# The law of the unconscious statistician

Bengt Ringnér[*][†]

January 16, 2009

## 1 A property of the Lebesgue integral

As opposed to the Riemann integral, Lebesgue integration consists in "chopping up" the ordinate axis rather than the abscissa. Specifically, the integral with respect to a measure $\mu$ is given by $\int g\, d\mu = \int g^+\, d\mu - \int g^-\, d\mu$, where $g^+ = \max(0, g)$,

$$\int g^+\, d\mu = \lim_{n\to\infty} \sum_{k=1}^{n^2-1} \frac{k}{n}\mu\{x : \frac{k}{n} < g^+(x) \leq \frac{k+1}{n}\} + n\mu\{x : g^+ > n\},$$

and similarly for $g^-$. This follows either by definition or through the limit theorems. The integral exists if and only if both limits on the right hand sides exist.

An implication important in probability theory is that the integral depends only on the functions

$$y \mapsto \mu\{x : g(x) > y\}, \quad y > 0$$

and

$$y \mapsto \mu\{x : g(x) < y\}, \quad y < 0.$$

If $\mu$ is finite, we can replace these functions by the cumulative distribution function of $g$

$$F_g(y) = \mu\{x : g(x) \leq y\}, \quad -\infty < y < \infty.$$

[*]Centre for Mathematical Sciences, Lund University, Lund, Sweden.
[†]Homepage: http://www.maths.lth.se/∼bengtr

If the integral happens to be over the real line, the Lebesgue-Stieltjes integral is defined as

$$\int g(x)\, dF(x) = \int g(x)\, \mu(x)$$

where $\mu$ is the measure defined by $\mu(]a,b]) = F(b) - F(a)$ if $F$ is right-contonuous. In probability, one regards $x$ as a function $x = X(\omega)$ of some underlying conceptual random element $\omega$ and writes $F_X$ instead of $F$.

## 2 Random variables

A probability space consists of a measure $\mathbb{P}$ defined on a $\sigma$-algebra of subsets of a universal set commonly called $\Omega$. One has $\mathbb{P}(\Omega) = 1$. A random varable, $X$, is formally defined as a real-valued function which is defined on a probability space, and is measurable in the sense that its cumulative distribution function,

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}\{\omega \in \Omega : X(\omega) \le x\},$$

is defined for all $x$.

In most cases one does not know the functional relation $\omega \mapsto X(\omega)$, only $F_X$. Often one is interested in some new random variable $Y$ given by $Y = g(X)$, that is $Y(\omega) = g(X(\omega)$, where $g$ is a given function. For instance, the distribution of $X$, the speed of a particle, might be given, and one wants to determine the distribution of its kinetic energy, $Y = mX^2/2$. Now

$$\begin{aligned} F_Y(y) = \mathbb{P}(Y \le y) &= \mathbb{P}\{\omega \in \Omega : Y(\omega) \le y\} \\ &= \mathbb{P}\{\omega \in \Omega : g(X(\omega)) \le y\} \\ &= \mathbb{P}\{\omega \in \Omega : X(\omega) \in A\} \\ &= \mathbb{P}(X \in A) \end{aligned}$$

where

$$A = \{x : g(x) \le y\}.$$

In the example with kinetic energy, we have

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(mX^2/2 \le y) = \mathbb{P}(X \le \sqrt{2y/m}) = F_X(\sqrt{2y/m}).$$

It is required that $g$ is such that $\mathbb{P}(A)$ is defined for all sets $A = \{x : g(x) \le y\}$. In other words, $g$ is a random variable defined on the real line equipped with a probability measure

$$\boldsymbol{P}_X(A) = \int_A dF_X$$

defined on the $\sigma$-algebra generated by intervals. It has the same distribution function as $Y$, since

$$F_g(y) = \boldsymbol{P}_X\{x : g(x) \leq y\} = F_Y(y).$$

## 3 Expectation

The expectation of a random variable $X$ is given by

$$\mathbb{E}(X) = \int X(\omega)\, d\mathbb{P}(\omega)$$

and is said to exist if and only if the integral exists. To give a formula for the expectation which does not involve the function $\omega \mapsto X(\omega)$, the following obvious consequence of the first section is used.

- If two random variables, which may be defined on different spaces, have the same distribution functions, then their expectations are equal.

**Theorem 1** *Let $g$ be a real-valued Borel function on the real line, that is all sets $\{x : g(x) \leq y\}$ belong to the $\sigma$-algebra generated by intervals. Define $Y = g(X)$. Then*

$$\mathbb{E}(Y) = \int g(x)\, dF_X(x).$$

PROOF As said in the previous section, $F_g = F_Y$, so $g$ and $Y$ have the same expectation. Therefore

$$\mathbb{E}(Y) = \int g(x)\, d\boldsymbol{P}_X(x),$$

since the right hand side is, by definition, the expectation of $g$. As noted towards the end of the first section,

$$\int g(x)\, d\boldsymbol{P}_X(x) = \int g(x)\, dF_X(x).$$

$\square$

The special case $g(x) = x$ gives

$$\mathbb{E}(X) = \int x\, dF_X(x),$$

which simplifies to

$$\mathbb{E}(X) = \begin{cases} \int x f_X(x)\, dx & \text{for a continuous} \quad X \\ \sum_x x p_X(x) & \text{for a discrete} \quad X, \end{cases}$$

which is the definition in elementary textbooks.