Lesson 14: Variance

Meike Niederhausen and Nicky Wakim

2025-11-12

Learning Objectives

1. Define and calculate the expected value for a function of discrete and continuous RVs
2. Define and calculate variance for a single random variable
3. Define and calculate variance for multiple random variables

Where are we?

Basics of probability

- Outcomes and events
- Sample space
- Simulations
- Probability axioms
- Probability properties
- Counting
- Rules of probability

Probability for random variables

- Probability functions: pmfs/pdfs
- Cumulative distribution functions (CDFs)
- Important distributions

Two random variables

- Transformations
- Joint distributions
- Independence and conditioning

Expected values

- Means / expected values
- Linear combinations and conditioning
- Variance

Advanced probability

- Central limit theorem
- Functions: moment generating functions

Learning Objectives

1. Define and calculate the expected value for a function of discrete and continuous RVs
2. Define and calculate variance for a single random variable
3. Define and calculate variance for multiple random variables

Let's start building the variance through expected values of functions

Example 1

Let g be a function and let $g(x) = ax + b$, for real-valued constants a and b. What is $\mathbb{E}[g(X)]$?

X is RV a & b constants

$g(X)$ is fn of RV X

$E(g(X)) = E[aX \pm b]$

$= E(aX) + E(b)$ b/c constant

$= E(a)E(X) + b$

$= aE(X) + b$

works b/c LINEAR fn

what about $g(X) = X^2$

$E(g(X)) = E(X^2)$

$\neq [E(X)]^2$

What is the expected value of a function?

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

Expected value of function of discrete RV

For any function g and discrete RV X, the expected value of $g(X)$ is

$$\mathbb{E}[g(X)] = \sum_{\{all\ x\}} g(x)p_X(x)$$

Expected value of function of continuous RV

For any function g and continuous RV X, the expected value of $g(X)$ is

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

- For example, if we have $g(x) = x^2$, then

$$\mathbb{E}[X^2] = \sum_{\{all\ x\}} x^2 p_X(x) \neq \left(\sum_{\{all\ x\}} x p_X(x)\right)^2 = (\mathbb{E}[X])^2$$

Let's revisit the card example (1/2)

Example 2

Suppose you draw 2 cards from a standard deck of cards *with* replacement. Let X be the number of hearts you draw.

1. Find $\mathbb{E}[X^2]$.

Recall Binomial RV with $n = 2$:

$$p_X(x) = \binom{2}{x} p^x (1-p)^{2-x} \text{ for } x = 0, 1, 2$$

$P(B) = p = 0.25$

$$E(g(X)) = E(X^2)$$

$$= \sum_{\{\text{all } x\}} x^2 \, p_X(x)$$

$$= \sum_{x=0}^{2} x^2 \left[\binom{2}{x} 0.25^x (1-0.25)^{2-x} \right]$$

$$= \cancel{0^2 \binom{2}{0} 0.25^0 \, 0.75^2}^{\,0} + 1^2 \binom{2}{1} 0.25 \cdot 0.75$$

$$+ 2^2 \binom{2}{2} 0.25^2 \, 0.75^0$$

$$E(X^2) = \frac{5}{8}$$

$$\rightarrow E(X^2) \neq E(X)^2 = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$$

Let's revisit the card example (2/2)

Example 2

Suppose you draw 2 cards from a standard deck of cards *with* replacement. Let X be the number of hearts you draw.

2. Find $\mathbb{E}\left[\left(X-\frac{1}{2}\right)^2\right]$.

Recall Binomial RV with $n=2$:

$$p_X(x) = \binom{2}{x} p^x (1-p)^{2-x} \text{ for } x = 0, 1, 2$$

$$g(X) = \left(X - \tfrac{1}{2}\right)^2$$

$$E(g(X)) = E\left[\left(X-\tfrac{1}{2}\right)^2\right] = \sum_{\{\text{all } x\}} \left(x-\tfrac{1}{2}\right)^2 p_X(x)$$

$$= \sum_{x=0}^{2} \left(x-\tfrac{1}{2}\right)^2 \binom{2}{x} 0.25^x \, 0.75^{2-x}$$

$$= \left(0-\tfrac{1}{2}\right)^2 \binom{2}{0} 0.25^0 \, 0.75^2 + \left(1-\tfrac{1}{2}\right)^2 \binom{2}{1} 0.25^1 \, 0.75^1$$

$$+ \left(2-\tfrac{1}{2}\right)^2 \binom{2}{2} 0.25^2 \, 0.75^0$$

$$E\left[\left(X-\tfrac{1}{2}\right)^2\right] = 3/8$$

ANOTHER WAY

prev slide　　last class

$$E\left(\left(X-\tfrac{1}{2}\right)^2\right) = E\left[X^2 - X + \tfrac{1}{4}\right] = E(X^2) - E(X) + \underbrace{E\left(\tfrac{1}{4}\right)}_{=\frac{1}{4}}$$

$$= \frac{5}{8} - \left(\frac{1}{2}\right) + \frac{1}{4} = \frac{3}{8}$$

Learning Objectives

1. Define and calculate the expected value for a function of discrete and continuous RVs
2. Define and calculate variance for a single random variable
3. Define and calculate variance for multiple random variables

Variance of a RV

in units of X^2

Definition: Variance of RV

The variance of a RV X, with (finite) expected value $\mu_X = \mathbb{E}[X]$ is

$$\sigma_X^2 = Var(X) = \mathbb{E}[(X - \mu_X)^2] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

$$g(X) = (X - \mu_X)^2$$
$$= (X - E(X))^2$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

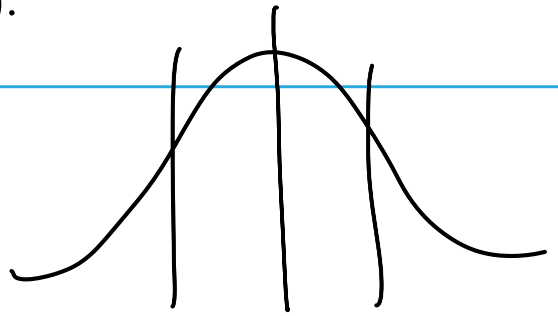
avg deviation squared

Definition: Standard deviation of RV

The standard deviation of a RV X is

$$\sigma_X = SD(X) = \sqrt{\sigma_X^2} = \sqrt{Var(X)}.$$

in the units of X



Variance of discrete and continuous RVs

$g(X) = (X - \mu_X)^2$

How do we calculate the variance of a discrete RV?

For discrete RVs:

$$
\begin{aligned}
Var(X) &= \\
&= \sum_{\{all\ x\}} \boxed{(x-\mu_x)^2} p_X(x) \\
&= \mathbb{E}[(X-\mu_X)^2] \\
&= \mathbb{E}[(X-\mathbb{E}[X])^2] \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2
\end{aligned}
$$

How do we calculate the variance of a continuous RV?

For continuous RVs:

$$
\begin{aligned}
Var(X) &= \int_{-\infty}^{\infty} \boxed{(x-\mu_X)^2} f_X(x)dx \\
&= \mathbb{E}[(X-\mu_X)^2] \\
&= \mathbb{E}[(X-\mathbb{E}[X])^2] \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2
\end{aligned}
$$

[var of X is exp val of sq − sq of exp val]

Let's calculate the variance and prove it! (1/2)

$\mu_X = E(X)$

Lemma 6: "Computation formula" for Variance

The variance of a RV X, can be computed as

$$\sigma_X^2 = Var(X) = \mathbb{E}[X^2] - \mu_X^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

$$var(X) = E((X-\mu_X)^2)$$

$$= E(X^2 - 2\mu_X X + \mu_X^2)$$

$$= \sum_{\{all\, x\}} (x^2 - 2\mu_X x + \mu_X^2) p_X(x)$$

$$\left\{ \int_{-\infty}^{\infty} (x^2 - 2\mu_X x + \mu_X^2) f_X(x)\, dx \right\}$$

$$= \sum_{\{all\, x\}} x^2 p_X(x) - 2\sum_{\{all\, x\}} \mu_X x p_X(x) + \sum_{\{all\, x\}} \mu_X^2 p_X(x)$$

$$\sum_{\{all\, x\}} x p_X(x)$$

$$g(X) = X^2$$

$$= \underbrace{\sum_{\{all\, x\}} x^2 p_X(x)}_{E(X^2)} - 2\mu_X \underbrace{\sum_{\{all\, x\}} x p_X(x)}_{E(X)} + \mu_X^2 \underbrace{\sum_{\{all\, x\}} p_X(x)}_{=1}$$

Let's calculate the variance and prove it! (2/2)

$$= E(X^2) - 2\mu_X E(X) + \mu_X^2$$

$\mu_X = E(X)$

$$= E(X^2) - 2\mu_X^2 + \mu_X^2$$

$$= E(X^2) - \mu_X^2$$

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

variance is the (exp val of x^2) − (the square of the exp val)

Variance of an Uniform distribution

Example 2

Let $f_X(x) = \frac{1}{b-a}$, for $a \leq x \leq b$. Find $Var[X]$.

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a}\,dx$$

$$= \frac{1}{3(b-a)} x^3 \Big|_{x=a}^{x=b}$$

$$= \frac{1}{3(b-a)}(b^3 - a^3)$$

$$= \frac{\cancel{(b-a)}(b^2+ab+a^2)}{3\cancel{(b-a)}}$$

$$= \frac{b^2+ab+a^2}{3}$$

last $E(X) = \frac{a+b}{2}$

could do:

$$Var(X) = \int_a^b (x-\mu_x)^2 \frac{1}{b-a}\,dx$$

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

$$= \frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{4(b^2+ab+a^2)}{4(3)} - \frac{(a^2+2ab+b^2)3}{(4)3}$$

$$= \frac{4b^2+4ab+4a^2-(3a^2+6ab+3b^2)}{12}$$

$$= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

Variance of exponential distribution

In the homework:

Example 3

Let $f_X(x) = \lambda e^{-\lambda x}$, for $x > 0$ and $\lambda > 0$. Find $Var[X]$.

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

$$E(X) \to \frac{1}{\lambda}$$

$$= \left(\int_0^\infty x^2 \underline{\lambda e^{-\lambda x}} dx\right) - \left(\frac{1}{\lambda}\right)^2$$

$\vdots$ int by parts $2x$

$$= \frac{1}{\lambda^2}$$

Learning Objectives

1. Define and calculate the expected value for a function of discrete and continuous RVs
2. Define and calculate variance for a single random variable
3. Define and calculate variance for multiple random variables

Variance of a function with a single RV

Lemma 7

For a RV X and constants a and b,

$g(X) = aX + b$

$$Var(aX + b) = a^2 Var(X).$$

Proof will be exercise in homework. It's fun! In a mathy kinda way.

Important results for *independent* RVs

Theorem 8

For independent RV's X and Y, and functions g and h,

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)].$$

* **Corollary 1**

For independent RV's X and Y,

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

if $A \perp B$,

$P(A \cap B) = P(A)P(B)$

$$E(g(X)h(Y)) = \sum_{\{\text{all } x\}} \sum_{\{\text{all } y\}} g(x)h(y) \underbrace{p_{X,Y}(x,y)}_{p_X(x)\,p_Y(y)} = E(g(X))E(h(Y))$$

Variance of sum of independent discrete RVs

Theorem 9: Variance of sum of independent discrete RV's

For independent ~~discrete~~ RV's X_i and constants a_i, $i = 1, 2, \ldots, n$,

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i).$$

Simpler version:

$$Var(a_1X + a_2Y) = Var(a_1X) + Var(a_2Y) = a_1^2 Var(X) + a_2^2 Var(Y)$$

linearity of variance
for IND RVs

Corollaries

Corollary 2

For independent ~~discrete~~ RV's $X_i, i = 1, 2, \ldots, n,$

Var of sum = sum of var

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i).$$

Corollary 3

For independent identically distributed (i.i.d.) ~~discrete~~ RV's $X_i, i = 1, 2, \ldots, n,$

$X_i \overset{iid}{\sim} N(\mu, \sigma)$

iid

$$Var\left(\sum_{i=1}^{n} X_i\right) = nVar(X_1).$$

Ghost problem: *with replacement*

Example 3.2

The ghost is trick-or-treating at a different house ~~now~~. In this case it is known that the bag of candy has 10 chocolates, 20 lollipops, and 30 laffy taffies. The ghost grabs a handful of five pieces of candy. What is the variance for the number of chocolates the ghost takes? Let's solve this for the cases *with* replacement.

$p = \frac{10}{60} = \frac{1}{6}$

Recall probability with replacement:

binomial

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

$X = \#$ chocolates (w/ rep)

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

$X = \sum_{i=1}^{5} Y_i$

$$\boxed{\begin{array}{l} Y_i \overset{iid}{\sim} \text{Bernoulli}(p = \frac{1}{6}) \\ P_Y(y) = p^y (1-p)^{1-y} \\ Y_i = \begin{cases} 1 & \text{choc} \\ 0 & \text{not choc} \end{cases} \end{array}}$$

$Var(X) = Var\left(\sum_{i=1}^{5} Y_i\right) = \sum_{i=1}^{5} Var(Y_i) = 5 Var(Y_1)$

$* \; Var(Y_1) = E(Y_1^2) - [E(Y)]^2$

$\hookrightarrow E(Y) = p$

$E(Y^2) = \sum_{y=0}^{1} y^2 P_Y(y) = 0^2 p^0 (1-p)^1 + 1^2 p (1-p)^0 = p$

$Var(Y_1) = p - p^2 = p(1-p)$

~~Find the mean and sd from word problem (1/2)~~

$$\begin{aligned} Var(X) &= 5\,Var(Y) \\ &= 5p\cdot(1-p) \\ &= 5\left(\tfrac{1}{6}\right)\left(1-\tfrac{1}{6}\right) \\ &= 0.6944 \end{aligned}$$

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

Var of binomial

= n · var of bern

Back to our hotel example from Lesson 13

$E(T) = \$6,600$

Example 4

var= $100

A tour group is planning a visit to the city of Minneapolis and needs to book 30 hotel rooms. The average price of a room is $200 **with standard deviation $10.** In addition, there is a 10% tourism tax for each room. What is the **standard deviation** of the cost for the 30 hotel rooms? Assume rooms are independent

~~iid~~

mispoke, not IID

Problem to do at home if we don't have enough time.

Let T = total cost of 30 rooms

C_i = cost of room i

$$T = \sum_{i=1}^{30} 1.1\, C_i$$

$$Var(T) = Var\left(\sum_{i=1}^{30} 1.1\, C_i\right)$$

$$= Var\left(1.1 \sum_{i=1}^{30} C_i\right)$$

$$= 1.1^2\, Var\left(\sum_{i=1}^{30} C_i\right)$$

$$= 1.1^2 \sum_{i=1}^{30} Var(C_i) = 1.1^2 \cdot 30\, Var(C_i)$$

$$= 1.1^2 \cdot 30 \cdot 100 = \$^2\, 3,600$$

WRONG

$$T = 30 \cdot 1.1 \cdot C_i$$

$$Var(30 \cdot 1.1\, C_i)$$

$$= (30 \cdot 1.1)^2\, Var(C_i)$$

even though not IID, $\sum var = 30 \cdot var$ b/c all same var

$$SD(T) = \sqrt{Var(T)} = \sqrt{\$^2\, 3600} = \$60$$

Find the mean and sd from word problem (1/2)

Example 4

A machine manufactures cubes with a side length that varies uniformly from 1 to 2 inches. Assume the sides of the base and height are equal. The cost to make a cube is 10 ¢ per cubic inch, and 5 ¢ cents for the general cost per cube. Find the mean and standard deviation of the cost to make 10 cubes.

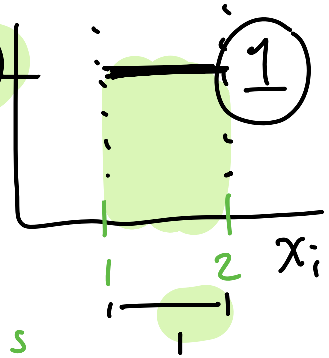
Model cost for 10 cubes:

let C = cost of 10 cubes (in cents)

C_i = cost of cube i ($i = 1, 2, \dots, 10$)

X_i = length of each side of cube i

$$C = \sum_{i=1}^{10} C_i \qquad C_i = 5 + 10(X_i^3)$$



MEAN

$$E(C) = E\left[\sum_{i=1}^{10}(5 + 10X_i^3)\right]$$

linear combos

$$= \sum_{i=1}^{10} E(5 + 10X_i^3)$$

constants 5 & 10

$$= \sum_{i=1}^{10}(5 + 10E(X_i^3))$$

$\neq [E(X_i)]^3$

$$= \sum_{i=1}^{10}\left(5 + 10\left(\frac{15}{4}\right)\right) = \sum_{i=1}^{10} 42.5 = 425 ¢$$

$E(C) = 425 ¢$

$4.25

$$E(X_i^3) = \int_1^2 x_i^3 \, 1 \, dx$$

g(x)

$f_X(x)$ (pdf)

$$= \frac{1}{4} x_i^4 \Big|_{x_i=1}^{x_i=2}$$

$$= \frac{1}{4}(2^4 - 1^4) = \frac{15}{4}$$

Find the mean and sd from word problem (1/2)

$sd(C) = \sqrt{Var(C)}$

3 X_is w/in cube i $\not\perp$

each cube $\perp$

$$Var(C) = Var\left[\sum_{i=1}^{10}(5+10X_i^3)\right] = \sum_{i=1}^{10} Var(5+10X_i^3)$$

each cube i is independent

$$= \sum_{i=1}^{10}\left[\underbrace{Var(5)}_{=0} + Var(10X_i^3)\right] = \sum_{i=1}^{10} 10^2\, Var(X_i^3)$$

10^2

$$= 10^2 \sum_{i=1}^{10}[4.0803]$$

$$= 100 \cdot 10 \cdot 4.0803$$

$$= 4080.36 ¢^2$$

$$sd(C) = \sqrt{4080.36¢^2} = 63.8776¢ \simeq 64¢$$

$sd(C)$

$$Var(X_i^3) = E[\underbrace{(X_i^3)^2}_{X_i^6}] - [E(X_i^3)]^2$$

$$E(X_i^6) = \int_1^2 x_i^6\, 1\, dx_i = \frac{1}{7}x_i^7 \Big|_{x_i=1}^{x_i=2}$$

$g(X) = X_i^6$

$f_X(x) = 1$

$$= \frac{1}{7}(2^7 - 1^7) = \frac{128}{7} - \frac{1}{7}$$

$$= \frac{127}{7}$$

$$Var(X_i^3) = \frac{127}{7} - \left[\frac{15}{4}\right]^2 = 4.0803$$