Integral Calculus: Probability Edition

1 Introduction

A **random variable** is a variable whose value is not deterministic. A frequentist interpretation of random variables is that as we repeat the same experiment many times, we will observe different values of the random variables. As we repeat more and more experiments, the number of times we see a certain value of the random variable is proportional to that value's probability (in discrete settings). For example, the up-facing side of a fair die when thrown is a random variable which can take integer values from 1 to 6.

A random variable is said to be "distributed according to" its probability distribution. A **probability distribution** is a function from a domain (e.g. a set of real numbers) to a range (the probability). Probabilities have to be non-negative.

Therefore, it is important to distinguish two quantities. The value of the variable is different from its probability distribution. This is demonstrated by the following plot. The value of the random variable is the X-axis, while its probability distribution is on the Y-axis. The value of the variable can be negative, whereas its probability cannot.

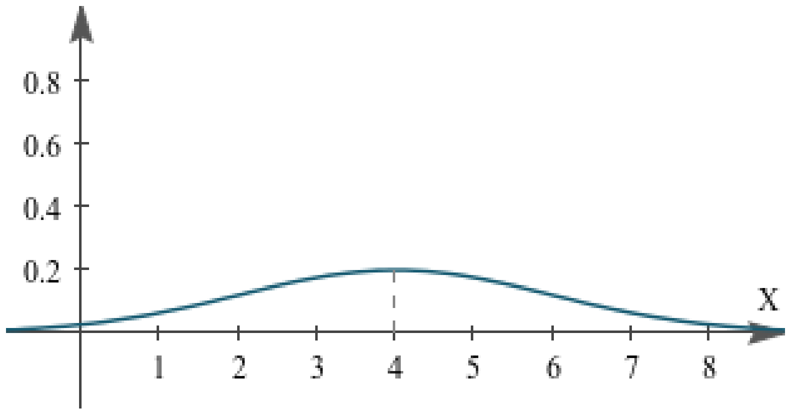

Figure 1. A continuous random variable and its probability distribution

Suppose a random variable can take values in the real numbers. Then generally it is not meaningful to speak of the probability of a single value. For example, the probability that a person is exactly 170.0 cm tall (with no measurement error) is 0. We usually speak of the probability that a continuous random variable X will take value within a certain range $[a, b]$. In this case, we sum up all the infinitesimal probabilities that X takes any value from a to b. This sum over all continuous values requires an integral.

$$p(a \leq X \leq b) = \int_a^b p(X)dX$$

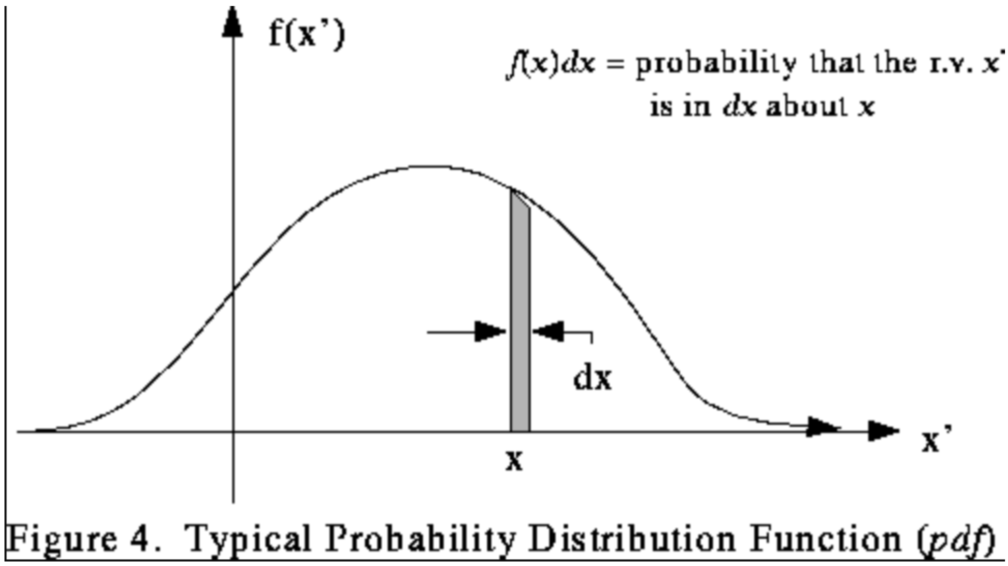

Figure 2. The probability of a continuous random variable is only meaningful within ranges.

2 Normalization Condition

In the discrete case, probability mass functions have to sum up to 1. This reflects the assumption that at the very least, something must happen. Similarly, in the continuous case, the probability that a univariate random variable takes a value between $-\infty$ and $+\infty$ is 1. This requirement is known as the **normalization condition**, expressed as an integral.

$$p(-\infty \leq X \leq +\infty) = \int_{-\infty}^{+\infty} p(X)dX = 1$$

Therefore, it is a requirement that if a function $p(X)$ describes a probability distribution, its integral must sum up to 1. This is a useful fact when working with probability distributions, since it allows you to simplify many integrals without actually doing the hard work of integrating.

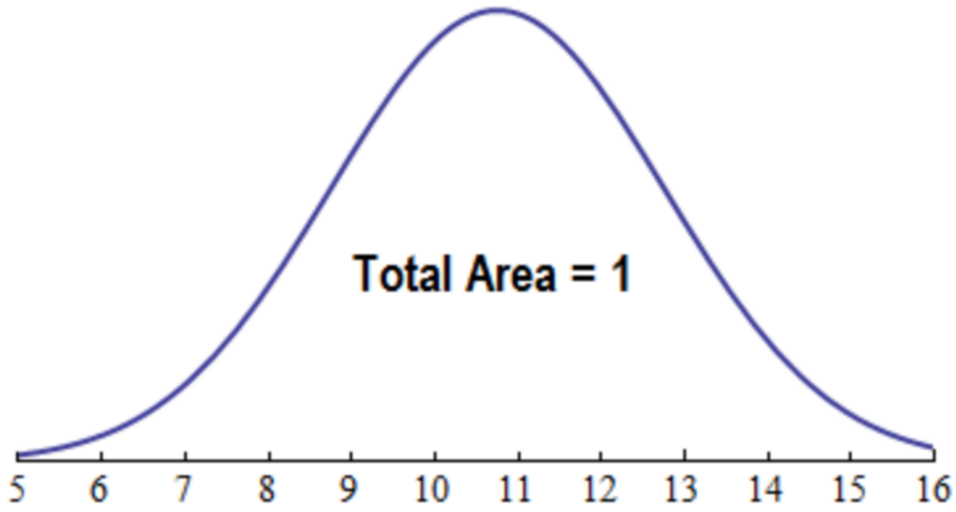

Figure 3. The normalization condition requires that every probability distribution sum up to 1.

3 Probability Distributions of Many Variables

So far, we have only considered probability distributions of one random variable. Two random variables X, Y can have a **joint probability distribution**, denoted $p(X, Y)$. The joint probability distribution $p(X = x, Y = y)$ describes the probability density that X takes value x **and** Y takes value y. Just like in the univariate case, the joint probability distribution satisfies the normalization condition.

$$p(-\infty \leq X \leq +\infty, -\infty \leq Y \leq +\infty) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(X, Y) dX dY = 1$$

Given a joint distribution $p(X, Y)$, the **marginal distribution** $p(X)$ involves fixing a particular value of $X = x$ but taking into account all possible values of Y. This procedure is called "integrating out" or "marginalizing out" Y since we are taking a sum over it, removing that variable from the expression. This is akin to saying that to find the probability that $X = x$, we find the joint probabilities that $X = x$ and Y taking all possible values, and sum all of those probabilities.

$$p(X = x) = \int_{-\infty}^{+\infty} p(X = x, Y)dY$$

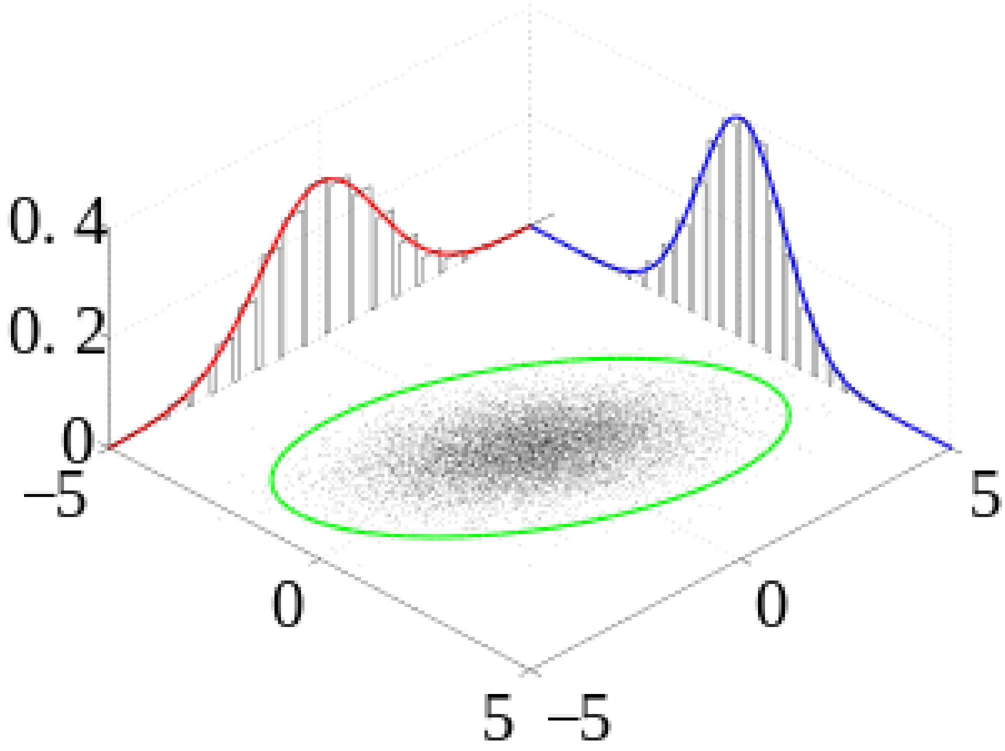

Figure 4. Joint and marginal probability distributions

The variable of integration is very important, since the integration implies that we sum the integrand over many values of the variable of integration. All multiplicative terms that do not depend on the variable of integration can be pulled out of the integral. This is similar to the distributive property, since an integral can be interpreted as a sum. If all terms of a sum share a common multiplicative term, the term can be pulled outside of the sum due to the distributive property. To demonstrate this in an equation, we have

$$\int f(x)g(y)dY = f(x)\int g(y)dY$$

since $f(x)$ does not depend on y.

4 An Example

To demonstrate these concepts in action, suppose we want to determine whether the product of two Gaussians is a valid probability distribution. A Gaussian $N(x|\mu, \sigma)$ is a probability distribution defined to be

$$N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

To determine whether the product $N(x|\mu_1, \sigma_1)N(x|\mu_2, \sigma_2)$ is a valid probability distribution, the normalization condition tells us that the integral of this product over all values of x must be 1.

$$\int_{-\infty}^{+\infty} N(x|\mu_1, \sigma_1^2)N(x|\mu_2, \sigma_2^2)dx = 1$$

You might know that products of Gaussians obey the following identity.

$$N(x \mid \mu_1, \sigma_1^2)N(x \mid \mu_2, \sigma_2^2) = N\left(\mu_1 \mid \mu_2, \sigma_1^2 + \sigma_2^2\right) N\left(x \,\middle|\, \frac{\mu_2\sigma_1^2 + \mu_1\sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}\right)$$

For a proof of the identity, read this link.

Substituting this back into the integral, we have

$$\begin{aligned}
&\int_{-\infty}^{+\infty} N(x|\mu_1, \sigma_1^2) N(x|\mu_2, \sigma_2^2) dx \\
&= \int_{-\infty}^{+\infty} N\left(\mu_1 \,|\, \mu_2, \sigma_1^2 + \sigma_2^2\right) N\left(x \,\middle|\, \frac{\mu_2\sigma_1^2 + \mu_1\sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}\right) dx \\
&= N\left(\mu_1 \,|\, \mu_2, \sigma_1^2 + \sigma_2^2\right) \int_{-\infty}^{+\infty} N\left(x \,\middle|\, \frac{\mu_2\sigma_1^2 + \mu_1\sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}\right) dx \qquad (1) \\
&= N\left(\mu_1 \,|\, \mu_2, \sigma_1^2 + \sigma_2^2\right) \qquad (2)
\end{aligned}$$

In equation (1) we pulled the first term outside of the integral because it does not depend on the variable of integration x. Note that we don't actually need to perform the Gaussian integral, which might be difficult! In equation (2) we note that $N\left(x \,\middle|\, \frac{\mu_2\sigma_1^2+\mu_1\sigma_2^2}{\sigma_1^2+\sigma_2^2}, \frac{1}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}\right)$ is a valid probability distribution (namely, a Gaussian) and therefore the integral must evaluate to 1. Since $N\left(\mu_1 \,|\, \mu_2, \sigma_1^2 + \sigma_2^2\right)$ is not always 1, a product of 2 Gaussians is generally **not** a valid probability distribution, since it violates the normalization condition, a necessary property for a function to be a probability distribution.

5 Summary

By the end of this tutorial, you should

1. Understand the difference between a random variable and its probability distribution
2. Understand the relationship between integration and the probability of a random variable within a range

3. State the normalization condition and apply it to simplify certain integrals

4. Apply the normalization condition in settings with many variables

5. Understand the idea of "integrating out" variables in a joint distribution to find the marginal distribution

6 Exercises

Solve the following exercises to check your understanding.

1. Find the values of k for which the given functions are probability density functions.

    (a) $f(x) = 2k$ on [1, 1]

    (b) $f(x) = k$ on [2, 0]

    (c) $f(x) = ke^{kx}$ on [0, 1]

    (d) $f(x) = kxe^{x2}$ on [0, 1]

2. Your friend thinks that if f is a probability density function for the continuous random variable X, then $f(a)$ is the probability that $X = a$. Explain to your friend why this is wrong.

3. Not satisfied with your explanation in the previous exercise, your friend then challenges you by asking, "If $f(a)$ is not the probability that $X = a$, then just what does $f(a)$ mean?" How would you respond?

4. Calculate the following integral $\int_0^{+\infty} x^4 e^{-5x} dx$. (Hint: What distribution

does this remind you of?)