1 Conditional Probability: Continuous Case

Conditional probability is slightly different in the "world" of continuous random variables. Most of the same definitions carry over from the discrete case, but since we're dealing with *PDFs*[1] $f(x)$ instead of PMFs $p(x) = \mathbb{P}(X = x)$ (which are probabilities, unlike PDFs).

Recall that if $\mathbb{P}(\omega)$ is a discrete probability distribution and $\mathbb{P}(B) > 0$, then the distribution $\mathbb{P}(\omega|B) = \frac{\mathbb{P}(\omega)}{\mathbb{P}(B)}$ is a probability distribution on B.

Definition 1.1

Let X be a continuous random variable with PDF f. Suppose that B is an event with $\mathbb{P}(B) > 0$. Then, the conditional PDF of X, given B, is

$$f(x|B) = \begin{cases} \frac{f(x)}{\mathbb{P}(B)} & x \in B \\ 0 & x \notin B \end{cases}.$$

We note that $f(x) \geq 0$ and

$$\int_{-\infty}^{\infty} f(x|B)dx = \frac{1}{\mathbb{P}(B)} \int_B f(x)dx = \frac{1}{\mathbb{P}(B)}\mathbb{P}(B) = 1.$$

Note that the integration over a region gives us the probability of being in that region, i.e. integrating $f(x)$ over the region B gives us just $\mathbb{P}(B)$. Therefore, $f(x|B)$ is indeed a probability density on B.

(Example.) Recall the spinner example. The location of the spinner, when it eventually comes to rest, is uniform on the unit circle; thus,

$$f(x) = 1 \text{ for } x \in [0, 1).$$

Suppose that the spinner comes to rest in the upper-half of the circle, i.e. in $[0, 1/2]$. What is the (conditional) probability with which it lands in the region $[1/6, 2/3]$?

The probability of B occurring (the event that the spinner lands in the upper-half of the circle) is $\frac{1}{2}$ since it is uniform from $[0, 1)$.

Normally, we would integrate from $\frac{1}{6}$ to $\frac{2}{3}$. However, since we know that the spinner lands in the region $[0, 1/2]$, we know that there is no density *outside* of this region. Since x is outside of the region for $\frac{1}{2} \leq x \leq \frac{2}{3}$, we don't need to consider it. Instead, we only need to consider the region $[1/6, 1/2]$. Hence,

$$\int_{\frac{1}{6}}^{\frac{1}{2}} \frac{1}{1/2}dx = \frac{2/6}{1/2} = \frac{2}{3}.$$

Recall that the CDF of a continuous random variable is the function

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(u)du.$$

[1] Remember, PDFs are not probabilities; they are densities. They allow us to get a probability by integrating, but they're not probabilities.

Definition 1.2

A sequence of random variables $X_1, \ldots, X_n$ is said to be **mutually independent** if their joint CDF is the product of the individual CDF. That is, if

$$F(x_1, \ldots, x_n) = \mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n) = \prod_{i=1}^{n} F_i(x_i)$$

for *all* $x_1, \ldots, x_n$.

By calculus, we can find the joint PDF of a sequence of continuous random variables by differentating:

$$f(x_1, \ldots, x_n) = \frac{\partial^n}{\partial x_1, \ldots, \partial x_n} F(x_1, \ldots, x_n).$$

From this, we have the following theorem:

Theorem 1.1

$X_1, \ldots, X_n$ are mutually independent if and only if their joint PDF is the product of the individual PDFs, $f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i)$.

Additionally, consider the following theorem:

Theorem 1.2

If $X_1, \ldots, X_n$ are mutually independent, and f is a continuous function, then $f(X_1), \ldots, f(X_n)$ are mutually independent.

(Example.) Suppose that X_1 and X_2 are independent exponential random variables with rates λ_1 and λ_2. Find the PDF of $M = \min\{X_1, X_2\}$.

Recall that if X is an exponential rate λ random variable if and only if its survival function is

$$S(x) = 1 - F(x) = P(X > x) = e^{-\lambda x}.$$

Since X_1, X_2 are independent, their joint PDF is the product (from differentating)

$$f(x_1, x_2) = f_1(x_1) f_2(x_2) = \lambda_1 e^{-\lambda_1 x_1} \cdot \lambda_2 e^{-\lambda_2 x_2}.$$

Also, note that $M = \min\{X_1, X_2\} > m$ if and only if $X_1 > m$ and $X_2 > m$. Hence,

$$\begin{aligned}
\mathbb{P}(M > m) &= \int_m^\infty \int_m^\infty f(x_1, x_2) dx_1 dx_2 \\
&= \left(\int_m^\infty \lambda_1 e^{-\lambda_1 x_1} dx_1 \right) \left(\int_m^\infty \lambda_2 e^{-\lambda_2 x_2} dx_2 \right) \\
&= \mathbb{P}(X_1 > m) \mathbb{P}(X_2 > m) \\
&= e^{-\lambda_1 m} e^{-\lambda_2 m} \\
&= e^{-(\lambda_1 + \lambda_2) m}.
\end{aligned}$$

Therefore, $M = \min\{X_1, X_2\}$ is an exponential random variable with rate $\lambda_1 + \lambda_2$ (the sum of the rates of X_1 and X_2).

1.1 Memoryless Property of Exponential Variables

Theorem 1.3

Suppose X is exponential with rate λ. Then, X has no memory, in the sense that for any $t, s > 0$,

$$\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t).$$

That is, given the *survival* of X to time s, it then behaves like a "brand new" exponential rate λ random variable.

Proof. Note that $\mathbb{P}(X > x) = e^{-\lambda x}$ for any $x > 0$. Hence,

$$\mathbb{P}(X > s + t | X > s) = \int_{s+t}^{\infty} \frac{\lambda e^{-\lambda x}}{e^{-\lambda s}} dx = e^{\lambda s}\mathbb{P}(X > s + t) = e^{\lambda s} e^{-\lambda(s+t)} = e^{-\lambda t}.$$

Hence, this is equal to $\mathbb{P}(X > t)$ as claimed. □

(Example: Beta Distribution.) The Beta(α, β) random variable, with parameters $\alpha, \beta > 0$, has PDF

$$f(x) = B(\alpha, \beta, x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

for $x \in [0, 1]$ (and $f(x) = 0$ otherwise), where

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$$

is the beta integral. Now, when α, β are integers, then

$$B(\alpha, \beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}.$$

Note that when $\alpha = \beta = 1$, then $B(\alpha, \beta, x) = 1$. Thus, Beta$(1, 1)$ is uniform on $[0, 1]$.

The beta distribution is used in **Bayesian statistics**.

(Example: Drug Testing.) Suppose that a drug has an unknown probability X of being effective. Therefore, the first time it is administered, it would be quite natural to assume that the distribution of X is uniform on $[0, 1]$. However, as time goes on, and more data is collected, we might want to update this distribution. For example, if it is successful in all except 6 of the first 100 trials, a uniform distribution would no longer seem appropriate.

Suppose that we given this drug to n patients. Assuming independence, we can model this as a series of n Bernoulli trials with (unknown) success probability X. If $X = x$, then the probability that the i trials will be successful is the binomial probability

$$b(n, x, i) = \binom{n}{i} x^i (1-x)^{n-i}.$$

Hence, the conditional PMF is $p(i|x) = b(n, x, i)$.

If X is uniform on $[0, 1]$, which has PDF of 1 on $[0, 1]$, then the i of the n trials will be successful with probability

$$\int_0^1 1 \cdot p(i|x)dx = \int_0^1 b(n, x, i)dx = \binom{n}{i} \int_0^1 x^i (1-x)^{n-i} dx.$$

We note that

$$\int_0^1 x^i (1-x)^{n-i} dx = B(i+1, n-i+1).$$

Hence, the PMF is given by

$$p(i) = \binom{n}{i} B(i+1, n-i+1).$$

Now, still assuming that X is uniform, note that the joint distribution is given by $f(x, i) = p(i|x) \cdot 1 = b(n, x, i)$. So, the conditional PDF is

$$f(x|i) = \frac{f(x, i)}{p(i)} = \frac{b(n, x, i)}{\binom{n}{i} B(i+1, n-i+1)} = \frac{x^i (1-x)^{n-i}}{B(i+1, n-i+1)},$$

which is the PDF of a $\mathrm{Beta}(i+1, n-i+1)$ random variable.

In fact, we could continue to update this distribution as we go along and continue to collect more data. So, for the first patient, we have seen 0 successes and 0 failures (no data yet). So, we assume that the probability of success on the first patient is distributed as a $\mathrm{Beta}(1, 1) = \mathrm{Uniform}[0, 1]$ random variable.