Expected Value, Variance and Covariance[1]

STA 256: Fall 2018

[1] This slide show is an open-source document. See last slide for copyright information.

Overview

Definition

The expected value of a discrete random variable is

$$E(X) = \sum_x x \, p_x(x)$$

- Provided $\sum_x |x| \, p_x(x) < \infty$. If the sum diverges, the expected value does not exist.
- Existence is only an issue for infinite sums (and integrals over infinite intervals).

Expected value is an average

- Imagine a very large jar full of balls. This is the population.
- The balls are numbered $x_1, \ldots, x_N$. These are measurements carried out on members of the population.
- Suppose for now that all the numbers are different.
- A ball is selected at random; all balls are equally likely to be chosen.
- Let X be the number on the ball selected.
- $P(X = x_i) = \frac{1}{N}$.

$$
\begin{aligned}
E(X) &= \sum_x x\, p_x(x) \\
&= \sum_{i=1}^{N} x_i \frac{1}{N} \\
&= \frac{\sum_{i=1}^{N} x_i}{N}
\end{aligned}
$$

For the jar full of numbered balls, $E(X) = \frac{\sum_{i=1}^{N} x_i}{N}$

- This is the common average, or arithmetic mean.
- Suppose there are ties.
- Unique values are v_i, for $i = 1, \ldots, n$.
- Say n_1 balls have value v_1, and n_2 balls have value v_2, and $\ldots n_n$ balls have value v_n.
- Note $n_1 + \cdots + n_n = N$, and $P(X = v_j) = \frac{n_j}{N}$.

$$
\begin{aligned}
E(X) &= \frac{\sum_{i=1}^{N} x_i}{N} \\
&= \sum_{j=1}^{n} n_j v_j \frac{1}{N} \\
&= \sum_{j=1}^{n} v_j \frac{n_j}{N} \\
&= \sum_{j=1}^{n} v_j P(X = v_j)
\end{aligned}
$$

Compare $E(X) = \sum_{j=1}^{n} v_j P(X = v_j)$ and $\sum_x x \, p_x(x)$

- Expected value is a generalization of the idea of an average, or mean.
- Specifically a *population* mean.
- It is often just called the "mean."

Definition

The expected value of a continuous random variable is

$$E(X) = \int_{-\infty}^{\infty} x \, f_x(x) \, dx$$

- Provided $\int_{-\infty}^{\infty} |x| \, f_x(x) \, dx < \infty$. If the integral diverges, the expected value does not exist.

The expected value is the physical balance point.

Sometimes the expected value does not exist

Need $\int_{-\infty}^{\infty} |x| \, f_x(x) \, dx < \infty$

For the Cauchy distribution, $f(x) = \frac{1}{\pi(1+x^2)}$.

$$
\begin{aligned}
E(|X|) &= \int_{-\infty}^{\infty} |x| \, \frac{1}{\pi(1+x^2)} \, dx \\
&= 2 \int_0^{\infty} \frac{x}{\pi(1+x^2)} \, dx \\
& \quad u = 1 + x^2, \; du = 2x \, dx \\
&= \frac{1}{\pi} \int_1^{\infty} \frac{1}{u} \, du \\
&= \ln u |_1^{\infty} \\
&= \infty - 0 = \infty
\end{aligned}
$$

So to speak. When we say an integral "equals" infinity, we just mean it is unbounded above.

Existence of expected values

- If it is not mentioned in a general problem, existence of expected values is assumed.
- Sometimes, the answer to a specific problem is “Oops! The expected value dies not exist.”
- You never need to show existence unless you are explicitly asked to do so.
- If you do need to deal with existence, Fubini’s Theorem can help with multiple sums or integrals.
    - Part One says that if the integrand is positive, the answer is the same when you switch order of integration, even when the answer is “‘∞.”
    - Part Two says that if the integral converges absolutely, you can switch order of integration. For us, absolute convergence just means that the expected value exists.

The change of variables formula for expected value

Theorems A and B in Chapter 4

Let X be a random variable and $Y = g(X)$. There are two ways to get $E(Y)$.

1. Derive the distribution of Y and compute

$$E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy$$

2. Use the distribution of X and calculate

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$$

Big theorem: These two expressions are equal.

The change of variables formula is very general

Including but not limited to

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$$

$$E(g(\mathbf{X})) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_p)\, f_{\mathbf{X}}(x_1, \ldots, x_p)\, dx_1 \ldots dx_p$$

$$E\left(g(X)\right) = \sum_x g(x) p_X(x)$$

$$E(g(\mathbf{X})) = \sum_{x_1} \cdots \sum_{x_p} g(x_1, \ldots, x_p)\, p_{\mathbf{X}}(x_1, \ldots, x_p)$$

Example: Let $Y = aX$. Find $E(Y)$.

$$\begin{aligned} E(aX) &= \sum_x ax\, p_X(x) \\ &= a \sum_x x\, p_X(x) \\ &= a\, E(X) \end{aligned}$$

So $E(aX) = aE(X)$.

Show that the expected value of a constant is the constant.

$$\begin{aligned} E(a) &= \sum_x a\, p_X(x) \\ &= a \sum_x p_X(x) \\ &= a \cdot 1 \\ &= a \end{aligned}$$

So $E(a) = a$.

$E(X+Y) = E(X) + E(Y)$

$$
\begin{aligned}
E(X+Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)\, f_{xy}(x,y)\, dx\, dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f_{xy}(x,y)\, dx\, dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{xy}(x,y)\, dx\, dy \\
&= E(X) + E(Y)
\end{aligned}
$$

Putting it together

$$E(a + bX + cY) = a + b\,E(X) + c\,E(Y)$$

And in fact,

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

You can move the expected value sign through summation signs and constants. Expected value is a linear transformation.

$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$, but in general

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

Unless $g(x)$ is a linear function. So for example,

$$E(\ln(X)) \neq \ln(E(X))$$

$$E(\tfrac{1}{X}) \neq \tfrac{1}{E(X)}$$

$$E(X^k) \neq (E(X))^k$$

That is, the statements are not true in general. They might be true for some distributions.

Variance of a random variable X

Let $E(X) = \mu$ (The Greek letter "mu").

$$Var(X) = E\left((X - \mu)^2\right)$$

- The average (squared) difference from the average.
- It's a measure of how spread out the distribution is.
- Another measure of spread is the standard deviation, the square root of the variance.

Variance rules

$$Var(a + bX) = b^2 Var(X)$$

$$Var(X) = E(X^2) - [E(X)]^2$$

Famous Russian Inequalities

Very useful later

Because the variance is a measure of spread or dispersion, it places limits on how much probability can be out in the tails of a probability density of probability mass function. To see this, we will

- Prove Markov's inequalty.
- Use Markov's inequality to prove Chebyshev's inequality.
- Look at some examples.

Markov's inequality

Let Y be a random variable with $P(Y \geq 0) = 1$ and $E(Y) = \mu$. Then for any $t > 0$, $P(Y \geq t) \leq E(Y)/t$. Proof:

$$
\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} y f(y)\, dy \\
&= \int_{-\infty}^{t} y f(y)\, dy + \int_{t}^{\infty} y f(y)\, dy \\
&\geq \int_{t}^{\infty} y f(y)\, dy \\
&\geq \int_{t}^{\infty} t f(y)\, dy \\
&= t \int_{t}^{\infty} f(y)\, dy \\
&= t\, P(Y \geq t)
\end{aligned}
$$

So $P(Y \geq t) \leq E(Y)/t$.

Chebyshev's inequality

Let X be a random variable with $E(X) = \mu$ and $Var(X) = \sigma^2$. Then for any $k > 0$,

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- We are measuring distance from the mean *in units of the standard deviation.*
- The probability of observing a value of X more than 2 standard deviations away from the mean cannot be more than one fourth.
- This is true for *any* random variable that has a standard deviation.
- For the normal distribution, $P(|X - \mu| \geq 2\sigma) \approx .0455 < \frac{1}{4}$.

Proof of Chebyshev's inequality

Markov says $P(Y \geq t) \leq E(Y)/t$

In Markov's inequality, let $Y = (X-\mu)^2$ and $t = k^2\sigma^2$. Then

$$
\begin{aligned}
P\left((X-\mu)^2 \geq k^2\sigma^2\right) &\leq \frac{E\left((X-\mu)^2\right)}{k^2\sigma^2} \\
&= \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}
\end{aligned}
$$

So $P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$.

Example

Chebyshev says $P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

Let X have density $f(x) = e^{-x}$ for $x \geq 0$ (standard exponential). We know $E(X) = Var(X) = 1$. Find $P(|X-\mu| \geq 3\sigma)$ and compare with Chebyshev's inequality.

$F(x) = 1 - e^{-x}$ for $x \geq 0$, so

$$\begin{aligned} P(|X-\mu| \geq 3\sigma) &= P(X < -2) + P(X > 4) \\ &= 1 - F(4) \\ &= e^{-4} \\ &\approx 0.01831564 \end{aligned}$$

Compared to $\frac{1}{3^2} = \frac{1}{9} = 0.11$.

Conditional Expectation
The idea

Consider jointly distributed random variables X and Y.

- For each possible value of X, there is a conditional distribution of Y.
- Each conditional distribution has an expected value (population mean).
- If you could estimate $E(Y|X=x)$, it would be a good way to predict Y from X.
- Estimation comes later (in STA260).

Definition of Conditional Expectation

If X and Y are discrete, the conditional expected value of Y given X is

$$E(Y|X=x) = \sum_y y\, p_{y|x}(y|x)$$

If X and Y are continuous,

$$E(Y|X=x) = \int_{-\infty}^{\infty} y\, f_{y|x}(y|x)\, dy$$

Double Expectation: $E(Y) = E[E(Y|X)]$

Theorem A on page 149

To make sense of this, note

- While $E(Y|X=x) = \int_{-\infty}^{\infty} y\, f_{y|x}(y|x)\, dy$ is a real-valued function of x,
- $E(Y|X)$ is a random variable, a function of the random variable X.
- $E(Y|X) = g(X) = \int_{-\infty}^{\infty} y\, f_{y|x}(y|X)\, dy$.
- So that in $E[E(Y|X)] = E[g(X)]$, the outer expected value is with respect to the probability distribution of X.

$$\begin{aligned} E[E(Y|X)] &= E[g(X)] \\ &= \int_{-\infty}^{\infty} g(x)\, f_x(x)\, dx \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y\, f_{y|x}(y|x)\, dy \right) f_x(x)\, dx \end{aligned}$$

Proof of the double expectation formula

Book calls it the "Law of Total Expectation"

$$
\begin{aligned}
E[E(Y|X)] &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y\, f_{y|x}(y|x)\, dy \right) f_x(x)\, dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\, \frac{f_{xy}(x,y)}{f_x(x)}\, dy\, f_x(x)\, dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\, f_{xy}(x,y)\, dy\, dx \\
&= E(Y)
\end{aligned}
$$

Definition of Covariance

Let X and Y be jointly distributed random variables with $E(X) = \mu_x$ and $E(Y) = \mu_y$. The *covariance* between X and Y is

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

- If values of X that are above average tend to go woth values of Y that are above average, the covariance will be positive.
- If values of X that are above average tend to go woth values of Y that are *below* average, the covariance will be negative.
- Covariance means they vary together.
- You could think of $Var(X) = E[(X - \mu_x)^2]$ as $Cov(X, X)$.

Properties of Covariance

$Cov(X, Y) = E(XY) - E(X)E(Y)$

If X and Y are independent, $Cov(X, Y) = 0$

If $Cov(X, Y) = 0$, it does *not* follow that X and Y are independent.

$Cov(a + X, b + Y) = Cov(X, Y)$

$Cov(aX, bY) = abCov(X, Y)$

$Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$

$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$

If $X_1, \ldots, X_n$ are ind. $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$

Correlation

$$Corr(X,Y) = \rho = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$$

- $-1 \leq \rho \leq 1$
- Scale free: $Corr(aX, bY) = Corr(X,Y)$

Copyright Information