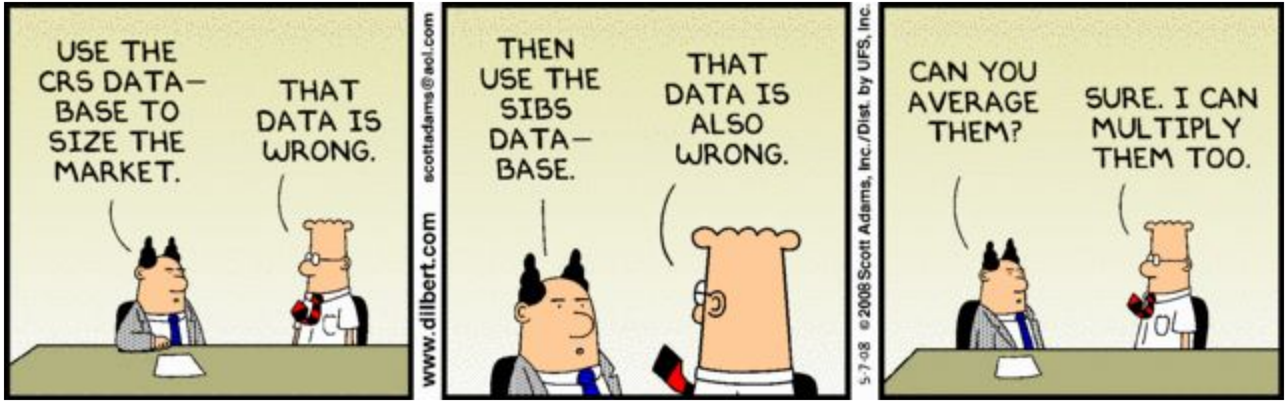
Lecture 11: Variance of Discrete and Continuous Random Variables



Announcements and reminders

Practicum 1 posted, due Monday 4 March at 11:59 PM.

→ Monday after your midterm. Plan ahead!

Midterm:

- Tuesday 26 February, 7-8:30 PM, HUMN 1B50
- Special accommodations: 6-? PM, HUMN 335
- Tell me as soon as possible about conflicts -- include documentation
- Concept guide on Piazza; material up through variance (today/Friday)
- 3”x5” notecard for cheat-sheet. Calculator is okay. Smart phone is **not**.
- Review in class on Monday 25 March (Q&A)
- Study from: old exams, lecture notes, homework, practicum, textbooks…

Previously, on CSCI 3022…

Definition: The **expectation** or **expected value** of a discrete random variable *X* that takes the values a_1, a_2, … and with pmf *p* is given by

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

Definition: The **expectation**, **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 $g : \mathbb{R} \to \mathbb{R}$ 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!

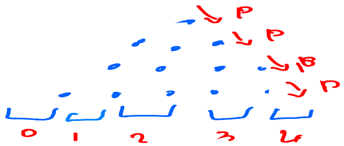
Let X be the random variable describing the result in each round of Plinko with n rows and probability p of moving to the right off of each peg. (Ignoring the edges for now.)

Question: What distribution does X follow?

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.)

Question: What distribution does X follow?



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

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

→ each row is a Bernoulli trial.   Call result of row *i* Y_i

→ entire thing is sum of Bernoulli trials: $X = Y_1 + Y_2 + \ldots + Y_n$

→ that makes $X \sim \text{Bin}(n, p)$

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.)

Question: What is the expected value of X?

Hint: *Remember that the expectation of a linear function is* E[aX+b] = a E[X] + b

$$E[X] = E[Y_1 + Y_2 + \dots Y_n] = \underbrace{E[Y_1] + E[Y_2] + \dots E[Y_n]}_{\text{n of these and each we'll assume is indep of the other } Y_i}$$

$$= nE[Y]$$

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.)

Question: What is the expected value of X?

Hint: *Remember that the expectation of a linear function is* E[aX+b] = a E[X] + b

$E[X] = E[Y_1 + Y_2 + \ldots + Y_n]$ ← E[*] is linear, so we can distribute it across the sum

$= E[Y_1] + E[Y_2] + \ldots + E[Y_n]$ ← each Y_i ~ Ber(n, p), so $E[Y_i]$ = p for i = 1, 2, … n

$= p + p + \ldots + p$ ← there are still n terms, so round them all up to find…

$E[Y] = \sum a_i P(a_i)$

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

$\boxed{E[X] = np}$ (Binomial random variable)

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.)

Question: What is the variance of X?

$$\mathrm{Var}(x) = \underbrace{\frac{1}{n-1}\sum_{i=1}^{n}}_{\text{Average}} (x_i - \bar{x})^2$$

Average [datum − mean]^2

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

some function

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.)

Question: What is the variance of X?

Actually… what ***is*** variance?

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.)

Question: What is the variance of X?

Actually… what ***is*** variance?

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

… which amounts to: Average[(datum - Average_of_Data)^2]

… or more formally: $E[(X - E[X])^2]$

Variance

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

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

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

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

How to compute:

- First, compute E[X]
- Then, use the definition of Variance and change-of-variables formula (w/ $g(x) = (x-E[X])^2$) to get Var(X):

$$\mathrm{Var}(X) = \sum_i (a_i - \mathrm{E}[X])^2\, p(a_i) \qquad \text{or} \qquad \mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2\, f(x)\, dx$$

Variance

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

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

Now hang on a second. There's a $(X\text{-}E[X])^2$ in there… can we FOIL that out and simplify?

Variance

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

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

Now hang on a second. There's a $(X\text{-}E[X])^2$ in there… can we FOIL that out and simplify?

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

Variance

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

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

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

Okay, back to 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.)

Question: What is the variance of $X \sim \mathrm{Bin}(n, p)$?

First step: What is the variance of each $Y \sim \mathrm{Ber}(p)$?

Okay, back to 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.)

Question: What is the variance of X ~ Bin(n, p) ?

First step: What is the variance of each Y ~ Ber(p) ?

Let's use $E[Y^2] - E[Y]^2$

$$\mathrm{E}[Y^2] = \sum_i a_i^2 P(X = a_i) = 1^2 \cdot p + 0^2 \cdot (1-p) = p$$

and $\mathrm{E}[Y]^2 = (p)^2 = p^2$

So: $\mathrm{Var}(Y) = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2 = p - p^2 = p(1-p)$

Quick summary

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

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

Okay, back to 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.)

Question: What is the variance of $X \sim \text{Bin}(n, p)$?

Fun fact: If X and Y are **independent**, then $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

Okay, back to 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.)

Question: What is the variance of $X \sim \mathrm{Bin}(n, p)$?

Fun fact: If X and Y are **independent**, then $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

$\rightarrow$ $\mathrm{Var}(X) = \mathrm{Var}(Y_1 + Y_2 + \dots + Y_n) = \mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) + \dots + \mathrm{Var}(Y_n)$

$= p(1-p) + p(1-p) + \dots + p(1-p)$

$= np(1-p)$

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?

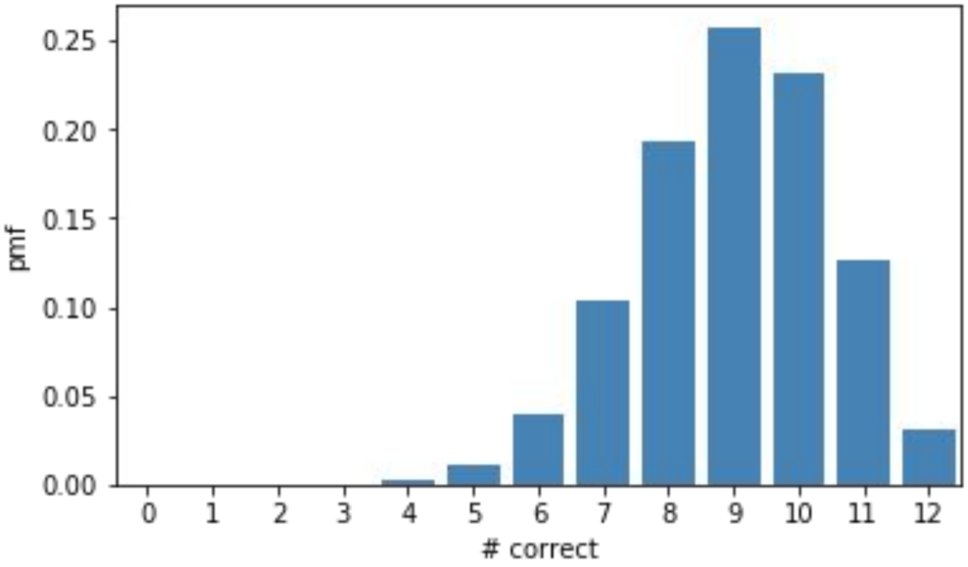
$X \sim Bin(n=12, p=0.75)$

→ $E[X] = np = 12 \cdot 0.75 = 9$

And

→ $Var[X] = np(1-p)$
$= 12 \cdot 0.75 \cdot 0.25 = 9/4$

→ $SD[X] = 1.5$



More Fun Facts about Variance!

Recall: Expectation is **linear**: E[aX+b] = a E[X] + b

So what about **variance**?

More Fun Facts about Variance!

Recall: Expectation is **linear**: $E[aX+b] = a\,E[X] + b$

So what about **variance**?

- What happens if we shift $X \rightarrow X+b$?

More Fun Facts about Variance!

Recall: Expectation is **linear**: $E[aX+b] = a\,E[X] + b$

So what about **variance**?

- What happens if we scale $X \to aX$?

More Fun Facts about Variance!

Recall: Expectation is **linear**: $E[aX+b] = a\,E[X] + b$

Conclusion: Variance is ***not*** linear: $\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$

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)?

The pdf of X is: $$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{if} \quad \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

So the expected value is: $$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_{\alpha}^{\beta} x \cdot \frac{1}{\beta - \alpha} dx$$

$$= \frac{x^2}{2(\beta - \alpha)} \bigg|_{\alpha}^{\beta} = \frac{\alpha + \beta}{2}$$

And the variance is $\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2$, so we'll calculate $E[X^2]$ next…

Mean and variance of a uniform random variable

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

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

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

So…

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \left(\frac{\alpha + \beta}{2}\right)^2$$

$$= \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \frac{\alpha^2 + 2\alpha\beta + \beta^2}{4} = \frac{4\beta^2 + 4\alpha\beta + 4\alpha^2 - 3\alpha^2 - 6\alpha\beta - 3\beta^2}{12}$$

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

So if $X \sim U[\alpha, \beta]$, then $\mathrm{E}[X] = \frac{1}{2}(\alpha + \beta)$ and $\mathrm{Var}(X) = \frac{1}{12}(\beta - \alpha)^2$

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 (or later this time, more likely) …

… notebook day!

Then …

… we review for the midterm exam!

- Q&A format, so you bring the *Q*s and I'll bring the *A*s
- Many study materials.
  … **HW/quizlet/in-class notebooks/lecture examples/old exams**