intro to data science with probability & statistics

CSCI 3022

Octoer 3, 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. Suggested **milestones**:
  - Probs 1, 2, 3 done before the end of the week.
  - Probs 4, 5 done next week.
- **Midterm** next week. Weds, 6:30-8:00 PM. SEE PIAZZA for details..
- Midterm review *in class* next Monday.

OH today until 4:30

Fri: ∅

M: 4-5

W: 4-6

Last time on CSCI 3022

values Pr(value)

- **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:

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

weighted average

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

value Pr(value)

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

- **Change of Variables**: Let X be a RV and let $g : \mathcal{R} \to \mathcal{R}$ 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 O'Flaherty Family

- Stella O'Flaherty is having her whole crazy extended family over for dinner. What a bunch! They are considering a random variable X:

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

Stella & the O'Flaherties

1. Is X a discrete or continuous random variable?

2. Compute $E[X]$

$$(-1)\frac{1}{5} + (0)\frac{1}{5} + (1)\frac{3}{5}$$

$$-\frac{1}{5} + 0 + \frac{3}{5} = \boxed{\frac{2}{5}}$$

The O'Flaherty Family

- Stella O'Flaherty is having her whole crazy extended family over for dinner. What a bunch! They are considering a random variable X:

$$\sum_i g(a_i)\,Pr(a_i)$$

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

3. **Oscar O'Flaherty** 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?

$P(Y=3)=\frac{1}{5}$ $\quad P(Y=2)=\frac{1}{5}$ $\quad P(Y=3)=\frac{3}{5}$

$$\boxed{\begin{array}{l} P(Y=2)=\frac{1}{5} \\ P(Y=3)=\frac{4}{5} \end{array}}$$

4. What is $E[Y]$?

$g(x)=x^2+2$

$$2\,\tfrac{1}{5} + 3\,\tfrac{4}{5} = \frac{2}{5}+\frac{12}{5} = \boxed{\frac{14}{5}} = E[g(x)]$$

Oscar O'Flaherty

Quinn O'Flaherty.

- It is early-October, and Quinn is swimming along the reef, watching the coral change colors. The water is cool. A bird above the water contemplates flying South for the winter. Is it too early? Who can say. Rain falls softly on the water above the reef at an average rate of 22 drops per minute. Quinn watches a drop fall and checks her watch. Let **X** be a random variable that represents the time she waits until the next drop lands. Which distribution does **X** follow?

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

#raindrops in a minute
↓
Poisson.

waiting time between events is exponential.

Quinn

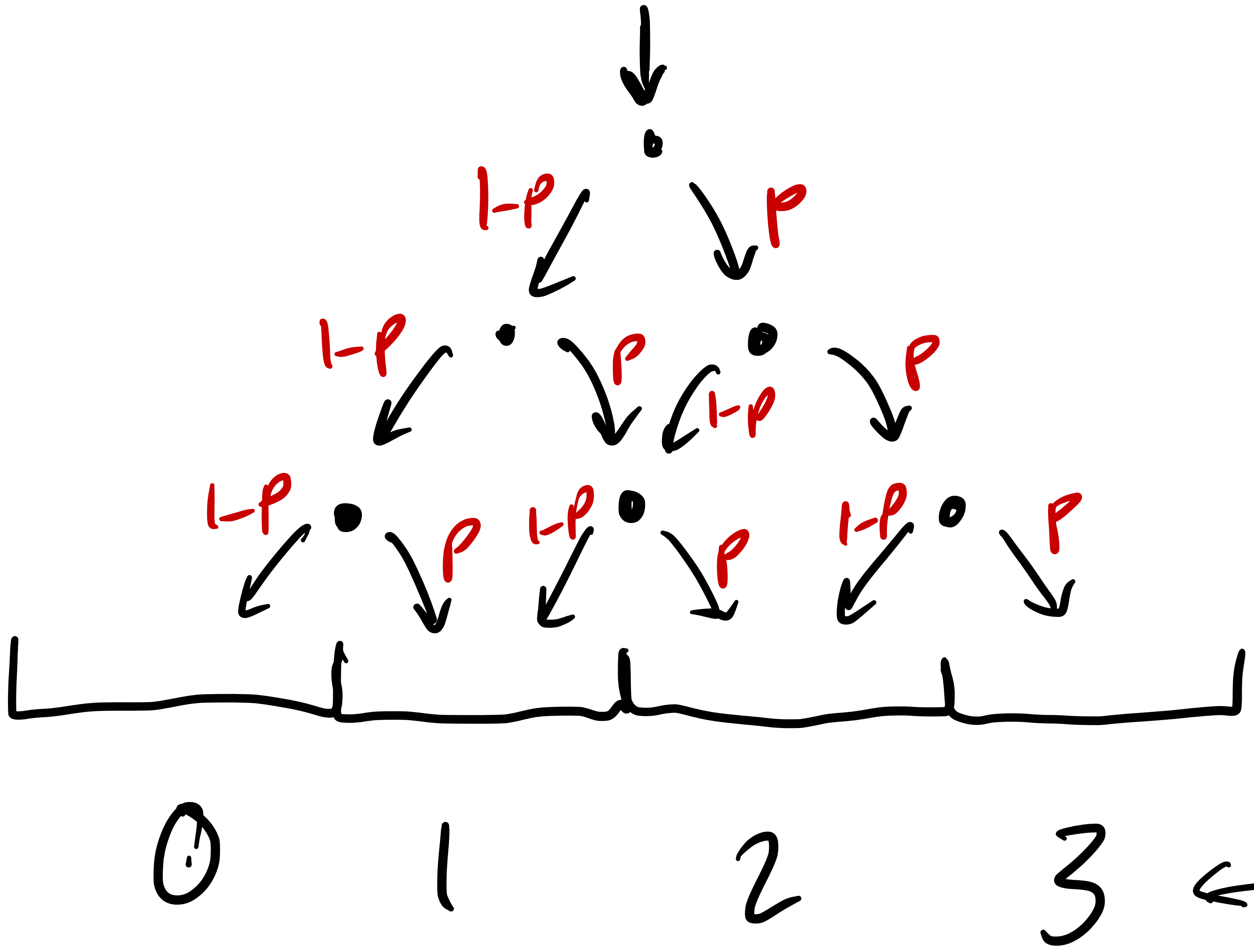
Plinko!

rewatch!

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?



Bin (n, p)

$$\boxed{\begin{aligned} P(X=\cancel{R}\,1) &= p \\ P(X=\cancel{L}\,0) &= 1-p \end{aligned}}$$

Ber(p)

Binomial

# of times plinko chip went R

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?

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

$\cancel{0}$ $\cancel{1}$ $\cancel{2}$ $\cancel{3}$

a_i : \$$\pi$ \$1000 \$100 \$12

$P(a_i)$: $P(X=0)$ $P(X=1)$ $P(X=2)$ $P(X=3)$

$\binom{3}{0}p^0(1-p)^3$ $\binom{3}{1}p^1(1-p)^2$ $\binom{3}{2}p^2(1-p)$ $\binom{3}{3}p^3(1-p)^0$

What is $E[X]$, when X is $Bin(n,p)$?

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?
- **Pro Tip**: expectation is a linear function! $\boxed{E[rX+s] = rE[X] + s}$

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

$$E[X] = E[Y_1 + Y_2 + Y_3 + \ldots + Y_n] = E[Y_1] + E[Y_2] + \ldots + E[Y_n] = n\,E[Ber(p)] = \boxed{np}$$

Bin(n,p)

n Ber(p)

$$E[Bin(n,p)] = np$$

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? $X \sim B(n,p)$

Defined sample variance : $\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$\{x_1, x_2, x_3, \ldots x_n\} \leftarrow$ think of these as draws from a R.V.

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

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

avg. of dataset.

like an expected value.

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

Variance

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

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

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

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 $\mu = \mathrm{E}[X]$ (from last time)
  - Second, use the change-of-variables formula with $g(x) = (x - \mu)^2$

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

Variance

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

foil

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

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

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

$E[4] = 4$

$E[E[X]] = E[X]$

$= E[X^2] + E[-2XE[X]] + E[E[X]^2]$

$= E[X^2] - 2E[X]E[X] + E[X]^2 = \boxed{E[X^2] - E[X]^2 = \mathrm{var}(X)}$

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?

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

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

- 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 $Y \sim Ber(p)$

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

Together: $p - p^2 = \boxed{p(1-p) = var(Ber(p))}$

1st $E[Y]^2 = (p)^2 = p^2$

Next $E[Y^2] \ldots \sum_i a_i \, p(a_i)$

Y^2: 0^2 w.p. $(1-p)$

1^2 w.p. p

$E[Y^2] = 0^2(1-p) + 1^2(p)$

$= p$

Independence & Variance

$$X \sim Bin(n,p) = \underbrace{Ber(p) + Ber(p) + \dots + Ber(p)}_{n}$$

- 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 $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

$$var(Bin(n,p)) = \overbrace{var(Ber(p)) + var(Ber(p)) + \dots + var(Ber(p))}^{n} = \boxed{np(1-p)}$$

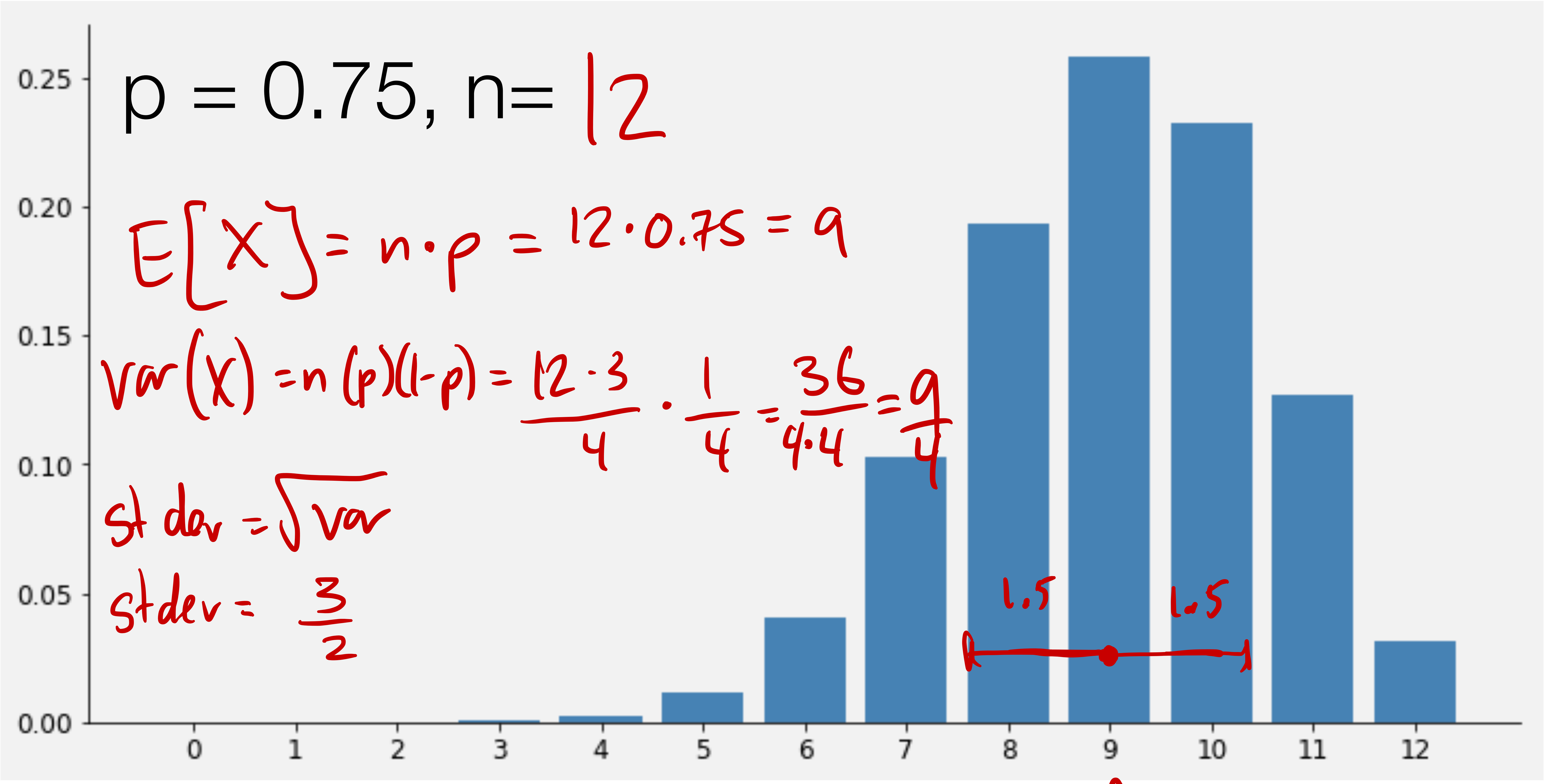
indep. of flips

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 $\mathrm{E}[X] = np \quad \text{and} \quad \mathrm{Var}(X) = np(1-p)$

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**?

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

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

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

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

$$\mathrm{var}(rX) = \ ?$$

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 $\mathrm{Var}(X)$.

Mean & Variance: Uniform

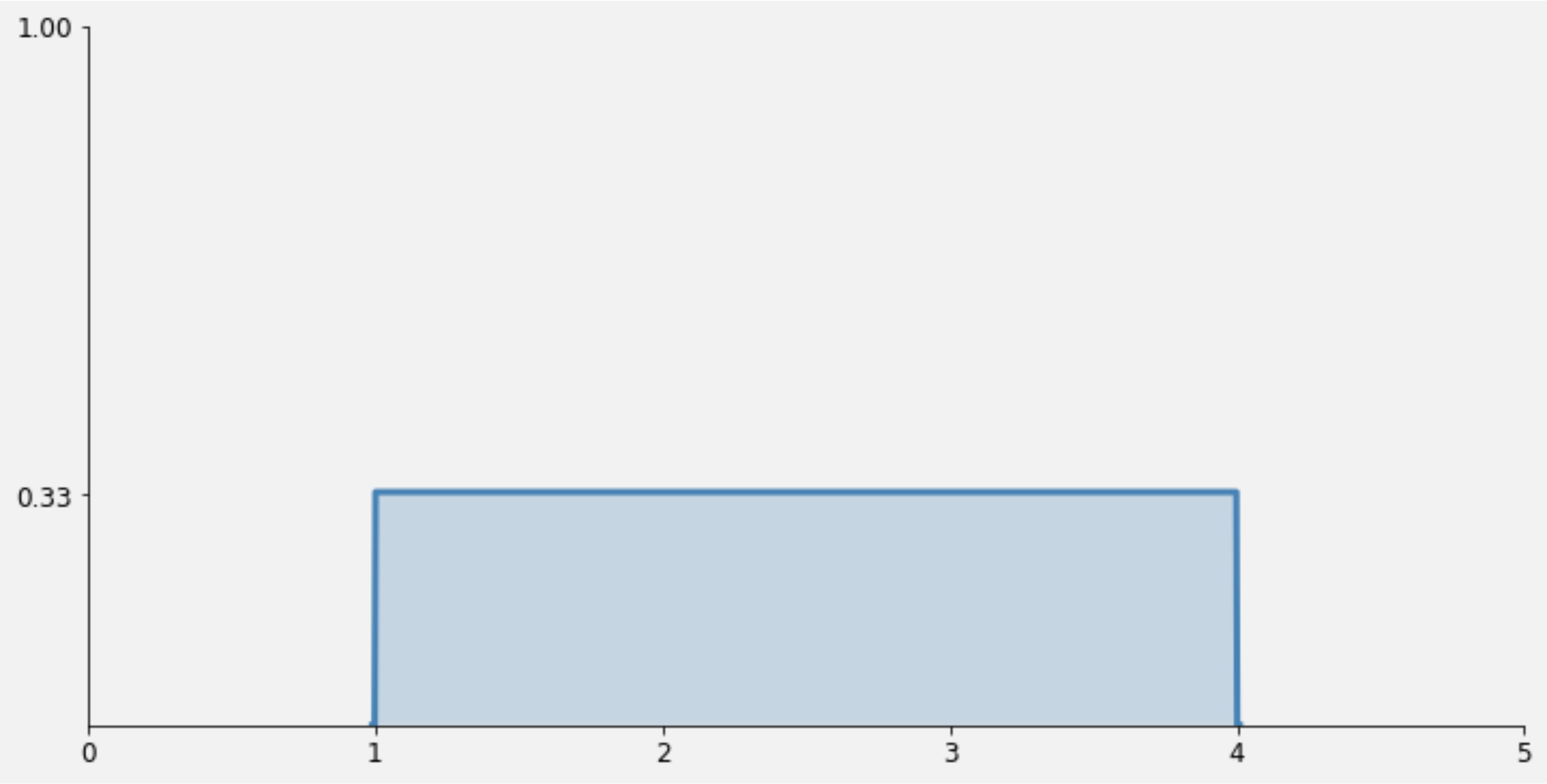
- **Suppose** $X \sim U[\alpha, \beta]$. Find $\mathrm{E}[X]$ and $\mathrm{Var}(X)$.

Mean & Variance: Uniform

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

Mean & Variance: Uniform

- **Theorem**: Let $X \sim U[\alpha, \beta]$. Then: $\mathrm{E}[X] = \dfrac{\alpha + \beta}{2}$ $\quad \mathrm{Var}(X) = \dfrac{(\beta - \alpha)^2}{12}$



Summary

$$X \sim Ber(p)$$

$$E[X] = p$$

$$Var(X) = p(1-p)$$

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

$$E[X] = np$$

$$Var(X) = np(1-p)$$

$$X \sim U(\alpha, \beta)$$

$$E[X] = \frac{1}{2}(\alpha + \beta)$$

$$Var(X) = \frac{1}{12}(\beta - \alpha)^2$$