Conditioning a random variable on an event

Let X be a continuous random variable and A be an event with $\mathrm{P}(A) > 0$.

Then the **conditional pdf** of X given A is defined as the non-negative function $f_{X|A}$ that obeys

$$\mathrm{P}(B|A) = \int_B f_{X|A}(x)\,\mathrm{d}x \quad \text{for all events } B \text{ defined by } X.$$

A special (and very important) case is when the event A explicitly involves the random variable X itself:

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(X \in B, X \in A)}{\mathrm{P}(X \in A)} = \frac{\int_{A \cap B} f_X(x)\,\mathrm{d}x}{\int_A f_X(x)\,\mathrm{d}x},$$

in which case

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{\mathrm{P}(A)}, & x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

This is simply a **rescaling** of the pdf of X over the set A.

Example. Let T be a random variable corresponding to the amount of time that a new light bulb takes to burn out. We model T using an *exponential* pdf

$$f_T(t) = \begin{cases} \lambda\,\mathrm{e}^{-\lambda t}, & t \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

In shorthand, we write this as $T \sim \mathrm{Exp}(\lambda)$.

Suppose we observe the light bulb at time t_0 and it has burned out. What is the conditional pdf for T?

Here, the event A is

$$A = \{T \leq t_0\},$$

and so

$$f_{T|A}(t) = \begin{cases} \frac{f_T(t)}{\mathrm{P}(A)} & 0 \leq t \leq t_0 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \frac{\lambda\,\mathrm{e}^{-\lambda t}}{1-\mathrm{e}^{-\lambda t_0}} & 0 \leq t \leq t_0 \\ 0 & \text{otherwise.} \end{cases}$$

Now suppose that we observe the light bulb at time t_0 and it *has not* burned out yet. Now what is the conditional pdf for T conditioned on $A = \{T \geq t_0\}$?

$$f_{T|A}(t) = \begin{cases} \frac{\lambda\,\mathrm{e}^{-\lambda t}}{\mathrm{e}^{-\lambda t_0}} & t \geq t_0 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \lambda \mathrm{e}^{-\lambda(t-t_0)}, & t \geq t_0 \\ 0 & \text{otherwise.} \end{cases}$$

Notice that if X is the *additional* amount of time *after* t_0, $X = T - t_0$, then $X \sim \mathrm{Exp}(\lambda)$, which is *exactly* the same pdf as T.

In this sense, exponential random variables are **memoryless**. It does not matter if it has been a second, a day, or a year, the amount of additional time always has the same conditional distribution.

Total probability theorem for pdfs

If $A_1, \ldots, A_n$ are events that *partition* the sample space Ω,

$$A_i \cap A_j = \emptyset, \quad \text{and}$$
$$\bigcup_{i=1}^{n} A_i = \Omega,$$

then we can break apart the pdf $f_X(x)$ for a random variable X as

$$f_X(x) \;=\; \sum_{i=1}^{n} \mathrm{P}\left(A_i\right)\, f_{X|A_i}(x)$$

Exercise:
Suppose that Sublime Doughnuts makes a fresh batch once every hour starting at 6am. You enter the store between 8:30am and 10:15am, with your arrival time being a uniform random variable over this interval. What is the pdf for how old the doughnuts are?

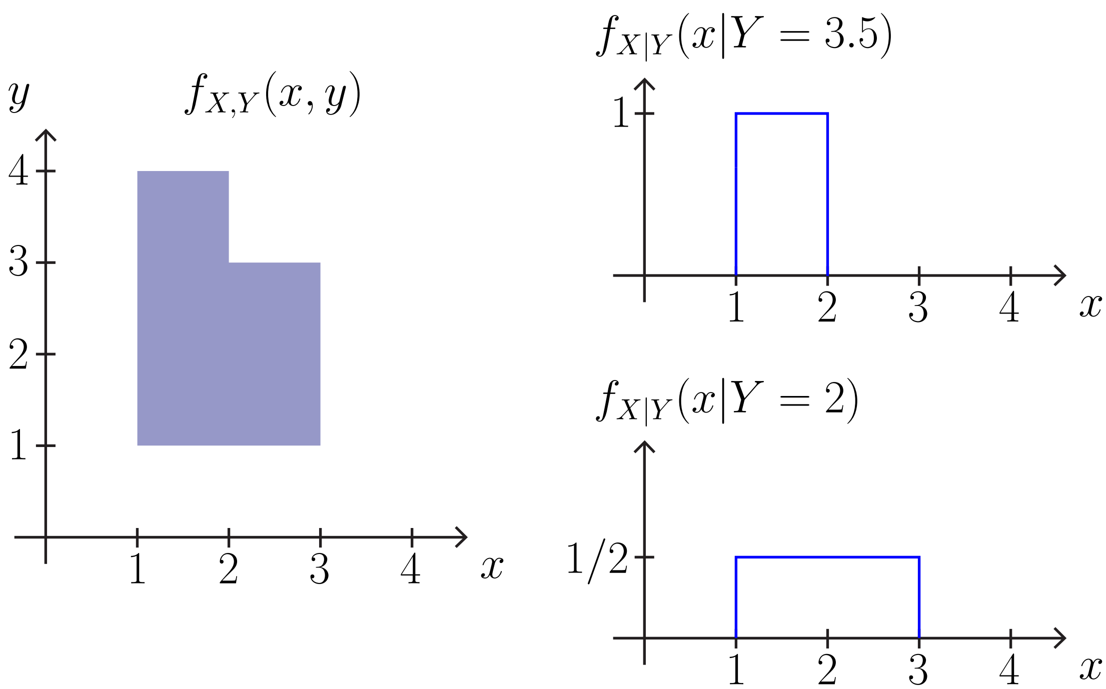
Conditioning one random variable on another

Let X, Y be continuous random variables with joint pdf $f_{X,Y}(x, y)$. For any y with $f_Y(y) > 0$, we can define the **conditional pdf**:

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

The conditional pdf is a **valid pdf** which reflects how our knowledge of X changes given a certain observation $Y = y$. For any fixed value of y, this is just a function of x, but it can be a *different function* for different values of y.

Example.



Example. Uniform pdf on a disc

You throw a dart at a circular target of radius r. We will assume you always hit the target, and each point of impact (x, y) is equally likely, so that

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi r^2}, & \text{if } x^2 + y^2 \leq r^2 \\ 0 & \text{otherwise.} \end{cases}$$

First, we calculate $f_Y(y)$:

$$f_Y(y) = \int f_{X,Y}(x, y) \, \mathrm{d}x$$

$$=$$

Now we have

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

$$=$$

Notice that our definition of conditional pdf gives us a general way of **factoring** the joint pdf:

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

or equivalently

$$f_{X,Y}(x, y) = f_{Y|X}(y|x) \, f_X(x).$$

Sometimes, it is more natural to build up a joint model using this factorization, as the next example illustrates.

Exercise:
The speed of a typical vehicle on I-285 can be modeled as an exponentially distributed random variable X with mean 65 miles per hour. Suppose that we (or a police officer) measure the speed Y of a randomly chosen vehicle using a radar gun, but our measurement has an error which is modeled as a normal random variable with zero mean and standard deviation equal to one tenth of the vehicle's speed. What is the joint pdf of X and Y?

Conditional expectation

Once we have the conditional density defined, the definition of conditional expectation is straightforward.

- If A is an event with $\mathrm{P}(A) > 0$, then

$$\mathrm{E}[X|A] = \int_{-\infty}^{\infty} x\, f_{X|A}(x)\ \mathrm{d}x$$

- If Y is a continuous random variable, then

$$\mathrm{E}[X|Y=y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\ \mathrm{d}x$$

- The definitions above extend to arbitrary functions $g(X)$ of X:

$$\mathrm{E}[g(X)|A] = \int_{-\infty}^{\infty} g(x)\ f_{X|A}(x)\ \mathrm{d}x$$

$$\mathrm{E}[g(X)|Y=y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\ \mathrm{d}x$$

- **Total expectation theorem**
  If $A_1, \ldots, A_n$ are disjoint events that partition the sample space, then

$$\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X|A_i]\ \mathrm{P}(A_i)$$

  Similarly, if Y is a continuous random variable, then

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} \mathrm{E}[X|Y=y]\, f_Y(y)\ \mathrm{d}y$$

- The definition of conditional expectation and the total expectation theorem extends to arbitrary functions $g(X, Y)$ of the random variables X, Y as well:

$$\mathrm{E}[g(X,Y)|Y=y] = \int g(x,y)\, f_{X|Y}(x|y)\, \mathrm{d}x$$

and

$$\mathrm{E}[g(X,Y)] = \int \mathrm{E}[g(X,Y)|Y=y]\, f_Y(y)\, \mathrm{d}y$$

or equivalently

$$\mathrm{E}[g(X,Y)] = \int \mathrm{E}[g(X,Y)|X=x]\, f_X(x)\, \mathrm{d}x$$

Exercise:
Suppose that the random variable X has the piecewise constant pdf

$$f_X(x) = \begin{cases} 2/3, & 0 \le x \le 1, \\ 1/3, & 1 < x \le 2, \\ 0, & \text{otherwise} \end{cases}$$

What is $\mathrm{E}[X]$ and $\mathrm{var}(X)$?

Iterated expectation

Here is an identity which seems a little weird at first, but is actually very useful:

$$\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|Y]]$$

This is called the law of iterated expectation, or **double expectation**. Don't worry if that expression looks confusing the first time you see it; everybody thinks that. Hopefully the explanation below will help you make some sense of it.

Let's see where the law of iterated expectation comes from. By now, we are comfortable with the notion of conditional expectation; if X and Y are related random variables, then observing $Y = y$ may change the distribution, and hence the expectation, of X. Thus

$$\mathrm{E}[X|Y = y] \quad \text{is a function of the observed value } y$$

Without observing Y (i.e. "pinning down" its value),

$$\mathrm{E}[X|Y] \quad \text{is a function of random variable } Y$$

and hence is itself a random variable. We might write

$$g(y) = \mathrm{E}[X|Y = y], \quad \text{and} g(Y) = \mathrm{E}[X|Y].$$

By the *total expectation theorem*

$$\begin{aligned}\mathrm{E}[X] &= \int \mathrm{E}[X|Y = y]\, f_Y(y)\, \mathrm{d}y = \int g(y)\, f_Y(y)\, \mathrm{d}y \\ &= \mathrm{E}[g(Y)] = \mathrm{E}[\mathrm{E}[X|Y]]\end{aligned}$$

Example. Suppose that a coin is potentially biased and that the probability of heads, denoted by P is itself random with a uniform distribution over $[0, 1]$. We toss the coin n times, and let X be the number of heads obtained. Then for any fixed value of $P = p$,

$$\mathrm{E}[X|P = p] = np,$$

and so $\mathrm{E}[X|P]$ is the random variable

$$\mathrm{E}[X|P] = nP.$$

Then the expected number of heads $\mathrm{E}[X]$ is

$$\begin{aligned}\mathrm{E}[X] &= \mathrm{E}[\mathrm{E}[X|P]] \\ &= \mathrm{E}[nP] \\ &= n\,\mathrm{E}[P] \\ &= n/2\end{aligned}$$

since $\mathrm{E}[P] = 1/2$.

Exercise:
You are holding a stick of length ℓ. You choose a point uniformly at random along the length of the stick and break it, keeping the piece in your left hand. You then repeat this process, breaking the (now smaller) stick at a randomly chosen location and then keeping the piece in your left hand.

What is the expected length of the piece we are left with after breaking twice?