CSCI 3022

intro to data science with probability & statistics

Lecture 11
February 19, 2018

1. Plinko!
2. Variance of discrete and continuous RVs

Department of Computer Science
UNIVERSITY OF COLORADO BOULDER

Dan Larremore

Stuff & Things

- **Homework 3** due next Friday, March 2. Suggested **milestones**:
  - Probs 1, and 2, done before the end of the week.
  - Probs 3 and 4 done next week
- **Midterm coming up** next week. Weds, Feb 28th, ~~7:30~~ 6:30-8:00 PM. Room TBD. Duane G130
- Midterm review *in class* next Monday, Feb 26.

Last time on CSCI 3022

- **Definition**: The expectation or expected value of a discrete random variable *X* that takes the values $a_1, a_2, \ldots$ and with PMF *p* is given by:

$$\mathrm{E}[X] = \sum_i a_i P(X = a_i) = \sum_i a_i p(a_i)$$

- **Definition**: The *expectation* or *expected value* of a continuous random variable *X* with PDF *f* is the number:

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$$

- **Change of Variables**: Let *X* be a RV and let $g : \mathcal{R} \rightarrow \mathcal{R}$ be a function

$$\mathrm{E}[g(X)] = \sum_i g(a_i) f(a_i) \qquad \mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

The Fergusons

- Maureen Ferguson is having her whole crazy extended family over for dinner. What a bunch! They are considering a random variable X:

$$P(X=-1)=\frac{1}{5}, \quad P(X=0)=\frac{1}{5}, \quad P(X=1)=\frac{3}{5}$$

maureen ferguson & family

a "raft"

1. Is X a discrete or continuous random variable?

2. Compute $E[X]$

The Fergusons

- Maureen Ferguson is having her whole crazy extended family over for dinner. What a bunch! They are considering a random variable X:

$$P(X = -1) = \frac{1}{5}, \quad P(X = 0) = \frac{1}{5}, \quad P(X = 1) = \frac{3}{5}$$

3. Tony Ferguson is sleepy from dinner, but considering a *new* random variable Y.
If $Y = X^2 + 2$, what is the probability distribution of Y?
Is it a PMF or a PDF?

tony ferguson

4. What is $E[Y]$?

Stacy.

- It is mid-November, and Stacy is floating in the bay, watching the pups learn to swim. A drizzle begins, and a bird on the shoreline contemplates flying South for the winter. Is it too early? No matter. Rain drops fall softly on Stacy's silky otter-coat at an average rate of 17 drops per minute. You watch a drop fall on Stacy's silky coat and check your watch. Let **X** be a random variable that represents the time you wait until the next drop lands (on Stacy's silky otter-coat). Which distribution does **X** follow?

A. Binomial
B. Geometric
C. Poisson
D. Uniform
E. Exponential
F. None of the Above

stacy

Plinko!

$31.5 K

https://www.youtube.com/watch?v=naUppHrHJpI

Plinko on paper

- Let X be the RV describing the winnings in each round of Plinko with n rows and a probability p of moving to the right off of each peg.
- **Question**: what distribution does X follow?

$\boxed{\text{Bin}(3, p) + 1}$

If indep, "AND" $\longrightarrow$ $\times$

If indep, "OR" $\longrightarrow$ $+$

$P(X=2) = 3\,(1-p)^2\,p^1$

orders

$\binom{3}{1}$

Bin(n, p)

$\boxed{\text{Binomial}}$

$P(X=1) = (1-p)^3$

$P(X=4) = p^3$

Plinko on paper

- Let X be the RV describing the winnings in each round of Plinko with n rows and a probability p of moving to the right off of each peg.
- **Question**: what is the expected value of X? What is $E[X]$, where $X \sim Bin(n,p)$

$$E[X] = \sum_i a_i \, p(a_i)$$

instead, use a trick!

Plinko on paper

$$E[Ber(p)] = \sum_i a_i \, p(a_i)$$
$$= 0(1-p) + 1(p) = p$$

- Let X be the RV describing the winnings in each round of Plinko with n rows and a probability p of moving to the right off of each peg.
- **Question**: what is the expected value of X?
- **Pro Tip**: expectation is a linear function! E[rX+s] = rE[X] + s

$$E[X+Y] = E[X] + E[Y]$$

$$E[X] = E[Y_1 + Y_2 + \dots + Y_n] = E[Y_1] + E[Y_2] + \dots + E[Y_n] = n\,E[Ber(p)] = \boxed{np}$$

X: Bin(n,p)

Y_i: each is Ber(p)

for example $E[5X + 2Y - e^{\pi}] = 5E[X] + 2E[Y] - e^{\pi}$

Plinko on paper

- Let X be the RV describing the winnings in each round of Plinko with n rows and a probability p of moving to the right off of each peg.
- **Question**: what is the variance of X?

draws (data)

Defined variance for $\{x_1, x_2, \ldots x_n\}$

$$\text{variance: } \frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \bar{x}\right)^2 \qquad \text{st. dev: } \sqrt{\text{variance}}$$

Plinko on paper

- Let X be the RV describing the winnings in each round of Plinko with n rows and a probability p of moving to the right off of each peg.
- **Question**: what is the variance of X?
- **Well ok but maybe first**: what is variance?

Variance

- Recall that the sample variance of data $x_1, x_2, \ldots, x_n$ is given by

$$\frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \bar{x}\right)^2$$

average of data

$$\text{average}\left(\left[\text{datum} - \text{average of data}\right]^2\right)$$

Variance

- **Definition**: the variance *Var(X)* of a random variable *X* is the number

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

$$E\left[\left(X - E[X]\right)^2\right]$$

Variance

- **Definition**: the variance *Var(X)* of a random variable *X* is the number

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

- **Definition**: the standard deviation of *X* is the sq. root of the variance: $\sqrt{\mathrm{Var}(X)}$
- How to compute:
  - First, compute $\mu = \mathrm{E}[X]$
  - Second, use the change-of-variables formula with $g(x) = (x - \mu)^2$ (From last time)

$$\mathrm{Var}(X) = \sum_i (a_i - \mu)^2 p(a_i) \qquad \mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Variance

- **Definition**: the variance *Var(X)* of a random variable *X* is the number

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

- **Even better**: Can we use E[rX+s] = rE[X]+s?

$$E[(X - E[X])^2] = E[X^2 - 2X\,E[X] + E[X]^2]$$

const const const

$$= E[X^2] - E[2X\,E[X]] + E[E[X]^2]$$

nice formula for variance

$$= E[X^2] - 2\underbrace{E[X]\,E[X]}_{E[X]^2} + E[X]^2 = \boxed{E[X^2] - E[X]^2}$$

Variance

- **Definition**: the variance *Var(X)* of a random variable *X* is the number

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

- **Even better**: Can we use E[rX+s] = rE[X]+s?

legible!

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

Binomial Variance

- Let X be the RV describing the winnings in each round of Plinko with n rows and a probability p of moving to the right off of each peg.
- **Question**: what is the variance of X, if $X \sim Bin(n, p)$

Bernoulli Variance

$E[g(X)] = \sum_i g(a_i)\, p(a_i)$

- Let X be the RV describing the winnings in each round of Plinko with n rows and a probability p of moving to the right off of each peg.
- **Question**: what is the variance of X, if $X \sim Bin(n, p)$
- How about: what is the variance of Y, if $Y \sim Ber(p)$

$Var(Y) = E[Y^2] - E[Y]^2$

$Y = 0$ with prob. $1-p$

$Y = 1$ with prob. p

1st: $E[Y^2] = \sum_i y_i^2\, p(y_i) = 0^2(1-p) + 1^2 p = p$

2nd: $E[Y]^2 = p^2$

$Var(Y) = p - p^2 = \boxed{p(1-p) = Var(Ber(p))}$

Independence & Variance

$X = Bin(n,p) = Y_1 + Y_2 + Y_3 + \ldots + Y_n$ where each Y_i is $Ber(p)$

- Let X be the RV describing the winnings in each round of Plinko with n rows and a probability p of moving to the right off of each peg.
- **Question**: what is the variance of X, if $X \sim Bin(n, p)$
- How about: what is the variance of Y, if $Y \sim Ber(p)$
- Fact: if X and Y are **independent**, then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

$$\mathrm{Var}(X) = \underbrace{\mathrm{Var}(Ber(p)) + \mathrm{Var}(Ber(p)) + \ldots}_{n \text{ of these.}}$$

$$= \boxed{n\, \underbrace{p(1-p)}_{\mathrm{Var}(Ber(p))}}$$

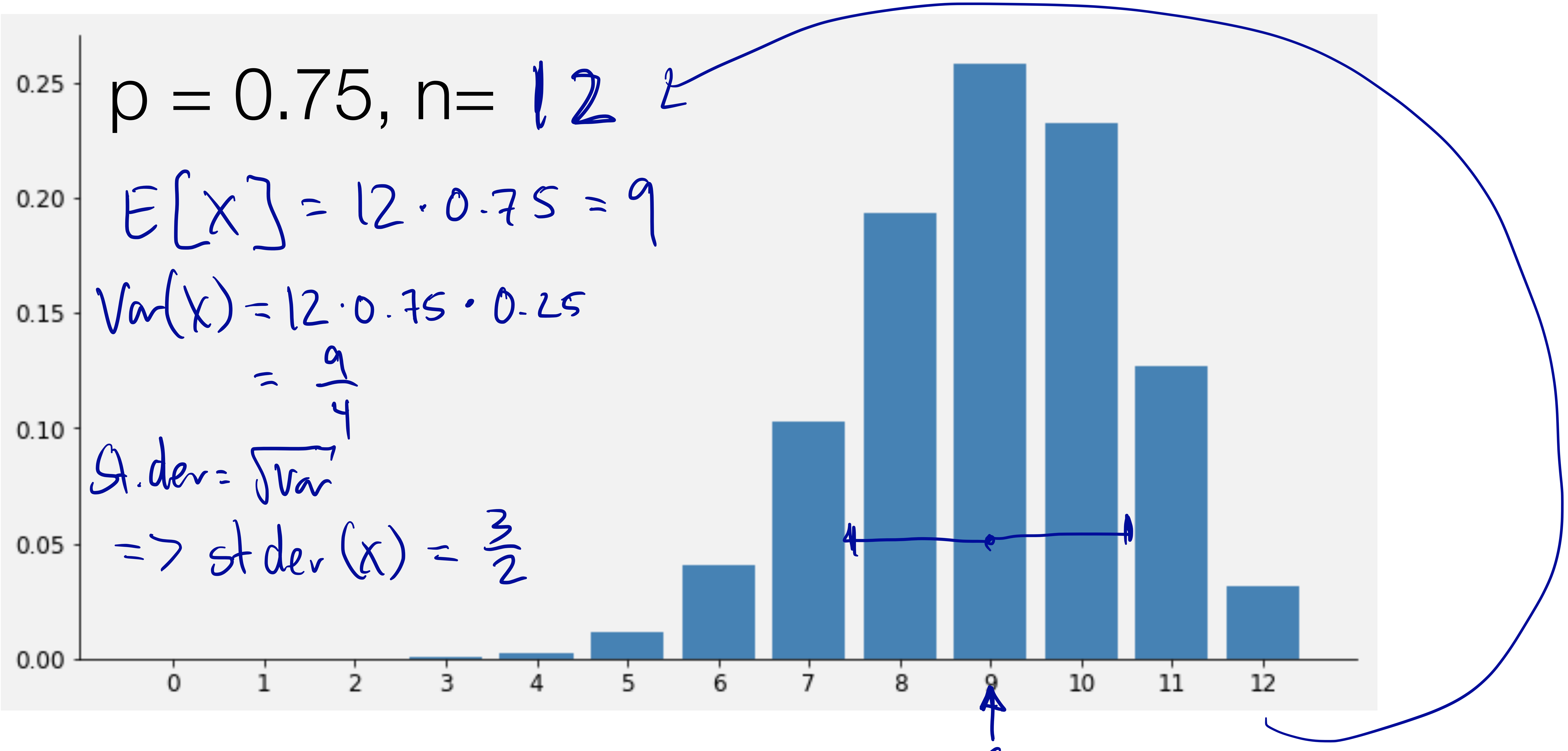
n of them

Binomial Variance

- Let *X* be the RV describing the winnings in each round of Plinko with *n* rows and a probability *p* of moving to the right off of each peg.
- **Question**: what is the variance of *X*, if $X \sim Bin(n, p)$
- How about: what is the variance of Y, if $Y \sim Ber(p)$
- Fact: if *X* and *Y* are **independent**, then $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$
- **Theorem**: Let $X \sim Bin(n, p)$. Then $\mathrm{E}[X] = np$ and $\mathrm{Var}(X) = np(1 - p)$

Plinko on paper

- **Theorem**: Let $X \sim Bin(n, p)$. Then $\mathrm{E}[X] = np$ and $\mathrm{Var}(X) = np(1-p)$



$E[X]$

Variance Facts!

- **Expectation** is linear: $E[rX+s] = rE[X] + s$
- What about **Variance**?

$$Var(X+s) = E\left[(X+s - E[X+s])^2\right]$$

def.

$$= E\left[(X+\cancel{s} - E[X] - \cancel{s})^2\right]$$

$$= E\left[(X - E[X])^2\right]$$

$$= Var(X)$$

$$Var(rX) = E[(rX)^2] - (E[rX])^2$$

other def'n!

$$= E[r^2X^2] - (rE[X])^2$$

$$= r^2E[X^2] - r^2E[X]^2$$

$$= r^2\left(E[X^2] - E[X]^2\right)$$

$$= r^2\,Var(X)$$

Variance Facts!

- **Expectation** is linear: $\mathrm{E}[rX + s] = r\mathrm{E}[X] + s$
- **Variance** is not linear: $\mathrm{Var}(rX + s) = r^2\mathrm{Var}(X)$

Mean & Variance: Uniform

- **Suppose** $X \sim U[\alpha, \beta]$. Find $\mathrm{E}[X]$ and $\mathrm{Var}(X)$.

$$E\left[\mathrm{Unif}[\alpha,\beta]\right] = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$= \int_{\alpha}^{\beta} x f(x)\,dx$$

$$= \int_{\alpha}^{\beta} x \frac{1}{\beta-\alpha}\,dx$$

$$\frac{x^2}{2}\Big|_{\alpha}^{\beta} \frac{1}{\beta-\alpha}$$

$$= \frac{\beta^2-\alpha^2}{2} \cdot \frac{1}{\beta-\alpha}$$

$$= \frac{\cancel{(\beta-\alpha)}(\beta+\alpha)}{2\cancel{(\beta-\alpha)}}$$

$$= \boxed{\frac{\beta+\alpha}{2}}$$

average of endpoints