07: Variance, Bernoulli, Binomial

Lisa Yan and Jerry Cain
September 28, 2020
Quick slide reference

3  Variance                      07a_variance_i
10 Properties of variance      07b_variance_ii
17 Bernoulli RV                07c_bernoulli
22 Binomial RV                 07d_binomial
34 Exercises                   LIVE
Variance
Average annual weather

Stanford, CA
$E[\text{high}] = 68^\circ F$
$E[\text{low}] = 52^\circ F$

Washington, DC
$E[\text{high}] = 67^\circ F$
$E[\text{low}] = 51^\circ F$

Is $E[X]$ enough?
Average annual weather

Stanford, CA

\[ E[\text{high}] = 68^\circ F \]
\[ E[\text{low}] = 52^\circ F \]

Stanford high temps

Washington, DC

\[ E[\text{high}] = 67^\circ F \]
\[ E[\text{low}] = 51^\circ F \]

Washington high temps

Normalized histograms are approximations of PMFs.
Variance = “spread”

Consider the following three distributions (PMFs):

- Expectation: \( E[X] = 3 \) for all distributions
- But the “spread” in the distributions is different!
- **Variance**, \( \text{Var}(X) \): a formal quantification of “spread”
Variance

The variance of a random variable $X$ with mean $E[X] = \mu$ is

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

- Also written as: $E[(X - E[X])^2]$
- Note: $\text{Var}(X) \geq 0$
- Other names: 2nd central moment, or square of the standard deviation

\[ \text{Var}(X) \quad \text{def} \quad \text{standard deviation} \quad \text{SD}(X) = \sqrt{\text{Var}(X)} \]
Variance of Stanford weather

Stanford, CA

\[ E[\text{high}] = 68^\circ F \]
\[ E[\text{low}] = 52^\circ F \]

Stanford high temps

\[ E[X] = \mu = 68 \]

\[
\begin{array}{cc}
X & (X - \mu)^2 \\
57^\circ F & 124 (^\circ F)^2 \\
71^\circ F & 9 (^\circ F)^2 \\
75^\circ F & 49 (^\circ F)^2 \\
69^\circ F & 1 (^\circ F)^2 \\
... & ...
\end{array}
\]

Variance \[ E[(X - \mu)^2] = 39 (^\circ F)^2 \]
Standard deviation \[ = 6.2^\circ F \]
Comparing variance

Stanford, CA
\[ E[\text{high}] = 68^\circ F \]

Washington, DC
\[ E[\text{high}] = 67^\circ F \]

Variances

Stanford high temps
\[ \text{Var}(X) = 39 \text{ (°F)}^2 \]

Washington high temps
\[ \text{Var}(X) = 248 \text{ (°F)}^2 \]
Properties of Variance
Properties of variance

Definition \[ \text{Var}(X) = E[(X - E[X])^2] \] Units of $X^2$

\textbf{def standard deviation} \[ \text{SD}(X) = \sqrt{\text{Var}(X)} \] Units of $X$

Property 1 \[ \text{Var}(X) = E[X^2] - (E[X])^2 \]

Property 2 \[ \text{Var}(aX + b) = a^2 \text{Var}(X) \]

- Property 1 is often easier to compute than the definition
- Unlike expectation, variance is not linear
Properties of variance

Definition

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

**def** standard deviation

\[ \text{SD}(X) = \sqrt{\text{Var}(X)} \]

Property 1

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

Property 2

\[ \text{Var}(aX + b) = a^2 \text{Var}(X) \]

- Property 1 is often easier to compute than the definition
- Unlike expectation, variance is not linear
Computing variance, a proof

\[
\text{Var}(X) = E[(X - E[X])^2] = E[(X - \mu)^2]
\]

\[
= \sum_x (x - \mu)^2 p(x)
\]

\[
= \sum_x (x^2 - 2\mu x + \mu^2) p(x)
\]

\[
= \sum_x x^2 p(x) - 2\mu \sum_x xp(x) + \mu^2 \sum_x p(x)
\]

\[
= E[X^2] - 2\mu E[X] + \mu^2 \cdot 1
\]

\[
= E[X^2] - 2\mu^2 + \mu^2
\]

\[
= E[X^2] - \mu^2
\]

\[
= E[X^2] - (E[X])^2
\]

Let \( E[X] = \mu \)

Everyone, please welcome the second moment!
Variance of a 6-sided die


1. Approach #1: Definition

\[ Var(Y) = \frac{1}{6} \left( 1 - \frac{7}{2} \right)^2 + \frac{1}{6} \left( 2 - \frac{7}{2} \right)^2 + \frac{1}{6} \left( 3 - \frac{7}{2} \right)^2 + \frac{1}{6} \left( 4 - \frac{7}{2} \right)^2 + \frac{1}{6} \left( 5 - \frac{7}{2} \right)^2 + \frac{1}{6} \left( 6 - \frac{7}{2} \right)^2 \]

\[ = \frac{35}{12} \]

2. Approach #2: A property

\[ E[Y^2] = \frac{1}{6} \left[ 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 \right] = \frac{91}{6} \]

\[ Var(Y) = \frac{91}{6} - \left( \frac{7}{2} \right)^2 = \frac{35}{12} \]
Properties of variance

Definition
\[ \text{Var}(X) = E[(X - E[X])^2] \]

Def standard deviation
\[ \text{SD}(X) = \sqrt{\text{Var}(X)} \]

Property 1
\[ \text{Var}(X) = E[X^2] - (E[X])^2 \]

Property 2
\[ \text{Var}(aX + b) = a^2 \text{Var}(X) \]

- Property 1 is often easier to compute than the definition
- Unlike expectation, variance is not linear
Property 2: A proof

Property 2 \( \text{Var}(aX + b) = a^2 \text{Var}(X) \)

Proof: \( \text{Var}(aX + b) \)

\[
\begin{align*}
    &= E[(aX + b)^2] - (E[aX + b])^2 \\
    &= E[a^2X^2 + 2abX + b^2] - (aE[X] + b)^2 \\
    &= a^2E[X^2] + 2abE[X] + b^2 - (a^2(E[X])^2 + 2abE[X] + b^2) \\
    &= a^2E[X^2] - a^2(E[X])^2 \\
    &= a^2(E[X^2] - (E[X])^2) \\
    &= a^2 \text{Var}(X)
\end{align*}
\]

Factoring/Linearity of Expectation
Bernoulli RV
Jacob Bernoulli (1654-1705), also known as “James”, was a Swiss mathematician. One of many mathematicians in Bernoulli family. The Bernoulli Random Variable is named for him. My academic great\textsuperscript{14} grandfather.
Bernoulli Random Variable

Consider an experiment with two outcomes: “success” and “failure.”

A Bernoulli random variable $X$ maps “success” to 1 and “failure” to 0. Other names: indicator random variable, boolean random variable

- **PMF**
  - $P(X = 1) = p(1) = p$
  - $P(X = 0) = p(0) = 1 - p$

- **Support**: $\{0, 1\}$

- **Expectation**
  - $E[X] = p$

- **Variance**
  - $\text{Var}(X) = p(1 - p)$

**Examples:**
- Coin flip
- Random binary digit
- Whether a disk drive crashed

Remember this nice property of expectation. It will come back!
Defining Bernoulli RVs

Run a program
- Crashes w.p. $p$
- Works w.p. $1 - p$

Let $X$: 1 if crash

$$X \sim \text{Ber}(p)$$

$$P(X = 1) = p$$

$$P(X = 0) = 1 - p$$

Serve an ad.
- User clicks w.p. 0.2
- Ignores otherwise

Let $X$: 1 if clicked

$$X \sim \text{Ber}(\___)$$

$$P(X = 1) = \___$$

$$P(X = 0) = \___$$

Roll two dice.
- Success: roll two 6’s
- Failure: anything else

Let $X$: 1 if success

$$X \sim \text{Ber}(\___)$$

$$E[X] = \___$$

$p_X(1) = p$

$p_X(0) = 1 - p$
Defining Bernoulli RVs

Run a program
• Crashes w.p. \( p \)
• Works w.p. \( 1 - p \)

Let \( X \): 1 if crash

\[
X \sim \text{Ber}(p)
\]

\[
P(X = 1) = p
\]

\[
P(X = 0) = 1 - p
\]

Serve an ad.
• User clicks w.p. 0.2
• Ignores otherwise

Let \( X \): 1 if clicked

\[
X \sim \text{Ber}(__)
\]

\[
P(X = 1) = __
\]

\[
P(X = 0) = __
\]

Roll two dice.
• Success: roll two 6’s
• Failure: anything else

Let \( X \): 1 if success

\[
X \sim \text{Ber}(__)
\]

\[
E[X] = __
\]
Binomial RV
Binomial Random Variable

Consider an experiment: $n$ independent trials of $\text{Ber}(p)$ random variables. A **Binomial** random variable $X$ is the number of successes in $n$ trials.

\[
X \sim \text{Bin}(n, p)
\]

**PMF**

\[
P(X = k) = p(k) = \binom{n}{k} p^k (1 - p)^{n-k}
\]

**Expectation**

\[
E[X] = np
\]

**Variance**

\[
\text{Var}(X) = np(1 - p)
\]

Examples:

- # heads in $n$ coin flips
- # of 1's in randomly generated length $n$ bit string
- # of disk drives crashed in 1000 computer cluster (assuming disks crash independently)
Reiterating notation

The parameters of a Binomial random variable:
- \( n \): number of independent trials
- \( p \): probability of success on each trial

\[ X \sim \text{Bin}(n, p) \]
Reiterating notation

\[ X \sim \text{Bin}(n, p) \]

If \( X \) is a binomial with parameters \( n \) and \( p \), the PMF of \( X \) is

\[ P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \]

- Probability that \( X \) takes on the value \( k \)
- **Probability Mass Function** for a Binomial
Three coin flips

Three fair (“heads” with $p = 0.5$) coins are flipped.

- $X$ is number of heads
- $X \sim \text{Bin}(3, 0.5)$

Compute the following event probabilities:

$P(X = 0)$

$P(X = 1)$

$P(X = 2)$

$P(X = 3)$

$P(X = 7)$

$P(\text{event})$
Three coin flips

Three fair (“heads” with $p = 0.5$) coins are flipped.

- $X$ is number of heads
- $X \sim \text{Bin}(3, 0.5)$

Compute the following event probabilities:

- $P(X = 0) = p(0) = \binom{3}{0} p^0 (1 - p)^3 = \frac{1}{8}$
- $P(X = 1) = p(1) = \binom{3}{1} p^1 (1 - p)^2 = \frac{3}{8}$
- $P(X = 2) = p(2) = \binom{3}{2} p^2 (1 - p)^1 = \frac{3}{8}$
- $P(X = 3) = p(3) = \binom{3}{3} p^3 (1 - p)^0 = \frac{1}{8}$
- $P(X = 7) = p(7) = 0$

P(event) PMF

Extra math note:
By Binomial Theorem, we can prove
$\sum_{k=0}^{n} P(X = k) = 1$
Binomial Random Variable

Consider an experiment: \( n \) independent trials of \( \text{Ber}(p) \) random variables.

**def** A **Binomial** random variable \( X \) is the number of successes in \( n \) trials.

\[
X \sim \text{Bin}(n, p)
\]

**PMF**

\[
P(X = k) = p(k) = \binom{n}{k} p^k (1 - p)^{n-k}
\]

**Range:** \( \{0, 1, \ldots, n\} \)

**Expectation**

\[
E[X] = np
\]

**Variance**

\[
\text{Var}(X) = np(1 - p)
\]

**Examples:**

- # heads in \( n \) coin flips
- # of 1’s in randomly generated length \( n \) bit string
- # of disk drives crashed in 1000 computer cluster (assuming disks crash independently)
Binomial RV is sum of Bernoulli RVs

Bernoulli
• $X \sim \text{Ber}(p)$

Binomial
• $Y \sim \text{Bin}(n, p)$
• The sum of $n$ independent Bernoulli RVs

\[ Y = \sum_{i=1}^{n} X_i, \quad X_i \sim \text{Ber}(p) \]

$\text{Ber}(p) = \text{Bin}(1, p)$
Consider an experiment: \( n \) independent trials of \( \text{Ber}(p) \) random variables. A **Binomial** random variable \( X \) is the number of successes in \( n \) trials.

\[
X \sim \text{Bin}(n, p)
\]

**PMF**

\[
P(X = k) = p(k) = \binom{n}{k} p^k (1 - p)^{n-k}
\]

**Expectation**

\[
E[X] = np
\]

**Variance**

\[
\text{Var}(X) = np(1 - p)
\]

**Examples:**

- # heads in \( n \) coin flips
- # of 1’s in randomly generated length \( n \) bit string
- # of disk drives crashed in 1000 computer cluster (assuming disks crash independently)

**Proof:**

\[
E[X] = np
\]

\[
\text{Var}(X) = np(1 - p)
\]
## Binomial Random Variable

Consider an experiment: \( n \) independent trials of \( \text{Ber}(p) \) random variables.

**def** A **Binomial** random variable \( X \) is the number of successes in \( n \) trials.

### \( X \sim \text{Bin}(n, p) \)

- **PMF**
  \[
P(X = k) = p(k) = \binom{n}{k} p^k (1 - p)^{n-k}
  \]

- **Expectation**
  \[
  E[X] = np
  \]

- **Variance**
  \[
  \text{Var}(X) = np(1 - p)
  \]

### Examples:
- # heads in \( n \) coin flips
- # of 1’s in randomly generated length \( n \) bit string
- # of disk drives crashed in 1000 computer cluster (assuming disks crash independently)

We’ll prove this later in the course.
To simplify the algebra a bit, let \( q = 1 - p \), so \( p + q = 1 \).

So:

\[
E(X^2) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k}
\]

\[
= \sum_{k=0}^{n} k n \binom{n-1}{k-1} p^k q^{n-k}
\]

\[
= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)}
\]

\[
= np \sum_{m=0}^{n} (j+1) \binom{m}{j} p^j q^{m-j}
\]

\[
= np \left( \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j} + \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j} \right)
\]

\[
= np \left( \sum_{j=0}^{m} \binom{m}{j} p^{j-1} q^{(m-1)-(j-1)} + \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j} \right)
\]

\[
= np \left( (n-1)p \sum_{j=1}^{m} \binom{m-1}{j-1} p^{j-1} q^{(m-1)-(j-1)} + \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j} \right)
\]

\[
= np \left( (n-1)p LP + q^{m-1} \right)
\]

\[
= np(n-1)p + (p+q)^m
\]

\[
= np(n-1)p + p + q
\]

\[
= n^2 p^2 + np(1-p)
\]

Then:

\[
\text{var}(X) = E(X^2) - (E(X))^2
\]

\[
= np(1-p) + n^2 p^2 - (np)^2
\]

\[
= np(1-p)
\]

as required.
07: Variance, Bernoulli, and Binomial

Lisa Yan and Jerry Cain
September 28, 2020
Our first common RVs

1. The random variable $X \sim \text{Ber}(p)$
2. is distributed as varies as a
3. Bernoulli
4. with parameter

Example: Heads in one coin flip, $P(\text{heads}) = 0.8 = p$

Example: # heads in 40 coin flips, $P(\text{heads}) = 0.8 = p$

otherwise

Identify PMF, or identify as a function of an existing random variable
Check out the questions on the next slide (Slide 37). Post any clarifications here!

https://us.edstem.org/courses/2678/discussion/134630

Breakout rooms: 5 min. Introduce yourself!
Statistics: Expectation and variance

1. a. Let \( X \) = the outcome of a fair 4-sided die roll. What is \( E[X] \)?
   
   b. Let \( Y \) = the sum of three rolls of a fair 4-sided die. What is \( E[Y] \)?

2. Let \( Z \) = \# of tails on 10 flips of a biased coin (w.p. 0.4 of heads). What is \( E[Z] \)?

3. Compare the variances of \( B_1 \sim \text{Ber}(0.1) \) and \( B_2 \sim \text{Ber}(0.5) \).
Statistics: Expectation and variance

1. a. Let $X = \text{the outcome of a fair 4-sided die roll. What is } E[X]?$
   b. Let $Y = \text{the sum of three rolls of a fair 4-sided die. What is } E[Y]?$

2. Let $Z = \# \text{ of tails on 10 flips of a biased coin (w.p. 0.4 of heads). What is } E[Z]?$

3. Compare the variances of $B_1 \sim \text{Ber}(0.1)$ and $B_2 \sim \text{Ber}(0.5)$.

If you can identify common RVs, just look up statistics instead of re-deriving from definitions.
Slide 40 has a matching question to go over by yourself. We’ll go over it together afterwards.

Post any clarifications here!

https://us.edstem.org/courses/2678/discussion/134630

Think by yourself: 1 min

Type your answer in Zoom chat but don’t press enter until time is up
Example: 1: A,   2: B,   3: C,   4: D
Visualizing Binomial PMFs

Match the distribution of $X$ to the graph:

1. Bin(10, 0.5)
2. Bin(10, 0.3)
3. Bin(10, 0.7)
4. Bin(5, 0.5)

Type your answer in Zoom chat but don’t press enter until time is up

Example: 1: A, 2: B, 3: C, 4: D

(by yourself)
Visualizing Binomial PMFs

Match the distribution of $X$ to the graph:
1. Bin(10, 0.5)
2. Bin(10, 0.3)
3. Bin(10, 0.7)
4. Bin(5, 0.5)
Binomial RV is sum of Bernoulli RVs

Bernoulli
• \( X \sim \text{Ber}(p) \)

Binomial
• \( Y \sim \text{Bin}(n, p) \)
• The sum of \( n \) independent Bernoulli RVs

\[
Y = \sum_{i=1}^{n} X_i, \quad X_i \sim \text{Ber}(p)
\]
Galton Board

http://web.stanford.edu/class/cs109/demos/galton.html
Slide 45 has a question to go over by yourself.

Post any clarifications here!

https://us.edstem.org/courses/2678/discussion/134630

Think by yourself: 1 min
When a marble hits a pin, it has an equal chance of going left or right. Let $B$ = the bucket index a ball drops into. What is the distribution of $B$?

$X \sim \text{Bin}(n, p) \quad p(k) = \binom{n}{k} p^k (1 - p)^{n-k}$

(Interpret: If $B$ is a common random variable, report it, otherwise report PMF)
A Galton Board is shown with 5 levels, and each level has a series of pins. When a marble hits a pin, it has an equal chance of going left or right. Let $B$ be the bucket index a ball drops into. What is the distribution of $B$?

- Each pin is an independent trial.
- One decision is made for level $i = 1, 2, \ldots, 5$.
- Consider a Bernoulli RV with success $R_i$ if the ball went right on level $i$.
- The bucket index $B = \#$ times the ball went right.

For $n = 5$, the distribution is $B \sim \text{Bin}(n = 5, p = 0.5)$.

Mathematically:
\[ X \sim \text{Bin}(n, p) \quad p(k) = \binom{n}{k} p^k (1 - p)^{n-k} \]
When a marble hits a pin, it has an equal chance of going left or right. Let $B = \text{the bucket index a ball drops into.}$ $B$ is distributed as a Binomial RV, 

$$B \sim \text{Bin}(n = 5, p = 0.5)$$

Calculate the probability of a ball landing in bucket $k$.

$$P(B = 0) = \binom{5}{0} 0.5^5 \approx 0.03$$

$$P(B = 1) = \binom{5}{1} 0.5^5 \approx 0.16$$

$$P(B = 2) = \binom{5}{2} 0.5^5 \approx 0.31$$
Interlude for jokes/announcements

Python tutorial #2
When: Wed 9/30 3:30-4:30pm PT
Recorded?: Yes
Covers: PS2 content
Notes: to be posted online
Think, then
Breakout Rooms

Check out the questions on the next slide (Slide 50). Post any clarifications here!

https://us.edstem.org/courses/2678/discussion/134630

By yourself: 2 min

Breakout rooms: 5 min.
Genetics and NBA Finals

1. Each person has 2 genes per trait (e.g., eye color).
   • Child receives 1 gene (equally likely) from each parent
   • Brown is “dominant”, blue is ”recessive”:
     • Child has brown eyes if either (or both) genes are brown
     • Blue eyes only if both genes are blue.
   • Parents each have 1 brown and 1 blue gene.

A family has 4 children. What is P(exactly 3 children with brown eyes)?

2. The LA Lakers are going to play the Miami Heat in a 7-game series during the 2019 NBA finals.
   • The Lakers have a probability of 58% of winning each game, independently.
   • A team wins the series if they win at least 4 games (we play all 7 games).

What is P(Lakers winning)?
Genetic inheritance

Each person has 2 genes per trait (e.g., eye color).

- Child receives 1 gene (equally likely) from each parent
- Brown is “dominant”, blue is ”recessive”:
  - Child has brown eyes if either (or both) genes are brown
  - Blue eyes only if both genes are blue.
- Parents each have 1 brown and 1 blue gene.

A family has 4 children. What is \( P(\text{exactly 3 children with brown eyes}) \)?

**Big Q:** Fixed parameter or random variable?

**Parameters**  What is **common** among all outcomes of our experiment?

**Random variable** What **differentiates** our event from the rest of the sample space?
Each person has 2 genes per trait (e.g., eye color).

- Child receives 1 gene (equally likely) from each parent
- Brown is “dominant”, blue is ”recessive”:
  - Child has brown eyes if either (or both) genes are brown
  - Blue eyes only if both genes are blue.
- Parents each have 1 brown and 1 blue gene.

A family has 4 children. What is $P$(exactly 3 children with brown eyes)?

1. Define events/ RVs & state goal
2. Identify known probabilities
3. Solve

$X$: # brown-eyed children,
$X \sim \text{Bin}(4, p)$

$p$: $P$(brown-eyed child)

Want: $P(X = 3)$
NBA Finals

The LA Lakers are going to play the Miami Heat in a 7-game series during the 2020 NBA finals.
- The Lakers have a probability of 58% of winning each game, independently.
- A team wins the series if they win at least 4 games (we play all 7 games).

What is $P(\text{Lakers winning})$?

1. Define events/RVs & state goal

   $X$: # games Lakers win
   $X \sim \text{Bin}(7, 0.58)$

   Want:

   **Big Q:** Fixed parameter or random variable?
   - **Parameters**
     - # of total games
     - prob. Lakers winning a game
   - **Random variable**
     - # of games Lakers win
   - **Event based on RV**
     - Lakers win 4 or more games
NBA Finals

The LA Lakers are going to play the Miami Heat in a 7-game series during the 2020 NBA finals.

- The Lakers have a probability of 58% of winning each game, independently.
- A team wins the series if they win at least 4 games (we play all 7 games).

What is \( P(\text{Lakers winning}) \)?

1. Define events/RVs & state goal
   - \( X \): # games Lakers win
   - \( X \sim \text{Bin}(7, 0.58) \)
   - Want: \( P(X \geq 4) \)

2. Solve
   \[
P(X \geq 4) = \sum_{k=4}^{7} P(X = k) = \sum_{k=4}^{7} \binom{7}{k} 0.58^k (0.42)^{7-k}
   \]

Cool Algebra/Probability Fact: this is identical to the probability of winning if we define winning = first to win 4 games.
See you next time