Jensen's Inequality

We fix a measure space $(\Omega, \mathcal{M}, \mu)$, and we define

$$L^+ = \text{the space of all measurable functions from } \Omega \text{ to } [0, \infty].$$

We need the following fact:

If $f \in L^+$, let $\lambda(E) = \int_E f\, d\mu$ for $E \in \mathcal{M}$. Then λ is a measure on $\mathcal{M}$. Furthermore, for any $g \in L^+$, $\int g\, d\lambda = \int fg d\mu$. (This is an exercise in Folland's *Real Analysis*, Second Edition, exercise 14 on page 52.)

This fact allows us to relate the "undergraduate" and "graduate"-level definitions of expected value. The undergraduate level definition of the expected value of a continuous random variable X with density function f is

$$E(X) = \int_0^\infty x f(x)\, dx.$$

The graduate-level definition (see Durrett, *Probability: Theory and Examples*, 5th edition, p.25) is this. Suppose that P is a probability measure. If $X \geq 0$ is a random variable on $(\Omega, \mathcal{M}, P)$, then we define its expected value to be

$$E(X) = \int X\, dP.$$

Suppose that there exists a measurable function f such that for all x, $P(X \leq x)$ has the form

$$P(X \leq x) = \int_{-\infty}^x f(y)\, dy.$$

Then we say that X has density function f (Durrett, p.10). In this case $P = \lambda$. We have $X = g$, and so by the fact,

$$E(X) = \int g\, dP = \int g \cdot f\, dx = \int_0^\infty x f(x)\, dx.$$

Now we discuss Jensen's inequality. Here is how it is stated in an undergraduate text (A Modern Introduction to Probability and Statistics). A twice differentiable function g is *convex* on an interval I if $g''(x) \geq 0$ for all $x \in I$. Let φ be a convex function, and let X be a random variable. Then

$$\varphi(E[X]) \leq E[\varphi(X)].$$

This is really a measure-theory fact, as we explain. Jensen's inequality says: Suppose φ is convex. If μ is a probability measure, and f and $\varphi(f)$ are integrable, then

$$\varphi\left(\int f\,d\mu\right) \le \int \varphi \circ f\,d\mu.$$

Now take f to be X. Then

$$\varphi\left(\int X\,dP\right) \le \int \varphi \circ X\,dP$$

which means that

$$\varphi(EX) \le E(\varphi \circ X).$$

There is a neat application of this. Consider the convex function $\varphi(x) = x^2$. By Jensen's inequality,

$$E[X]^2 \le E[X^2].$$

The variance of X is equal to $E[X^2] - E[X]^2$, and now we see that this is non-negative, as expected. (The other way is to define the variance of X to be $E(X - EX)^2$, which is clearly non-negative, and check that this is equal to the other expression.)

July 1, 2025