mid X}(n \mid x) := \lim_{\varepsilon \searrow 0} \frac{\mathbf{P}\{N = n,\ X \in (x,\, x + \varepsilon)\}}{\mathbf{P}\{X \in (x,\, x + \varepsilon)\}} = \lim_{\varepsilon \searrow 0} \frac{\mathbf{P}\{N = n,\ X \in (x - \varepsilon,\, x)\}}{\mathbf{P}\{X \in (x - \varepsilon,\, x)\}},$$

  if the two limits exist.

- The conditional distribution function of X, given the condition $N = n$:

$$F_{X \mid N}(x \mid n) = \frac{\mathbf{P}\{X \le x,\ N = n\}}{\mathbf{P}\{N = n\}}.$$

- The conditional density of X, given the condition $N = n$:

$$\begin{aligned} f_{X \mid N}(x \mid n) &= \lim_{\varepsilon \searrow 0} \frac{F_{X \mid N}(x + \varepsilon \mid n) - F_{X \mid N}(x \mid n)}{\varepsilon} = \lim_{\varepsilon \searrow 0} \frac{\mathbf{P}\{N = n,\ X \in (x,\, x + \varepsilon]\}/\varepsilon}{\mathbf{P}\{N = n\}} = \\ &= \lim_{\varepsilon \searrow 0} \frac{\mathbf{P}\{N = n \mid X \in (x,\, x + \varepsilon]\}}{\mathbf{P}\{N = n\}} \cdot \frac{\mathbf{P}\{X \in (x,\, x + \varepsilon]\}}{\varepsilon} = \frac{p_{N \mid X}(n \mid x)}{p_N(n)} \cdot f_X(x), \end{aligned} \tag{1}$$

  if the limit as $\varepsilon \nearrow$ agrees with this, and also the conditional mass function exists.

As this latter quantity was defined via differentiation, re-integrating it gives

$$1 = \int\limits_{-\infty}^{\infty} \frac{p_{N \mid X}(n \mid x)}{p_N(n)} \cdot f_X(x)\, \mathrm{d}x$$

$$p_N(n) = \int\limits_{-\infty}^{\infty} p_{N \mid X}(n \mid x) \cdot f_X(x)\, \mathrm{d}x. \tag{2}$$

We shall see this identity later as well, using conditional expectations. This formula is actually a continuous version of the Law of Total Probability (aka. Partition Theorem). With the help of this, we can rewrite (1) as

$$f_{X\,|\,N}(x\,|\,n) = \frac{p_{N\,|\,X}(n\,|\,x)\cdot f_X(x)}{\int\limits_{-\infty}^{\infty} p_{N\,|\,X}(n\,|\,\widehat{x})\cdot f_X(\widehat{x})\,\mathrm{d}\widehat{x}}.$$

What we see here is a continuous version of Bayes' Theorem (c.f. the original version).

2 Conditional expectations

Discrete conditional expectations have been dealt with in Probability 1. Here we concentrate on the jointly continuous case.

Definition 4 *Let X and Y be discrete or jointly continuous random variables, and let y be such that $p_Y(y) > 0$ or $f_Y(y) > 0$ and f_Y is continuous in y. Then $p_{X\,|\,Y}(\cdot\,|\,y)$ is a proper probability mass function, or $f_{X\,|\,Y}(\cdot\,|\,y)$ is a proper density function. We can therefore define the* conditional expectation

$$\mathbf{E}(X\,|\,Y=y) := \begin{cases} \sum\limits_i x_i\cdot p_{X\,|\,Y}(x_i\,|\,y), & \textit{(discrete case)}\\ \int\limits_{-\infty}^{\infty} x\cdot f_{X\,|\,Y}(x\,|\,y)\,\mathrm{d}x, & \textit{(continuous case)}\end{cases}$$

(if it exists, which we'll assume from now on).

Example 5 *In Example 2 we have determined the marginal distribution of Y and the conditional distribution $(X\,|\,Y=y)\sim Exp(\frac{1}{y})$ (equivalently, $(X\,|\,Y)\sim Exp(\frac{1}{Y})$). Based on this, $\mathbf{E}(X\,|\,Y=y)=y$. In an equivalent way, we can also write $\mathbf{E}(X\,|\,Y)=Y$ for this.*

We have seen the Tower rule (aka. Law of total expectation) in Probability 1:

Proposition 6 (Tower rule) *Let X and Y be random variables, and g a function of two variables. Then*

$$\mathbf{E}g(X,\,Y) = \mathbf{E}\mathbf{E}(g(X,\,Y)\,|\,Y),$$

where the outer $\mathbf{E}$ refers to the expected value of the random variable $\mathbf{E}(g(X,\,Y)\,|\,Y)$, which is actually a function of Y.

Proof. We prove for the jointly continuous case, the discrete case has been covered in Probability 1. This statement is true in great generality.

$$\mathbf{E}\mathbf{E}(g(X,\,Y)\,|\,Y) = \int\limits_{-\infty}^{\infty}\mathbf{E}(g(X,\,Y)\,|\,Y=y)\cdot f_Y(y)\,\mathrm{d}y = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} g(x,\,y)\cdot f_{X\,|\,Y}(x\,|\,y)\,\mathrm{d}x\cdot f_Y(y)\,\mathrm{d}y =$$
$$= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} g(x,\,y)\cdot f_{X\,|\,Y}(x\,|\,y)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} g(x,\,y)\cdot f(x,\,y)\,\mathrm{d}x\,\mathrm{d}y = \mathbf{E}g(X,\,Y).$$

□

The Tower rule actually generalises the Law of Total Probability:

Corollary 7 *Let X be a random variable, A an event, and $I=\mathbf{1}\{A\}$ its indicator. Then*

$$\mathbf{P}\{A\} = \mathbf{E}I = \mathbf{E}\mathbf{E}(I\,|\,X) = \mathbf{E}\mathbf{P}\{A\,|\,X\},$$

where $\mathbf{P}\{A\,|\,X\}$ can be defined similarly to the above. If X is discrete with possible values x_i, then the above display expands to

$$\mathbf{P}\{A\} = \sum_i \mathbf{P}\{A\,|\,X=x_i\}\cdot p_X(x_i).$$

If X is continuous,

$$\mathbf{P}\{A\} = \int\limits_{-\infty}^{\infty}\mathbf{P}\{A\,|\,X=x\}\cdot f_X(x)\,\mathrm{d}x.$$

Example 8 *Let N be an integer valued, X a continuous random variable. By the above,*

$$p_N(n) = \mathbf{P}\{N = n\} = \int_{-\infty}^{\infty} \mathbf{P}\{N = n \,|\, X = x\} \cdot f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} p_{N\,|\,X}(n\,|\,x) \cdot f_X(x)\,\mathrm{d}x,$$

which is identity (2) *from above.*

Example 9 *Consider the joint distribution of Example 3. Determine the quantities* $\mathbf{E}Y$, $\mathbf{Var}Y$, $\mathbf{E}(X\,|\,Y)$.

Start with using the definitions and the marginal that has been computed:

$$\begin{aligned}
\mathbf{E}Y &= \int_{-\infty}^{\infty} y f_Y(y)\,\mathrm{d}y = \int_0^1 -y\ln y\,\mathrm{d}y = \Bigl[-\frac{y^2}{2}\ln y\Bigr]_0^1 + \int_0^1 \frac{y}{2}\,\mathrm{d}y = \frac{1}{4},\\
\mathbf{E}Y^2 &= \int_{-\infty}^{\infty} y^2 f_Y(y)\,\mathrm{d}y = \int_0^1 -y^2\ln y\,\mathrm{d}y = \Bigl[-\frac{y^3}{3}\ln y\Bigr]_0^1 + \int_0^1 \frac{y^2}{3}\,\mathrm{d}y = \frac{1}{9},\\
\mathbf{Var}Y &= \frac{1}{9} - \frac{1}{4^2} = \frac{7}{144}.
\end{aligned}$$

Now use the tower rule and the conditional distribution $(Y\,|\,X) \sim \mathrm{U}(0,\,X)$ given to us:

$$\begin{aligned}
\mathbf{E}Y &= \mathbf{E}\mathbf{E}(Y\,|\,X) = \mathbf{E}\frac{X}{2} = \frac{1}{2}\cdot\mathbf{E}X = \frac{1}{4},\\
\mathbf{Var}Y &= \mathbf{E}\mathbf{Var}(Y\,|\,X) + \mathbf{Var}\mathbf{E}(Y\,|\,X) = \mathbf{E}\frac{X^2}{12} + \mathbf{Var}\frac{X}{2} =\\
&= \frac{1}{12}\cdot(\mathbf{Var}X + [\mathbf{E}X]^2) + \frac{1}{4}\mathbf{Var}X = \frac{1}{12}\cdot\Bigl(\frac{1}{12} + \frac{1}{2^2}\Bigr) + \frac{1}{4}\cdot\frac{1}{12} = \frac{7}{144}.
\end{aligned}$$

To answer the last question, with the help of the conditional distribution from before, we have for $0 < y < 1$

$$\begin{aligned}
\mathbf{E}(X\,|\,Y = y) &= \int_{-\infty}^{\infty} x\cdot f_{X\,|\,Y}(x\,|\,y)\,\mathrm{d}x = \int_y^1 x\cdot\frac{1/x}{-\ln y}\,\mathrm{d}x = \frac{y-1}{\ln y}, \qquad \text{that is}\\
\mathbf{E}(X\,|\,Y) &= \frac{Y-1}{\ln Y}.
\end{aligned}$$

Notice that this converges to zero as $Y \to 0$. It is less surprising that the limit is 1 as $Y \to 1$. As a verification,

$$\mathbf{E}X = \mathbf{E}\mathbf{E}(X\,|\,Y) = \mathbf{E}\Bigl(\frac{Y-1}{\ln Y}\Bigr) = \int_{-\infty}^{\infty} \frac{y-1}{\ln y}\cdot f_Y(y)\,\mathrm{d}y = \int_0^1 \frac{y-1}{\ln y}\cdot(-\ln y)\,\mathrm{d}y = \frac{1}{2},$$

as it should be due to $X \sim \mathrm{U}(0,\,1)$.

Example 10 *A signal $S \sim \mathcal{N}(\mu,\,\sigma^2)$ is sent through a network. Due to noise, the received signal R has conditional distribution $(R\,|\,S = s) \sim \mathcal{N}(s,\,1)$ given that $S = s$ was sent. What is our best guess for the value of what was sent, if received is $R = r$?*

The best guess (in the least square sense, see in another part) is $\mathbf{E}(S\,|\,R = r)$. To determine this, we need the conditional distribution $(S\,|\,R = r)$. Computing the conditional density:

$$f_{S\,|\,R}(s\,|\,r) = \frac{f(s,\,r)}{f_R(r)} = \frac{f_{R\,|\,S}(r\,|\,s)\cdot f_S(s)}{f_R(r)} = \frac{1}{f_R(r)}\cdot\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-(r-s)^2/2}\cdot\frac{1}{\sqrt{2\pi}\,\sigma}\mathrm{e}^{-(s-\mu)^2/2\sigma^2}.$$

With some work we could get $f_R(r)$ too, but this is not needed if we are only after the s-dependence of the above conditional density. Omitting the factor $-1/2$, we expand the exponent as

$$(r-s)^2 + \frac{(s-\mu)^2}{\sigma^2} = r^2 - 2rs + s^2 + \frac{s^2}{\sigma^2} - 2\frac{s\mu}{\sigma^2} + \frac{\mu^2}{\sigma^2} = \frac{\sigma^2+1}{\sigma^2}\cdot\Bigl(s - r\cdot\frac{\sigma^2}{\sigma^2+1} - \frac{\mu}{\sigma^2+1}\Bigr)^2 + C(r),$$

where $C(r)$ does not depend on s. As $f_R(r)$ is also independent of s, it follows that

$$(S\,|\,R = r) \sim \mathcal{N}\Bigl(r\cdot\frac{\sigma^2}{\sigma^2+1} + \mu\cdot\frac{1}{\sigma^2+1},\,\frac{\sigma^2}{\sigma^2+1}\Bigr), \qquad \text{hence} \qquad \mathbf{E}(S\,|\,R = r) = r\cdot\frac{\sigma^2}{\sigma^2+1} + \mu\cdot\frac{1}{\sigma^2+1}$$

is the best guess. This is a weighted average of the received value r and the mean μ of the sent values. When σ is large, that is, the noise is small relative to the fluctuations of the sent signal, then we mostly take r into account. If, on the other hand, σ is small, the noise is relatively large, then the received signal does not matter too much, and the best guess will be close to μ.