Introduction to Data Science With Probability and Statistics

Lecture 13: Variance of Discrete and Continuous Random Variables

CSCI 3022 - Summer 2020

Sourav Chakraborty
Dept. of Computer Science
University of Colorado Boulder

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!

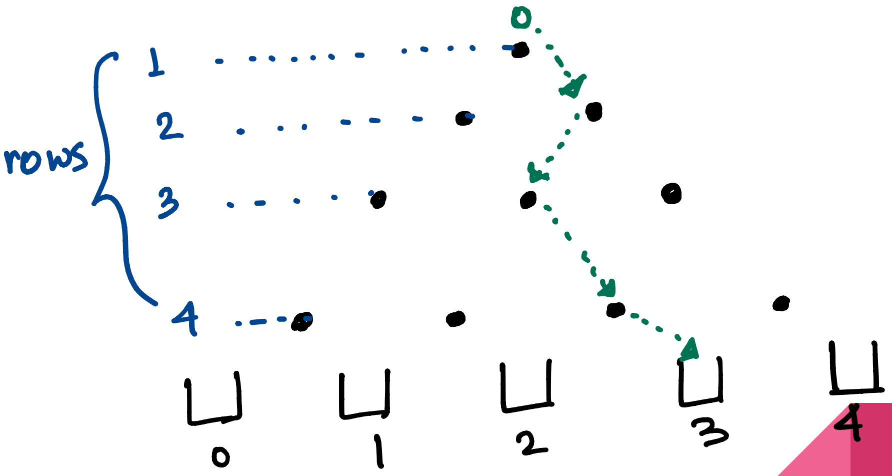
Let X be the random variable describing the result in each round of Plinko with n rows and probability p of moving to the right off of each peg. (Ignoring the edges for now)

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)

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



Plinko!

Let X be the random variable which denotes the final bin on which the ball will eventually end up. At each row, the ball is equally likely to go left or right.

What is the distribution that X follow?

for every row i, Let $y_i \sim Ber(p)$. $= \begin{cases} 1 \text{ (moving right)}, & p. \\ 0 \text{ (moving left)}, & 1-p \end{cases}$

$$\left.\begin{array}{l} X = y_1 + y_2 + y_3 + y_4 = \Sigma y_i \\ \therefore \; X \sim Bin(4, p) \end{array}\right\} \text{for the simplified example.}$$

$X \sim Bin(n, p)$ } General case with 'n' rows.

Plinko!

Let X be the random variable which denotes the final bin on which the ball will eventually end up. At each row, the ball is equally likely to go left or right.

What is the expected value of X?

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

$$E[x] = E\left[\sum y_i\right] = E[y_1 + y_2 + \cdots + y_n]$$

$$= E[y_1] + E[y_2] + \cdots + E[y_n] \quad \left\{ \because E[ax+b] = aE[x] + b \right.$$

$$= p + p + \cdots p = n \cdot p. \quad \left\{ \begin{array}{l} \because E[y_i] = 0\cdot(1-p) + 1\cdot p \\ \qquad\qquad = p \\ \text{if } y_i \sim Ber(p) \end{array} \right.$$

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

Plinko!

Let X be the random variable which denotes the final bin on which the ball will eventually end up. At each row, the ball is equally likely to go left or right.

What is the **variance** of X?

what have we seen before?

- Given $x_1, x_2, \ldots x_n$ their sample variance is $\frac{1}{n-1}\sum_{k=1}^{n}(x_k-\bar{x})^2$

which is basically : Avg of squared differences from the mean.

: $E[(X-E[X])^2]$ //

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:

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

Variance

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

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

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

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

$= E[X^2] + E[-2 \times \underbrace{E[X]}_{\mu}] + E[(\underbrace{E[X]}_{\mu})^2]$

μ : constant value.

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

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

Plinko!

Let X be the random variable which denotes the final bin on which the ball will eventually end up. At each row, the ball is equally likely to go left or right.

What is the **variance** of X?

Plinko!

Let X be the random variable which denotes the final bin on which the ball will eventually end up. At each row, the ball is equally likely to go left or right.

What is the **variance** of X?

$$Var[X] = Var[y_1 + y_2 + y_3 + \cdots + y_n]$$

$$Var[y_i] = E[y^2] - E[y]^2$$

$$= (0^2 \cdot (1-p) + 1^2 p) - p^2$$

$$= p - p^2 = p(1-p)$$

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

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

So, $Var(X) = Var\left(\sum_{i}^{n} y_i\right) = Var(y_1) + Var(y_2) + \dots Var(y_n)$

$$\boxed{= n \cdot p(1-p)}$$

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?

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 : # correct answers.

$\therefore \; X \sim Bin(n = 12, \; p = 0.75)$

$E[X] = n \cdot p = 12 \times 0.75 = 9$

$Var(X) = n \cdot p \cdot (1-p) = 12 \times 0.75 \times 0.25 = 2.25$

$S.D(X) = 1.5$

More Facts about Variance

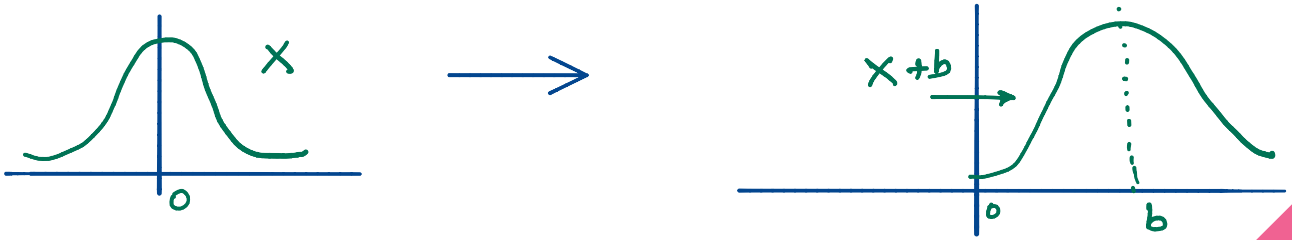
Expectation is linear: $E[aX + b] = aE[X] + b$

What about variance?

Consider 2 cases :

1). Effect of adding a constant ($X + \underline{\underline{b}}$)

1).



$$\boxed{\therefore \quad Var(X+b) = Var(X)}$$

More Facts about Variance

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

What about variance?

2). Effect Multiplying a constant to the random variable. ($\underline{\underline{a}}x$)

$$\mathrm{Var}(ax) = E[(ax)^2] - (E[ax])^2 = E[a^2x^2] - (aE[x])^2$$

$$= a^2 E[x^2] - a^2 E[x]^2$$

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

$$\therefore \boxed{\mathrm{Var}(ax) = a^2 \mathrm{Var}(x)}$$

Mean and Variance of a Uniform Random Variable

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

Mean and Variance of a Uniform Random Variable

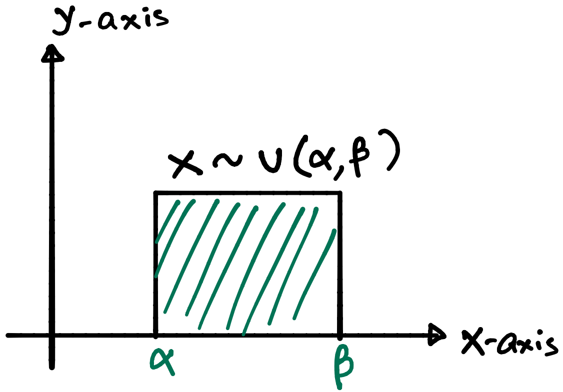
Example: Let $X \sim U[\alpha, \beta\,]$. What are E[X] and Var(X)?

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

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

$$= \frac{1}{\beta-\alpha}\int_{\alpha}^{\beta} x\,dx = \frac{1}{\beta-\alpha} \cdot \left.\frac{x^2}{2}\right|_{\alpha}^{\beta}$$

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



Mean and Variance of a Uniform Random Variable

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

$$Var(x) = \underbrace{E[x^2]} - \underbrace{E[x]^2}_{\text{already done.}}$$

$$E[x^2] = \int_{-\infty}^{\infty} x^2 \cdot f(x) = \int_{\alpha}^{\beta} x^2 \cdot \frac{1}{\beta-\alpha} \cdot dx = \frac{1}{\beta-\alpha} \cdot \frac{x^3}{3}\Big|_{\alpha}^{\beta} = \frac{1}{\beta-\alpha} \cdot \frac{1}{3} \cdot \beta^3 - \alpha^3$$

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

$$\therefore Var(x) = \frac{1}{3} \cdot (\alpha^2 + \beta^2 + \alpha\beta) - \left[\frac{1}{2}(\alpha+\beta)\right]^2 = \boxed{\frac{(\beta-\alpha)^2}{12}}$$

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 ~ $U[\alpha, \beta]$, then:

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

Next Time:

- ❖ The Normal Distribution !