# Chapter 12: Variance of Discrete RVs

Meike Niederhausen and Nicky Wakim

2023-10-23

# Table of contents

# Learning Objectives

1. Calculate the variance and standard deviation of discrete random variables
2. Calculate the variance of sums of discrete random variables
3. Calculate the variance of functions of discrete random variables

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

**Example 1**

small $x$

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

$X$ is a RV

a & b are constants

$g(x)$ is function of RV

$$E[g(X)] = E[aX + b]$$

big $X$

exp val of function is of RV

linearity of exp value

$$= E[aX] + E(b)$$

a & b are constants

exp value of constant = constant

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

exp value of const. times RV is const. times exp RV

if NOT linear function: $E[g(X)] = E[X^2]$

$g(x) = x^2$

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

# What is the expected value of a function?

**Definition: Expected value of function of RV**

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

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

if $g(x) = x^2$ :

$$E(X^2) = \underbrace{\sum_{\{\text{all } x\}} x^2 p_X(x)}_{E(X^2)} \neq \Big[ \underbrace{\sum_{\{\text{all } x\}} x p_X(x)}_{E(X)} \Big]^2$$

sq.

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

$p(\heartsuit) = \frac{13}{52} = \frac{1}{4}$

**Example 2**

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

$g(x) = x^2$

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

Recall Binomial RV with $n = 2$:

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

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

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

$$= \sum_{x=0}^{2} x^2 \binom{2}{x} p^x (1-p)^{2-x}$$

$$= 0^2 \binom{2}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^2 + 1^2\binom{2}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^1 + 2^2\binom{2}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^0$$

$$E[X^2] = 5/8$$

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

getting one step closer to fn that defines variance

**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 $n = 2$:

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

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

↳ chp 11, ex 1

$$E(X) = \tfrac{1}{2}$$

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

$$= \sum_{\{\text{all } x\}} \left(x - \tfrac{1}{2}\right)^2 \binom{2}{x} \left(\tfrac{1}{4}\right)^x \left(\tfrac{3}{4}\right)^{2-x}$$

$$= \left(0 - \tfrac{1}{2}\right)^2 \binom{2}{0} \left(\tfrac{1}{4}\right)^0 \left(\tfrac{3}{4}\right)^2 + \left(1 - \tfrac{1}{2}\right)^2 \binom{2}{1} \left(\tfrac{1}{4}\right)^1 \left(\tfrac{3}{4}\right)^1$$

$$+ \left(2 - \tfrac{1}{2}\right)^2 \binom{2}{2} \left(\tfrac{1}{4}\right)^2 \left(\tfrac{3}{4}\right)^0 = 3/8$$

ANOTHER WAY:

$$E\left[\left(X - \tfrac{1}{2}\right)^2\right] = E\left[X^2 - X + \tfrac{1}{4}\right] = \underbrace{E(X^2)}_{\text{prev slide}} - \underbrace{E(X)}_{\text{chp 11, ex 1}} + \underbrace{E\left(\tfrac{1}{4}\right)}_{1/4} = \tfrac{5}{8} - \tfrac{1}{2} + \tfrac{1}{4}$$

$$= 3/8$$

# Variance of a RV

$$Var(X) = \sum_{\{all\, x\}} (x-\mu_x)^2 \cdot p_X(x)$$

## Definition: Variance of RV

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

$$g(x) = (x-\mu_x)^2$$

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

## Definition: Standard deviation of RV

The standard deviation of a r.v. $X$ is $\sigma_X$

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

measure of spread: what are other measures?

$$E[X-\mu_X] = 0$$

$$E[|X-\mu_X|]$$

# Let's calculate the variance and prove it!

**Lemma 6: "Computation formula" for Variance**

The variance of a r.v. $X$, can be computed as

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

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

$$= E[X^2 - 2X\mu_X + \mu_X^2]$$

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

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

$$\left(\sum x\, p_X(x)\right)^2$$

$$= E(X^2) - 2\sum_{\{\text{all } x\}} x \left(\boxed{\sum_{\{\text{all } y\}} x\, p_X(x)}\right) p_X(x) + \sum_{\{\text{all } x\}} \mu_x^2\, p_X(x)$$

$$= E(X^2) - 2\sum_{\{all\,x\}} x \left(\boxed{\sum_{\{all\,x\}} x\, P_X(x)}\right) P_X(x)$$

$$+ \sum_{\{all\,x\}} \mu_x^2\, P_X(x)$$

$$= E(X^2) - 2\sum_{\{all\,x\}} x\, P_X(x) \sum_{\{all\,x\}} x\, P_X(x) + \sum_{\{all\,x\}} \mu_x^2\, P_X(x)$$

$$= E(X^2) - 2\left[E(X)\right]^2 + \mu_x^2 \underbrace{\sum_{\{all\,x\}} P_x(x)}_{\text{must equal 1}}$$

$$= E(X^2) - 2\left[E(X)\right]^2 + \underbrace{\mu_x^2}_{\left[E(X)\right]^2}$$

$$= E(X^2) - \left[E(X)\right]^2$$

# (break) Some Important Variance and Expected Values Results

# Variance of a function with a single RV

**Lemma 7**

For a r.v. $X$ and constants a and b,

$g(x) = ax + b$

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

shift of b: all X goes to L or R, so does affect our variance.

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

# Important results for *independent* RVs

**Theorem 8**

For independent r.v.'s $X$ and $Y$, and functions g and h,

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

rooted in fact that

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

**Corollary 1**

For independent r.v.'s $X$ and $Y$,

$g(x) = x \quad h(y) = y$

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

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

# Variance of sum of independent discrete RVs

**Theorem 9: Variance of sum of independent discrete r.v.'s**

For independent discrete r.v.'s $X_i$ and constants $a_i, i = 1, 2, \ldots, n$,

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

$$Var(a_1 X + a_2 Y) = var(a_1 X) + var(a_2 Y)$$

$$= a_1^2 var(X) + a_2^2 var(Y)$$

# Corollaries

**Corollary 2**

For independent discrete r.v.'s $X_i, i = 1, 2, \ldots, n,$

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

**Corollary 3**

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

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

# Let's revisit our ghost problems *without replacement*

$E(X) = \frac{5}{6}$

★ hypergeometric

**Example 3.1**

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 ***without*** replacement.

→ Recall probability without replacement:

$$\rightarrow p_X(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

$X$ = # chocolates (out of 5) w/out replacement

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

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

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

$$= 0^2 \frac{\binom{10}{0}\binom{50}{5-0}}{\binom{60}{5}} + 1^2 \frac{\binom{10}{1}\binom{50}{4}}{\binom{60}{5}} + 2^2 \frac{\binom{10}{2}\binom{50}{3}}{\binom{60}{5}}$$

$$+ 3^2 \frac{\binom{10}{3}\binom{50}{2}}{\binom{60}{5}} + 4^2 \frac{\binom{10}{4}\binom{50}{1}}{\binom{60}{5}} + 5^2 \frac{\binom{10}{5}\binom{50}{0}}{\binom{60}{5}}$$

$$= \frac{475}{354} = 1.3418$$

$$E[X^2] - [E(X)]^2 = 1.3418 - \left(\frac{5}{6}\right)^2 = 0.6474$$

$$5\left(\frac{10}{60}\right)\left(\frac{50}{60}\right)\left(\frac{60-5}{60-1}\right)$$ we'll see ↙

# Let's revisit our ghost problems *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.

Recall probability with replacement:

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

Bern: $p_Y(y) = p^y (1-p)^{1-y}$

$X$ = # chocolates (w/ rep)

$X = \sum_{i=1}^{5} Y_i \rightarrow Y_i = \begin{cases} 1 & i\text{th candy is choc. (w/ prob } p) \\ 0 & \text{else} \end{cases}$

$Y_i$'s are ind & identically dist. (iid)

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

linearity

$$= \sum_{i=1}^{5} \text{var}(Y_i)$$

iid

$$= \boxed{5}\ \boxed{\text{var}(Y_1)} \rightarrow \boxed{5}\ \boxed{[p(1-p)]}$$

$\frac{10}{60}$ $\frac{50}{60}$

$$\text{var}(Y_1) = E(Y_1^2) - [\underbrace{E(Y_1)}_{p}]^2$$

$$E(Y_1^2) = \sum_{y=0,1} y^2 p_Y(y) = 0^2 p^0 (1-p)^1 + 1^2 p^1 (1-p)^0 = p$$

$$\text{var}(Y_1) = p - (p)^2 = \boxed{p(1-p)}$$

# Back to our hotel example from Chapter 11

**Example 4**

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?

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