CSCI 3022: Intro to Data Science

Lecture 11: Variance of Discrete and Continuous Random Variables

Rachel Cox

Department of Computer Science

Announcements & Reminders

- ☐ Practicum1 Due Monday, at 11:59 pm
- ☐ HW3 – to be posted
  - due end of week 9

What will we learn today?

- ❑ Variance

- ❑ *A Modern Introduction to Probability and Statistics, Chapter 7*

Review from Last Time

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:

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

The **expectation** or **expected value** or **mean** of a continuous random variable X with probability density function f is:

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

Change-of-Variables Formula: Let X be a random variable and $g\colon \mathbb{R} \to \mathbb{R}$ be a function. Then:

$$E[g(x)] = \sum_i g(a_i) P(X = a_i) \quad \text{and} \quad E[g(x)] = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$$

Plinko!

Let X be the random variable describing the result in each round of Plinko with n rows and probability p of moving to the right off of each peg. (Ignoring the edges for now.)

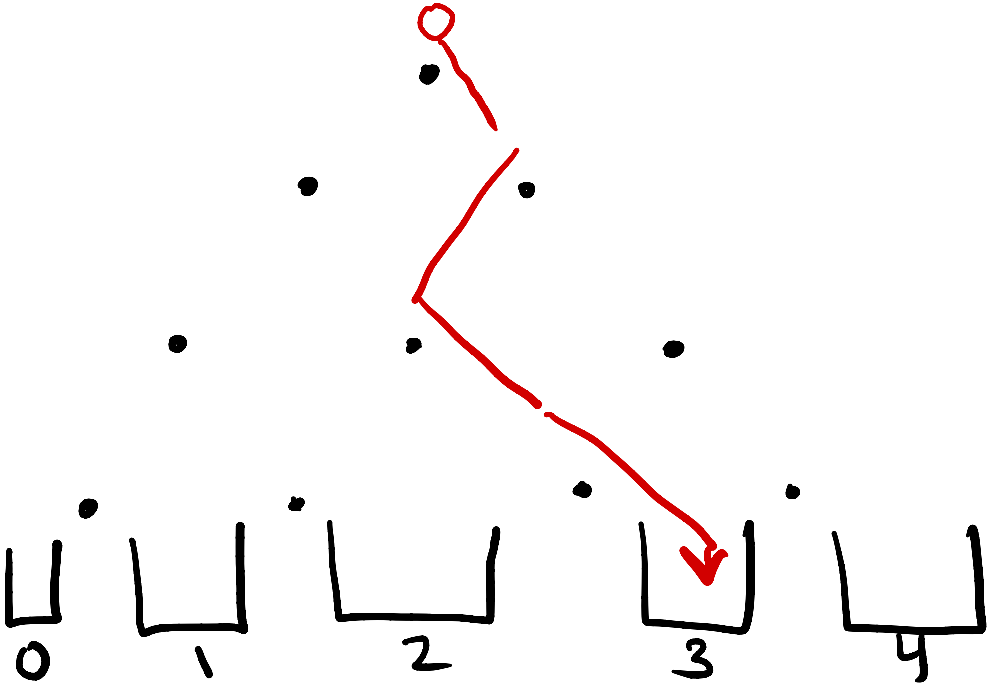
https://www.youtube.com/watch?v=naUppHrHJpI

Plinko!

Let X be the random variable describing the result in each round of Plinko with n rows and probability p of moving to the right off of each peg. (Ignoring the edges for now.)

What distribution does X follow?

- Each row results in either a move right (w/ prob p) or left (w/ prob (1-p))
- We have n rows…or trials



X = bin number that Plinko disk ends up in.

X = is a sum of Bernoulli trials.

let $Y_i \sim Ber(p)$ for each row i

$X = Y_1 + Y_2 + Y_3 + Y_4$

$X \sim Bin(n, p)$

Plinko!

Let X be the random variable describing the result in each round of Plinko with n rows and probability p of moving to the right off of each peg. (Ignoring the edges for now.)

What is the expected value of X?

Bin(n, p)

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

$$p(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$E[X] = E[Y_1 + Y_2 + Y_3 + \cdots + Y_n]$$

$$= E[Y_1] + E[Y_2] + E[Y_3] + \cdots + E[Y_n]$$

Recall: $\left.\begin{array}{l} E[Y] = 0 \cdot (1-p) + 1 \cdot p \\ \quad = p \end{array}\right\}$ when $Y \sim \text{Ber}(p)$

$$E[x] = n E[Y]$$

$$\boxed{E[x] = np}$$

Plinko!

Let X be the random variable describing the result in each round of Plinko with n rows and probability p of moving to the right off of each peg. (Ignoring the edges for now.)

What is the **variance** of X?

Given data $x_1, x_2, \dots, x_n$ their sample variance is $\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2$ talked about before.

Which amounts to: Average[(datum – Average_of_Data)^2]

Or more formally: $E\left[\left(X - E(X)\right)^2\right]$

Variance

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

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

The **standard deviation** of a random variable X is the square root of the variance:

$$SD(X) = \sqrt{Var(X)}$$

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

$$= E[x^2 - 2x E[x] + (E[x])^2]$$

$$= E[x^2] + E[-2x E[x]] + E[(E[x])^2]$$

$$= E[x^2] - 2 E[x] \cdot E[x] + (E[x])^2$$

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

Variance

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

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

Alternatively: $Var(X) = E[X^2] - E[X]^2$

Variance

Let X be the random variable describing the result in each round of Plinko with n rows and probability p of moving to the right off of each peg. (Ignoring the edges for now.)

What is the **variance** of X? First step is finding the variance of each $Y \sim Ber(p)$

$$X = Y_1 + Y_2 + Y_3 + \cdots + Y_n$$

$Y \sim Ber(p)$

$$\begin{aligned}
Var(Y) &= E[Y^2] - (E[Y])^2 \\
&= 0^2(1-p) + 1^2 \cdot p - (p)^2 \\
&= p - p^2 \\
Var(Y) &= p(1-p)
\end{aligned}$$

Because the Y_i are independent.

$$\begin{aligned}
Var(X) &= Var(Y_1 + Y_2 + \cdots + Y_n) \\
&= Var(Y_1) + Var(Y_2) + \cdots + Var(Y_n) \boxed{= np(1-p)}
\end{aligned}$$

Variance

Let X be the random variable describing the result in each round of Plinko with n rows and probability p of moving to the right off of each peg. (Ignoring the edges for now.)

What is the **variance** of $X \sim Bin(n, p)$? — $Var(X) = np(1-p)$

If X and Y are **independent**, then Var(X+Y) = Var(X) + Var(Y)

Key

Quick Summary

If $X \sim Ber(p)$, then:

- $E[X] = p$
- $Var(X) = p(1-p)$

If $X \sim Bin(n,\ p)$, then:

- $E[X] = np$
- $Var(X) = np(1-p)$

The Binomial Distribution

Example: You are taking a 12-question quiz and believe there to be about a 75% chance that any one of your answers is correct. What is your expected grade on the quiz? What is the variance in your quiz grade?

Let X be the total number of questions you answer correctly.

$X \sim \text{Bin}(n = 12, p = \frac{3}{4})$

$E[X] = np$
$Var(X) = np(1-p)$

$E[X] = 12 \cdot \frac{3}{4} = 9$

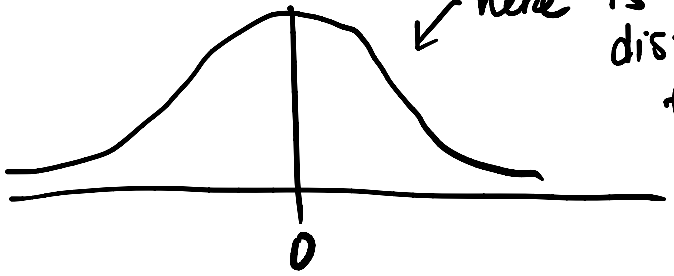
$Var(X) = 12 \cdot \frac{3}{4} \cdot \frac{1}{4} = 9/4$

$SD(X) = 3/2$

More Facts about Variance

Expectation is linear: $E[aX + b] = aE[X] + b$

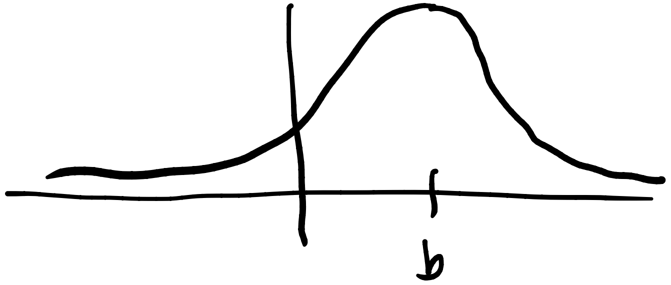
What about variance?



here is a distribution for X

$\rightarrow$

what about X+b



Takeaway: ★ $\mathrm{Var}(X+b) = \mathrm{Var}(X)$

Now consider:

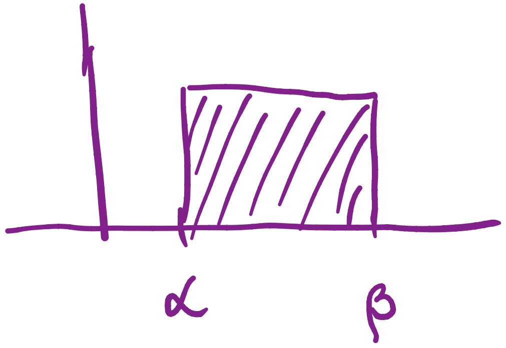
$$\begin{aligned}\mathrm{Var}(aX) &= E[(aX)^2] - (E[aX])^2 \\ &= E[a^2X^2] - (aE[X])^2 \\ &= a^2E[X^2] - a^2(E[X])^2 = a^2(E[X^2] - (E[X])^2) \\ &= a^2\,\mathrm{Var}(X)\end{aligned}$$

Result: ★ $\boxed{\mathrm{Var}(aX) = a^2\,\mathrm{Var}(X)}$

not linear.

Mean and Variance of a Uniform Random Variable

Example: Let $X \sim U[\alpha, \beta\,]$. What are E[X] and Var(X)?



$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \alpha \le x \le \beta \\ 0 & \text{elsewhere} \end{cases}$$

$$E[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{\beta-\alpha}\, dx$$

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

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

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

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

$$\boxed{E[X] = \frac{\beta+\alpha}{2}}$$

Mean and Variance of a Uniform Random Variable

$$(a^3-b^3) = (a-b)(a^2+ab+b^2)$$

"SOAP"

Example (continued): Let $X \sim U[\alpha, \beta\ ]$. What are E[X] and Var(X)?

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

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

$$= \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{\beta-\alpha}\, dx - \left(\frac{\beta+\alpha}{2}\right)^2$$

$$= \frac{1}{\beta-\alpha} \cdot \frac{x^3}{3}\Big|_{\alpha}^{\beta} - \left(\frac{\beta+\alpha}{2}\right)^2$$

$$= \frac{1}{\beta-\alpha} \cdot \frac{1}{3}(\beta^3-\alpha^3) - \left(\frac{\beta+\alpha}{2}\right)^2$$

$$= \frac{1}{\beta-\alpha} \cdot \frac{1}{3} \cdot (\beta-\alpha)(\beta^2+\alpha\beta+\alpha^2) - \left(\frac{\beta+\alpha}{2}\right)^2$$

$$= \frac{1}{3}(\beta^2+\alpha\beta+\alpha^2) - \frac{\beta^2+2\alpha\beta+\alpha^2}{4}.$$

Mean and Variance of a Uniform Random Variable

Example (continued): Let $X \sim U[\alpha, \beta\,]$. What are E[X] and Var(X)?

... continuing

$$Var(x) = \frac{4(\beta^2 + \alpha\beta + \alpha^2) - 3(\beta^2 + 2\alpha\beta + \alpha^2)}{12}$$

$$= \frac{\beta^2 - 2\alpha\beta + \alpha^2}{12}$$

$$\boxed{Var(x) = \frac{(\beta - \alpha)^2}{12}}$$

Quick Summary

If X ~ Ber(p), then:

- E[X] = p
- Var(X) = p(1-p)

If X ~ Bin(n, p), then:

- E[X] = np
- Var(X) = np(1-p)

If X ~ *U*[α, β], then:

- $\mathrm{E}[X] = \frac{1}{2}(\alpha + \beta)$
- $\mathrm{Var}(X) = \frac{1}{12}(\beta - \alpha)^2$

Next Time:

- Normal Distribution!