JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Lecture 3

1. Define expected values
2. Properties of expected values
3. Unbiasedness of the sample mean
4. Define variances
5. Define the standard deviation
6. Calculate Bernoulli variance

Expected values

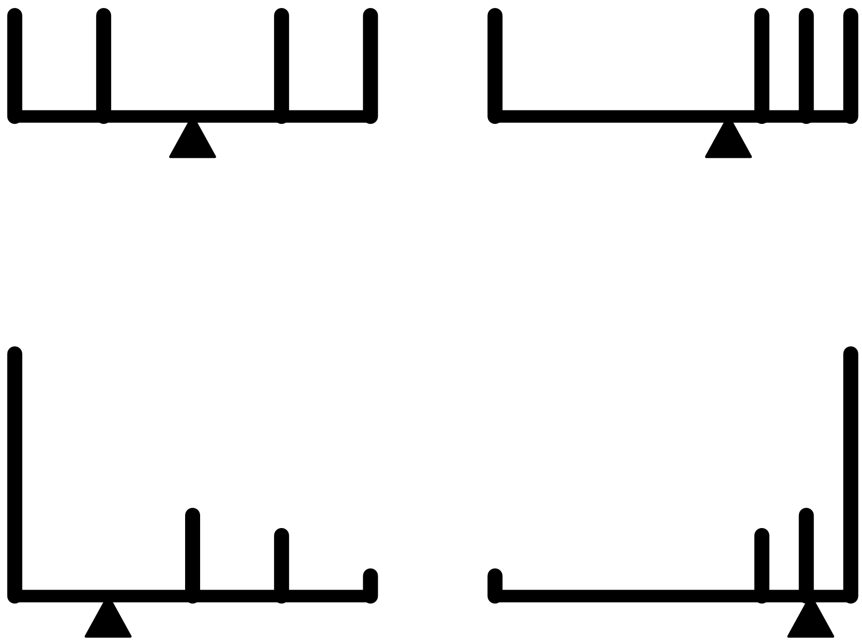
The **expected value** or **mean** of a random variable is the center of its distribution

For discrete random variable X with PMF $p(x)$, it is defnied as follows

$$E[X] = \sum_x xp(x).$$

where the sum is taken over the possible values of x

$E[X]$ represents the center of mass of a collection of locations and weights, $\{x, p(x)\}$



Example

- Suppose a coin is flipped and X is declared 0 or 1 corresponding to a head or a tail, respectively
- What is the expected value of X?

$$E[X] = .5 \times 0 + .5 \times 1 = .5.$$

Example

Suppose that a die is tossed and X is the number face up

What is the expected value of X?

$$E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5.$$

Continous random variables

- For a continuous random variable, X, with density, f, the expected value is defined as follows
$$E[X] = \int_{-\infty}^{\infty} t f(t) dt.$$
- This definition borrows from the definition of center of mass for a continous body

Example

- Consider a density where $f(x) = 1$ for x between zero and one
- (Is this a valid density?)
- Suppose that X follows this density; what is its expected value?

$$E[X] = \int_0^1 x dx = \frac{x^2}{2}\Big|_0^1 = 1/2.$$

Rules about expected values

- The expected value is a linear operator
- If a and b are not random and X and Y are two random variables then
  - $E[aX + b] = aE[X] + b$
  - $E[X + Y] = E[X] + E[Y]$
- *In general* if g is a function that is not linear,

$$E[g(X)] \neq g(E[X])$$

- For example, in general, $E[X^2] \neq E[X]^2$

Example

You flip a coin, X and simulate a uniform random number Y, what is the expected value of their sum?

$$E[X+Y] = E[X] + E[Y] = .5 + .5 = 1$$

Another example, you roll a coin twice. What is the expected value of the average?

Let X_1 and X_2 be the results of the two rolls

$$E[(X_1 + X_2)/2] = \frac{1}{2}(E[X_1] + E[X_2]) = \frac{1}{2}(3.5 + 3.5) = 3.5$$

Example

1. Let X_i for $i = 1, \ldots, n$ be a collection of random variables, each from a distribution with mean μ
2. Calculate the expected value of the sample average of the X_i.

$$
\begin{aligned}
E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] &= \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mu = \mu.
\end{aligned}
$$

Remark

Therefore, the expected value of the **sample mean** is the **population mean** that it's trying to estimate

When the expected value of an estimator is what its trying to estimate, we say that the estimator is **unbiased**

The variance

- The variance of a random variable is a measure of *spread*.
- If X is a random variable with mean μ, the variance of X is defined as

$$\mathrm{Var}(X) = E[(X - \mu)^2].$$

  the expected (squared) distance from the mean
- Densities with a higher variance are more spread out than densities with a lower variance

- Convenient computational form

$$\mathrm{Var}(X) = E[X^2] - E[X]^2.$$

- If a is constant then $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$
- The square root of the variance is called the **standard deviation**
- The standard deviation has the same units as X

Example

What's the sample variance from the result of a toss of a die?

$$E[X] = 3.5$$

$$E[X^2] = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6} = 15.17$$

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 \approx 2.92$$

Example

What's the sample variance from the result of the toss of a coin with probability of heads (1) of p?

$$E[X] = 0 \times (1-p) + 1 \times p = p$$

$$E[X^2] = E[X] = p$$

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = p - p^2 = p(1-p)$$

Example

- Suppose that a random variable is such that $0 \leq X \leq 1$ and $E[X] = p$
- Note $X^2 \leq X$ so that $E[X^2] \leq E[X] = p$
- $\mathrm{Var}(X) = E[X^2] - E[X]^2 \leq E[X] - E[X]^2 = p(1-p)$
- Therefore the Bernoulli variance is the largest possible for random variables bounded between 0 and 1