STA 623 – Fall 2013 – Dr. Charnigo

Section 2.2: Expected values

Expected value of a discrete random variable. Let X be a discrete random variable with probability mass function $f_X(x)$ and support $\mathcal{X}$. Let $g(x)$ be a function such that $g^{-1}(B) \in \mathcal{B}^1$ for any $B \in \mathcal{B}^1$. We define the expected value of $g(X)$, also referred to as the mean of $g(X)$, as

$$E[g(X)] := \sum_{x \in \mathcal{X}} g(x) f_X(x) = \sum_{x \in \mathcal{X}} g(x) P(X = x),$$

provided that the sum is absolutely convergent. If the sum is not absolutely convergent, then we say that $E[g(X)]$ does not exist as a finite number. If the sum is not absolutely convergent and $g(x) \geq 0$ for all but finitely many $x \in \mathcal{X}$, then we may also say that $E[g(X)] = \infty$. Note that the expected value of $g(X)$, when it does exist as a finite number, is just a weighted average of all values of $g(X)$ that occur with nonzero probabilities, the weights being the probabilities themselves.

Example (expected value of a discrete random variable). Let X be the number of flips required to get your first tails on a fair coin. Then X has probability mass function $f_X(x) := (1/2)^x$ for $x \in \{1, 2, \ldots\}$. Putting $g(x) := x$, we have

$$E[X] = \sum_{x=1}^{\infty} x(1/2)^x = \sum_{x=1}^{\infty} (1/2)^x + \sum_{x=2}^{\infty} (1/2)^x + \sum_{x=3}^{\infty} (1/2)^x + \cdots$$

$$=$$

This computation suggests an answer if you were asked, before you flipped the coin, what you expected X to be. However, despite what we saw in this example, we are not generally guaranteed that $E[X] \in \mathcal{X}$, even when $E[X]$ exists as a finite number. (Can you construct another example in which $E[X]$ exists as a finite number but does not belong to $\mathcal{X}$?)

Again, let X be the number of flips required to get your first tails on a fair coin. Putting $g(x) := 2^x - C$ for some positive constant C, we have

$$E[g(X)] = \sum_{x=1}^{\infty}\{(2^x - C)(1/2)^x\} = \sum_{x=1}^{\infty}\{1 - C(1/2)^x\} = \infty.$$

To see the last equality above, let x^* be the smallest positive integer greater than $-\log C/\log(1/2)+1$. Then, for every positive integer $x \geq x^*$, we have $C(1/2)^x < (1/2)$. Hence,

$$\sum_{x=x^*}^{\infty}\{1 - C(1/2)^x\} = \lim_{n\to\infty}\sum_{x=x^*}^{n}\{1 - C(1/2)^x\} \geq \lim_{n\to\infty}\sum_{x=x^*}^{n}(1/2)$$
$$= \lim_{n\to\infty}(n - x^* + 1)(1/2) = \infty.$$

Since $\sum_{x=1}^{x^*-1}\{1 - C(1/2)^x\}$ is finite, we also have $\sum_{x=1}^{\infty}\{1 - C(1/2)^x\} = \infty$.

Here is an interpretation. I sell you a fair coin for C dollars and promise to pay you 2^X dollars if you get your first tails on flip X. Your net winnings are $g(X)$. If $C < 2$, then your net winnings are positive with probability 1 and you will surely choose to play (assuming you are a rational human being). However, the calculations above show that your expected net winnings are positive infinity regardless of the price C.

So, I have a quarter in my briefcase. Does anyone want to buy it for $C = 1,000,000,000$ dollars? This is called the St. Petersburg Paradox and nicely demonstrates that, in real life situations that can be modeled probabilistically, our behaviors are not necessarily governed by expected values. (If they were, then nobody would play the lottery.)

Let us do one last example. Say that X has probability mass function $f_X(x) := \exp[-\lambda]\lambda^x/x!$ for $x \in \{0, 1, 2, \ldots\}$, where λ is a positive real. Putting $g(x) := x$, we have

$$E[X] = \sum_{x=0}^{\infty} x\exp[-\lambda]\lambda^x/x! = \sum_{x=1}^{\infty} x\exp[-\lambda]\lambda^x/x! = \lambda\sum_{x=1}^{\infty}\exp[-\lambda]\lambda^{x-1}/(x-1)!$$
$$=$$

Expected value of a continuous random variable. Let X be a continuous random variable with probability density function $f_X(x)$ and support $\mathcal{X}$. Let $g(x)$ be a function such that $g^{-1}(B) \in \mathcal{B}^1$ for any $B \in \mathcal{B}^1$. We define the expected value of $g(X)$, also referred to as the mean of $g(X)$, as

$$E[g(X)] := \int_{\mathcal{X}} g(x) f_X(x)\, dx,$$

provided that the integral is absolutely convergent. If the integral is not absolutely convergent, then we say that $E[g(X)]$ does not exist as a finite number. If the integral is not absolutely convergent and $g(x) \geq 0$ for all $x \in \mathcal{X}$, then we may also say that $E[g(X)] = \infty$.

Example (expected value of a continuous random variable). Let α be a positive real. The gamma function is $\Gamma[\alpha] := \int_0^\infty x^{\alpha-1} \exp[-x]\, dx$. Since $n! = \Gamma[n+1]$ for any positive integer n and $\Gamma[\alpha+1] = \alpha\Gamma[\alpha]$ for any positive real α, the gamma function may be viewed as an extension of the factorial function from the positive integers to the positive reals.

Let X have probability density function

$$f_X(x) := \frac{\lambda^\alpha}{\Gamma[\alpha]} x^{\alpha-1} \exp[-\lambda x] 1_{\{x>0\}}$$

for some positive reals α and λ. Putting $g(x) := x^\beta$ for $x \in \mathcal{X}$ (and defining $g(x)$ to be, say, 0 for $x \in \mathcal{X}^c$), where β is a positive real, we have

$$E[X^\beta] = \int_0^\infty \frac{\lambda^\alpha}{\Gamma[\alpha]} x^{\alpha+\beta-1} \exp[-\lambda x]\, dx$$

$$=$$

Some special cases:

- If $\alpha = 1$, then $E[X^\beta] =$
- If $\beta \in \{1, 2, \ldots\}$, then $E[X^\beta] =$

Suppose that $\mathcal{X} = [0, \infty)$, so that X is a nonnegative random variable. Suppose also that X has probability density function $f_X(x)$, continuous on $\mathcal{X}$, and cumulative distribution function $F_X(x)$. Then integrating by parts with $u := x$, $dv := f_X(x)\ dx$, $v := -[1 - F_X(x)]$, and $du := dx$ yields

$$E[X] = \int_0^\infty x f_X(x)\ dx = \lim_{M\to\infty} \int_0^M x f_X(x)\ dx$$

$$= \lim_{M\to\infty} \left\{ -M[1 - F_X(M)] + \int_0^M [1 - F_X(x)]\ dx \right\}.$$

If $\lim_{M\to\infty} M[1 - F_X(M)] = 0$, then we obtain

$$E[X] = \lim_{M\to\infty} \int_0^M [1 - F_X(x)]\ dx = \int_0^\infty [1 - F_X(x)]\ dx.$$

To illustrate, let X have cumulative distribution function $F_X(x) := 1 - \exp[-x^2]$ for $x \geq 0$. Then $F'_X(x) = 2x \exp[-x^2]$ for $x > 0$ and we can put $f_X(x) := 2x \exp[-x^2]$ for $x \geq 0$. Since

$$\lim_{M\to\infty} M[1 - F_X(M)] = \lim_{M\to\infty} M \exp[-M^2] = 0,$$

we can calculate $E[X]$ by evaluating

$$\int_0^\infty [1 - F_X(x)]\ dx = \int_0^\infty \exp[-x^2]\ dx = (1/2) \int_{-\infty}^\infty \exp[-x^2]\ dx =: I.$$

Squaring this quantity and switching to polar coordinates, we have

$$I^2 = (1/4) \int_{-\infty}^\infty \int_{-\infty}^\infty \exp[-(x^2+y^2)]\ dx\ dy = (1/4) \int_0^\infty \int_0^{2\pi} \exp[-r^2]\ r\ dr\ d\theta$$

$$= (1/4) \int_0^\infty 2\pi \exp[-r^2]\ r\ dr =$$

from which we conclude that $I = E[X] = \sqrt{\pi}/2$.

Of course, another option would be to evaluate

$$E[X] = \int_0^\infty x f_X(x)\ dx = \int_0^\infty 2x^2 \exp[-x^2]\ dx.$$

Substituting $u := x^2$ and $du := 2x\ dx$, we would obtain

$$E[X] = \int_0^\infty u^{1/2} \exp[-u]\ du = \Gamma[3/2].$$

If you knew that $\Gamma[1/2] = \sqrt{\pi}$, you could conclude that $E[X] = \sqrt{\pi}/2$.

Let us do one last example. Say that X has probability density function $f_X(x) := (\alpha - 1)x^{-\alpha} 1_{\{x \geq 1\}}$ for some $\alpha > 1$. Let β be a positive real, and take $g(x) := x^\beta$ for $x \in [1, \infty)$. We have

$$E[X^\beta] = \int_1^\infty (\alpha - 1)x^{\beta-\alpha}\ dx = \lim_{M\to\infty} \int_1^M (\alpha - 1)x^{\beta-\alpha}\ dx.$$

If $\beta = \alpha - 1$, then we have

$$E[X^\beta] =$$

If $\beta > \alpha - 1$, then we have

$$E[X^\beta] = \lim_{M\to\infty} \frac{(\alpha - 1)}{(\beta - \alpha + 1)}(M^{\beta-\alpha+1} - 1) = \infty.$$

If $\beta < \alpha - 1$, then we have

$$E[X^\beta] = \lim_{M\to\infty} \frac{(\alpha - 1)}{(\beta - \alpha + 1)}(M^{\beta-\alpha+1} - 1) = \frac{(1 - \alpha)}{(\beta - \alpha + 1)}.$$

Linearity and monotonicity of expectation. We have

$$E[c_1 g_1(X) + c_2 g_2(X)] = c_1 E[g_1(X)] + c_2 E[g_2(X)]$$

whenever all of the expectations exist as finite numbers, due to the linearity of summation (for discrete X) and integration (for continuous X). Moreover, if the last two expectations exist as finite numbers, then so does the first.

If $g_1(x) \leq g_2(x) \leq g_3(x)$ for all $x \in \mathcal{X}$, then we have

$$E[g_1(X)] \leq E[g_2(X)] \leq E[g_3(X)].$$

In particular, if $E[g_1(X)]$ and $E[g_3(X)]$ exist as finite numbers, then so does $E[g_2(X)]$.

Example (linearity and monotonicity of expectation). Here is a problem from the 2009 comprehensive examination. Suppose that X is a nonnegative random variable for which $E[X^{2009}]$ exists as a finite number. Prove that $E[X^c]$ exists as a finite number for any $c \in (0, 2009)$.

Our strategy will be to set $g_2(x) := x^c$ and then exhibit $g_1(x)$ and $g_3(x)$ with $g_1(x) \leq g_2(x) \leq g_3(x)$ such that $E[g_1(X)]$ and $E[g_3(X)]$ exist as finite numbers. An obvious choice for $g_1(x)$ is   yielding $E[g_1(X)] =$ We would like to choose $g_3(x) := x^{2009}$, but unfortunately $x^{2009} < x^c$ when $x \in (0, 1)$. However, $x^c \leq 1$ when $x \in (0, 1)$, which gives us the idea to take $g_3(x) := 1 + x^{2009}$. Then by linearity we have $E[g_3(X)] = E[1] + E[X^{2009}] = 1 + E[X^{2009}]$, which is finite since $E[X^{2009}]$ was finite. Therefore, we conclude that $E[g_2(X)] = E[X^c]$ is finite.

Minimizing distance. Suppose that $E[X^2]$ exists as a finite number. Then we can show that $E[X]$ also exists as a finite number (by mimicking the argument above with 2 in place of 2009), so $E[(X - b)^2]$ exists as a finite number and is equal to $E[X^2] - 2E[X]b + b^2$ for any real b.

The authors of your textbook pose the following question: For what choice of b is $E[(X - b)^2]$, the average squared distance of X from b, minimized?

The authors of your textbook prove that the answer is $b := E[X]$ without using calculus, by writing $(X - b)^2$ as $(\{X - E[X]\} + \{E[X] - b\})^2$ and expanding the square. Below is a second proof that does use calculus.

Put $h(b) := E[X^2] - 2E[X]b + b^2$. Then $h'(b) = -2E[X] + 2b$, which is negative for $b \in (-\infty, E[X])$ and positive for $b \in (E[X], \infty)$, showing that $h(b)$ is decreasing for $b \in (-\infty, E[X])$ and increasing for $b \in (E[X], \infty)$. This implies that $h(b)$ is minimized at $b := E[X]$.