Some Examples of Borel's "Paradox"

by Dennis Cox

Here, we introduce Borel's paradox with a simple example, and provide a couple of exercises to elucidate it further.

First, a brief introduction to conditional probability and conditional distributions. With conditional probability, we restrict the sample space to a collection of outcomes of interest, and re-adjust the probabilities. The basic idea is that we are given some information (that our outcome is in some subset) and we want to update the probabilities to account for this. The formula is

$$P(A|B) \;=\; \frac{P(A \cap B)}{P(B)}, \quad \text{provided } P(B) > 0, \tag{1}$$

where A and B are two events (subsets of the sample space), and B is the event that has been given to have occurred. Thus if we know B has occurred, then we update our probability that A occurs by looking only at the outcomes in A that are in B (i.e., $A \cap B$), and we renormalize so that all probabilities total up to 1 by dividing by the probability of B.

Here is a simple example. Suppose we flip a coin twice. The probability of two heads is

$$P(\{HH\}) \;=\; 1/4,$$

since there are 4 equally likely outcomes:

$$\text{Sample space } \Omega \ = \ \{TT, TH, HT, HH\}.$$

However suppose we are given the additional information that at least one flip is a head. Then

$$\begin{aligned} P(\text{ 2 Heads } \mid \text{ at least 1 Head }) \quad &= \quad \frac{P(\{HH\})}{P(\{TH, HT, HH\})} \\ &= \quad 1/3. \end{aligned}$$

Intuitively, this makes sense: if we know at least one head has occurred, our chance of 2 heads should go up. (Note: Many people feel intuitively that the chance of 2 heads given at least 1 head should be 1/2. If you feel this way, you should ask yourself why this is incorrect).

This extends to computing conditional distributions of discrete random variables. Continuing with the coin flipping, suppose we flip 3 times and define random variables

$$\begin{aligned} Y \quad &= \quad \text{total number of heads in first two flips} \\ X \quad &= \quad \text{total number of heads in second two flips.} \end{aligned}$$

Then the conditional distribution of X given the value of Y will depend on the value of Y. Denoting the conditional probability mass function as $f_{X|Y}(x|y)$, we have given $Y = 0$:

$$\begin{aligned} f_{X|Y}(0|0) \quad &= \quad P[X = 0|Y = 0] \ = \ P(\{TTT\})/P(\{TTT, TTH\}) \ = \ 1/2, \\ f_{X|Y}(1|0) \quad &= \quad P[X = 1|Y = 0] \ = \ P(\{TTH\})/P(\{TTT, TTH\}) \ = \ 1/2, \\ f_{X|Y}(2|0) \quad &= \quad 0. \end{aligned}$$

Note that for the latter, even though it is possible to get 2 heads in second 2 flips, it can't happen unless we get a H on the second flip. In a similar way, we can compute $f_{X|Y}(x|1)$:

$$\begin{aligned}
f_{X|Y}(0|1) &= P[X=0|Y=1] \;=\; P(\{HTT\})/P(\{THT, THH, HTT, HTH\}) \;=\; 1/4, \\
f_{X|Y}(1|1) &= P(\{HTH, THT\})/P(\{THT, THH, HTT, HTH\}) \;=\; 1/2, \\
f_{X|Y}(2|1) &= P(\{THH\})/P(\{THT, THH, HTT, HTH\}) \;=\; 1/4.
\end{aligned}$$

This all works as long as $P(B) > 0$ in (1). As events of probability 0 shouldn't occur "very" often, we might be tempted to say that we can ignore this possibility, but when we look at conditional distributions given continuous random variables, we are confronted with the problem. Of course, continuous random variables don't really happen in the real world (or if they did, we wouldn't be able to record them anyway), but they arise theoretically as limits of discrete random variables. Think of a continuous probability density function as approximating the probability mass function for a rounded off random variable obtained by some measurement process - e.g., draw a person at random and measure his/her height. We will have to record it to some finite number of decimal places, so the best we can say is the true height is in some range $x \pm \Delta x/2$. Assuming the probability density function for height of a random person is $f(x)$, and X represents the height of the random person, we have

$$P[X \in (x-\Delta x/2, x+\Delta x/2)] \;=\; \int_{x-\Delta x/2}^{x+\Delta x/2} f(\xi) d\xi \;=\; f(x)\Delta x + o(\Delta x) \;\approx\; f(x)\Delta x,$$

provided f is continuous at x.

Now suppose we measure both height and weight of our random person, denoted X and Y, respectively. Then, we have a bivariate probability density function f_{XY} which (assuming continuity again) satisfies

$$\begin{aligned} & P[X \in (x - \Delta x/2, x + \Delta x/2) \text{ and } Y \in (y - \Delta y/2, y + \Delta y/2)] \\ = & \quad = \int_{x-\Delta x/2}^{x+\Delta x/2} = \int_{y-\Delta y/2}^{y+\Delta y/2} f_{XY}(\xi, \eta) d\eta d\xi \\ \approx & \quad f_{XY}(x, y)\Delta x \Delta y. \end{aligned}$$

As long as $P[Y \in (y - \Delta y/2, y + \Delta y/2)] > 0$, we can compute

$$\begin{aligned} & P[X \in (x - \Delta x/2, x + \Delta x/2) \,|\, Y \in (y - \Delta y/2, y + \Delta y/2)] \\ = & \quad \frac{P[X \in (x - \Delta x/2, x + \Delta x/2) \text{ and } Y \in (y - \Delta y/2, y + \Delta y/2)]}{P[Y \in (y - \Delta y/2, y + \Delta y/2)]} \quad (2) \\ \approx & \quad \frac{f_{XY}(x, y)\Delta x \Delta y}{f_Y(y)\Delta y} \quad (3) \\ = & \quad f_{X|Y}(x|y)\Delta x. \quad (4) \end{aligned}$$

Here, the conditional probability density function

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

is defined in a similar way to the conditional probability mass function.

Note that in the expression in (2) we will need some additional assumption like that f_Y is bounded away from zero in an appropriate neighborhood of y in order that the remainder term is $o(\Delta x)$.

With this formal definition of conditional densities, we can proceed to compute conditional densities, but often we can reason what the conditional density should be without resorting to long calculations. For example, suppose we have an explorer who goes to a "random" spot on the globe - say that the probability she is in any particular region is proportional to the area of the region. (We shall suppose the earth is a perfect sphere for this exercise.) Thus, her location is "uniformly" distributed on the earth. (If this seems too flippant, consider the problem of an object from outter space striking the earth - it would be reasonable a priori to assume a uniform distribution for the strike point). Now, suppose we are given the additional information that she is on the equator (which is 0 degrees latitude). Given this new information, what is the conditional distribution of her position (as would be determined by her longitude, with is between -180 and +180 degrees)?

Intuitively, it would seem that she is equally likely to be anywhere along the equator, so, for instance, the probability that she is in South America is the probability that her longitude is between -49 and -80 degrees (approximately; check it out on the globe), which is $31/360 \doteq 0.08$.

How can we verify this formally? We will parameterize the earth's surface: let (θ, λ) be longitude and latitude, respectively (in radians), where $0 \leq \theta < 2\pi$ and $-\pi/2 \leq \lambda \leq \pi/2$. Note that $\lambda = \phi - \pi/2$ where ϕ is the angle of a point from the North Pole (when you consider the rays from the center of the earth through the North Pole and the center to the given point on the surface of the earth), and we usually use ϕ in spherical coordinates. Then the element of surface area is

$$dA(\theta, \lambda) \;=\; \cos(\lambda) d\theta d\phi.$$

Since the total area (in solid angle) is 4π, we have the joint density

$$f(\theta, \lambda) \;=\; \begin{cases} \frac{1}{4\pi}\cos(\lambda) & \text{if } 0 \leq \theta < 2\pi \text{ and } -\pi/2 \leq \lambda \leq \pi/2, \\ 0 & \text{otherwise.} \end{cases}$$

This is the density for the uniform distribution on the sphere. The (marginal) density for λ is

$$\begin{aligned} f(\lambda) &= \int_0^{2\pi} f(\theta, \lambda) d\theta \\ &= \frac{1}{2}\cos(\lambda), \quad -\pi/2 \le \lambda \le \pi/2. \end{aligned}$$

Thus, we obtain the conditional distribution for longitude given a value of latitude to be

$$\begin{aligned} f(\theta|\lambda) &= \frac{\frac{1}{4\pi}\cos(\lambda)}{\frac{1}{2}\cos(\lambda)} \\ &= \frac{1}{2\pi}, \quad 0 \le \theta < 2\pi, \end{aligned}$$

and so, the conditional distribution of longitude given latitude is a uniform distribution, no matter what the given value of latitude is, which seems intuitively correct.

However, now let's consider the conditional distribution of latitude given a value of longitude. First, the marginal distribution of longitude is uniform (in fact, latitude and longitude are independent as their joint density factors into the product of their marginal densities):

$$f(\theta) = \frac{1}{2\pi}, \quad 0 \le \theta < 2\pi.$$

Hence,

$$f(\lambda|\theta) = \frac{1}{2}\cos(\lambda), \quad -\pi/2 \le \lambda \le \pi/2.$$

Here is the "paradox": Suppose we are given the information that our random explorer is on the prime meridian (0 degrees longitude), which is 1/2 of a great circle (the equator is an entire great circle). What is the conditional distribution of the latitude given that longitude is 0? It has this cosine distribution, which means there is less chance of finding her in a neighborhood of the poles than in a corresponding neighborhood of the equator. But it seems like geometrically, this should be the same as in the equator case - uniform. In fact, if we are given that she is at 0 or 180 degrees, which forms a great circle, it should be uniform, shouldn't it?

Exercises:

1. Of course, we can't know the explorer's position exactly - there is always some measurement error. Consider the problem of conditioning on latitude or longitude 0 degrees when we simply know the location up to the nearest degree (i.e., within $\pm 1/2$ degree) for both coordinates. Argue that the formal result give approximately correct values in that case, although there are issues near the poles.

2. Here is another example: Let (U, V) be uniformly distributed on the unit square, i.e. they have joint (Lebesgue) probability density

$$f_{UV}(u, v) \;=\; 1, \quad 0 < u < 1,\; 0 < v < 1.$$

(a) Let $X = V - U$. Find the joint density of U and X and the conditional density of U given $X = x$.

(b) Let $Y = V/U$. Find the joint density of U and Y and the conditional density of U given $Y = y$.

(c) Note that the events $[U = V]$ and $[X = 0]$ and $[Y = 1]$ are all the same. However,

$$f_{U|X}(u|0) \;\neq\; f_{U|Y}(u|1).$$

Explain.