Expected Values, Covariance, Variance and Tail Inequalities

Andreas Klappenecker

Texas A&M University

Definition

Let X be a discrete random variable over the probability space $(\Omega, \mathcal{F}, \Pr)$. The **expectation value** of X is defined to be

$$\mathrm{E}[X] = \sum_{\alpha \in X(\Omega)} \alpha \Pr[X = \alpha],$$

when this sum is unconditionally convergent in $\overline{\mathbf{R}}$, the extended real numbers.

The expectation value is also called the **mean** of X.

Simplest Example

Example

Let $(\Omega, \mathcal{F}, \Pr)$ denote a probability space. Let I_A denote the indicator random variable of the event $A \in \mathcal{F}$. Then

$$E[I_A] = 0 \Pr[I_A = 0] + 1 \Pr[I_A = 1] = \Pr[A].$$

Example

Exercise

Suppose you have a weighted coin

$$\Pr[\textbf{heads}] = \frac{3}{4} \quad \textit{and} \quad \Pr[\textbf{tails}] = \frac{1}{4}.$$

If you flip heads, you win \$2, but if you flip tails, you lose \$1. What is the expected value of a coin flip?

Solution

Solution

Let X denote the random variable that gives the win/loss for each coin toss. Then

$$\begin{aligned}
\mathrm{E}[X] &= \Pr[X = 2] \cdot 2 + \Pr[X = -1](-1) \\
&= \frac{3}{4}2 - \frac{1}{4}1 \\
&= \$1.25
\end{aligned}$$

Expectation Value of Nonnegative Integral Random Variables

Proposition

If X is a random variable with nonnegative integer values, then the expectation can be calculated by

$$\mathrm{E}[X] = \sum_{x=1}^{\infty} \Pr[X \geqslant x],$$

which is often convenient.

Proof.

Writing $\Pr[X \geqslant x] = \sum_{k=x}^{\infty} \Pr[X = k]$, we can express the expectation in the form

$$\begin{array}{rcllll}
\sum_{x=1}^{\infty} \Pr[X \geqslant x] & = & \Pr[X=1] & + \Pr[X=2] & + \Pr[X=3] + \cdots \\
 & & & + \Pr[X=2] & + \Pr[X=3] + \cdots \\
 & & & & + \Pr[X=3] + \cdots \\
 & \vdots & & & \ddots \\
 & = & 1 \Pr[X=1] & +2 \Pr[X=2] & +3 \Pr[X=3] \\
 & = & \mathrm{E}[X]. & &
\end{array}$$

□

The equalities are justified, since $\mathrm{E}[X] = \sum_{x=1}^{\infty} x \Pr[X = x]$ is unconditionally convergent.

Linearity of Expectation

Linearity of Expectation

Proposition

If X and Y are two arbitrary discrete random variables, then

$$\mathrm{E}[aX + bY] = a\mathrm{E}[X] + b\mathrm{E}[Y],$$

that is, the expectation operator is linear. This is an extremely useful result.

This follows directly from the definition.

Example: Hat Check Girl

Problem

Suppose that n persons give their hats to the hat check girl. She is upset because her goat has just passed away, and is handing the hats back at random. We want to answer the following question: On average, how many persons get their own hat back?

Example

The **sample space** is $\Omega = \{1, \ldots, n\}$. The **$\sigma$-algebra** is $\mathcal{F} = 2^{\Omega}$.

An **event** $S \in \mathcal{F} = 2^{\Omega}$ means that each person $k \in S$ received her own hat. For example, $\{1, 3, 7\}$ means that the first, third, and seventh person received their own hat back.

Example: Hat Check Girl

Example

We want to count how many persons receive their own hat back. It is convenient to introduce the **indicator random variable** X_k of the event $\{k\}$ that the k-th person received her own hat back.

Then $X_k \colon \Omega \to \{0, 1\}$ is the function given by

$$X_k(n) = \begin{cases} 1 & \text{if } n = k, \\ 0 & \text{if } n \neq k. \end{cases}$$

Example: Hat Check Girl

Example

Since the hat check girl hands out the hats uniformly at random, we have

$$\Pr[k\text{-th person receives her own hat back}] = \frac{1}{n}.$$

Thus,

$$\Pr[X_k = 1] = \frac{1}{n}.$$

Example: Hat Check Girl

Example

Let $X = X_1 + \cdots + X_n$ denote the number of persons receiving their own hats. By linearity of expectation, we get

$$E[X] = \sum_{k=1}^{n} E[X_k] = \sum_{k=1}^{n} 1 \cdot \Pr[X_k = 1] = n(1/n) = 1.$$

Thus, on average one person receives her own hat back.

Covariance and Products

Definition

The **covariance** of two random variables is defined as

$$\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$

when this expression makes sense.

Covariance

Proposition

$$\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y].$$

Proof.

$$\begin{aligned}
\mathrm{Cov}[X, Y] &= \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] \\
&= \mathrm{E}[XY - X\mathrm{E}[Y] - \mathrm{E}[X]Y + \mathrm{E}[X]\mathrm{E}[Y]] \\
&= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y].
\end{aligned}$$

□

The Expected Value of Products

Proposition

If X and Y are random variables, then

$$\mathrm{E}[XY] = \mathrm{Cov}[X, Y] + \mathrm{E}[X]\mathrm{E}[Y].$$

Proof.

This follows from the previous proposition. $\square$

Expected Value of Products

Proposition

If X and Y are **independent** *discrete random variables, then*

$$E[XY] = E[X]\,E[Y].$$

Caveat: If X and Y are not independent, then this is in general false.

Proof

We prove it form simple random variables X and Y. We can write X and Y in the form

$$X = \sum_{i=1}^{m} a_i I_{A_i} \quad \text{and} \quad Y = \sum_{j=1}^{n} b_j I_{B_j},$$

where

1. A_i denotes the event $X = a_i$,
2. B_j denotes the event $Y = b_j$.

Since the random variables are independent, we have

$$\mathrm{E}[I_{A_i}]\mathrm{E}[I_{B_j}] = \Pr[A_i]\Pr[B_j] = \Pr[A_i \cap B_j] = \mathrm{E}[I_{A_i \cap B_j}] = \mathrm{E}[I_{A_i} I_{B_j}].$$

Proof (Continued)

It follows that

$$\begin{aligned}
\mathrm{E}[X]\mathrm{E}[Y] &= \mathrm{E}\left[\sum_{i=1}^{m} a_i I_{A_i}\right] \mathrm{E}\left[\sum_{j=1}^{n} b_j I_{B_j}\right] \\
&= \left(\sum_{i=1}^{m} a_i \mathrm{E}[I_{A_i}]\right)\left(\sum_{j=1}^{n} b_j \mathrm{E}[I_{B_j}]\right) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j E[I_{A_i} I_{B_j}]
\end{aligned}$$

Proof (Continued)

It follows that

$$\begin{aligned}
\mathrm{E}[X]\mathrm{E}[Y] &= \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j E[I_{A_i} I_{B_j}] \\
&= \mathrm{E}\left[\left(\sum_{i=1}^{m} a_i I_{A_i}\right)\left(\sum_{j=1}^{n} b_j I_{B_j}\right)\right] \\
&= \mathrm{E}[XY]
\end{aligned}$$

Consequence

Corollary

If X and Y are independent random variables, then

$$\mathrm{Cov}[X, Y] = 0.$$

Tail Inequalities

Markov's Inequality

The expectation can be used to bound probabilities, as the following simple, but fundamental, result shows:

Theorem (Markov's Inequality)

If X is a nonnegative random variable and t a positive real number, then

$$\Pr[X \geqslant t] \leqslant \frac{\mathrm{E}[X]}{t}.$$

Proof.

Let Y denote the indicator random variable of the event $X \geqslant t$, so

$$Y(\omega) = \begin{cases} 1 & \text{if } X(\omega) \geqslant t, \\ 0 & \text{if } X(\omega) < t. \end{cases}$$

The expectation value of X satisfies

$$\mathrm{E}[X] \geqslant \mathrm{E}[tY] = t\,\mathrm{E}[Y] = t\,\Pr[X \geqslant t],$$

which proves the claim. $\square$

Markov's Inequality: What goes wrong if X is not nonnegative?

Counterexample

Suppose that X is a random variable with values in $\{-2, 2\}$ such that

$$\Pr[X = -2] = \frac{1}{2}, \quad \Pr[X = 2] = \frac{1}{2}.$$

Thus, X is **not** *a nonnegative random variable. Then*

$$\Pr[X \geqslant 1] = \frac{1}{2} \not\leqslant \frac{\mathrm{E}[X]}{1} = 0.$$

So Markov's inequality does not hold here!

Exercise

Let X be a discrete random variable and let $h\colon \mathbf{R} \to \mathbf{R}$ be a nonnegative function. Show that for all positive real numbers t, we have

$$\Pr[h(X) \geqslant t] \leqslant \frac{E[h(X)]}{t}.$$

Solution

If X is a discrete random variable, then $h(X)$ is a nonnegative discrete random variable. Define Y by

$$Y(\omega) = \begin{cases} 0 & \text{if } h(X)(\omega) < t, \\ 1 & \text{if } h(X)(\omega) \geqslant t, \end{cases}$$

hence $Y = 1$ denotes the event $h(X) \geqslant t$, and $Y = 0$ the event $h(X) < t$. Thus, Y is an indicator random variable. We have

$$\mathrm{E}[h(X)] \geqslant \mathrm{E}[tY] = t\mathrm{E}[Y] = t\Pr[h(X) \geqslant t],$$

and the claim follows.

Variance

Variance

Definition

The **variance** $\mathrm{Var}[X]$ of a discrete random variable X is defined by

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

whenever this expression is well-defined. The variance measures the squared deviation from the expected value $\mathrm{E}[X]$.

Caveats

Nonlinear!

The variance is **not** a linear operator, since

$$\mathrm{Var}[X + X] = 4\mathrm{Var}[X]$$

holds, to mention just one example.

In general

$\mathrm{Var}[aX + b] = a^2\mathrm{Var}[X]$ for all $a, b \in \mathbf{R}$.

Independent Random Variables

Proposition

If X and Y are independent random variables, then the variance satisfies

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]. \tag{1}$$

The random variable X will rarely deviate from the expectation value if the variance is small. This is a consequence of the Chebychev's useful inequality:

Theorem (Chebychev's inequality)

If X is a random variable, then

$$\Pr[(X - E[X])^2 \geqslant \beta] \leqslant \frac{\mathrm{Var}[X]}{\beta}. \tag{2}$$

Proof.

Given the random variable X, we can define the new random variables $Y = X - \mathrm{E}[X]$ and Y^2. Since Y^2 is a nonnegative random variable, Markov's inequality shows that

$$\Pr[Y^2 \geqslant \beta] \leqslant \frac{\mathrm{E}[Y^2]}{\beta}.$$

Since $\mathrm{E}[Y^2] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{Var}(X)$, we have

$$\Pr[(X - \mathrm{E}[X])^2 \geqslant \beta] = \Pr[Y^2 \geqslant \beta] \leqslant \frac{\mathrm{E}[Y^2]}{\beta} = \frac{\mathrm{Var}[X]}{\beta}. \quad \square$$

Standard Deviation

The square root of the variance, $\sigma = \sqrt{\mathrm{Var}[X]}$, is called the *standard deviation* of the random variable X.

Exercise

Show that if X is a random variable with standard deviation σ, then

$$\Pr[\,|X - \mathrm{E}[X]| \geqslant c\sigma] \leqslant \frac{1}{c^2}$$

for any positive constant $c \in \mathbf{R}$. This formula is often also called Chebychev's inequality. Can you explain why?

Solution

We have

$$\Pr[|X - \mathrm{E}[X]| \geqslant c\sigma] = \Pr[(X - \mathrm{E}[X])^2 \geqslant (c\sigma)^2] \leqslant \frac{\mathrm{Var}[X]}{(c\sigma)^2} = \frac{1}{c^2},$$

which proves the claim. The proof shows that this inequality is really equivalent to Chebychev's inequality.

In general, we cannot improve upon Chebychev's inequality.

Example

Let c be a real constant, $c \geqslant 1$. Let X be a random variable with the probability distribution

$$\Pr[X = -c] = \frac{1}{2c^2}, \quad \Pr[X = 0] = 1 - \frac{1}{c^2}, \quad \Pr[X = c] = \frac{1}{2c^2}.$$

Then the expected value is given by

$$E[X] = \Pr[X = -c](-c) + \Pr[X = 0] \cdot 0 + \Pr[X = c](c) = 0.$$

Example (Continued)

Recall that

$$\Pr[X = -c] = \frac{1}{2c^2}, \quad \Pr[X = 0] = 1 - \frac{1}{c^2}, \quad \Pr[X = c] = \frac{1}{2c^2}.$$

The variance of X is given by

$$\begin{aligned}
\mathrm{Var}[X] &= \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2] \\
&= \Pr[X = -c]c^2 + \Pr[X = 0] \cdot 0 + \Pr[X = c]c^2 \\
&= \frac{1}{2} + 0 + \frac{1}{2} = 1.
\end{aligned}$$

Example (Continued)

By definition of the random variable X, we have

$$\Pr[|X - \mathrm{E}[X]| \geqslant c\sigma] = \Pr[|X| \geqslant c] = \Pr[X = -c] + \Pr[X = c] = \frac{1}{c^2}.$$

Since the standard deviation $\sigma = 1$, the bound given by Chebychev's inequality is tight, since

$$\Pr[|X - \mathrm{E}[X]| \geqslant c\sigma] \leqslant \frac{1}{c^2}.$$

Careful!!!

Chebychev's inequality

If X is a random variable, then

$$\Pr[(X - E[X])^2 \geqslant \beta] \leqslant \frac{\mathrm{Var}[X]}{\beta}. \tag{3}$$

Chebychev's inequality

If X is a random variable, then

$$\Pr[|X - E[X]| \geqslant \beta] \leqslant \frac{\mathrm{Var}[X]}{\beta^2}. \tag{4}$$