THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 321: Introduction to Probability and Statistics

Lecture 9: Discrete random variables

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

www.cs.cmu.edu/∼psarkar/teaching

The Geometric random variable

- The Bernoulli PMF describes the probability of success/failure in a single trial.
- The Binomial PMF describes the probability of k successes out of n trials.
- Sometimes we may also be interested in doing trials until we see a success.

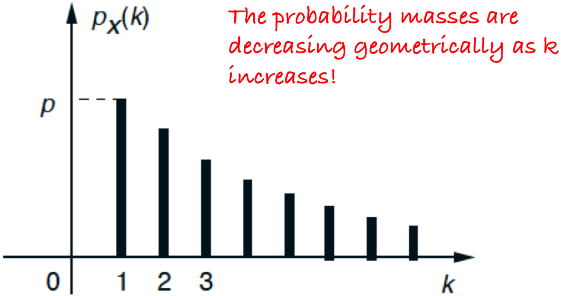
- Alice resolves to keep buying lottery tickets until he wins a hundred million dollars. She is interested in the random variable "number of lottery tickets bought until he wins the 100M$ lottery".
- Annie is trying to catch a taxi. How many occupied taxis will drive pass before she finds one that is taking passengers?

- The number of trials required to get a single success is a **Geometric Random Variable**

The geometric random variable

We repeatedly toss a biased coin ($P(\{H\}) = p$). The geometric random variable is the number X of tosses to get a head.

- X can take any integral value.
- $P(X = k) = P(\{\underbrace{TT \ldots T}_{k-1} H\}) = (1-p)^{k-1} p$.
- $\sum_k P(X = k) = 1$ (why?)



The geometric random variable

What is $P(X \geq k)$? What is $P(X > k)$?

The geometric random variable

What is $P(X \geq k)$? What is $P(X > k)$?

- $P(X \geq k) = \sum_{i=k}^{\infty} p(1-p)^{i-1} = (1-p)^{k-1}$

The geometric random variable

What is $P(X \geq k)$? What is $P(X > k)$?

- $P(X \geq k) = \sum_{i=k}^{\infty} p(1-p)^{i-1} = (1-p)^{k-1}$
- Intuitively, this is asking for the probability that the first $k-1$ tosses are tails.
- This probability is $P(X \geq k) = (1-p)^{k-1}$

The geometric random variable

What is $P(X \geq k)$? What is $P(X > k)$?

- $P(X \geq k) = \sum_{i=k}^{\infty} p(1-p)^{i-1} = (1-p)^{k-1}$
- Intuitively, this is asking for the probability that the first $k-1$ tosses are tails.
- This probability is $P(X \geq k) = (1-p)^{k-1}$
- $X > k$ is the event that $X \geq k+1$, and so $P(X > k) = (1-p)^{k}$

The memoryless property

What is $P(X = a + b | X > a)$?

The memoryless property

What is $P(X = a + b | X > a)$?

$$P(X = a + b | X > a) = \frac{P(X = a + b)}{P(X > a)}$$

- $$= \frac{p(1-p)^{a+b-1}}{(1-p)^a}$$
$$= p(1-p)^{b-1} = P(X = b)$$
- You forgot about $X > a$ and started the clock afresh!

The memoryless property

What is $P(X > a + b | X > a)$?

The memoryless property

What is $P(X > a + b|X > a)$?

$$P(X > a + b|X > a) = \frac{P(X > a + b)}{P(X > a)}$$

- $$= \frac{(1-p)^{a+b}}{(1-p)^a} = (1-p)^b$$
  $$= P(X > b)$$
- You forgot about $X > a$ and started the clock afresh!

The memoryless property

What is $P(X \leq a + b | X > a)$?

The memoryless property

What is $P(X \leq a+b | X > a)$?

- $$\begin{aligned} P(X \leq a+b | X > a) &= \frac{P(a < X \leq a+b)}{P(X > a)} \\ &= \frac{P(X > a) - P(X > a+b)}{(1-p)^a} \\ &= \frac{(1-p)^a - (1-p)^{a+b}}{(1-p)^a} \\ &= 1 - (1-p)^b = P(X \leq b) \end{aligned}$$
- You forgot about $X > a$ and started the clock afresh!

The Poisson random variable

I have a book with 10000 words. Probability that a word has a typo is 1/1000. I am interested in how many misprints can be there on average? So a Poisson often shows up when you have a Binomial random variable with very large n and very small p but $n \times p$ is moderate. Here $np = 10$.

Our random variable might be:

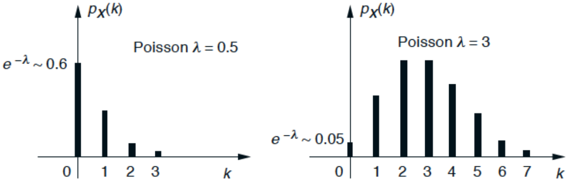
- The number of car crashes in a given day.
- The number of buses arriving within a given time period.
- The number of mutations on a strand of DNA.

We can describe such situations using a **Poisson random variable**.

The Poisson random variable

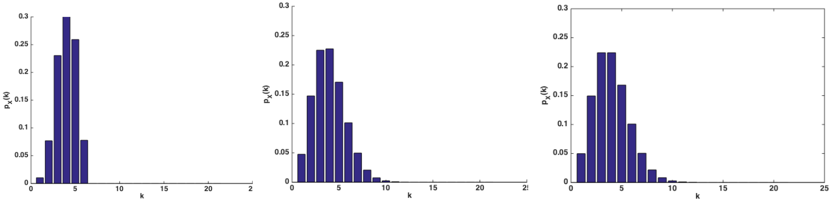
- A Poisson random variable takes non-negative integers as values. It has a nonnegative parameter λ.
- $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, for $k = 0, 1, 2 \ldots$.
- $\sum_{k=0}^{\infty} P(X = k) = e^{-\lambda}(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots) = 1$. (Exponential series!)

The PMF is monotonically decreasing for λ=0.5



The PMF is increasing and then decreasing for λ=3

Poisson random variable



Binomial(5,0.6) Binomial(100,0.03) Poisson(3)

- When n is very large and p is very small, a binomial random variable can be well approximated by a Poisson with $\lambda = np$.
- In the above figure we increased n and decreased p so that $np = 3$.
- See how close the PMF's of the Binomial(100,0.03) and Poisson(3) are!
- More formally, we see that $\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{e^{-\lambda}\lambda^k}{k!}$ when n is large, k is fixed, and p is small and $\lambda = np$.

Example

Assume that on a given day 1000 cars are out in Austin. On average, three out of 1000 cars run into a traffic accident per day.

1. What is the probability that we see at least two accidents in a day?

4. If you know there is at least one accident, what is the probability that the total number of accidents is at least two?

Example

Assume that on a given day 1000 cars are out in Austin. On average, three out of 1000 cars run into a traffic accident per day.

1. What is the probability that we see at least two accidents in a day?

2. Use poisson approximation!

4. If you know there is at least one accident, what is the probability that the total number of accidents is at least two?

Example

Assume that on a given day 1000 cars are out in Austin. On average, three out of 1000 cars run into a traffic accident per day.

1. What is the probability that we see at least two accidents in a day?
2. Use poisson approximation!
3. $P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-3}(1 + 3) = 0.8$
4. If you know there is at least one accident, what is the probability that the total number of accidents is at least two?

Example

Assume that on a given day 1000 cars are out in Austin. On average, three out of 1000 cars run into a traffic accident per day.

1. What is the probability that we see at least two accidents in a day?
2. Use poisson approximation!
3. $P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-3}(1 + 3) = 0.8$
4. If you know there is at least one accident, what is the probability that the total number of accidents is at least two?
5. $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-3} = 0.950$.
   $P(X \geq 2 | X \geq 1) = P(X \geq 2)/P(X \geq 1) = 0.8/0.950 = 0.84$

Mean

You want to calculate average grade points from hw1. You know that 20 students got 30/30, 30 students got 25/30, and 50 students got 20/30. Whats the average?

- The average grade point is

$$\frac{30 \times 20 + 25 \times 30 + 20 \times 50}{100} = 30 \times 0.2 + 25 \times 0.3 + 20 \times 0.5$$

- Let X be a random variable which represents grade points of hw1.
- How will you calculate $P(X = 30)$?
  - See how many out of 100 students got 30 out of 30 points.
  - $P(X = 30) \approx 0.2$
  - $P(X = 25) \approx 0.3$
  - $P(X = 20) \approx 0.5$
- So roughly speaking,
  average grade $\approx 30 \times P(X = 30) + 25 \times P(X = 25) + 20 \times P(X = 20)$

Expectation

We define the expected value (or expectation or mean) of a discrete random variable X by

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

- X is a Bernoulli random variable with the following PMF:

$$P(X = x) = \begin{cases} p & X = 1 \\ 1 - p & X = 0 \end{cases}$$

So $E[X] =$.

Expectation

We define the expected value (or expectation or mean) of a discrete random variable X by

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

- X is a Bernoulli random variable with the following PMF:

$$P(X = x) = \begin{cases} p & X = 1 \\ 1 - p & X = 0 \end{cases}$$

  So $E[X] = 1 \times p + 0 \times (1 - p) = p$.
- Expectation of a Bernoulli random variable is just the probability that it is one.
- You will also see notation like μ_X.

Expectation: example

You are tossing 4 fair coins independently. Let X denote the number of heads. What is $E[X]$?

- Any guesses? Well, on an average we should see about 2 coin tosses. No?
- Lets write down the PMF first.

Expectation: example

You are tossing 4 fair coins independently. Let X denote the number of heads. What is $E[X]$?

- Any guesses? Well, on an average we should see about 2 coin tosses. No?
- Lets write down the PMF first.
- $P(X = x) = \begin{cases} 1/2^4 & X = 0 \\ 4/2^4 & X = 1 \\ 6/2^4 & X = 2 \\ 4/2^4 & X = 3 \\ 1/2^4 & X = 4 \end{cases}$

Expectation: example

You are tossing 4 fair coins independently. Let X denote the number of heads. What is $E[X]$?

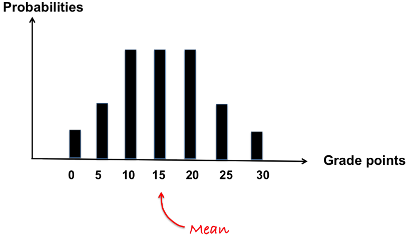
- Any guesses? Well, on an average we should see about 2 coin tosses. No?
- Lets write down the PMF first.
- $P(X=x)=\begin{cases}1/2^4 & X=0\\ 4/2^4 & X=1\\ 6/2^4 & X=2\\ 4/2^4 & X=3\\ 1/2^4 & X=4\end{cases}$
- So $E[X]=\frac{4}{2^4}+2\frac{6}{2^4}+3\frac{4}{2^4}+4\frac{1}{2^4}=\frac{32}{16}=2.$

Expectation of a function of a random variable

Lets say you want to compute $E[g(X)]$. Example, I know average temperature in Fahrenheit, but I now want it in Celsius.

- $E[g(X)] = \sum_x g(x)P(X = x)$.
- Follows from the definition of PMF of functions of random variables.
- Look at page 15 of Bersekas-Tsitsiklis and derive it at home!
- So $E[X^2] = \sum_x x^2 P(X = x)$. **Second moment of X**
- So $E[X^3] = \sum_x x^3 P(X = x)$. **Third moment of X**
- So $E[X^k] = \sum_x x^k P(X = x)$. k^{th} **moment of X**
- We are assuming "under the rugs" that all these expectations are well defined.

Expectation



- Think of expectation as center of gravity of the PMF or a representative value of X.
- How about the spread of the distribution? Is there a number for it?

Variance

Often, you may want to know the spread or variation of the grade points for homework1.

- If everyone got the same grade point, then variation is?
- If there is high variation, then we know that many students got grade points very different from the average grade point in class.
- Formally we measure this using variance of a random variable X.
- $\mathrm{var}(X) = E[(X - E[X])^2]$
- The standard deviation of X is given by $\sigma_X = \sqrt{\mathrm{var} X}$.
- Its easier to think about σ_X, since its on the same scale.
- The grade points have average 20 out of 30 with a standard deviation of 5 grade points. Roughly this means, most of the students have grade points within $[20 - 5, 20 + 5]$.

Variance

Often, you may want to know the spread or variation of the grade points for homework1.

- If everyone got the same grade point, then variation is? 0
- If there is high variation, then we know that many students got grade points very different from the average grade point in class.
- Formally we measure this using variance of a random variable X.
- $\mathrm{var}(X) = E[(X - E[X])^2]$
- The standard deviation of X is given by $\sigma_X = \sqrt{\mathrm{var} X}$.
- Its easier to think about σ_X, since its on the same scale.
- The grade points have average 20 out of 30 with a standard deviation of 5 grade points. Roughly this means, most of the students have grade points within $[20 - 5, 20 + 5]$.

Computing the variance

- $\mathrm{var}(X) = E[(X - E[X])^2] = \sum_x (x - E[X])^2 \, P(X = x)$
- Always remember! $E[X]$ or $E[g(X)]$ **do not depend on any particular value of** x. You can treat it as a constant. It only depends on the PMF of X.

- This can actually be made simpler.
- $\mathrm{var}(X) = E[X]^2 - (E[X])^2$.

- So you can calculate $E[X^2]$ (second moment) and then subtract the square of $E[X]$ to get the variance!

A tiny bit of algebra

$$\mathrm{var}(X) =$$

A tiny bit of algebra

$$\mathrm{var}(X) =$$

$$\sum_x (x - E[X])^2 \, P(X = x) = \sum_x \left(x^2 + (E[X])^2 - 2xE[X]\right) P(X = x)$$

A tiny bit of algebra

$$\mathrm{var}(X) =$$

$$\sum_x (x - E[X])^2 \, P(X = x) = \sum_x \left(x^2 + (E[X])^2 - 2xE[X]\right) P(X = x)$$

$$= \sum_x x^2 P(X = x) + \sum_x (E[X])^2 P(X = x) - 2 \sum_x x E[X] P(X = x)$$

A tiny bit of algebra

$$\mathrm{var}(X) =$$

$$\sum_x (x - E[X])^2 \, P(X = x) = \sum_x \left(x^2 + (E[X])^2 - 2xE[X]\right) P(X = x)$$

$$= \sum_x x^2 P(X = x) + \sum_x (E[X])^2 P(X = x) - 2 \sum_x xE[X] P(X = x)$$

$$= \sum_x x^2 P(X = x) + \underbrace{(E[X])^2 \sum_x P(X = x)}_{(E[X])^2 \text{ does not depend on } x} - 2E[X] \underbrace{\sum_x xP(X = x)}_{\text{This is } E[X]}$$

A tiny bit of algebra

$\mathrm{var}(X) =$

$$\sum_x (x - E[X])^2 \, P(X = x) = \sum_x \left(x^2 + (E[X])^2 - 2xE[X]\right) P(X = x)$$

$$= \sum_x x^2 P(X = x) + \sum_x (E[X])^2 P(X = x) - 2 \sum_x xE[X]P(X = x)$$

$$= \sum_x x^2 P(X = x) + \underbrace{(E[X])^2 \sum_x P(X = x)}_{(E[X])^2 \text{ does not depend on } x} \quad -2E[X] \underbrace{\sum_x xP(X = x)}_{\text{This is } E[X]}$$

$$= \sum_x x^2 P(X = x) + (E[X])^2 - 2(E[X])^2 = E[X^2] - (E[X])^2$$

Some simple rules– Expectation

Say you are looking at a linear function (or transformation) of your random variable X.

- $Y = aX + b$. Remember celsius to fahrenheit conversions? They are linear too!
- $E[Y] = E[aX + b] = aE[X] + b$, as simple as that! why?
  - $$\begin{aligned} E[aX + b] &= \sum_x (ax + b)P(X = x) \\ &= a\sum_x xP(X = x) + b\sum_x P(X = x) \end{aligned}$$

Some simple rules– Expectation

Say you are looking at a linear function (or transformation) of your random variable X.

- $Y = aX + b$. Remember celsius to fahrenheit conversions? They are linear too!
- $E[Y] = E[aX + b] = aE[X] + b$, as simple as that! why?
  - $$\begin{aligned} E[aX + b] &= \sum_x (ax + b)P(X = x) \\ &= a\sum_x xP(X = x) + b\sum_x P(X = x) \\ &= aE[X] + b \end{aligned}$$

Some simple rules– Expectation

How about $E[Y]$ for $Y = aX^2 + bX + c$?

- $E[Y] = E[aX^2 + bX + c] = aE[X^2] + bE[X] + c$, as simple as that! why?
  - $E[aX^2 + bX + c] = \sum_x (ax^2 + bx + c)P(X = x)$

    $= a\sum_x x^2 P(X = x) + b\sum_x xP(X = x) + c\sum_x P(X = x)$

Some simple rules– Expectation

How about $E[Y]$ for $Y = aX^2 + bX + c$?

- $E[Y] = E[aX^2 + bX + c] = aE[X^2] + bE[X] + c$, as simple as that! why?
  - $E[aX^2 + bX + c] = \sum_x (ax^2 + bx + c)P(X = x)$

$$= a\sum_x x^2 P(X = x) + b\sum_x xP(X = x) + c\sum_x P(X = x)$$

$$= aE[X^2] + bE[X] + c$$

Some simple rules– Expectation

How about $E[Y]$ for $Y = aX^2 + bX + c$?

- $E[Y] = E[aX^2 + bX + c] = aE[X^2] + bE[X] + c$, as simple as that! why?
    - $$E[aX^2 + bX + c] = \sum_x (ax^2 + bx + c)P(X = x)$$
      $$= a\sum_x x^2 P(X = x) + b\sum_x xP(X = x) + c\sum_x P(X = x)$$
      $$= aE[X^2] + bE[X] + c$$

- $Y = aX^3 + bX^2 + cX + d$. Can you guess what $E[Y]$ is?

Some simple rules– Expectation

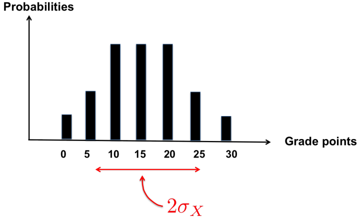
How about $E[Y]$ for $Y = aX^2 + bX + c$?

- $E[Y] = E[aX^2 + bX + c] = aE[X^2] + bE[X] + c$, as simple as that! why?
  - $$E[aX^2 + bX + c] = \sum_x (ax^2 + bx + c)P(X = x)$$
    $$= a\sum_x x^2 P(X = x) + b\sum_x xP(X = x) + c\sum_x P(X = x)$$
    $$= aE[X^2] + bE[X] + c$$
- $Y = aX^3 + bX^2 + cX + d$. Can you guess what $E[Y]$ is?

  $E[Y] = aE[X^3] + bE[X^2] + cE[X] + d$.

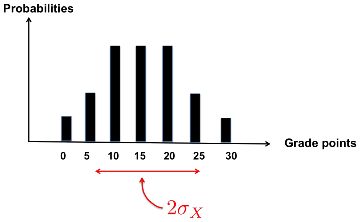
Some simple rules– Variance

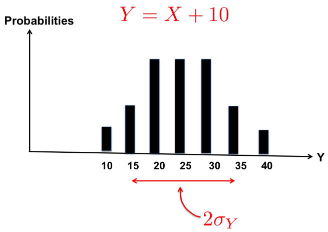
Let $Y = X + b$. What is $\text{var}(Y)$?



Some simple rules– Variance

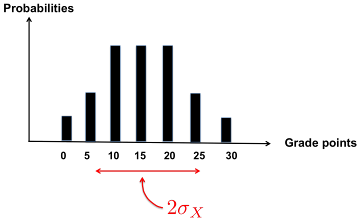
Let $Y = X + b$. What is $\mathrm{var}(Y)$?

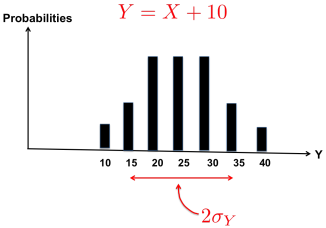




Some simple rules– Variance

Let $Y = X + b$. What is $\mathrm{var}(Y)$?





- Intuitively? Well you are just shifting everything by the same number.
- So? the spread of the numbers should stay the same!
- Prove it at home.

Some simple rules– Variance

Let $Y = X + b$. What is $\mathrm{var}(Y)$?

Some simple rules– Variance

Let $Y = X + b$. What is $\mathrm{var}(Y)$?

Some simple rules– Variance

Let $Y = X + b$. What is $\mathrm{var}(Y)$?

- Proof:
- $\mathrm{var}(X + b) = E\left[(X + b)^2\right] - (E[X + b])^2$

Some simple rules– Variance

Let $Y = X + b$. What is $\mathrm{var}(Y)$?

- Proof:
- $\mathrm{var}(X + b) = E\left[(X + b)^2\right] - (E[X + b])^2$
  $= E\left[X^2 + 2bX + b^2\right] - (E[X] + b)^2$

Some simple rules– Variance

Let $Y = X + b$. What is $\mathrm{var}(Y)$?

- Proof:
- $$\begin{aligned} \mathrm{var}(X+b) &= E\left[(X+b)^2\right] - (E[X+b])^2 \\ &= E\left[X^2 + 2bX + b^2\right] - (E[X]+b)^2 \\ &= E[X^2] + 2bE[X] + b^2 - ((E[X])^2 + 2bE[X] + b^2) \end{aligned}$$

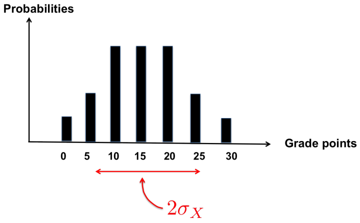
Some simple rules– Variance

Let $Y = X + b$. What is $\mathrm{var}(Y)$?

- Proof:
- $$\begin{aligned}\mathrm{var}(X+b) &= E\left[(X+b)^2\right] - (E[X+b])^2 \\ &= E\left[X^2 + 2bX + b^2\right] - (E[X]+b)^2 \\ &= E[X^2] + 2bE[X] + b^2 - ((E[X])^2 + 2bE[X] + b^2) \\ &= E[X^2] - (E[X])^2 = \mathrm{var}(X)\end{aligned}$$

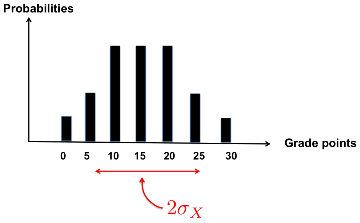
Some simple rules– Variance

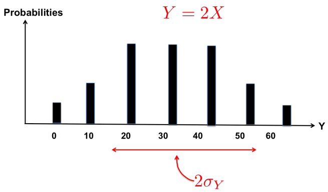
Let $Y = aX$. What is $\mathrm{var}(Y)$?



Some simple rules– Variance

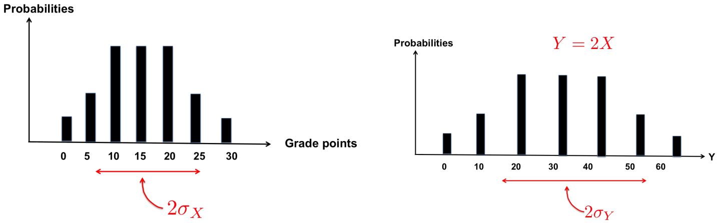
Let $Y = aX$. What is $\mathrm{var}(Y)$?





Some simple rules– Variance

Let $Y = aX$. What is $\mathrm{var}(Y)$?



- Intuitively? Well you are just scaling everything by the same number.
- So? the spread should increase if $a > 1$!

Some simple rules– Variance

Let $Y = aX$. Turns out $\mathrm{var}(Y) = a^2 \mathrm{var}(X)$.

Some simple rules– Variance

Let $Y = aX$. Turns out $\mathrm{var}(Y) = a^2 \mathrm{var}(X)$.

- Proof:
-

Some simple rules– Variance

Let $Y = aX$. Turns out $\mathrm{var}(Y) = a^2\mathrm{var}(X)$.

- Proof:
- $\mathrm{var}(aX) = E\left[(aX)^2\right] - (E[aX])^2$

Some simple rules– Variance

Let $Y = aX$. Turns out $\mathrm{var}(Y) = a^2\mathrm{var}(X)$.

- Proof:
- $$\begin{aligned}\mathrm{var}(aX) &= E\left[(aX)^2\right] - (E[aX])^2 \\ &= E\left[a^2X^2\right] - (aE[X])^2\end{aligned}$$

Some simple rules– Variance

Let $Y = aX$. Turns out $\mathrm{var}(Y) = a^2\mathrm{var}(X)$.

- Proof:
- $$\begin{aligned}\mathrm{var}(aX) &= E\left[(aX)^2\right] - (E[aX])^2 \\ &= E\left[a^2X^2\right] - (aE[X])^2 \\ &= a^2E\left[X^2\right] - a^2(E[X])^2\end{aligned}$$

Some simple rules– Variance

Let $Y = aX$. Turns out $\mathrm{var}(Y) = a^2 \mathrm{var}(X)$.

- Proof:
- $$\begin{aligned} \mathrm{var}(aX) &= E\left[(aX)^2\right] - (E[aX])^2 \\ &= E\left[a^2 X^2\right] - (aE[X])^2 \\ &= a^2 E\left[X^2\right] - a^2 (E[X])^2 \\ &= a^2 (E[X^2] - (E[X])^2) = a^2 \mathrm{var}(X) \end{aligned}$$

Some simple rules– Variance

Let $Y = aX$. Turns out $\mathrm{var}(Y) = a^2\mathrm{var}(X)$.

- Proof:
- $$\begin{aligned}\mathrm{var}(aX) &= E\left[(aX)^2\right] - (E[aX])^2 \\ &= E\left[a^2X^2\right] - (aE[X])^2 \\ &= a^2E\left[X^2\right] - a^2(E[X])^2 \\ &= a^2(E[X^2] - (E[X])^2) = a^2\mathrm{var}(X)\end{aligned}$$
- In general we can show that $\mathrm{var}(aX + b) = a^2\mathrm{var}(X)$.

Mean and Variance of Bernoulli

X is a Bernoulli random variable wit $P(X = 1) = p$. We saw that $E[X] = p$. What is $\mathrm{var}(X)$?

- First lets get $E[X^2]$. This is

$$E[X^2] = (1^2 \times P(X = 1) + 0^2 \times P(X = 0)) = p$$

  We see that $E[X^2] = E[X]$. Is this surprising?

- Well, what is the PMF of X^2?
  - X^2 can take two values:

.

Mean and Variance of Bernoulli

X is a Bernoulli random variable wit $P(X = 1) = p$. We saw that $E[X] = p$. What is $\mathrm{var}(X)$?

- First lets get $E[X^2]$. This is

$$E[X^2] = (1^2 \times P(X = 1) + 0^2 \times P(X = 0)) = p$$

  We see that $E[X^2] = E[X]$. Is this surprising?

- Well, what is the PMF of X^2?
  - X^2 can take two values: 0 and 1
  - $P(X^2 = 1)$.

Mean and Variance of Bernoulli

X is a Bernoulli random variable wit $P(X = 1) = p$. We saw that $E[X] = p$. What is $\mathrm{var}(X)$?

- First lets get $E[X^2]$. This is

  $$E[X^2] = (1^2 \times P(X = 1) + 0^2 \times P(X = 0)) = p$$

  We see that $E[X^2] = E[X]$. Is this surprising?

- Well, what is the PMF of X^2?
  - X^2 can take two values: 0 and 1
  - $P(X^2 = 1) = P(X = 1) = p$.

.

Mean and Variance of Bernoulli

X is a Bernoulli random variable wit $P(X = 1) = p$. We saw that $E[X] = p$. What is $\mathrm{var}(X)$?

- First lets get $E[X^2]$. This is

  $$E[X^2] = (1^2 \times P(X = 1) + 0^2 \times P(X = 0)) = p$$

  We see that $E[X^2] = E[X]$. Is this surprising?

- Well, what is the PMF of X^2?
  - X^2 can take two values: 0 and 1
  - $P(X^2 = 1) = P(X = 1) = p$. $P(X^2 = 0) = P(X = 0) = 1 - p$.

Mean and Variance of Bernoulli

X is a Bernoulli random variable wit $P(X = 1) = p$. We saw that $E[X] = p$. What is $\mathrm{var}(X)$?

- First lets get $E[X^2]$. This is

$$E[X^2] = (1^2 \times P(X = 1) + 0^2 \times P(X = 0)) = p$$

  We see that $E[X^2] = E[X]$. Is this surprising?

- Well, what is the PMF of X^2?
  - X^2 can take two values: 0 and 1
  - $P(X^2 = 1) = P(X = 1) = p$. $P(X^2 = 0) = P(X = 0) = 1 - p$.
  - X and X^2 have identical PMF's! **They are identically distributed.**

Mean and Variance of Bernoulli

X is a Bernoulli random variable wit $P(X = 1) = p$. We saw that $E[X] = p$. What is $\mathrm{var}(X)$?

- First lets get $E[X^2]$. This is

  $$E[X^2] = (1^2 \times P(X = 1) + 0^2 \times P(X = 0)) = p$$

  We see that $E[X^2] = E[X]$. Is this surprising?

- Well, what is the PMF of X^2?
  - X^2 can take two values: 0 and 1
  - $P(X^2 = 1) = P(X = 1) = p$. $P(X^2 = 0) = P(X = 0) = 1 - p$.
  - X and X^2 have identical PMF's! **They are identically distributed.**

- $\mathrm{var}(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p)$.

Mean and Variance of a Binomial

Let $X \sim Bin(n, p)$.

- $E[X] = np$ and $\mathrm{var}(X) = np(1 - p)$.
- We will derive these in the next class.

Mean and Variance of a Poisson

X has a Poisson(λ) distribution. What is its mean and variance?

- One can use algebra to show that $E[X] = \lambda$ and also $\mathrm{var}(X) = \lambda$.
- How do you remember this?
- Hint: mean and variance of the Binomial approach that of a Poisson when n is large and p is small, such that $np \approx \lambda$? Anything yet?

Mean and variance of a geometric

- The PMF of a geometric distribution is $P(X = k) = (1 - p)^{k-1} p$.
  - $E[X] = 1/p$
  - $\mathrm{var}(X) = (1 - p)/p^2$
  - We will also prove this later.