Integrals are not Scarey

Paul E. Johnson

July 15, 2011

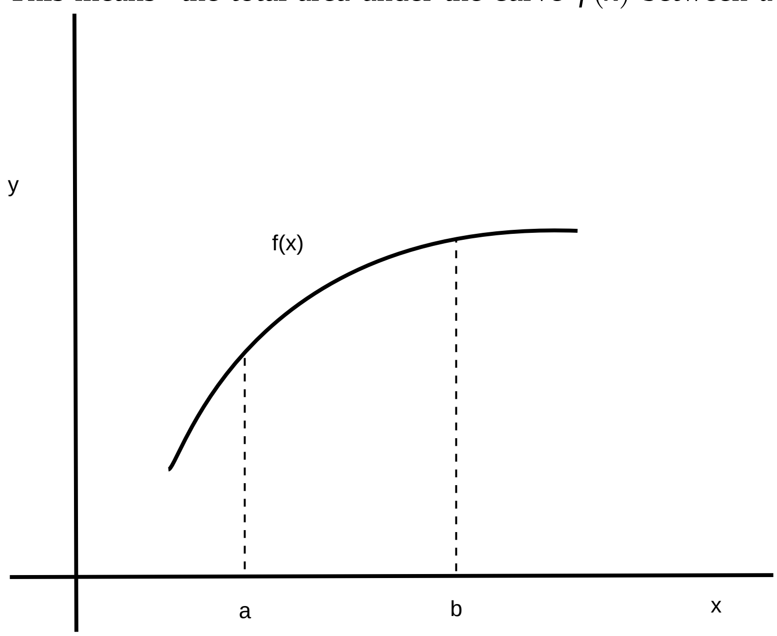
1 The Integral

Even if you never took a calculus class, you may still have to read a book that uses integrals. The concept is not complicated, and as long as you don't actually have to solve some integrals, I expect a conceptual understanding is good enough for most people.

The elongated S symbol represents integration.

$$\int_a^b f(x)dx$$

This means "the total area under the curve $f(x)$ between a and b." Consider this.



The symbol dx represents the "dummy variable of integration." It is a signal that you are supposed to move along the x axis when you sum up from a to b.

Sometimes people will describe the integral as a sum of really small slices out of $f(x)$.

2 Integrals & Probability

If you are studying a continuous random variable, x, it means you are studying a variable that can take on real values in some domain, X. Suppose the "endpoints" of X are $left$ and $right$, where $left$ and $right$ can take on any real value, as well as infinity.

Real number digression: The symbol for the "real number line" is $\mathbb{R}$. The set of all real numbers can be formally defined, but most of the time I just think of it as "all numbers that can be written down as numbers with decimals, including infinite digits after the decimal point."

If $f(x)$ is a **probability density function** (pdf) representing a probability distribution, it means these 3 conditions are true:

1. $f(x) \geq 0$ for all $x \in X$
2. $f(x) \leq 1$ for all $x \in X$
3. $\int_{left}^{right} f(x)dx = 1$.

There is always some "tricky business" about the probability of a particular, individual point. The probability that a particular point will occur is 0 because a point is a thing with no "width." The probability that you could observe a particular outcome c is

$$\int_{c}^{c} f(x)dx$$

And that is always 0 by definition. So we are restricted to talking about outcomes in a particular range, say between c and d.

$$\int_{c}^{d} f(x)dx$$

The **cumulative distribution function** (cdf) is the probability that the outcome of a random draw will be smaller than some given value, say k. It is often symbolized by a capital letter corresponding to the pdf, in this case $F(k)$. Formally, it is the integral from $left$ up to the point k.

$$F(k) = \int_{left}^{k} f(x)dx$$

3 Integrals for Multivariate Probability

If the domain of outcomes is 2 or more dimensional, then the integral extends to represent it. For example, with 2 dimensions, $X = \{left_x, right_x\}$ and $Y = \{left_y, right_y\}$. The pdf is $f(x,y)$ and the probability that an observation will occur in a region in which $x \in \{c_x, d_x\}$ and $y \in \{c_y, d_y\}$ is represented as

$$\int_{c_x}^{d_x} \int_{c_y}^{d_y} f(x,y)dxdy$$

There's a theorem that say the order of integration does not affect the value, so you can swap the x and y things in there.

Generally, it is difficult to solve multivariate probability distributions, so we go searching about for simplifying results, such as independence,

$$f(x,y) = g(x) \cdot h(y)$$

This means that the joint observation of the pair (x,y) is just as likely as the separate observations of x and y .

[stopped here, will work on more next season]