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 $X : \Omega \to \Re$.

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

Definitions

(a) For $a \in \Re$, one defines the **event**

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

(b) For $A \subset \Re$, 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

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

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

An Example

Flip a fair coin three times.

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

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

- 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 $g : \Re^3 \to \Re$ 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:
$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \to \{3, 1, 1, -1, 1, -1, -1, -3\}$.

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

Can you ever win 0?

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

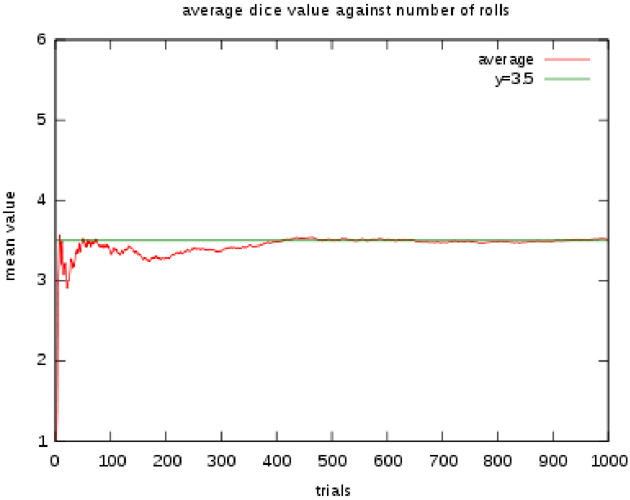
The expected value of X is not the value that you expect!
It is the average value per experiment, if you perform the experiment many times. Let X_1 be your winnings the first time you play the game, X_2 are your winnings the second time you play the game, and so on. (Notice that X_i's have the same distribution!) When $n \gg 1$:

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

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

$$E[a_1X_1+\cdots+a_nX_n] = a_1E[X_1]+\cdots+a_nE[X_n].$$

Proof:

$$\begin{aligned}
&E[a_1X_1+\cdots+a_nX_n] \\
&\quad = \sum_{\omega}(a_1X_1+\cdots+a_nX_n)(\omega)Pr[\omega] \\
&\quad = \sum_{\omega}(a_1X_1(\omega)+\cdots+a_nX_n(\omega))Pr[\omega] \\
&\quad = a_1\sum_{\omega}X_1(\omega)Pr[\omega]+\cdots+a_n\sum_{\omega}X_n(\omega)Pr[\omega] \\
&\quad = a_1E[X_1]+\cdots+a_nE[X_n].
\end{aligned}$$

$\square$

Note: If we had defined $Y = a_1X_1+\cdots+a_nX_n$ has had tried to compute $E[Y] = \sum_y yPr[Y=y]$, 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 Distibution: $Pr[X=i]$, 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 $X = X_1 + \cdots X_n$ and

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

Using Linearity - 3

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:

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

This is typically rather tedious!

Method 2: We use the following result.

Theorem:

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

Proof:

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

$\square$

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,

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

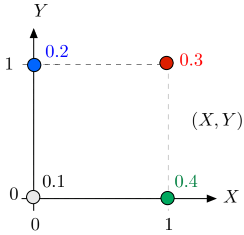
Calculating $E[g(X, Y, Z)]$

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:

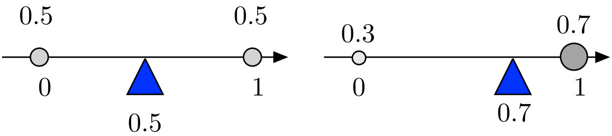


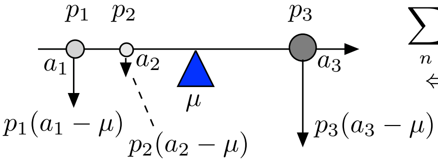
$$(X, Y) = \begin{cases} (0, 0), & \text{w.p. } 0.1 \\ (1, 0), & \text{w.p. } 0.4 \\ (0, 1), & \text{w.p. } 0.2 \\ (1, 1), & \text{w.p. } 0.3 \end{cases}$$

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

Center of Mass

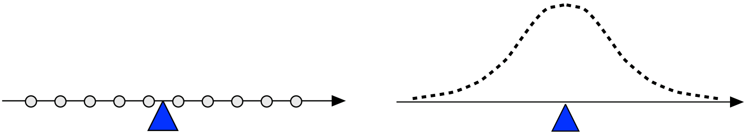
The expected value has a *center of mass* interpretation:





$$\sum_n p_n(a_n-\mu)=0$$

$$\Leftrightarrow \mu = \sum_n a_n p_n = E[X]$$



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 Hs in first three flips, Y = number of Hs 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

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

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 ($\Rightarrow$):

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

$\square$

Functions of Independent random Variables

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

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

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,

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

$\square$

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

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

(2) Let X, Y be independent and takes values from $\{1, 2, \ldots n\}$ 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 $A, B, C, \ldots$ are pairwise (resp. mutually) independent iff the random variables $1_A, 1_B, 1_C, \ldots$ are pairwise (resp. mutually) independent.
Proof:

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

$\square$

Functions of pairwise independent RVs

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

$$f(X) \text{ and } g(Y, Z) \text{ are not independent.}$$

Example 1: 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.

Example 2: Let A, B, C be pairwise but not mutually independent in a way that A and $B \cap C$ are not independent. Let $X = 1_A, Y = 1_B, Z = 1_C$. Choose $f(X) = X, g(Y, Z) = YZ$.

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_2. Then

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

$\square$

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.

Proof:

$1_{A \Delta B} = f(1_A, 1_B)$ where
$f(0,0) = 0, f(1,0) = 1, f(0,1) = 1, f(1,1) = 0$

$1_{C \setminus D} = g(1_C, 1_D)$ where
$g(0,0) = 0, g(1,0) = 1, g(0,1) = 0, g(1,1) = 0$

$1_{\bar{E}} = h(1_E)$ where
$h(0) = 1$ and $h(1) = 0$.

Hence, $1_{A \Delta B}, 1_{C \setminus D}, 1_{\bar{E}}$ are functions of mutually independent RVs. Thus, those RVs are mutually independent. Consequently, the events of which they are indicators are mutually independent. $\square$

Product of mutually independent RVs

Theorem

Let $X_1, \ldots, X_n$ be mutually independent RVs. Then,

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

Proof:

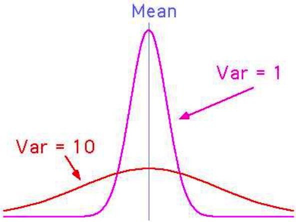
Assume that the result is true for n. (It is true for $n = 2$.)

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

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

□

Variance



The variance measures the deviation from the mean value.

Definition: The variance of X is

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

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

Variance and Standard Deviation

Fact:

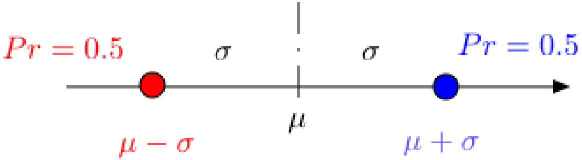
$$var[X] = E[X^2] - E[X]^2.$$

Indeed:

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

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,

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

Example

Consider X with

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

Then

$$\begin{aligned} E[X] &= -1 \times 0.99 + 99 \times 0.01 = 0. \\ E[X^2] &= 1 \times 0.99 + (99)^2 \times 0.01 \approx 100. \\ Var(X) &\approx 100 \implies \sigma(X) \approx 10. \end{aligned}$$

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 $\frac{\sigma(X)}{E[|X-E[X]|]}$?

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

Summary

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