Chapter 10

Conditioning on a random variable with a continuous distribution

At this point in the course, I hope you understand the importance of the conditioning formula

$$\mathbb{E}\left(Y \mid \text{ info}\right) = \sum\nolimits_i \mathbb{P}\left(F_i \mid \text{ info}\right) \mathbb{E}\left(Y \mid F_i, \text{ info}\right)$$

for finite or countably infinite collections of disjoint events F_i for which $\Omega = \cup_i F_i$. As a particular case, if X is a random variable that takes only a discrete set of values $\{x_1, x_2, \ldots\}$ then

$$\mathbb{E}\left(Y \mid \text{ info}\right) = \sum\nolimits_i \mathbb{P}\{X = x_i \mid \text{ info}\} \mathbb{E}\left(Y \mid X = x_i, \text{ info}\right).$$

This formula can be simplified by the introduction of the function

$$h(x) = \mathbb{E}\left(Y \mid X = x, \text{ info}\right).$$

For then

$$\mathbb{E}\left(Y \mid \text{ info}\right) = \sum\nolimits_i \mathbb{P}\{X = x_i \mid \text{ info}\} h(x_i) = \mathbb{E}\left(h(X) \mid \text{ info}\right).$$

In this Chapter, I want to persuade you that a similar formula applies when X has a continuous distribution, with density function f (given the info):

$$(*) \qquad \mathbb{E}\left(Y \mid \text{ info}\right) = \mathbb{E}\left(h(X) \mid \text{ info}\right) = \int_{-\infty}^{\infty} h(x) f(x)\, dx.$$

As a special case, when Y equals the indicator function of an event B, the formula reduces to

$$\mathbb{P}B = \int_{-\infty}^{\infty} \mathbb{P}(B \mid X = x) f(x)\, dx.$$

From now on, I will omit explicit mention of the conditioning information "info", writing $h(x)$ for $\mathbb{E}(Y \mid X = x)$.

There are several ways to arrive at formula $(*)$. The most direct relies on the plausible assertion that

$$\mathbb{E}\left(Y \mid X \in J\right) \approx h(x) \qquad \text{if } J \text{ is a small interval with } x \in J.$$

The error of approximation should disappear as J shrinks to the point x. Split $\mathbb{R}$ into a union of disjoint, small intervals $J_i = [x_i, x_{i+1})$, where $x_{i+1} = x_i + \delta$, then condition:

$$\mathbb{E}Y = \sum\nolimits_i \mathbb{P}\{X \in J_i\} \mathbb{E}(Y \mid X \in J_i\} \approx \sum\nolimits_i \delta f(x_i) h(x_i) \approx \int_{-\infty}^{\infty} h(x) f(x)\, dx.$$

The combined errors of all the approximation should disappear in the limit as δ tends to zero.

Alternatively, we could start from a slightly less intuitive assumption that $\mathbb{E}Y$ should be nonnnegative if $\mathbb{E}(Y \mid X = x) \geq 0$ for every x. If we replace Y by $Y - h(X)$ then we have

$$\mathbb{E}\left(Y - h(X) \mid X = x\right) = \mathbb{E}\left(Y \mid X = x\right) - h(x) = 0,$$

which gives $\mathbb{E}(Y - h(X)) \geq 0$. A similar argument applied to $h(X) - Y$ gives $\mathbb{E}\left(h(X) - Y\right) \geq 0$. Equality $(*)$ follows.

REMARK. Notice that formula $(*)$ also implies that

$(**)$, $$\mathbb{E}\left(Yg(X)\right) = \mathbb{E}\left(g(X)h(X)\right) \qquad \text{at least for bounded functions } g$$

because $\mathbb{E}\left(Yg(X) \mid X = x\right) = g(x)h(x)$. In advanced probability theory, the treatment of conditional expectations becomes most abstract. Formula $(**)$ is used to define the conditional expectation $h(x) = \mathbb{E}(Y \mid X = x)$. One needs to show that there exists a random variable of the form $h(X)$, which is uniquely determined up to trivial changes on sets of zero probability, for which

$$\mathbb{E}g(X)\left(Y - h(X)\right) = 0 \qquad \text{for every bounded } g.$$

Essentially $h(X)$ is the best approximation to Y using only information given by X.

With this abstract approach, one then needs to show that conditional expectations have the properties that I have taken as axiomatic for Stat 241.

Example <10.1>: The convolution formula for densities derived from $(*)$.

The Poisson process is often used to model the arrivals of customers in a waiting line, or the arrival of telephone calls at an exchange. The underlying idea is that of a large population of potential customers, each of whom acts independently of all the others.

Example <10.2>: A queuing problem with a surprising solution (can be skipped)

EXAMPLES FOR CHAPTER 10

<10.1> **Example.** Suppose X and Y are independent random variables with continuous distributions. If X has density f and Y has density g then (see Chapter 7) the random variable $Z = X + Y$ has density

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

The same formula can be derived from the formula

$$\mathbb{P}B = \int_{-\infty}^{\infty} \mathbb{P}(B \mid X = x) f(x)\, dx,$$

applied with $B = \{z \leq Z \leq z + \delta\}$ for a small, positive δ. Note that

$$\begin{aligned}\mathbb{P}(z \leq Z \leq z + \delta \mid X = x) &= \mathbb{P}(z - x \leq Y \leq z - x + \delta \mid X = x) \\ &= \mathbb{P}(z - x \leq Y \leq z - x + \delta) \qquad \text{because } X, Y \text{ independent} \\ &\approx \delta g(z - x).\end{aligned}$$

Invoke the conditioning formula.

$$\mathbb{P}\{z \leq Z \leq z + \delta\} \approx \int_{-\infty}^{\infty} \delta g(z - x) f(x)\, dx,$$

which leads us back to the convolution formula. □

<10.2> **Example.** Suppose an office receives two different types of inquiry: persons who walk in off the street, and persons who call by telephone. Suppose the two types of arrival are described by independent Poisson processes, with rate λ_w for the walk-ins, and rate λ_c for

the callers. What is the distribution of the number of telephone calls received before the first walk-in customer?

Write T for the arrival time of the first walk-in, and let N be the number of calls in $[0, T)$. The time T has a continuous distribution, with the exponential density $f(t) = \lambda_w e^{-\lambda_w t}$ for $t > 0$. We need to calculate $\mathbb{P}\{N = i\}$ for $i = 0, 1, 2, \ldots$. From formula $(*)$, with A equal to $\{N = i\}$,

$$\mathbb{P}\{N = i\} = \int_0^\infty \mathbb{P}\{N = i \mid T = t\} f(t)\, dt.$$

The conditional distribution of N is affected by the walk-in process only insofar as that process determines the length of the time interval over which N counts. Given $T = t$, the random variable N has a Poisson$(\lambda_c t)$ conditional distribution. Thus

$$\begin{aligned}
\mathbb{P}\{N = i\} &= \int_0^\infty \frac{e^{-\lambda_c t}(\lambda_c t)^i}{i!} \lambda_w e^{-\lambda_w t}\, dt \\
&= \lambda_w \frac{\lambda_c^i}{i!} \int_0^\infty \left(\frac{x}{\lambda_c + \lambda_w}\right)^i e^{-x} \frac{dx}{\lambda_c + \lambda_w} \qquad \text{putting } x = (\lambda_c + \lambda_w)t \\
&= \frac{\lambda_w}{\lambda_c + \lambda_w} \left(\frac{\lambda_c}{\lambda_c + \lambda_w}\right)^i \frac{1}{i!} \int_0^\infty x^i e^{-x}\, dx
\end{aligned}$$

The $1/i!$ and the last integral cancel. (Compare with $\Gamma(i + 1)$.) Writing p for $\lambda_w/(\lambda_c + \lambda_w)$ we have

$$\mathbb{P}\{N = i\} = p(1 - p)^i \qquad \text{for } i = 0, 1, 2, \ldots$$

That is, $1 + N$ has a geometric(p) distribution. The random variable N has the distribution of the number of tails tossed before the first head, for independent tosses of a coin that lands heads with probability p.

Such a nice clean result couldn't happen just by accident. Maybe we don't need all the Calculus to arrive at the distribution for N. In fact, the properties of the Poisson distribution and Problem 8.1 show what is going on, as I will now explain.

Consider the process of all inquiries, both walk-ins and calls. In an interval of length t, the total number of inquiries is the sum of a Poisson$(\lambda_w t)$ distributed random variable and an independent Poisson$(\lambda_c t)$ distributed random variable; the total has a Poisson$(\lambda_w t + \lambda_c t)$ distribution. Both walk-ins and calls contribute independent counts to disjoint intervals; the total counts for disjoint intervals are independent random variables. It follows that the process of all arrivals is a Poisson process with rate $\lambda_w + \lambda_c$.

Now consider an interval of length t in which there are X walk-ins and Y calls. From Problem 8.1, given that $X + Y = n$, the conditional distribution of X is Bin(n, p), where

$$p = \frac{\lambda_w t}{\lambda_w t + \lambda_c t} = \frac{\lambda_w}{\lambda_w + \lambda_c}$$

That is, X has the conditional distribution that would be generated by the following mechanism:

(1) Generate inquiries as a Poisson process with rate $\lambda_w + \lambda_c$.

(2) For each inquiry, toss a coin that lands heads with probability $p = \lambda_w/(\lambda_w + \lambda_c)$. For a head, declare the arrival to be a walk-in, for a tail declare it to be a call.

A formal proof that this two-step mechanism does generate a pair of independent Poisson processes, with rates λ_w and λ_c, would involve:

(1′) Prove independence between disjoint intervals. (Easy)

(2′) If step 2 generates X walk-ins and Y calls in an interval of length t, show that

$$\mathbb{P}\{X = i, Y = j\} = \mathbb{P}\{X = i\}\mathbb{P}\{Y = j\}$$
$$X \sim \text{Poisson}(\lambda_w t) \qquad \text{and} \qquad Y \sim \text{Poisson}(\lambda_c t)$$

You should be able to write out the necessary conditioning argument for $(2')$.

The two-step mechanism explains the appearance of the geometric distribution in the problem posed at the start of the Example. The classification of each inquiry as either a walk-in or a call is effectively carried out by a sequence of independent coin tosses, with probability p of a head (= a walk-in). The problem asks for the distribution of the number of tails before the first head. The embedding of the inquiries into continuous time is irrelevant. □

