Relating two definitions of expectation

John D. Cook

June 26, 2008

The expected value of a random variable has two different definitions, and it's not obvious that they're equivalent. Here I present both definitions and show why they are in fact equivalent.

An elementary probability or statistics book might say that a random variable X is continuous if there exists a function $f_X : \mathbb{R} \to \mathbb{R}$, the probability density function (PDF) of X, such that

$$P(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

In that case they would define the expectation of X to be

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx.$$

An advanced (*i.e.*, measure-theoretic) probability book would define $E(X)$ differently. Let $X : \Omega \to \mathbb{R}$ be a random variable. (This implies (Ω, P) is a measure space, $P(\Omega) = 1$, and X is an itegrable function.) The expectation of X is simply its integral:

$$E(X) = \int_{\Omega} X\, dP.$$

Note that while both definitions involve integrals, they integrate different functions over different spaces. In the elementary definition, a real-valued function is being integrated over the real line. In the advanced definition, the random variable itself is being integrated over a probability space.

In the elementary theory, you never see Ω, the domain of X. And you never work with X itself, only with its associated PDF. The advanced theory focuses on X and its domain. Perhaps the biggest source of confusion in theoretical probability is failure to distinguish X and f_X.

Now we start to show how the advanced definition relates to the elementary one.

The random variable X induces a probability measure μ_X on $\mathbb{R}$ by

$$\mu_X(A) = P\{\omega \in \Omega : X(\omega) \in A\}.$$

If μ_X is absolutely continuous with respect to Lebesgue measure then by the Radon-Nikodym theorem there exists a function $f_X : \mathbb{R} \to \mathbb{R}$ such that $\mu_X(A) = \int_A f_X \, dx$. Sometimes this result is written as

$$d\mu_X = f_X \, dx.$$

The function f_X is a PDF for X.

The following theorem is what is sometimes called the "rule of the unconscious statistician" or the RUS. The justification for the name is that the theorem is applied so frequently that it is done almost unconsciously.

Theorem 1 (RUS) *If $g : \mathbb{R} \to \mathbb{R}$ is μ_X-measurable, then $g(X)$ is P-measurable and*

$$\int_\Omega g(X) \, dP = \int_{\mathbb{R}} g(x) \, f_X(x) \, dx.$$

If X has a distribution function f_X,

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) \, dx.$$

The special case $g(x) = x$ shows that

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) \, dx.$$

We outline a proof the theorem above. Let 1_A be the indicator function of a Borel set $A \subseteq \mathbb{R}$. Then

$$\int_{\mathbb{R}} 1_A(x) f_X(x) \, dx = \mu_X(A) = P\{X^{-1}(A)\} = \int_\Omega 1_A(X(\omega)) \, dP.$$

By linearity, this proves the theorem for all functions g taking only a finite number of values. Since every measurable function is a monotone limit of such functions, the theorem follows from the monotone convergence theorem.