Expected value

Lecture 4b (STAT 24400 F24)

Expected value - definition

The **expected value** of a random variable X is the value we expect to get, on average, if we were able to repeat the random experiment or random process that generates X many times.

It is written as $\mathbb{E}(X)$, μ_X, or just μ.

It may also be called the "mean" of X or the "expectation" of X.

Expected value - calculation

For a discrete random variable it is calculated as

$$\mathbb{E}(X) = \sum_x x \cdot p_X(x)$$

which we can interpret as a "long-run average":
$p_X(x) = \mathbb{P}(X = x)$ is the proportion of times that we'll get the value x.

For a continuous random variable,

$$\mathbb{E}(X) = \int_{x=-\infty}^{\infty} x \cdot f_X(x)\, \mathrm{d}x$$

Expected value - assumptions

In all cases (discrete/continuous/etc), we need to be careful to check that this expected value exists.

- If X is discrete, the summation $\sum_x x \cdot p_X(x)$ exists as long as

  $$\sum_x |x| \cdot p_X(x) < \infty$$

  (If X has only finitely many values, this condition is always satisfied)

- If X is continuous, the integral $\int_{x=-\infty}^{\infty} x \cdot f_X(x)\, \mathrm{d}x$ exists

  $$\int_{x=-\infty}^{\infty} |x| \cdot f_X(x)\, \mathrm{d}x < \infty$$

- If the appropriate condition is not satisfied, then we say that the expectation does not exist for r.v. X.

Examples of calculating $\mathbb{E}(x)$

<u>Example</u> Suppose $X \sim$Bernoulli(p). What is $\mathbb{E}(X)$?

The possible values of X are 0 and 1, so

$$\mathbb{E}(X) = \sum_x x \cdot p_X(x) = 0 \cdot p_X(0) + 1 \cdot p_X(1) = 0 \cdot (1-p) + 1 \cdot p = p$$

In particular, if $X = \mathbb{1}_A$ then $\mathbb{E}(X) = \mathbb{P}(A)$.

<u>Example</u> Let $X \sim$Exponential(λ). What is $\mathbb{E}(X)$?

$$\mathbb{E}(X) = \int_{x=-\infty}^{\infty} x \cdot f_X(x)\,\mathrm{d}x = \int_{x=0}^{\infty} x \cdot \lambda e^{-\lambda x}\,\mathrm{d}x = \left[-xe^{-\lambda x} - \frac{1}{\lambda}e^{-\lambda x}\right]_{x=0}^{\infty} = \frac{1}{\lambda}$$

<u>Note</u>: $1/\lambda$ is often called the <u>scale</u>, λ the <u>rate</u>.

Properties of $\mathbb{E}(X)$

1. **Linearity.**

   Let $X_1, \ldots, X_n$ be random variables.

   Let $Y = a + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n$. Then

$$\mathbb{E}(Y) = a + b_1 \mathbb{E}(X_1) + b_2 \mathbb{E}(X_2) + \cdots + b_n \mathbb{E}(X_n)$$

Note that this property holds without assuming anything about the relationship among the X_i's, e.g. they do not need to be independent.

Caution: this property of linearity doesn't generalize to other functions!

For example,

$$\mathbb{E}(a\,X) = a\,\mathbb{E}(X) \quad \text{but} \quad \mathbb{E}(X^a) \neq \mathbb{E}(X)^a \quad (\text{for } a \neq 1)$$

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y) \quad \text{but} \quad \mathbb{E}(X/Y) \neq \mathbb{E}(X)/\mathbb{E}(Y)$$

Properties of $\mathbb{E}(X)$

2. **Monotonicity.**

   Let X and Y be two random variables such that $X \leq Y$ <u>almost surely</u>

   this means $\mathbb{P}(X \leq Y) = 1$
   (implicitly, X & Y defined on same sample space)

   Then $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

Properties of $\mathbb{E}(X)$

3. **Transformations.**

   Let X be a r.v. and let $Y = g(X)$, where g is any function.

   Then if X is discrete,

$$\mathbb{E}(Y) = \sum_x g(x) \cdot p_X(x),$$

   while if X is continuous,

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

Note: equivalently, we could calculate the PMF or density of $Y = g(X)$, and then calculate $\mathbb{E}(Y)$ directly (which could be more complicated).

Markov's inequality

Let X be any random variable supported on $[0, \infty)$ with mean μ.

Then for any $t > 0$,

$$\mathbb{P}(X \geq t) \leq \frac{\mu}{t} \qquad \text{(Markov's inequality)}$$

Intuitively, the Markov's inequality gives an upper bound of the tail behavior by the expected value $\mu = \mathbb{E}(X)$.

Markov's inequality - Proof

To prove the Markov's inequality $\mathbb{P}(X \geq t) \leq \frac{\mu}{t}$ for $t > 0$,

let $Y = t \cdot \mathbb{1}\{X \geq t\}$. Then $Y \leq X$ always.

- If $X < t$ then $Y = 0 \leq X$
- If $X \geq t$ then $Y = t \leq X$

By monotonicity, we know that $\mathbb{E}(Y) \leq \mathbb{E}(X) = \mu$.
To derive $\mathbb{E}(Y)$:

$$\begin{aligned}\mathbb{E}(Y) &= \textstyle\sum_y y \cdot p_Y(y) \\ &= 0 \cdot \mathbb{P}(Y = 0) + t \cdot \mathbb{P}(Y = t) \;=\; t \cdot \mathbb{P}(X \geq t)\end{aligned}$$

Now $\mathbb{E}(Y) \leq \mathbb{E}(X) = \mu$ implies $\mathbb{P}(X \geq t) \leq \frac{\mu}{t}$.

Example - Binomial

Example 1

Suppose $X \sim$Binomial(n, p). What is $\mathbb{E}(X)$?

$$\mathbb{E}(X) = \sum_{k=0}^{n} k \cdot p_X(k) = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k}$$

A simpler calculation:

Let $X = \mathbb{1}_{A_1} + \mathbb{1}_{A_2} + \cdots + \mathbb{1}_{A_n}$, where A_i is the event that the ith trial succeeds.

Then

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{1}_{A_1}) + \cdots + \mathbb{E}(\mathbb{1}_{A_n}) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n) = n \cdot p.$$

Examples - Expectation of sums

Example 2

26 cards labeled with the letters A through Z are randomly shuffled.
What's the expected value of L, the number of letters we need to draw to reach the letter A?

The possible values of L are 1, 2, 3, ..., 26, and each is equally likely.

$$\mathbb{E}(L) = 1 \cdot \frac{1}{26} + 2 \cdot \frac{1}{26} + 3 \cdot \frac{1}{26} + \cdots + 26 \cdot \frac{1}{26} = 13.5.$$

Remarks Although the calculation look like the binomial case in Example 1, here the draws are not independent, thus not of binomial distribution.

Examples - Expectation of sums (cont.)

Example 3

26 cards labeled with the letters A through Z are randomly shuffled.

What's the expected value of V, the number of vowels (A, E, I, O, or U) that we see in the first 10 draws?

Let A_i be the event that the ith draw is a vowel, for each $i = 1, \ldots, 10$. Then

$$V = \mathbb{1}_{A_1} + \mathbb{1}_{A_2} + \cdots + \mathbb{1}_{A_{10}}$$

and so by linearity,

$$\mathbb{E}(V) = \mathbb{E}(\mathbb{1}_{A_1}) + \cdots + \mathbb{E}(\mathbb{1}_{A_{10}}) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_{10}).$$

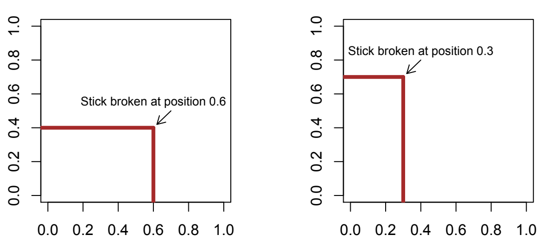
For each i, $\mathbb{P}(A_i) = \frac{5}{26} \rightsquigarrow \mathbb{E}(V) = 10 \cdot \frac{5}{26} = 1.92.$

Note:
This calculation looks exactly the same as for the Binomial distribution, but this example is not Binomial (because the trials are **not independent**.)

Example - Expectation of functions of r.v

Example 4

Suppose we break a stick of length 1 at a uniformly random location, and use the stick to block off a rectangle:



What is the expected value of A, the area of the rectangle?

We can write $A = g(U) = U(1-U)$, where $U \sim \text{Uniform}[0,1]$.
U has density $f(x) = 1$ on $x \in [0,1]$, so:

$$\mathbb{E}(A) = \int_{x=0}^{1} g(x) \cdot f(x)\, \mathrm{d}x = \int_{x=0}^{1} x(1-x) \cdot 1\, \mathrm{d}x = \frac{1}{6}.$$

Remarks on $\mathbb{E}(X)$

The examples show the usefulness of indicator variables and the surprising easiness of calculating expected values of functions of random variables.

For example, in Example 4, alternatively, we may derive f_A, the density function of A, from $A = U(1-U)$, then calculate $\mathbb{E}(A)$ from integral via f_A directly.

However, since this is not a monotone transformation, it takes a few steps to derive f_A, and the integration is not as easy.