Bayes rule for random variables

There are many situations where we want to know X, but can only measure a related random variable Y or observe a related event A. Bayes gives us a systematic way to *update* the pdf for X given this observation.

We will look at four different versions of Bayes rule for random variables. They all say essentially the same thing, but are tailored to situations where we are observing or inferring a mixture of continuous random variables and discrete random variables or events.

Bayes rule for continuous random variables

If X and Y are both continuous random variables with joint pdf $f_{X,Y}(x, y)$, we know that

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

Thus we can turn a conditional pdf in y, $f_{Y|X}(y|x)$ into one for X using

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

For a fixed observation of $Y = y$,

- $f_X(x)$ is a function of x (a pdf),
- $f_{Y|X}(y|x)$ is a particular function of x determined by y (although not a pdf),
- $f_Y(y)$ is a number.

Using the fact that

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|u)\, f_X(u)\, \mathrm{d}u,$$

we will often find it useful to rewrite the denominator above to get

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)\, f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|u)\, f_X(u)\, \mathrm{d}u}$$

Example. iPhones are known to have an exponentially distributed lifetime Y. However, the manufacturing plant has had some quality control problems lately. On any given day, the parameter λ of the pdf of Y is itself a random variable uniformly distributed on $[1/2, 1]$. We test an iPhone and record its lifetime. What can we say about the underlying parameter λ?

We have

$$f_\Lambda(\lambda) = 2, \quad \text{for } \frac{1}{2} \leq \lambda \leq 1,$$

and

$$f_{Y|\Lambda}(y|\lambda) = \lambda\, \mathrm{e}^{-\lambda y}, \quad y \geq 0.$$

Given a particular observation $Y = y$, we update the distribution for λ as

$$f_{\Lambda|Y}(\lambda|y) =$$

Bayes rule for discrete random variables

If X and Y are both discrete random variables, then we can simply replace the pdfs above with pmfs,

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)\, p_X(x)}{p_Y(y)}.$$

This really just follows from Bayes rule for events A and B (which we looked at in Section I of the notes),

$$\mathrm{P}\left(A|B\right) = \frac{\mathrm{P}\left(B|A\right)\mathrm{P}\left(A\right)}{\mathrm{P}\left(B\right)},$$

where A is the event $\{X = x\}$ and B is the event $\{Y = y\}$.

Again, using the law of total probability,

$$p_Y(y) = \sum_k p_{Y|X}(y|k)\, p_X(k)$$

we can rewrite the denominator above to get this version of Bayes rule:

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)\, p_X(x)}{\sum_k p_{Y|X}(y|k)\, p_X(k)}$$

Exercise:
Suppose that X is the number of 1s that appear in a binary string of length L; each bit in the string is equal to zero or one with probability $1/2$, and the bits are independent. Given $L = \ell$, we know that X has the binomial distribution when $k \leq \ell$ (the probability is zero otherwise):

$$p_{X|L}(k|\ell) = \binom{\ell}{k}(0.5)^k(0.5)^{\ell-k} = \binom{\ell}{k}2^{-\ell}.$$

Suppose that the length of the string is also random and uniformly distributed between 1 and 10:

$$p_L(\ell) = \begin{cases} \frac{1}{10}, & \ell = 1, \ldots, 10 \\ 0, & \text{otherwise.} \end{cases}$$

We learn that the binary string contains 4 ones. How can we use this information to update the pmf for the length of the string L?

Inference about a discrete event/random variable from a continuous observation

What does the observation of a random variable tell us about whether an event has occurred?

Recall that if A and B are events, then

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A)\ \mathrm{P}(A)}{\mathrm{P}(B)}, \quad \text{for } \mathrm{P}(B) > 0.$$

Now suppose that Y is a continuous random variable, and we observe $Y = y$. What can we say about $\mathrm{P}(A|Y = y)$?

We have to be a little careful here, since strictly speaking $\mathrm{P}(Y = y) = 0$ (since Y is continuous). But using densities for Y instead of probabilities, we can arrive at (see the next page)

$$\mathrm{P}(A|Y = y) = \frac{f_{Y|A}(y)\ \mathrm{P}(A)}{f_Y(y)}$$

where $f_{Y|A}(y)$ is a conditional density for Y evaluated at y, and Y is the pdf for Y evaluated at y.

Again, we can use a version of the law of total probability to expand the denominator above:

$$f_Y(y) = f_{Y|A}(y)\,\mathrm{P}(A) + f_{Y|A^c}(y)\,\mathrm{P}(A^c)$$

and so

$$\mathrm{P}(A|Y = y) = \frac{f_{Y|A}(y)\ \mathrm{P}(A)}{f_{Y|A}(y)\,\mathrm{P}(A) + f_{Y|A^c}(y)\,\mathrm{P}(A^c)}$$

We can derive this expression using Bayes rule for events, and then taking limits. For any $\delta > 0$, the event $B = \{y \leq Y \leq y + \delta\}$ is well-defined, and assuming that $f_Y(y) > 0$, $\mathrm{P}(B)$ will be positive. Then using Bayes rule for events

$$\begin{aligned}
\mathrm{P}(A|y \leq Y \leq y+\delta) &= \frac{\mathrm{P}(y \leq Y \leq y+\delta|A)\ \mathrm{P}(A)}{\mathrm{P}(y \leq Y \leq y+\delta)} \\
&= \frac{\left(\int_y^{y+\delta} f_{Y|A}(y')\ \mathrm{d}y'\right)\ \mathrm{P}(A)}{\int_y^{y+\delta} f_Y(y')\ \mathrm{d}y'}.
\end{aligned}$$

As $\delta \to 0$,

$$\begin{aligned}
\mathrm{P}(A|y \leq Y \leq y+\delta) &\to\ \mathrm{P}(A|Y=y), \\
\int_y^{y+\delta} f_{Y|A}(y')\ \mathrm{d}y' &\to\ f_{Y|A}(y)\delta, \\
\int_y^{y+\delta} f_Y(y')\ \mathrm{d}y' &\to\ f_Y(y)\delta,
\end{aligned}$$

and so taking the limit as $\delta \to 0$ on both sides above, the Bayes expression becomes

$$\mathrm{P}(A|Y=y) = \frac{f_{Y|A}(y)\ \mathrm{P}(A)}{f_Y(y)}.$$

The Bayes expression on the previous two pages easily extends to the case where X is a discrete random variable, as events of the form $\{X = x\}$ are well defined. Thus a mixed continuous-discrete version of Bayes rule is

$$\begin{aligned} p_{X|Y}(x|y) &= \mathrm{P}\left(X = x | Y = y\right) \\ &= \frac{f_{Y|X}(y|x)\, p_X(x)}{f_Y(y)} \\ &= \frac{f_{Y|X}(y|x)\, p_X(x)}{\sum_k f_{Y|X}(y|k)\, p_X(k)} \end{aligned}$$

Exercise:
S is a binary signal, with $\mathrm{P}\left(S = 1\right) = p$ and $\mathrm{P}\left(S = -1\right) = 1 - p$. Suppose that we transmit S and the received signal is $Y = S + N$, where N is normal noise with zero mean and unit variance, $N \sim \mathrm{Normal}(0, 1)$, and is independent of S. What is the probability that $S = 1$ as a function of the observed value y of Y?

Inference about a continuous random variable based on discrete observations

What does the observation of a discrete event tell us about a related continuous random variable?

We have observed that an event A has occurred, and want to use this information to update our probability model for a continuous random variable Y. In a manner similar to what we did in the previous section, we can derive the following version of Bayes rule that mixes continuous random variables and discrete events:

$$\begin{aligned} f_{Y|A}(y) &= \frac{\mathrm{P}\left(A|Y=y\right)\ f_Y(y)}{\mathrm{P}\left(A\right)} \\ &= \frac{\mathrm{P}\left(A|Y=y\right)\ f_Y(y)}{\int_{-\infty}^{\infty} \mathrm{P}\left(A|Y=u\right)\ f_Y(u)\ \mathrm{d}u} \end{aligned}$$

Exercise:
A bag of candy is filled with red and white jelly beans. The color of each jelly bean is independent, and the probability of pulling out a red jelly bean is p, while the probability of pulling out a white jelly bean is $1 - p$.

We have no idea what p is, so we model it as a uniformly distributed random variable P,

$$f_P(p) = \begin{cases} 1, & 0 \leq p \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

We pull out a jelly bean and see that it is red. How does this observation change the pdf for P?