Chapter 4
Expectation of Discrete R.V.s

Expectation

Defn: The **expectation of a discrete r.v.** X, written $\boldsymbol{E}[X]$, is the sum of the possible values of X, each weighted by its probability:

$$\boldsymbol{E}[X] = \sum_{x} x \cdot \boldsymbol{P}\{X = x\}$$

$\boldsymbol{E}[X]$ also represents the **mean of the distribution** from which X is drawn.

Average Cost of Lunch

MON	TUES	WED	THUR	FRI	SAT	SUN
$7	$7	$12	$12	$12	$0	$9

$$\text{Average Cost} = \frac{7+7+12+12+12+0+9}{7}$$

$$\parallel\!|$$

$$\boldsymbol{E}[\text{Cost}] = 7 \cdot \frac{2}{7} + 12 \cdot \frac{3}{7} + 9 \cdot \frac{1}{7} + 0 \cdot \frac{1}{7}$$

Expectation of Bernoulli($\boldsymbol{p}$)

X = value of the coin flip

Probability p of heads

Q: What is $\boldsymbol{E}[X]$?

$$\boldsymbol{E}[X] = 1 \cdot p + 0 \cdot (1 - p) = p$$

Remember! Mean of Bernoulli(p) is p.

Expected Time Until Disk Fails

Disk has probability $\frac{1}{3}$ of failing each year.

Q: On average, how many years will it be until the disk fails?

$X \sim Geometric(p)$ where $p = \frac{1}{3}$

$$\boxed{q = 1 - p}$$

$$\boldsymbol{E}[X] = \sum_{n=1}^{\infty} n\,(1-p)^{n-1} p$$

$$= p \sum_{n=1}^{\infty} n\, q^{n-1}$$

Remember me from Chapter 1?

$$= p \cdot \boxed{\left(1 + 2q + 3q^2 + 4q^3 + \cdots\right)} = p \cdot \frac{1}{(1-q)^2} = \frac{1}{p}$$

Remember! Mean of Geometric(p) is $\frac{1}{p}$.

Expectation of Poisson(λ)

$X \sim Poisson(\lambda)$

Q: What is $\boldsymbol{E}[X]$?

$$p_X(i) = \frac{e^{-\lambda}\lambda^i}{i!}, \ i = 0, 1, 2, \ldots$$

$$\boldsymbol{E}[X] = \sum_{i=0}^{\infty} i \cdot \frac{e^{-\lambda}\lambda^i}{i!} = \sum_{i=1}^{\infty} i \cdot \frac{e^{-\lambda}\lambda^i}{i!}$$

$$= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!}$$

Remember me from Chpt 1?

$$= \lambda e^{-\lambda} \boxed{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}} = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda$$

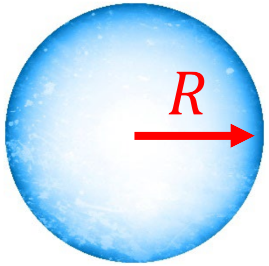
Remember! Mean of Poisson(λ) is λ.

Expectation of a Function of a R.V.

Defn: A **expectation of a function** $g(\cdot)$ **of a discrete r.v.** X is defined as follows:

$$\boldsymbol{E}[g(X)] = \sum_{x} g(x) \cdot p_X(x)$$

Consider a sphere, whose radius is a random variable R:



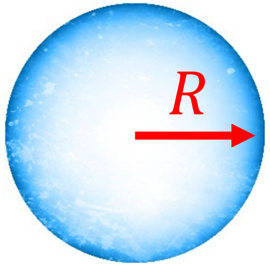
$$R = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 2 & \text{w.p. } \frac{1}{3} \\ 3 & \text{w.p. } \frac{1}{3} \end{cases}$$

Q: What is the expected volume of the sphere?

Expectation of a Function of a R.V.

<ins>Defn</ins>: A **expectation of a function $g(\cdot)$ of a discrete r.v.** X is defined as follows:

$$\boldsymbol{E}[g(X)] = \sum_x g(x) \cdot p_X(x)$$



$$\boldsymbol{E}[\text{Volume}] = \boldsymbol{E}\left[\frac{4}{3}\pi R^3\right]$$

$$= \frac{4}{3}\pi \cdot 1^3 \cdot \frac{1}{3} + \frac{4}{3}\pi \cdot 2^3 \cdot \frac{1}{3} + \frac{4}{3}\pi \cdot 3^3 \cdot \frac{1}{3}$$

$$= 16\,\pi$$

$$R = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 2 & \text{w.p. } \frac{1}{3} \\ 3 & \text{w.p. } \frac{1}{3} \end{cases}$$

Q: Is $\boldsymbol{E}[R^3] = (\boldsymbol{E}[R]^3)$?

Expectation of a Product

<u>Defn</u>: Let X and Y be r.v.s. The **expectation of the product** XY is defined as follows:

$$\mathbf{E}[XY] = \sum_x \sum_y xy \cdot p_{X,Y}(x,y)$$

$$\mathbf{E}[g(X)f(Y)] = \sum_x \sum_y g(x)f(y) \cdot p_{X,Y}(x,y)$$

where $p_{X,Y}(x,y) = \mathbf{P}\{X = x \,\&\, Y = y\}$.

Expectation of Product under Independence

Theorem 4.8: (Expectation of a product) If $X \perp Y$, then $\boldsymbol{E}[XY] = \boldsymbol{E}[X] \cdot \boldsymbol{E}[Y]$.

Proof:

$$\begin{aligned}\boldsymbol{E}[\mathrm{XY}] &= \sum_x \sum_y xy \cdot \boldsymbol{P}\{X = x, Y = y\} \\ &= \sum_x \sum_y xy \cdot \boldsymbol{P}\{X = x\} \cdot \boldsymbol{P}\{Y = y\} \\ &= \sum_x x \cdot \boldsymbol{P}\{X = x\} \sum_y y \cdot \boldsymbol{P}\{Y = y\} \\ &= \boldsymbol{E}[X] \cdot \boldsymbol{E}[Y]\end{aligned}$$

Via the same proof: If $X \perp Y$, then $\boldsymbol{E}[g(X)f(Y)] = \boldsymbol{E}[g(X)] \cdot \boldsymbol{E}[f(Y)]$.

Alternative Definition of Expectation

Theorem 4.9: (Alternative Definition of Expectation) Let X be a non-negative, discrete, integer-valued random variable. Then

$$\boldsymbol{E}[X] = \sum_{x=0}^{\infty} \boldsymbol{P}\{X > x\}.$$

This alternative formulation can be very useful.

Proof: See exercise in textbook. Hint: Rewrite the inside probability as a sum.

Expectation Property

Theorem: For random variable X and real number a,

$$\boldsymbol{E}[aX] = a\boldsymbol{E}[X]$$

Proof:

$$\boldsymbol{E}[aX] = \sum_x a\, x\, p_X(x) = a \sum_x x\, p_X(x) = a\boldsymbol{E}[X]$$

Linearity of Expectation

The following theorem greatly simplifies the computation of an expectation by breaking up the random variable into smaller pieces.

Theorem 4.10: [Linearity of Expectation] For random variables X and Y,

$$\boldsymbol{E}[X+Y] = \boldsymbol{E}[X] + \boldsymbol{E}[Y]$$

Proof: First try proving this yourself.
It's similar to the $\boldsymbol{E}[XY]$ derivation,
but you aren't allowed to split
the $p_{X,Y}(x,y)$…

Note this theorem
does not require
$X \perp Y$

Linearity of Expectation

Theorem 4.10: [Linearity of Expectation] For random variables X and Y,

$$\boldsymbol{E}[X+Y] = \boldsymbol{E}[X] + \boldsymbol{E}[Y]$$

Proof:

$$\boldsymbol{E}[\mathrm{X}+\mathrm{Y}] = \sum_y \sum_x (x+y) \cdot p_X(x,y)$$

$$= \sum_y \sum_x x \cdot p_X(x,y) + \sum_y \sum_x y \cdot p_X(x,y)$$

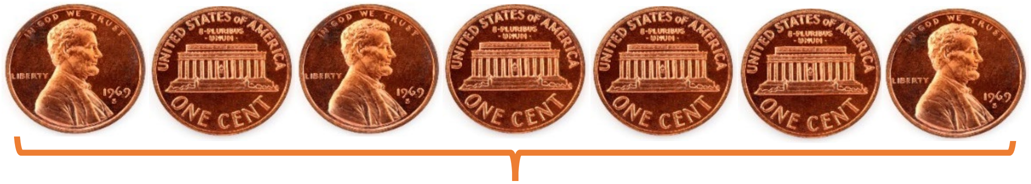
$$= \sum_x \sum_y x \cdot p_X(x,y) + \sum_y \sum_x y \cdot p_X(x,y)$$

$$= \sum_x x \sum_y p_X(x,y) + \sum_y y \sum_x p_X(x,y) \quad = \sum_x x p_X(x) + \sum_y y p_Y(y) \quad = \boldsymbol{E}[X] + \boldsymbol{E}[Y]$$

Expectation of Binomial(n, p)

Experiment: Flip a coin, with probability p of Heads, n times

Random Variable X = number of heads



n

Key Observation:

$$X = \mathrm{X}_1 + \mathrm{X}_2 + \cdots + X_n \text{ , where } X_i \sim \text{Bernoulli}(p)$$

What is $\boldsymbol{E}[X_i]$?

Applying Linearity of Expectation:

$$\boldsymbol{E}[X] = \boldsymbol{E}[\mathrm{X}_1] + \boldsymbol{E}[\mathrm{X}_2] + \cdots + \boldsymbol{E}[X_n]$$

$$= p + p + \cdots + p \ = np$$

Should make intuitive sense

Remember! Mean of Binomial(n, p) is np.

Expectation of Binomial(n, p)

Experiment: Flip a coin, with probability p of Heads, n times

Random Variable X = number of heads

n

Key Observation:

$$X = X_1 + X_2 + \cdots + X_n \text{ , where } X_i \sim \text{Bernoulli}(p)$$

Defn: The X_i here are called **indicator r.v.s,** because they take on values of 1 or 0.

Applying Linearity of Expectation:

$$\boldsymbol{E}[X] = \boldsymbol{E}[X_1] + \boldsymbol{E}[X_2] + \cdots + \boldsymbol{E}[X_n]$$

$$= p + p + \cdots + p = np$$

Q: Were the $X_i's$ indpt in here?

Drinking from your own Cup

At a party, n people put their drink on a table. Later that night, no one can remember which cup is theirs, so they simply each grab any cup at random.

Let $X =$ number of people who get back their own cup.

Q: What is $\boldsymbol{E}[X]$? Is it increasing with n ?

Drinking from your own Cup

At a party, n people put their drink on a table. Later that night, no one can remember which cup is theirs, so they simply each grab any cup at random.

Let X = number of people who get back their own cup.

Q: What is $\boldsymbol{E}[X]$? Is it increasing with n ?

Idea: $X = \mathrm{X}_1 + \mathrm{X}_2 + \cdots + X_n$

Q: What do the $\mathrm{X_i}$ represent?

Drinking from your own Cup

At a party, n people put their drink on a table. Later that night, no one can remember which cup is theirs, so they simply each grab any cup at random.

Let $X =$ number of people who get back their own cup.

Q: What is $\boldsymbol{E}[X]$?

Idea: $X = \mathrm{X}_1 + \mathrm{X}_2 + \cdots + X_n$

$$X_i = \begin{cases} 1 & \text{w.p. } 1/n \\ 0 & \text{o.w.} \end{cases}$$

$$\boxed{X_i = 1 \Leftrightarrow \text{person } i \text{ got back their own cup}}$$

Q: Are the X_i independent Bernoulli distributions? If so, is X Binomially distributed?

Drinking from your own Cup

At a party, n people put their drink on a table. Later that night, no one can remember which cup is theirs, so they simply each grab any cup at random.

Let X = number of people who get back their own cup.

Q: What is $\boldsymbol{E}[X]$?

Idea: $X = \mathrm{X}_1 + \mathrm{X}_2 + \cdots + X_n$

$$X_i = \begin{cases} 1 & \text{w.p. } 1/n \\ 0 & \text{o.w.} \end{cases}$$

$\boxed{X_i = 1 \Leftrightarrow \text{person } i \text{ got back their own cup}}$

A: The X_i 's are NOT independent. Nevertheless, Linearity of Expectation applies:

$$\boldsymbol{E}[X] = \boldsymbol{E}[\mathrm{X}_1] + \boldsymbol{E}[\mathrm{X}_2] + \cdots + \boldsymbol{E}[X_n] = n\boldsymbol{E}[X_i] = n \cdot \frac{1}{n} = 1.$$

regardless of n

Coupon Collector

There are n coupons we're trying to collect.
Each draw we get a random coupon
(sampling with replacement).

Let D = number of draws until we've collected all the coupons.

Q: What is $\boldsymbol{E}[D]$?

Coupon Collector

There are n coupons we're trying to collect. Each draw we get a random coupon (sampling with replacement).

Let D = number of draws until we've collected all the coupons.

Q: What is $\boldsymbol{E}[D]$?

Idea: $D = D_1 + D_2 + D_3 + \cdots + D_n$

Q: But what do the D_i represent?

What's wrong with letting D_i represent number of draws to get coupon i ?

Coupon Collector

There are n coupons we're trying to collect.
Each draw we get a random coupon
(sampling with replacement).

Let D = number of draws until we've collected all the coupons.

Q: What is $\boldsymbol{E}[D]$?

D_i = number of draws needed to get ith distinct coupon, given already have $i-1$ distinct coupons

Idea: $D = D_1 + D_2 + D_3 + \cdots + D_n$

draws to get 1st distinct coupon

Additional draws needed to reach 2 distinct coupons

Additional draws needed to reach 3 distinct coupons

Coupon Collector

There are n coupons we're trying to collect.
Each draw we get a random coupon
(sampling with replacement).

Let D = number of draws until we've collected all the coupons.

Q: What is $\boldsymbol{E}[D]$?

D_i = number of draws needed to get ith distinct coupon, given already have $i-1$ distinct coupons

Idea: $D = D_1 + D_2 + D_3 + \cdots + D_n$

draws to get 1st distinct coupon

Additional draws needed to reach 2 distinct coupons

Additional draws needed to reach 3 distinct coupons

$D_1 \sim Geometric(1)$

$D_2 \sim Geometric\left(\frac{n-1}{n}\right)$

$D_3 \sim Geometric\left(\frac{n-2}{n}\right)$

$D_n \sim Geometric\left(\frac{1}{n}\right)$

Coupon Collector

There are n coupons we're trying to collect.
Each draw we get a random coupon
(sampling with replacement).

Let $D =$ number of draws until we've collected all the coupons.

Q: What is $\boldsymbol{E}[D]$?

Idea: $D = D_1 + D_2 + D_3 + \cdots + D_n$

$$\boldsymbol{E}[D] = \boldsymbol{E}[D_1] + \boldsymbol{E}[D_2] + \boldsymbol{E}[D_3] + \cdots + \boldsymbol{E}[D_n]$$

$$= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + n$$

$$= n \cdot \left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + 1 \right)$$

$$D_1 \sim Geometric(1)$$

$$D_2 \sim Geometric\left(\frac{n-1}{n}\right)$$

$$D_3 \sim Geometric\left(\frac{n-2}{n}\right)$$

$$D_n \sim Geometric\left(\frac{1}{n}\right)$$

Coupon Collector

There are n coupons we're trying to collect.
Each draw we get a random coupon
(sampling with replacement).

Let D = number of draws until we've collected all the coupons.

Q: What is $\boldsymbol{E}[D]$?

Idea: $D = D_1 + D_2 + D_3 + \cdots + D_n$

$$\boldsymbol{E}[D] = \boldsymbol{E}[D_1] + \boldsymbol{E}[D_2] + \boldsymbol{E}[D_3] + \cdots + \boldsymbol{E}[D_n]$$

$$= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + n$$

$$= n \cdot \left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + 1 \right) = n \cdot H_n \approx n \ln(n)$$

Conditional p.m.f.

We often want the expected value of a r.v. X conditioned on some event, A, e.g.

$$\boldsymbol{E}[\text{Price of hotel room} \mid \text{Month is March}]$$

To define $\boldsymbol{E}[X|A]$ we will need to define a conditional p.m.f., $p_{X|A}(x)$.

Defn 4.14: Let X be a discrete r.v. with p.m.f. $p_X(x)$.
Let A be an event s.t. $\boldsymbol{P}\{A\} > 0$.
Then $p_{X|A}(x)$ is the **conditional p.m.f. of X given event A** where:

$$p_{X|A}(x) = \boldsymbol{P}\{X = x \mid A\} = \frac{\boldsymbol{P}\{(X = x) \cap A\}}{\boldsymbol{P}\{A\}}$$

Conditioning on an Event

Let r.v. X denote the size of a job:

$$X = \begin{cases} 1 & \text{w.p. } 0.1 \\ 2 & \text{w.p. } 0.2 \\ 3 & \text{w.p. } 0.3 \\ 4 & \text{w.p. } 0.2 \\ 5 & \text{w.p. } 0.2 \end{cases}$$

Let A denote the event that the job is “small,” meaning its size is ≤ 3.

Q: What is $p_{X|A}(1)$? How does this compare with $p_X(1)$?

Conditioning on an Event

Let r.v. X denote the size of a job:

$$X = \begin{cases} 1 & \text{w.p. } 0.1 \\ 2 & \text{w.p. } 0.2 \\ 3 & \text{w.p. } 0.3 \\ 4 & \text{w.p. } 0.2 \\ 5 & \text{w.p. } 0.2 \end{cases}$$

Let A denote the event that the job is "small," meaning its size is ≤ 3.

Q: What is $p_{X|A}(1)$? How does this compare with $p_X(1)$?

A: $p_{X|A}(1) = \boldsymbol{P}\{X = 1 \mid A\} = \dfrac{\boldsymbol{P}\{X = 1 \ \& \ A\}}{\boldsymbol{P}\{A\}} = \dfrac{\boldsymbol{P}\{X = 1\}}{\boldsymbol{P}\{A\}} = \dfrac{\frac{1}{10}}{\frac{6}{10}} = \dfrac{1}{6}$

Conditioning on an Event

Let r.v. X denote the size of a job:

$$X = \begin{cases} \boxed{\begin{array}{ll} 1 & \text{w.p. } 0.1 \\ 2 & \text{w.p. } 0.2 \\ 3 & \text{w.p. } 0.3 \end{array}} \\ 4 \quad \text{w.p. } 0.2 \\ 5 \quad \text{w.p. } 0.2 \end{cases}$$

Let A denote the event that the job is "small," meaning its size is ≤ 3.

Q: What is $p_{X|A}(x)$ if $x \notin A$? **Answer:** 0

Lemma 4.16: A conditional p.m.f. is a p.m.f., i.e.,

$$\sum_{x} p_{X|A}(x) = \sum_{x \in A} p_{X|A}(x) = 1$$

Example: Conditioning on an Event

Table shows $p_{X,Y}(x,y)$

	$X=0$	$X=1$	$X=2$
$Y=0$	1/6	1/8	0
$Y=1$	1/8	1/6	1/8
$Y=2$	0	1/6	1/8

Q: What is $p_{X|Y=2}(1)$?

A: $p_{X|Y=2}(1) = \boldsymbol{P}\{X=1 \mid Y=2\} = \dfrac{\boldsymbol{P}\{X=1\ \&\ Y=2\}}{\boldsymbol{P}\{Y=2\}} = \dfrac{\frac{1}{6}}{\frac{1}{6}+\frac{1}{8}} = \dfrac{4}{7}$

Conditional Expectation

The *conditional* expectation, $\boldsymbol{E}[X|A]$, is based on the *conditional* p.m.f., $p_{X|A}(x)$.

Defn: Let X be a discrete r.v.
The **conditional expectation of $\boldsymbol{X}$ given event $\boldsymbol{A}$** is defined as:

$$\boldsymbol{E}[X|A] = \sum_x x \cdot p_{X|A}(x) = \sum_x x \cdot \frac{\boldsymbol{P}\{(X = x) \cap A\}}{\boldsymbol{P}\{A\}}$$

Conditional Expectation Example

Let r.v. X denote the size of a job:

$$X = \begin{cases} \boxed{\begin{matrix} 1 & \text{w.p. } 0.1 \\ 2 & \text{w.p. } 0.2 \\ 3 & \text{w.p. } 0.3 \end{matrix}} \\ 4 \quad \text{w.p. } 0.2 \\ 5 \quad \text{w.p. } 0.2 \end{cases}$$

Let A denote the event that the job is “small,” meaning its size is ≤ 3.

Q: What is $\boldsymbol{E}[X|A]$?

Conditional Expectation Example

Let r.v. X denote the size of a job:

$$X = \begin{cases} \boxed{\begin{array}{ll} 1 & \text{w.p. } 0.1 \\ 2 & \text{w.p. } 0.2 \\ 3 & \text{w.p. } 0.3 \end{array}} \\ \begin{array}{ll} 4 & \text{w.p. } 0.2 \\ 5 & \text{w.p. } 0.2 \end{array} \end{cases}$$

Let A denote the event that the job is "small," meaning its size is ≤ 3.

Q: What is $\boldsymbol{E}[X|A]$?

A: $$\boldsymbol{E}[X|A] = 1 \cdot p_{X|A}(1) + 2 \cdot p_{X|A}(2) + 3 \cdot p_{X|A}(3)$$
$$= 1 \cdot \frac{1}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{3}{6} = \frac{14}{6}$$

Conditional Expectation Example

Table shows $p_{X,Y}(x,y)$

	$X = 0$	$X = 1$	$X = 2$
$Y = 0$	1/6	1/8	0
$Y = 1$	1/8	1/6	1/8
$Y = 2$	0	1/6	1/8

Q: What $\boldsymbol{E}[X \mid Y = 2]$?

Conditional Expectation Example

Table shows $p_{X,Y}(x,y)$

	$X = 0$	$X = 1$	$X = 2$
$Y = 0$	$1/6$	$1/8$	0
$Y = 1$	$1/8$	$1/6$	$1/8$
$Y = 2$	0	$1/6$	$1/8$

Q: What $\boldsymbol{E}[X \mid Y = 2]$?

A: $$\boldsymbol{E}[X \mid Y = 2] = 0 \cdot p_{X|Y=2}(0) + 1 \cdot p_{X|Y=2}(1) + 2 \cdot p_{X|Y=2}(2)$$

$$= 1 \cdot \frac{4}{7} + 2 \cdot \frac{3}{7} = \frac{10}{7}$$

Computing Expectations via Conditioning

Theorem 4.22**:** Let X be a discrete r.v.
Let events $F_1, F_2, \ldots, F_n$ partition the space Ω. Then

$$\boldsymbol{E}[X] = \sum_{i=1}^{n} \boldsymbol{E}[X \mid F_i] \cdot \boldsymbol{P}\{F_i\}$$

For a discrete r.v. Y :

$$\boldsymbol{E}[X] = \sum_{y} \boldsymbol{E}[X \mid Y = y] \cdot \boldsymbol{P}\{Y = y\}$$

Expected Value of Geometric, Revisited

$X \sim Geometric(p)$. Derive $\boldsymbol{E}[X]$ by conditioning.

Q: What should we condition on?

Expected Value of Geometric, Revisited

$X \sim Geometric(p)$. Derive $\boldsymbol{E}[X]$ by conditioning.

Q: What should we condition on?

A: Condition on the value of the first flip, Y

$\boldsymbol{E}[X] = \boldsymbol{E}[X \mid Y = 1] \cdot \boldsymbol{P}\{Y = 1\} + \boldsymbol{E}[X$ … $\boldsymbol{P}\{Y = 0\}$

What is this?

$$= \boldsymbol{E}[X \mid Y = 1] \cdot p + \boxed{\boldsymbol{E}[X \mid Y = 0]} \cdot (1 - p)$$

Expected Value of Geometric, Revisited

$X \sim Geometric(p)$. Derive $\boldsymbol{E}[X]$ by conditioning.

Q: What should we condition on?

A: Condition on the value of the first flip, Y.

$$\boldsymbol{E}[X] = \boldsymbol{E}[X \mid Y = 1] \cdot \boldsymbol{P}\{Y = 1\} + \boldsymbol{E}[X \mid Y = 0] \cdot \boldsymbol{P}\{Y = 0\}$$

$$= \boldsymbol{E}[X \mid Y = 1] \cdot p + \boxed{\boldsymbol{E}[X \mid Y = 0]} \cdot (1 - p)$$

$$= 1 \cdot p + (1 + \boldsymbol{E}[X]) \cdot (1 - p)$$

$$\Rightarrow \boldsymbol{E}[X] = \frac{1}{p}$$

Simpson's Paradox

Consider two treatments for kidney stones**:** **Treatment A** and **Treatment B**

- **Treatment A** is more effective on small kidney stones
- **Treatment A** is also more effective on large kidney stones

But if we ignore the type of stones, **Treatment B** is more effective!

Simpson's Paradox

Q: How is this possible?

	Treatment A	Treatment B
small stones	**90% effective**	**80% effective**
large stones	**60% effective**	**50% effective**
aggregate mix	**63% effective**	**77% effective**

Simpson's Paradox

	Treatment A	Treatment B
small stones	**90% effective** (successful on 90 out of 100)	**80% effective** (successful on 800 out of 1000)
large stones	**60% effective** (successful on 600 out of 1000)	**50% effective** (successful on 50 out of 100)
aggregate mix	**63% effective** (successful on 690 out of 1100)	**77% effective** (successful on 850 out of 1100)

	Treatment A	Treatment B
small stones	**90% effective** (successful on 90 out of 100)	**80% effective** (successful on 800 out of 1000)
large stones	**60% effective** (successful on 600 out of 1000)	**50% effective** (successful on 50 out of 100)
aggregate mix	**63% effective** (successful on 690 out of 1100)	**77% effective** (successful on 850 out of 1100)

Is treatment B better?

- No! Treatment A is better on both small stones and on large stones. It is the better treatment!
- But because A is better, it is given more “hard cases” – the large stone cases – and hence has lower average scores.