

Alex Psomas: Lecture 17.

Random Variables: Expectation, Variance

1. Random Variables, Expectation: Brief Review
2. Independent Random Variables.
3. Variance

Random Variables: Definitions

Definition

A **random variable**, X , for a random experiment with sample space Ω is a variable that takes as value one of the random samples.

NO!

Random Variables: Definitions

Definition

A **random variable**, X , for a random experiment with sample space Ω is a **function** $X : \Omega \rightarrow \mathfrak{R}$.

Thus, $X(\cdot)$ assigns a real number $X(\omega)$ to each $\omega \in \Omega$.

Definitions

(a) For $a \in \mathfrak{R}$, one defines the **event**

$$X^{-1}(a) := \{\omega \in \Omega \mid X(\omega) = a\}.$$

(b) For $A \subset \mathfrak{R}$, one defines the **event**

$$X^{-1}(A) := \{\omega \in \Omega \mid X(\omega) \in A\}.$$

(c) The probability that $X = a$ is defined as

$$Pr[X = a] = Pr[X^{-1}(a)].$$

(d) The probability that $X \in A$ is defined as

$$Pr[X \in A] = Pr[X^{-1}(A)].$$

(e) The **distribution** of a random variable X , is

$$\{(a, Pr[X = a]) : a \in \mathcal{A}\},$$

where \mathcal{A} is the *range* of X . That is, $\mathcal{A} = \{X(\omega), \omega \in \Omega\}$.

An Example

Flip a fair coin three times.

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

$$X = \text{number of } H\text{'s: } \{3, 2, 2, 2, 1, 1, 1, 0\}.$$

- ▶ Range of X ? $\{0, 1, 2, 3\}$. All the values X can take.
- ▶ $X^{-1}(2)$? $X^{-1}(2) = \{HHT, HTH, THH\}$. All the **outcomes** ω such that $X(\omega) = 2$.
- ▶ Is $X^{-1}(1)$ an event? **YES**. It's a subset of the outcomes.
- ▶ $Pr[X]$? This doesn't make any sense bro....
- ▶ $Pr[X = 2]$?

$$Pr[X = 2] = Pr[X^{-1}(2)] = Pr[\{HHT, HTH, THH\}]$$

$$= Pr[\{HHT\}] + Pr[\{HTH\}] + Pr[\{THH\}] = \frac{3}{8}$$

Random Variables: Definitions

Definition

Let X, Y, Z be random variables on Ω and $g : \Re^3 \rightarrow \Re$ a function. Then $g(X, Y, Z)$ is the random variable that assigns the value $g(X(\omega), Y(\omega), Z(\omega))$ to ω .

Thus, if $V = g(X, Y, Z)$, then $V(\omega) := g(X(\omega), Y(\omega), Z(\omega))$.

Examples:

- ▶ X^k
- ▶ $(X - a)^2$
- ▶ $a + bX + cX^2 + (Y - Z)^2$
- ▶ $(X - Y)^2$
- ▶ $X \cos(2\pi Y + Z)$.

Expectation - Definition

Definition: The **expected value** (or mean, or expectation) of a random variable X is

$$E[X] = \sum_a a \times Pr[X = a].$$

Theorem:

$$E[X] = \sum_{\omega} X(\omega) \times Pr[\omega].$$

An Example

Flip a fair coin three times.

$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$. $X =$ number of H 's: $\{3, 2, 2, 2, 1, 1, 1, 0\}$. Thus,

$$\sum_{\omega} X(\omega)Pr[\omega] = 3\frac{1}{8} + 2\frac{1}{8} + 2\frac{1}{8} + 2\frac{1}{8} + 1\frac{1}{8} + 1\frac{1}{8} + 1\frac{1}{8} + 0\frac{1}{8}.$$

Also,

$$\sum_a a \times Pr[X = a] = 3\frac{1}{8} + 2\frac{3}{8} + 1\frac{3}{8} + 0\frac{1}{8}.$$

Win or Lose.

Expected winnings for heads/tails games, with 3 flips?

Recall the definition of the random variable X :

$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \rightarrow \{3, 1, 1, -1, 1, -1, -1, -3\}$.

$$E[X] = 3\frac{1}{8} + 1\frac{3}{8} - 1\frac{3}{8} - 3\frac{1}{8} = 0.$$

Can you ever win 0?

Apparently: Expected value is not a common value. It doesn't have to be in the range of X .

The expected value of X is not the value that you expect!

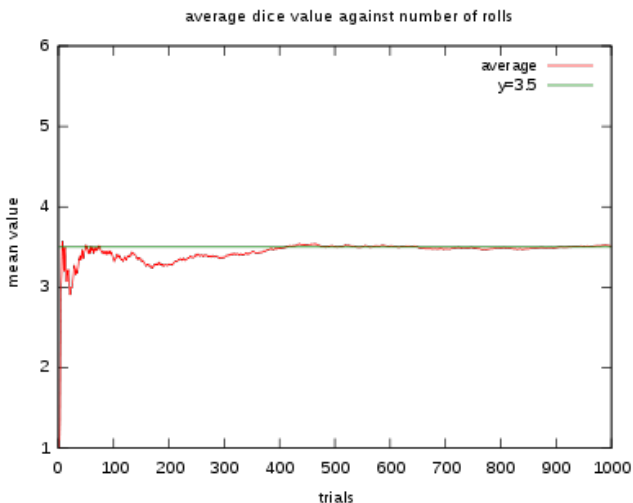
It is the average value per experiment, if you perform the experiment many times. Let X_1 be your winnings the first time you play the game, X_2 are your winnings the second time you play the game, and so on. (Notice that X_i 's have the same distribution!) When $n \gg 1$:

$$\frac{X_1 + \cdots + X_n}{n} \rightarrow 0$$

The fact that this average converges to $E[X]$ is a theorem: the [Law of Large Numbers](#). (See later.)

Law of Large Numbers

An Illustration: Rolling Dice



Indicators

Definition

Let A be an event. The random variable X defined by

$$X(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \notin A \end{cases}$$

is called the **indicator** of the event A .

Note that $Pr[X = 1] = Pr[A]$ and $Pr[X = 0] = 1 - Pr[A]$.

Hence,

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

This random variable $X(\omega)$ is sometimes written as

$$1\{\omega \in A\} \text{ or } 1_A(\omega).$$

Thus, we will write $X = 1_A$.

Linearity of Expectation

Theorem: Expectation is linear

$$E[a_1 X_1 + \cdots + a_n X_n] = a_1 E[X_1] + \cdots + a_n E[X_n].$$

Proof:

$$\begin{aligned} E[a_1 X_1 + \cdots + a_n X_n] &= \sum_{\omega} (a_1 X_1 + \cdots + a_n X_n)(\omega) Pr[\omega] \\ &= \sum_{\omega} (a_1 X_1(\omega) + \cdots + a_n X_n(\omega)) Pr[\omega] \\ &= a_1 \sum_{\omega} X_1(\omega) Pr[\omega] + \cdots + a_n \sum_{\omega} X_n(\omega) Pr[\omega] \\ &= a_1 E[X_1] + \cdots + a_n E[X_n]. \end{aligned}$$



Note: If we had defined $Y = a_1 X_1 + \cdots + a_n X_n$ and had tried to compute $E[Y] = \sum_y y Pr[Y = y]$, we would have been in trouble!

Using Linearity - 1: Dots on dice

Roll a die n times.

X_m = number of dots on roll m .

$X = X_1 + \cdots + X_n$ = total number of pips in n rolls.

$$\begin{aligned} E[X] &= E[X_1 + \cdots + X_n] \\ &= E[X_1] + \cdots + E[X_n], \text{ by linearity} \\ &= nE[X_1], \text{ because the } X_m \text{ have the same distribution} \end{aligned}$$

Now,

$$E[X_1] = 1 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \frac{6 \times 7}{2} \times \frac{1}{6} = \frac{7}{2}.$$

Hence,

$$E[X] = \frac{7n}{2}.$$

Note: Computing $\sum_x xPr[X = x]$ directly is not easy!

Using Linearity - 2: Binomial Distribution.

Flip n coins with heads probability p . X - number of heads

Binomial Distribution: $Pr[X = i]$, for each i .

$$Pr[X = i] = \binom{n}{i} p^i (1-p)^{n-i}.$$

$$E[X] = \sum_i i \times Pr[X = i] = \sum_i i \times \binom{n}{i} p^i (1-p)^{n-i}.$$

Better approach: Let

$$X_i = \begin{cases} 1 & \text{if } i\text{th flip is heads} \\ 0 & \text{otherwise} \end{cases}$$

$$E[X_i] = 1 \times Pr[\text{"heads"}] + 0 \times Pr[\text{"tails"}] = p.$$

Moreover $X = X_1 + \dots + X_n$ and

$$E[X] = E[X_1] + E[X_2] + \dots + E[X_n] = n \times E[X_i] = np.$$

Using Linearity - 3: The birthday paradox

Let X be the random variable indicating the number of pairs of people, in a group of k people, sharing the same birthday. What's $E(X)$?

Let $X_{i,j}$ be the indicator random variable for the event that two people i and j have the same birthday. $X = \sum_{i,j} X_{i,j}$.

$$\begin{aligned} E[X] &= E\left[\sum_{i,j} X_{i,j}\right] \\ &= \sum_{i,j} E[X_{i,j}] \\ &= \sum_{i,j} \Pr[X_{i,j}] \\ &= \sum_{i,j} \frac{1}{365} = \binom{k}{2} \frac{1}{365} = \frac{k(k-1)}{365} \end{aligned}$$

For a group of 28 it's about 1. For 100 it's 13.5. For 280 it's 107.

Calculating $E[g(X)]$

Let $Y = g(X)$. Assume that we know the distribution of X .

We want to calculate $E[Y]$.

Method 1: We calculate the distribution of Y :

$$Pr[Y = y] = Pr[X \in g^{-1}(y)] \text{ where } g^{-1}(x) = \{x \in \mathfrak{X} : g(x) = y\}.$$

This is typically rather tedious!

Method 2: We use the following result.

Theorem:

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

Proof:

$$\begin{aligned} E[g(X)] &= \sum_{\omega} g(X(\omega)) Pr[\omega] = \sum_x \sum_{\omega \in X^{-1}(x)} g(X(\omega)) Pr[\omega] \\ &= \sum_x \sum_{\omega \in X^{-1}(x)} g(x) Pr[\omega] = \sum_x g(x) \sum_{\omega \in X^{-1}(x)} Pr[\omega] \\ &= \sum_x g(x) Pr[X = x]. \end{aligned}$$



An Example

Let X be uniform in $\{-2, -1, 0, 1, 2, 3\}$.

Let also $g(X) = X^2$. Then (method 2)

$$\begin{aligned} E[g(X)] &= \sum_{x=-2}^3 x^2 \frac{1}{6} \\ &= \{4 + 1 + 0 + 1 + 4 + 9\} \frac{1}{6} = \frac{19}{6}. \end{aligned}$$

Method 1 - We find the distribution of $Y = X^2$:

$$Y = \begin{cases} 4, & \text{w.p. } \frac{2}{6} \\ 1, & \text{w.p. } \frac{2}{6} \\ 0, & \text{w.p. } \frac{1}{6} \\ 9, & \text{w.p. } \frac{1}{6}. \end{cases}$$

Thus,

$$E[Y] = 4 \frac{2}{6} + 1 \frac{2}{6} + 0 \frac{1}{6} + 9 \frac{1}{6} = \frac{19}{6}.$$

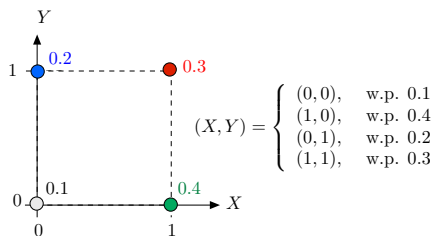
Calculating $E[g(X, Y, Z)]$

We have seen that $E[g(X)] = \sum_x g(x)Pr[X = x]$.

Using a similar derivation, one can show that

$$E[g(X, Y, Z)] = \sum_{x,y,z} g(x, y, z)Pr[X = x, Y = y, Z = z].$$

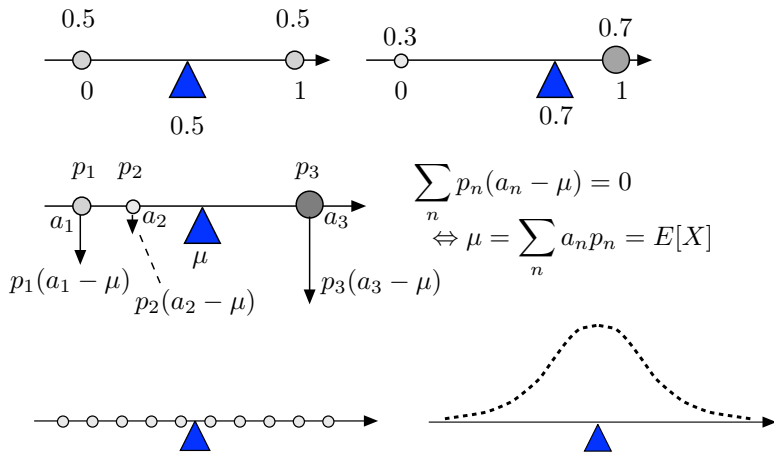
An Example. Let X, Y be as shown below:



$$\begin{aligned} E[\cos(2\pi X + \pi Y)] &= 0.1 \cos(0) + 0.4 \cos(2\pi) + 0.2 \cos(\pi) + 0.3 \cos(3\pi) \\ &= 0.1 \times 1 + 0.4 \times 1 + 0.2 \times (-1) + 0.3 \times (-1) = 0. \end{aligned}$$

Center of Mass

The expected value has a *center of mass* interpretation:



Best Guess: Least Squares

If you only know the distribution of X , it seems that $E[X]$ is a 'good guess' for X .

The following result makes that idea precise.

Theorem

The value of a that minimizes $E[(X - a)^2]$ is $a = E[X]$.

Unfortunately, we won't talk about this in this class...

Independent Random Variables.

Definition: Independence

The random variables X and Y are **independent** if and only if

$$Pr[Y = b|X = a] = Pr[Y = b], \text{ for all } a \text{ and } b.$$

Fact:

X, Y are independent if and only if

$$Pr[X = a, Y = b] = Pr[X = a]Pr[Y = b], \text{ for all } a \text{ and } b.$$

Obvious.

Independence: Examples

Example 1

Roll two die. X = number of dots on the first one, Y = number of dots on the other one. X, Y are independent.

Indeed: $Pr[X = a, Y = b] = \frac{1}{36}, Pr[X = a] = Pr[Y = b] = \frac{1}{6}$.

Example 2

Roll two die. X = total number of dots, Y = number of dots on die 1 minus number on die 2. X and Y are not independent.

Indeed: $Pr[X = 12, Y = 1] = 0 \neq Pr[X = 12]Pr[Y = 1] > 0$.

Example 3

Flip a fair coin five times, X = number of H s in first three flips, Y = number of H s in last two flips. X and Y are independent.

Indeed:

$$Pr[X = a, Y = b] = \binom{3}{a} \binom{2}{b} 2^{-5} = \binom{3}{a} 2^{-3} \times \binom{2}{b} 2^{-2} = Pr[X = a] Pr[Y = b].$$

A useful observation about independence

Theorem

X and Y are independent if and only if

$$Pr[X \in A, Y \in B] = Pr[X \in A]Pr[Y \in B] \text{ for all } A, B \subset \mathfrak{R}.$$

Proof:

If (\Leftarrow): Choose $A = \{a\}$ and $B = \{b\}$.

This shows that $Pr[X = a, Y = b] = Pr[X = a]Pr[Y = b]$.

Only if (\Rightarrow):

$$\begin{aligned} Pr[X \in A, Y \in B] &= \sum_{a \in A} \sum_{b \in B} Pr[X = a, Y = b] = \sum_{a \in A} \sum_{b \in B} Pr[X = a]Pr[Y = b] \\ &= \sum_{a \in A} \left[\sum_{b \in B} Pr[X = a]Pr[Y = b] \right] = \sum_{a \in A} Pr[X = a] \left[\sum_{b \in B} Pr[Y = b] \right] \\ &= \sum_{a \in A} Pr[X = a]Pr[Y \in B] = Pr[X \in A]Pr[Y \in B]. \end{aligned}$$



Functions of Independent random Variables

Theorem Functions of independent RVs are independent
Let X, Y be independent RV. Then

$f(X)$ and $g(Y)$ are independent, for all $f(\cdot), g(\cdot)$.

Mean of product of independent RV

Theorem

Let X, Y be independent RVs. Then

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

Proof:

Recall that $E[g(X, Y)] = \sum_{x,y} g(x, y)Pr[X = x, Y = y]$. Hence,

$$\begin{aligned} E[XY] &= \sum_{x,y} xyPr[X = x, Y = y] = \sum_{x,y} xyPr[X = x]Pr[Y = y], \text{ by ind.} \\ &= \sum_x \left[\sum_y xyPr[X = x]Pr[Y = y] \right] = \sum_x \left[xPr[X = x] \left(\sum_y yPr[Y = y] \right) \right] \\ &= \sum_x \left[xPr[X = x]E[Y] \right] = E[X]E[Y]. \end{aligned}$$



Examples

(1) Assume that X, Y, Z are (pairwise) independent, with $E[X] = E[Y] = E[Z] = 0$ and $E[X^2] = E[Y^2] = E[Z^2] = 1$.

Wait. Isn't X independent with itself? No. If I tell you the value of X , then you know the value of X .

Then

$$\begin{aligned} E[(X + 2Y + 3Z)^2] &= E[X^2 + 4Y^2 + 9Z^2 + 4XY + 12YZ + 6XZ] \\ &= 1 + 4 + 9 + 4 \times 0 + 12 \times 0 + 6 \times 0 \\ &= 14. \end{aligned}$$

(2) Let X, Y be independent and takes values from $\{1, 2, \dots, n\}$ uniformly at random. Then

$$\begin{aligned} E[(X - Y)^2] &= E[X^2 + Y^2 - 2XY] = 2E[X^2] - 2E[X]^2 \\ &= \frac{1 + 3n + 2n^2}{3} - \frac{(n+1)^2}{2}. \end{aligned}$$

Mutually Independent Random Variables

Definition

X, Y, Z are mutually independent if

$$Pr[X = x, Y = y, Z = z] = Pr[X = x]Pr[Y = y]Pr[Z = z], \text{ for all } x, y, z.$$

Theorem

The events A, B, C, \dots are pairwise (resp. mutually) independent iff the random variables $1_A, 1_B, 1_C, \dots$ are pairwise (resp. mutually) independent.

Proof:

$$Pr[1_A = 1, 1_B = 1, 1_C = 1] = Pr[A \cap B \cap C], \dots$$



Functions of pairwise independent RVs

If X, Y, Z are pairwise independent, but not mutually independent, it may be that

$f(X)$ and $g(Y, Z)$ are not independent.

Example: Flip two fair coins,
 $X = 1\{\text{coin 1 is } H\}$, $Y = 1\{\text{coin 2 is } H\}$, $Z = X \oplus Y$. Then,
 X, Y, Z are pairwise independent. Let $g(Y, Z) = Y \oplus Z$. Then
 $g(Y, Z) = X$ is not independent of X .

Functions of mutually independent RVs

One has the following result:

Theorem

Functions of disjoint collections of mutually independent random variables are mutually independent.

Example:

Let $\{X_n, n \geq 1\}$ be mutually independent. Then,

$Y_1 := X_1 X_2 (X_3 + X_4)^2$, $Y_2 := \max\{X_5, X_6\} - \min\{X_7, X_8\}$, $Y_3 := X_9 \cos(X_{10} + X_{11})$ are mutually independent.

Proof:

Let $B_1 := \{(x_1, x_2, x_3, x_4) \mid x_1 x_2 (x_3 + x_4)^2 \in A_1\}$. Similarly for B_2, B_3 . Then

$$\begin{aligned} &Pr[Y_1 \in A_1, Y_2 \in A_2, Y_3 \in A_3] \\ &= Pr[(X_1, \dots, X_4) \in B_1, (X_5, \dots, X_8) \in B_2, (X_9, \dots, X_{11}) \in B_3] \\ &= Pr[(X_1, \dots, X_4) \in B_1] Pr[(X_5, \dots, X_8) \in B_2] Pr[(X_9, \dots, X_{11}) \in B_3] \\ &= Pr[Y_1 \in A_1] Pr[Y_2 \in A_2] Pr[Y_3 \in A_3] \end{aligned}$$



Operations on Mutually Independent Events

Theorem

Operations on disjoint collections of mutually independent events produce mutually independent events.

For instance, if A, B, C, D, E are mutually independent, then $A \Delta B, C \setminus D, \bar{E}$ are mutually independent.

Product of mutually independent RVs

Theorem

Let X_1, \dots, X_n be mutually independent RVs. Then,

$$E[X_1 X_2 \cdots X_n] = E[X_1] E[X_2] \cdots E[X_n].$$

Proof:

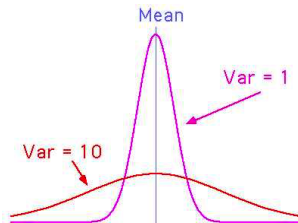
Assume that the result is true for n . (It is true for $n = 2$.)

Then, with $Y = X_1 \cdots X_n$, one has

$$\begin{aligned} E[X_1 \cdots X_n X_{n+1}] &= E[Y X_{n+1}], \\ &= E[Y] E[X_{n+1}], \\ &\quad \text{because } Y, X_{n+1} \text{ are independent} \\ &= E[X_1] \cdots E[X_n] E[X_{n+1}]. \end{aligned}$$



Variance



The variance measures the deviation from the mean value.

Definition: The **variance** of X is

$$\sigma^2(X) := \text{var}[X] = E[(X - E[X])^2].$$

$\sigma(X)$ is called the **standard deviation** of X .

Variance and Standard Deviation

Fact:

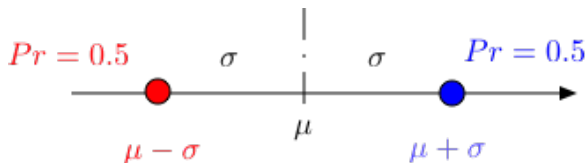
$$\text{var}[X] = E[X^2] - E[X]^2.$$

Indeed:

$$\begin{aligned}\text{var}(X) &= E[(X - E[X])^2] \\ &= E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2, \text{ by linearity} \\ &= E[X^2] - E[X]^2.\end{aligned}$$

A simple example

This example illustrates the term ‘standard deviation.’



Consider the random variable X such that

$$X = \begin{cases} \mu - \sigma, & \text{w.p. } 1/2 \\ \mu + \sigma, & \text{w.p. } 1/2. \end{cases}$$

Then, $E[X] = \mu$ and $(X - E[X])^2 = \sigma^2$. Hence,

$$\text{var}(X) = \sigma^2 \text{ and } \sigma(X) = \sigma.$$

Example

Consider X with

$$X = \begin{cases} -1, & \text{w. p. } 0.99 \\ 99, & \text{w. p. } 0.01. \end{cases}$$

Then

$$E[X] = -1 \times 0.99 + 99 \times 0.01 = 0.$$

$$E[X^2] = 1 \times 0.99 + (99)^2 \times 0.01 \approx 100.$$

$$\text{Var}(X) \approx 100 \implies \sigma(X) \approx 10.$$

Also,

$$E(|X|) = 1 \times 0.99 + 99 \times 0.01 = 1.98.$$

Thus, $\sigma(X) \neq E[|X - E[X]|]$!

Exercise: How big can you make $\frac{\sigma(X)}{E[|X - E[X]|]}$?

Properties of variance.

1. $\text{Var}(cX) = c^2 \text{Var}(X)$, where c is a constant.
Scales by c^2 .
2. $\text{Var}(X + c) = \text{Var}(X)$, where c is a constant.
Shifts center.

Proof:

$$\begin{aligned}\text{Var}(cX) &= E((cX)^2) - (E(cX))^2 \\ &= c^2 E(X^2) - c^2 (E(X))^2 = c^2 (E(X^2) - E(X)^2) \\ &= c^2 \text{Var}(X)\end{aligned}$$

$$\begin{aligned}\text{Var}(X + c) &= E((X + c - E(X + c))^2) \\ &= E((X + c - E(X) - c)^2) \\ &= E((X - E(X))^2) = \text{Var}(X)\end{aligned}$$



Variance of sum of two independent random variables

Theorem:

If X and Y are independent, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Proof:

Since shifting the random variables does not change their variance, let us subtract their means.

That is, we assume that $E(X) = 0$ and $E(Y) = 0$.

Then, by independence,

$$E(XY) = E(X)E(Y) = 0.$$

Hence,

$$\begin{aligned}\text{var}(X + Y) &= E((X + Y)^2) = E(X^2 + 2XY + Y^2) \\ &= E(X^2) + 2E(XY) + E(Y^2) = E(X^2) + E(Y^2) \\ &= \text{var}(X) + \text{var}(Y).\end{aligned}$$

Variance of sum of independent random variables

Theorem:

If X, Y, Z, \dots are pairwise independent, then

$$\text{var}(X + Y + Z + \dots) = \text{var}(X) + \text{var}(Y) + \text{var}(Z) + \dots.$$

Proof:

Since shifting the random variables does not change their variance, let us subtract their means.

That is, we assume that $E[X] = E[Y] = \dots = 0$.

Then, by independence,

$$E[XY] = E[X]E[Y] = 0. \text{ Also, } E[XZ] = E[YZ] = \dots = 0.$$

Hence,

$$\begin{aligned} \text{var}(X + Y + Z + \dots) &= E((X + Y + Z + \dots)^2) \\ &= E(X^2 + Y^2 + Z^2 + \dots + 2XY + 2XZ + 2YZ + \dots) \\ &= E(X^2) + E(Y^2) + E(Z^2) + \dots + 0 + \dots + 0 \\ &= \text{var}(X) + \text{var}(Y) + \text{var}(Z) + \dots. \end{aligned}$$



Today's gig: Lies!

Gig's so far:

1. How to tell random from human.
2. Monty Hall.
3. Birthday Paradox.
4. St. Petersburg paradox

Today: Simpson's paradox.

How come this show is still around?



Wait... Wrong Simpson.

The paradox

- ▶ A random variable X is a function $X : \Omega \rightarrow \mathfrak{R}$.
- ▶ $Pr[X = a] := Pr[X^{-1}(a)] = Pr[\{\omega \mid X(\omega) = a\}]$.
- ▶ $Pr[X \in A] := Pr[X^{-1}(A)]$.
- ▶ The distribution of X is the list of possible values and their probability: $\{(a, Pr[X = a]), a \in \mathcal{A}\}$.
- ▶ $g(X, Y, Z)$ assigns the value
- ▶ $E[X] := \sum_a a Pr[X = a]$.
- ▶ Expectation is Linear.
- ▶ Independent Random Variables.
- ▶ Variance.