# Chapter 12: Variance of Discrete RVs

Meike Niederhausen and Nicky Wakim

2024-10-30

# 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

# Where are we?

## Basics of probability

- Outcomes and events
- Sample space
- Probability axioms
- Probability properties
- Counting
- Independence
- Conditional probability
- Bayes' Theorem
- Random Variables

## Probability for discrete random variables

- Functions: pmfs/CDFs
- Important distributions
- Joint distributions
- Expected values and variance

## Probability for continuous random variables

- Calculus
- Functions: pdfs/CDFs
- Important distributions
- Joint distributions
- Expected values and variance

## Advanced probability

- Central limit theorem
- Functions: moment generating functions

# 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 a RV $a$ & $b$ constants

$g(x)$ in a fn of RV

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

$= aE(X) + b$

$E[aX] + E[b]$

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

$= aE(X) + b$

$g(x) = x^2$

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

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

take a step back to def'n of expected val

# 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

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

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

note $\neq \left( \sum_{\{all\, x\}} x\, p_X(x) \right)^2$

$[E(X)]^2$

# 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.

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

Recall Binomial RV with $n = 2$:

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

$p \rightarrow p(\heartsuit)$

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

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

$$= \sum_{x=0}^{2} x^2 \left[ \binom{2}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{2-x} \right]$$

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

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

$$E(X^2) = 5/8$$

# 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 $n=2$:

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

from chp 11, ex 1: $E(X)=\frac{1}{2}$

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

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

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

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

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

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

ANOTHER WAY:

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

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

prev slide

know is np from known binom pmf

$$=5/8-\tfrac{1}{2}+\tfrac{1}{4}=\tfrac{3}{8}$$

# Variance of a RV

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

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

avg deviation squared

## Definition: Variance of RV

$$E[g(X)] = E[(X-\mu)^2]$$

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

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

## Definition: Standard deviation of RV

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

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

measures of spread: $E(X-\mu) = 0$

$$E(|X-\mu|)$$

# 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

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

var is the (expected val of $x^2$) − (the square of the exp. val)

$$\sum_{\{all\,x\}} \left( \sum_{\{all\,x\}} x \right) = \sum_{\{all\,x\}} x$$

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

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

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

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

multiply

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

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

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

$$= E(X^2) - 2\underline{\mu_X \mu_X} + \underline{\mu_X{}^2}(1)$$

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

$$\mu_X = E(X)$$

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

# (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 $b$,

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

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

# Important results for *independent* RVs

**Theorem 8**

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

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

**Corollary 1**

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

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

rooted in fact that if $A \perp B$

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

$$E[g(X)h(Y)] = \sum_{\{all\ x\}} \sum_{\{all\ y\}} g(x)h(y) \underbrace{P_{X,Y}(x,y)}_{P_X(x)P_Y(y)}$$

# Variance of sum of independent discrete RVs

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

if $X_i$'s independ:

$Var(\sum X_i) = \sum var(X_i)$

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

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

Simpler version:

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

# Corollaries

**Corollary 2**

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

not nec true that $var(X_1) = var(X_2)$

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i). = var(X_1) + var(X_2) + \cdots + var(X_n)$$

**Corollary 3**

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

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

binomial $(n, p)$

exact same $n$ & $p$

same $\mu, \sigma$

# Let's look at a ghost problem *with replacement*

$n=5$

$X = \#$ chocolates (w/ rep) $\quad p = \frac{10}{60} = \frac{1}{6}$

**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 taffy 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.

$Var(X) = \underline{E(X^2)} - (E(X))^2$

$\underline{X = \sum_{i=1}^{5} Y_i} \qquad Y_i \overset{iid}{\sim} Bern\left(p = \frac{1}{6}\right)$

$\hookrightarrow Y_i = \begin{cases} 1 & \text{choc} \\ 0 & \text{~~not choc~~ all else} \end{cases}$

$Y_i$'s are iid (independent & identically dist)

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

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

Recall probability with replacement:

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

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

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

0

$$Var(Y_1) = E(Y_1^2) - [\underbrace{E(Y_1)}]^2$$

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

$$Var(X) = 5\,var(Y_1)$$

$$= 5p(1-p) \rightarrow var(x) = np(1-p)$$

for $X \sim binom(n,p)$

$$= 5\left(\frac{1}{6}\right)\left(\frac{5}{6}\right) = 0.6944$$

# Back to our hotel example from Chapter 11

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

**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? Assume rooms are independent.

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

Let: $T$ = total cost of 30 rooms

$C_i$ = cost of room $i$

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

X WRONG!

$T = 30 \cdot 1.1 C_i$

$Var(30 \cdot 1.1 C_i)$

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

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

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

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

$$= 1.1^2 \sum_{i=1}^{30} Var(C_i) \rightarrow 10^2 = 100$$

$$= 1.1^2 \sum_{i=1}^{30} (100) = 1.1^2 (30)(100)$$

$$= \$^2\, 3{,}600$$

$$SD = \sqrt{var} = \sqrt{\$^2\, 3{,}600} = \$60.25$$