Random variables (continuous)

Saad Mneimneh

1 Defining probability density

A random variable X is a continuous random variable if its domain, i.e. the set of values x that X can take, is continuous. In this case, and analogous to the discrete case, we define a probability density function $f_X(x)$ (as opposed to a probability mass function), which is a continuous function of x that captures the notion of probability in the following way:

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

We will drop the subscript X from f_X whenever it is clear from the context. Accordingly,

$$P(X = x) = \int_x^x f(y)dy = 0$$

for any single value x. This is because the domain of the random variable is now uncountable, and for the total probability to be 1, each single value must have 0 probability.

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Note that 0 is not really 0 in the sense that $X = x$ could occur. For instance,

$$\frac{P(X = a)}{P(X = b)} = \frac{f(a)}{f(b)}$$

This can be seen from the following approximation. Let δ be a small increment, then:

$$P(x \leq X \leq x + \delta) \approx f(x)\delta$$

Therefore, $P(a \leq X \leq a + \delta)/P(b \leq X \leq b + \delta) \approx \frac{f(a)\delta}{f(b)\delta}$. Taking the limit as δ goes to 0, we have $P(X = a)/P(X = b) = f(a)/f(b)$.

All the concepts that we have seen in the discrete case, carry over to the continuous case:

- Expectation: $E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx$
- Independence: X and Y are independent $\Leftrightarrow f_{X,Y}(x,y) = f_X(x)f_Y(y)$
- Conditional density: $f_{X|Y}(x) = f_{X,Y}(x,y)/f_Y(y)$

To see why conditional probability is consistent with the new definition, consider $P(x \leq X \leq x+\delta | y \leq Y \leq y+\delta)$ for a small increment δ. We would like this probability to be approximately $f_{X|Y}(x)\delta$. But,

$$P(x \leq X \leq x+\delta | y \leq Y \leq y+\delta) = P(x \leq X \leq x+\delta, y \leq Y \leq y+\delta)/P(y \leq Y \leq y+\delta)$$

Therefore,

$$f_{X|Y}(x)\delta = \frac{f_{X,Y}(x,y)\delta^2}{f_Y(y)\delta}$$

Taking the limit as δ goes to 0 gives the desired definition for $f_{X|Y}(x)$.

With this definition for conditional probability, Bayes' rule for a continuous random variable becomes:

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

Note $f_Y(y)$ can be expressed as an integral in the same way $P(Y = y)$ can be expressed as a sum $\sum_x P(Y = y|X = x)P(X = x)$ in the discrete case. Therefore,

$$f_Y(y) = \int f_{Y|X}(y)f_X(x)dx$$

Rewriting Bayes' rule, we have:

$$f_{X|Y}(x) = \frac{f_{Y|X}(y)f_X(x)}{\int f_{Y|X}(y)f_X(x)dx}$$

We often use the simplified notation $f(y|x)$ to mean $f_{Y|X}(y)$. In this way, we can drop the subscript from f as it can be inferred from the its argument. Therefore,

$$f(x|y) = \frac{f(x,y)}{f(y)} = \frac{f(y|x)f(x)}{\int f(y|x)f(x)dx}$$

2 Example of breaking a stick twice

The (continuous) uniform distribution is given by the following density function.

```
         |
         |
1/(b-a) -|-    --------------------------
         |    |                         |
         |    |                         |
       --|----|-------------------------|-----
              a                         b
```

$$\int_a^b \frac{1}{b-a} dx = \frac{x}{b-a}|_a^b = \frac{b}{b-a} - \frac{a}{b-a} = 1$$

Hold a stick of length l and break it uniformly at random, then repeat for the remaining part. Let X represent the first breaking point, then $f(x) = 1/l$ (uniform from 0 to l).

```
   f(x)


     |
     |
1/l--|------------------------------
     |                             |
     |                             |
   --|-----------------------------|-----
     0                             l
```

Similarly, $f(y|x) = 1/x$ (also uniform from 0 to $x <= l$).

```
 f(y|x)


     |
     |
1/x--|-----------------
     |                |
     |                |
   --|----------------|-----
     0                x
```

What is $f(x, y)$? From the definition:

$$f(x, y) = f(x)f(y|x) = \frac{1}{lx}, \qquad 0 \le y \le x \le l$$

Note that

$$\int_0^l \int_0^x \frac{1}{lx} dydx = \int_0^l \frac{1}{l} dx = 1$$

We could have also evaluated the double integral in the other order:

$$\int_0^l \int_y^l \frac{1}{lx} dxdy = \int_0^1 \frac{1}{l} \log x|_y^l dy = \int_0^l \frac{1}{l} \log(l/y) dy = 1 \text{ (not showing the work)}$$

Note how the bounds of the inner integral change depending on which variable is integrated first. Note also that after integrating once, in the first case, we obtain $f(x)$, and in the second case we obtain (what should be) $f(y)$. This is always true: $f(x, y)$ is called the joint density, and $f(x)$ and $f(y)$ are called the marginal densities. Note that $\int f(x, y)dx$ is nothing but $\int f(y|x)f(x)dx = f(y)$. As we have seen before, this phenomenon is also true in the discrete case but is exhibited as a sum instead.

We conclude that

$$f(y) = \int_y^l \frac{1}{lx} dx = \frac{1}{l} \log(l/y), \qquad 0 \leq y \leq l$$

Computing expectations, $E[X] = l/2$ (easy). However, computing $E[Y]$ involves computing the integral $\int_0^l (y/l) \log(l/y)dy$, which is not very difficult, but let's use conditional expectation instead: $E[Y] = E_x[E_y[Y|X = x]] = E_x[x/2] = 1/2 \cdot E_x[x] = l/4$ (much easier to compute).

For another example of expectation, let us compute $E[XY]$.

$$E[XY] = \int_0^l \int_0^x xyf(x, y)dydx = \int_0^l \int_0^x \frac{y}{l} dydx = l^2/6$$

We can also reach the same result using nested expectation.

$$E[XY] = E[E[XY|X]] = E[XE[Y|X]] = E[X^2/2] = \int_0^l \frac{x^2}{2l} dx$$

The second equality follows from the fact that conditioned on X, $E[XY|X] = XE[Y|X]$ because X acts as a constant.

Finally, in the spirit of Bayes, we can now ask the following? What is $f(x|y)$? The importance of this lies in the following: given that we observe what is left of the stick, what can we say about how it was broken first?

$$f(x|y) = \frac{f(y|x)f(x)}{f(y)} = \frac{1/lx}{1/l \cdot \log(l/y)} = \frac{1}{x \log(l/y)}, \qquad y \leq x \leq l$$

Let us interpret this result. If $y = l$, then $f(x|y) = 1/(0 \cdot x)$ for $l \leq x \leq l$. This means the entire mass is located at $x = l$. This is obviously true since that's the only way we could have ended up with the entire stick. If $y = 0$, then $f(x|y) = 0/x$ for $0 \leq x \leq l$, which is 0 for every x except when $x = 0$. This means that the entire mass is located at $x = 0$, which is again the only way we could have ended up with no stick at all. Note that in both cases, the integral of this mass is still 1 at the limit.