Conditional densities

12.1 Overview

Density functions determine continuous distributions. If a continuous distribution is calculated conditionally on some information, then the density is called a **conditional density**. When the conditioning information involves another random variable with a continuous distribution, the conditional density can be calculated from the joint density for the two random variables.

Suppose $X$ and $Y$ have a jointly continuous distribution with joint density $f(x, y)$. From Chapter 11, you know that the marginal distribution of $X$ is continuous with density

$$g(y) = \int_{-\infty}^{\infty} f(x, y) \, dx.$$  

The conditional distribution for $Y$ given $X = x$ has a (conditional) density, which I will denote by $h(y \mid X = x)$, or just $h(y \mid x)$ if the conditioning variable is unambiguous, for which

$$P\{y \leq Y \leq y + \delta \mid X = x\} \approx \delta h(y \mid X = x), \quad \text{for small } \delta > 0.$$  

Conditioning on $X = x$ should be almost the same as conditioning on the event $\{x \leq X \leq x + \epsilon\}$ for a very small $\epsilon > 0$. That is, provided $g(x) > 0$,

$$P\{y \leq Y \leq y + \delta \mid X = x\} \approx \frac{P\{y \leq Y \leq y + \delta, x \leq X \leq x + \epsilon\}}{P\{x \leq X \leq x + \epsilon\}} \approx \frac{\delta \epsilon f(x, y)}{\epsilon g(x)}.$$  

Statistics 241/541 fall 2014 © David Pollard, 18 Nov 2014
In the limit, as $\epsilon$ tends to zero, we are left with $\delta \approx \delta f(x,y)/g(x)$. That is,
\[ h(y \mid X = x) = f(x,y)/g(x) \quad \text{for each } x \text{ with } g(x) > 0. \]

Less formally, the conditional density is
\[ h(y \mid X = x) = \frac{\text{joint } (X,Y) \text{ density at } (x,y)}{\text{marginal } X \text{ density at } x} \]

The first Example illustrates two ways to find a conditional density: first by calculation of a joint density followed by an appeal to the formula for the conditional density; and then by a sneakier method where all the random variables are built directly using polar coordinates.

**Example** <12.1> Let $X$ and $Y$ be independent random variables, each distributed $N(0,1)$. Define $R = \sqrt{X^2 + Y^2}$. Show that, for each $r > 0$, the conditional distribution of $X$ given $R = r$ has density
\[ h(x \mid R = r) = \frac{1}{\pi \sqrt{r^2 - x^2}} \quad \text{for } r > 0. \]

The most famous example of a continuous condition distribution comes from pairs of random variables that have a bivariate normal distribution. For each constant $\rho \in (-1, +1)$, the **standard bivariate normal with correlation $\rho$** is defined as the joint distribution of a pair of random variables constructed from independent random variables $X$ and $Y$, each distributed $N(0,1)$. Define $Z = \rho X + \sqrt{1 - \rho^2}Y$. The pair $X,Y$ has a jointly continuous distribution with density $f(x,y) = (2\pi)^{-1} \exp \left( -\frac{x^2 + y^2}{2} \right)$. Apply the result from Example <11.4> with $$(X,Z) = (X,Y)A \quad \text{where } A = \begin{pmatrix} 1 & \rho \\ 0 & \sqrt{1 - \rho^2} \end{pmatrix}$$

to deduce that $X,Z$ have joint density
\[ f_\rho(x,z) = \frac{1}{\sqrt{1 - \rho^2}} \exp \left( -\frac{x^2 - 2\rho xz + z^2}{1 - \rho^2} \right). \]

Notice the symmetry in $x$ and $z$. The $X$ and $Z$ marginals must be the same. Thus $Z \sim N(0,1)$. Also
\[ \text{cov}(X,Z) = \text{cov}(X,\rho X + \sqrt{1 - \rho^2}Y) = \rho \text{cov}(X,X) + \sqrt{1 - \rho^2} \text{cov}(X,Y) = \rho. \]
Remark. The **correlation** between two random variables $S$ and $T$ is defined as
\[
\text{corr}(S, T) = \frac{\text{cov}(S, T)}{\sqrt{\text{var}(S)\text{var}(T)}}.
\]

If $\text{var}(S) = \text{var}(T) = 1$ the correlation reduces to the covariance.

By construction, the conditional distribution of $Z$ given $X = x$ is just the conditional distribution of $\rho x + \sqrt{1 - \rho^2} Y$ given $X = x$. Independence of $X$ and $Y$ then shows that
\[
Z \mid X = x \sim N(\rho x, 1 - \rho^2).
\]

In particular, $\mathbb{E}(Z \mid X = x) = \rho x$. By symmetry of $f_\rho$, we also have $X \mid Z = z \sim N(\rho z, 1 - \rho^2)$, a fact that you could check by dividing $f_\rho(x, z)$ by the standard normal density for $Z$.

**Example <12.2>** Let $S$ denote the height (in inches) of a randomly chosen father, and $T$ denote the height (in inches) of his son at maturity. Suppose each of $S$ and $T$ has a $N(\mu, \sigma^2)$ distribution with $\mu = 69$ and $\sigma = 2$. Suppose also that the standardized variables $(S - \mu)/\sigma$ and $(T - \mu)/\sigma$ have a standard bivariate normal distribution with correlation $\rho = .3$.

If Sam has a height of $S = 74$ inches, what would one predict about the ultimate height $T$ of his young son Tom?

For the standard bivariate normal, if the variables are uncorrelated (that is, if $\rho = 0$) then the joint density factorizes into the product of two $N(0, 1)$ densities, which implies that the variables are independent. This situation is one of the few where a zero covariance (zero correlation) implies independence.

The final Example demonstrates yet another connection between Poisson processes and order statistics from a uniform distribution. The arguments make use of the obvious generalizations of joint densities and conditional densities to more than two dimensions.

**Definition.** Say that random variables $X, Y, Z$ have a jointly continuous distribution with joint density $f(x, y, z)$ if
\[
\mathbb{P}\{(X, Y, Z) \in A\} = \iiint_A f(x, y, z) \, dx \, dy \, dz \quad \text{for each } A \subseteq \mathbb{R}^3.
\]
As in one and two dimensions, joint densities are typically calculated by looking at small regions: for a small region $\Delta$ around $(x_0, y_0, z_0)$

$$P\{(X, Y, Z) \in \Delta\} \approx \text{(volume of } \Delta) \times f(x_0, y_0, z_0).$$

Similarly, the joint density for $(X, Y)$ conditional on $Z = z$ is defined as the function $h(x, y \mid Z = z)$ for which

$$P\{(X, Y) \in B \mid Z = z\} = \iiint I\{(x, y) \in B\} h(x, y \mid Z = z) \, dx \, dy$$

for each subset $B$ of $\mathbb{R}^2$. It can be calculated, at $z$ values where the marginal density for $Z$, $g(z) = \iint f(x, y, z) \, dx \, dy$, is strictly positive, by yet another small-region calculation. If $\Delta$ is a small subset containing $(x_0, y_0)$ then, for small $\epsilon > 0$,

$$P\{(X, Y) \in \Delta \mid Z = z_0\} \approx P\{(X, Y) \in \Delta, z_0 \leq Z \leq z_0 + \epsilon\} \approx \frac{\iint f(x, y, z) \, dx \, dy \, \epsilon g(z_0)}{g(z_0)}.$$

Remark. Notice the identification of the set of points $(x, y, z)$ in $\mathbb{R}^3$ for which $(x, y) \in \Delta$ and $z_0 \leq z \leq z_0 + \epsilon$ as a small region with volume equal to $\text{(area of } \Delta) \times \epsilon$.

That is, the conditional (joint) distribution of $(X, Y)$ given $Z = z$ has density

$$h(x, y \mid Z = z) = \frac{f(x, y, z)}{g(z)} \quad \text{provided } g(z) > 0.$$

Remark. Many authors (including me) like to abbreviate $h(x, y \mid Z = z)$ to $h(x, y \mid z)$. Many others run out of symbols and write $f(x, y \mid z)$ for the conditional (joint) density of $(X, Y)$ given $Z = z$. This notation is defensible if one can somehow tell which values are being conditioned on. In a problem with lots of conditioning it can get confusing to remember which $f$ is the joint density and which is conditional on something. To avoid confusion, some authors write things like $f_{X, Y \mid Z}(x, y \mid z)$ for the conditional density and $f_X(x)$ for the $X$-marginal density, at the cost of more cumbersome notation.
Example <12.3> Let $T_i$ denote the time to the $i$th point in a Poisson process with rate $\lambda$ on $[0, \infty)$. Find the joint distribution of $(T_1, T_2)$ conditional on $T_3$.

From the result in the previous Example, you should be able to deduce that, conditional on $T_3 = t_3$ for a given $t_3 > 0$, the random variables $(T_1/T_3, T_2/T_3)$ are uniformly distributed over the triangular region

$\{(u_1 u_2) \in \mathbb{R}^2 : 0 < u_1 < u_2 < 1\}$.

HW11 will step you through an analogous result for order statistics.

12.2 Examples for Chapter 12

Example. Let $X$ and $Y$ be independent random variables, each distributed $N(0, 1)$. Define $R = \sqrt{X^2 + Y^2}$. For each $r > 0$, find the density for the conditional distribution of $X$ given $R = r$.

The joint density for $(X, Y)$ equals $f(x, y) = (2\pi)^{-1} \exp(-\frac{x^2 + y^2}{2})$.

To find the conditional density for $X$ given $R = r$, first I’ll find the joint density $\psi$ for $X$ and $R$, then I’ll calculate its $X$ marginal, and then I’ll divide to get the conditional density. A simpler method is described at the end of the Example.

We need to calculate $P\{x_0 \leq X \leq x_0 + \delta, r_0 \leq R \leq r_0 + \epsilon\}$ for small, positive $\delta$ and $\epsilon$. For $|x_0| < r_0$, the event corresponds to the two small regions in the $(X, Y)$-plane lying between the lines $x = x_0$ and $x = x_0 + \delta$, and between the circles centered at the origin with radii $r_0$ and $r_0 + \epsilon$.

By symmetry, both regions contribute the same probability. Consider the upper region. For small $\delta$ and $\epsilon$, the region is approximately a parallelogram.
with side length $\eta = \sqrt{(r_0 + \epsilon)^2 - x_0^2} - \sqrt{r_0^2 - x_0^2}$ and width $\delta$. We could expand the expression for $\eta$ as a power series in $\epsilon$ by multiple applications of Taylor’s theorem. It is easier to argue less directly, starting from the equalities

$$x_0^2 + (y_0 + \eta)^2 = (r_0 + \epsilon)^2 \quad \text{AND} \quad x_0^2 + y_0^2 = r_0^2.$$

Take differences to deduce that $2y_0 \eta + \eta^2 = 2r_0 \epsilon + \epsilon^2$. Ignore the lower order terms $\eta^2$ and $\epsilon^2$ to conclude that $\eta \approx (r_0 \epsilon / y_0)$. The upper region has approximate area $r_0 \epsilon \delta / y_0$, which implies

$$P\{x_0 \leq X \leq x_0 + \delta, r_0 \leq R \leq r_0 + \epsilon\} \approx 2r_0 \epsilon \delta \approx 2r_0 \sqrt{r_0^2 - x_0^2} \exp\left(-r_0^2/2\right) \pi \epsilon \delta.$$

Thus the random variables $X$ and $R$ have joint density

$$\psi(x, r) = \frac{r \exp(-r^2/2)}{\pi \sqrt{r^2 - x^2}} 1\{|x| < r, 0 < r\}.$$

Once again I have omitted the subscript on the dummy variables, to indicate that the argument works for every $x, r$ in the specified range.

For $r > 0$, the random variable $R$ has marginal density

$$g(r) = \int_{-r}^{r} \psi(x, r) \, dx = \frac{r \exp(-r^2/2)}{\pi} \int_{-r}^{r} \frac{dx}{\sqrt{r^2 - x^2}} \quad \text{put } x = r \cos \theta$$

$$= \frac{r \exp(-r^2/2)}{\pi} \int_{\pi}^{0} -r \sin \theta \, d\theta = r \exp(-r^2/2).$$

The conditional density for $X$ given $R = r$ equals

$$h(x \mid R = r) = \frac{\psi(x, r)}{g(r)} = \frac{1}{\pi \sqrt{r^2 - x^2}} \quad \text{for } |x| < r \text{ and } r > 0.$$

A goodly amount of work.

The calculation is easier when expressed in polar coordinates. From example <11.7> you know how to construct independent $N(0, 1)$ distributed

---

*Statistics 241/541 fall 2014 ©David Pollard, 18 Nov 2014*
random variables by starting with independent random variables \( \tilde{R} \) with density
\[
g(r) = r \exp(-r^2/2)1\{r > 0\},
\]
and \( U \sim \text{Uniform}(0, 2\pi) \): define \( X = \tilde{R} \cos(U) \) and \( Y = \tilde{R} \sin(U) \).

If we start with \( X \) and \( Y \) constructed in this way then
\[
R = \sqrt{X^2 + Y^2} = \tilde{R}
\]
and the conditional density \( h(x \mid R = r) \) is given, for \( |x| < r \) by
\[
\delta h(x \mid R = r) 
\approx \mathbb{P}\{x \leq R \cos(U) \leq x + \delta \mid R = r\} 
= \mathbb{P}\{x \leq r \cos(U) \leq x + \delta\} \quad \text{by independence of } R \text{ and } U 
= \mathbb{P}\{\theta_0 - \epsilon \leq U \leq \theta_0\} + \mathbb{P}\{\theta_0 - \epsilon + \pi \leq U \leq \theta_0 + \pi\}
\]
where \( \theta_0 \) is the unique value in \([0, \pi]\) for which
\[
x/r = \cos(\theta_0) \quad \text{AND} \quad (x + \delta)/r = \cos(\theta_0 - \epsilon) \approx \cos(\theta_0) + \epsilon \sin(\theta_0).
\]
Solve (approximately) for \( \epsilon \) then substitute into the expression for the conditional density:
\[
\delta h(x \mid R = r) \approx \frac{2\epsilon}{2\pi} \approx \frac{\delta}{\pi r \sin(\theta_0)} = \frac{\delta}{\pi r \sqrt{1 - (x/r)^2}}, \quad \text{for } |x| < r,
\]
the same as before. \(\square\)

**Example.** Let \( S \) denote the height (in inches) of a randomly chosen father, and \( T \) denote the height (in inches) of his son at maturity. Suppose each of \( S \) and \( T \) has a \( N(\mu, \sigma^2) \) distribution with \( \mu = 69 \) and \( \sigma = 2 \). Suppose also that the standardized variables \( (S - \mu)/\sigma \) and \( (T - \mu)/\sigma \) have a standard bivariate normal distribution with correlation \( \rho = 0.3 \).

If Sam has a height of \( S = 74 \) inches, what would one predict about the ultimate height \( T \) of his young son Tom?

In standardized units, Sam has height \( X = (S - \mu)/\sigma \), which we are given to equal 2.5. Tom’s ultimate standardized height is \( Y = (T - \mu)/\sigma \). By assumption, before the value of \( X \) was known, the pair \( (X, Y) \) has a standard bivariate normal distribution with correlation \( \rho \). The conditional distribution of \( Y \) given that \( X = 2.5 \) is
\[
Y \mid X = 2.5 \sim N(2.5\rho, 1 - \rho^2)
\]
In the original units, the conditional distribution of $T$ given $S = 74$ is normal with mean $\mu + 2.5\rho \sigma$ and variance $(1 - \rho^2)\sigma^2$, that is,

$$
\text{Tom’s ultimate height | Sam’s height } = 74 \text{ inches } \sim N(70.5, 3.64)
$$

If I had to make a guess, I would predict that Tom would ultimately reach a height of 70.5 inches. □

**Remark.** Notice that Tom expected height (given that Sam is 74 inches) is less than his father’s height. This fact is an example of a general phenomenon called “regression towards the mean”. The term *regression*, as a synonym for conditional expectation, has become commonplace in Statistics.

**Example.** Let $T_i$ denote the time to the $i$th point in a Poisson process with rate $\lambda$ on $[0, \infty)$. Find the joint distribution of $(T_1, T_2)$ conditional on $T_3$.

For fixed $0 < t_1 < t_2 < t_3 < \infty$ and suitably small positive $\delta_1, \delta_2, \delta_3$ define disjoint intervals

$$
I_1 = [0, t_1) \quad I_2 = [t_1, t_1 + \delta_1] \quad I_3 = (t_1 + \delta_1, t_2),
$$

$$
I_4 = [t_2, t_2 + \delta_2], \quad I_5 = (t_2 + \delta_2, t_3), \quad I_6 = [t_3, t_3 + \delta_3].
$$

Write $N_j$ for the number of points landing in $I_j$, for $j = 1, \ldots, 6$. The random variables $N_1, \ldots, N_6$ are independent Poissons, with expected values

$$
\lambda t_1, \quad \lambda \delta_1, \quad \lambda (t_2 - t_1 - \delta_1), \quad \lambda \delta_2, \quad \lambda (t_3 - t_2 - \delta_2), \quad \lambda \delta_3.
$$

To calculate the joint density for $(T_1, T_2, T_3)$ start from

$$
P\{t_1 \leq T_1 \leq t_1 + \delta_1, t_2 \leq T_2 \leq t_2 + \delta_2, t_3 \leq T_3 \leq t_3 + \delta_3\}
\begin{align*}
= & \, P\{N_1 = 0, N_2 = 1, N_3 = 0, N_4 = 1, N_5 = 0, N_6 = 1\} \\
+ & \text{ smaller order terms.}
\end{align*}
$$

Here the “smaller order terms” involve probabilities of subsets of events such as $\{N_2 \geq 2, N_4 \geq 1, N_6 \geq 1\}$, which has very small probability:

$$
P\{N_2 \geq 2\}P\{N_4 \geq 1\}P\{N_6 \geq 1\} = o(\delta_1 \delta_2 \delta_3).
$$

Independence also gives a factorization of the main contribution:

$$
P\{N_1 = 0, N_2 = 1, N_3 = 0, N_4 = 1, N_5 = 0, N_6 = 1\}
\begin{align*}
= & \, P\{N_1 = 0\}P\{N_2 = 1\}P\{N_3 = 0\}P\{N_4 = 1\}P\{N_5 = 0\}P\{N_6 = 1\} \\
= & \, e^{-\lambda t_1}[\lambda \delta_1 + o(\delta_1)]e^{-\lambda (t_2 - t_1 - \delta_1)} \times \\
& \times [\lambda \delta_2 + o(\delta_2)]e^{-\lambda (t_3 - t_2 - \delta_2)}[\lambda \delta_3 + o(\delta_3)] \\
= & \, \lambda^3 \delta_1 \delta_2 \delta_3 e^{-\lambda t_3} + o(\delta_1 \delta_2 \delta_3)
\end{align*}
$$
If you think of $\Delta$ as a small shoebox (hyperrectangle) with sides $\delta_1$, $\delta_2$, and $\delta_3$, with all three $\delta_j$’s of comparable magnitude (you could even take $\delta_1 = \delta_2 = \delta_3$), the preceding calculations reduce to

$$P \{(T_1, T_2, T_3) \in \Delta\} = (\text{volume of } \Delta) \lambda^3 e^{-\lambda t_3} + \text{smaller order terms}$$

where the “smaller order terms” are small relative to the volume of $\Delta$. Thus the joint density for $(T_1, T_2, T_3)$ is

$$f(t_1, t_2, t_3) = \lambda^3 e^{-\lambda t_3} I \{0 < t_1 < t_2 < t_3\}.$$

**Remark.** The indicator function is very important. Without it you would be unpleasantly surprised to find $\iiint_{\mathbb{R}^3} f = \infty$.

Just as a check, calculate the marginal density for $T_3$ as

$$g(t_3) = \int\int_{\mathbb{R}^2} f(t_1, t_2, t_3) \, dt_1 \, dt_2$$

$$= \lambda^3 e^{-\lambda t_3} \int\int I \{0 < t_1 < t_2 < t_3\} \, dt_1 \, dt_2.$$

The double integral equals

$$\int I \{0 < t_2 < t_3\} \left( \int_{t_0}^{t_2} \, dt_1 \right) = \int_{t_0}^{t_3} t_2 \, dt_2 = \frac{1}{2} t_3^2.$$

That is, $T_3$ has marginal density

$$g(t_3) = \frac{1}{2} \lambda^3 t_3^2 e^{-\lambda t_3} I \{t_3 > 0\},$$

which agrees with the result calculated in Example <10.1>.

Calculate the conditional density for a given $t_3 > 0$ as

$$h(t_1, t_2 \mid T_3 = t_3) = \frac{f(t_1, t_2, t_3)}{g(t_3)}$$

$$= \frac{\lambda^3 e^{-\lambda t_3} I \{0 < t_1 < t_2 < t_3\}}{\frac{1}{2} \lambda^3 t_3^2 e^{-\lambda t_3}}$$

$$= \frac{2}{t_3^2} I \{0 < t_1 < t_2 < t_3\}.$$

That is, conditional on $T_3 = t_3$, the pair $(T_1, T_2)$ is uniformly distributed in a triangular region of area $\frac{t_3^2}{2}$.

\[\square\]