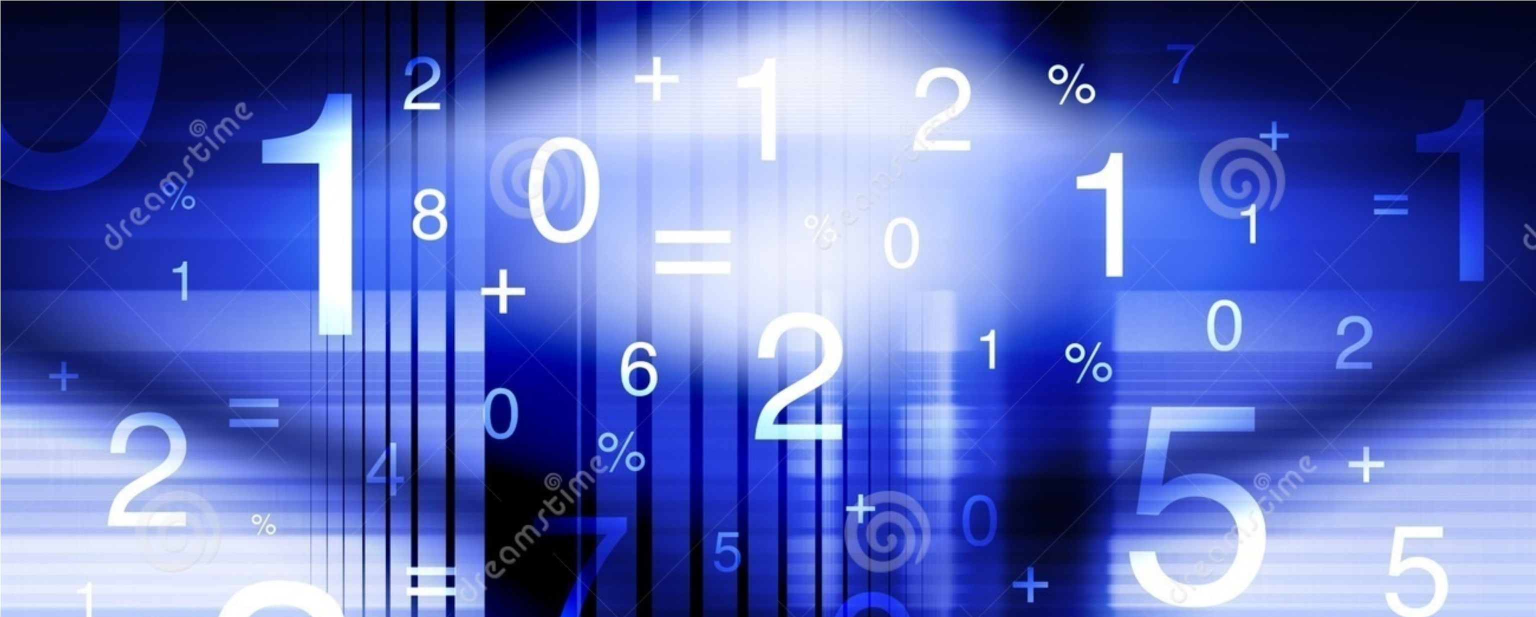
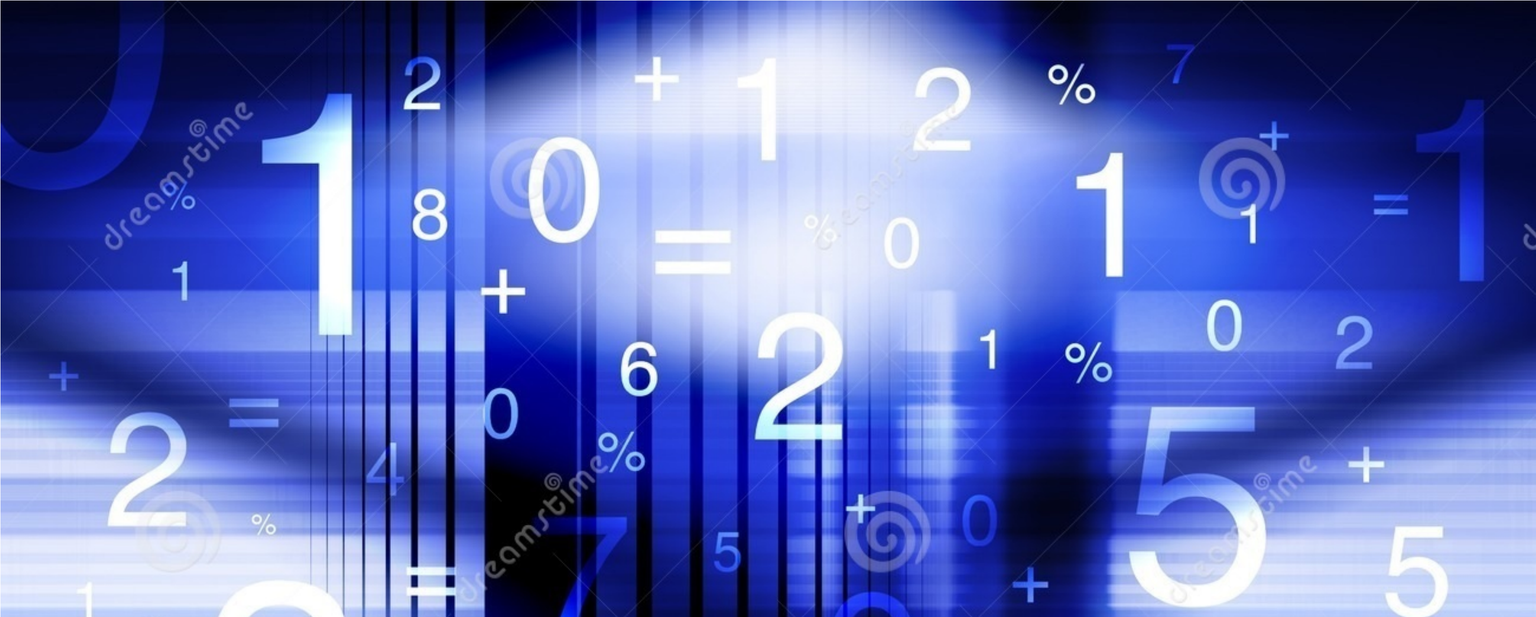
Notes on Probability and Statistics



Contentsmkboth CONTENTSCONTENTS

Probability



2. Random Variables

Often in engineering or the natural sciences, outcomes of random experiments are numbers associated with some physical quantities. Such outcomes, called random variables, will be denoted by **capital letters**, e.g.

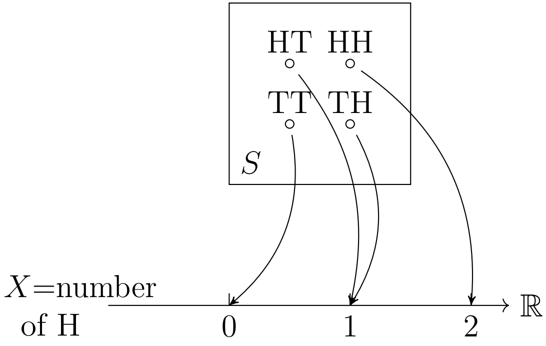
X=“time between Tech Trolleys at the CRC stop” or
Y=“number of customers per day an Uber driver serves”,

and a particular realizations of a random variable by **lowercase letters**, e.g.

$x = 11.4$ min, or $y = 6$ customers.

Random variable: the results of an experiment expressed as a number. Mathematically, a function $X : S \to \mathbb{R}$ that maps points from the sample space to the real line is called a random variable.

There are two types of random variables:
A **discrete** random variable is a rv which takes a finite or countable number of values.
A **continuous** random variable is a rv which takes values in (an interval of) the real line.



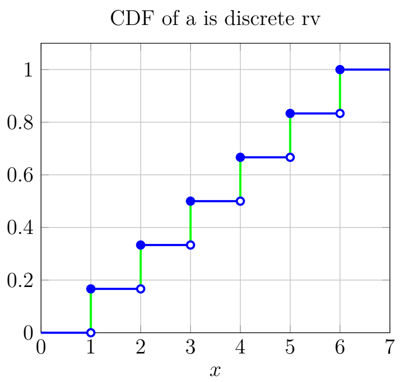
Events are statements of the type "$X \leq x$", "$X > x$", or "$a < X < b$". In general, an event is "$X \in A$" where $A \subseteq \mathbb{R}$.

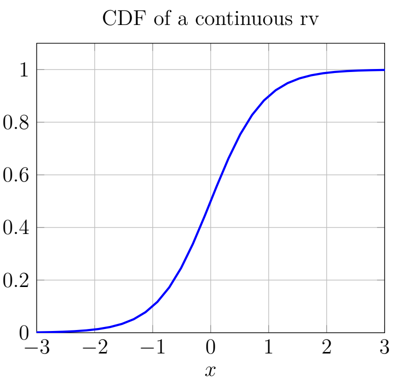
2.1 Probability distribution function

The function

$$F_X(x) = \mathrm{P}(X \leq x), \quad x \in \mathbb{R},$$

is called the **probability distribution**, **cumulative distribution function**, or **CDF** for short.

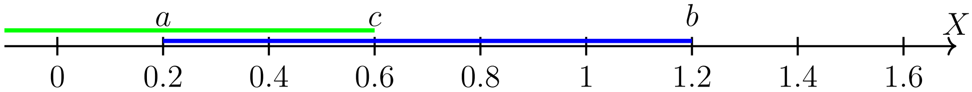




The probability of any statement about the X is computable when $F_X(x)$ is known, e.g.:

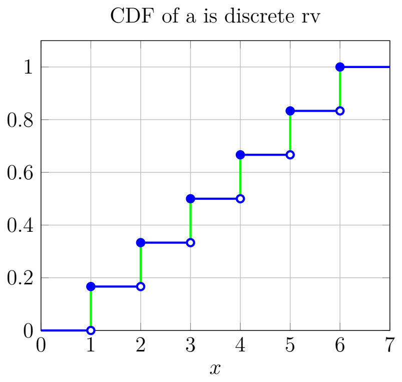
- The complement rule: $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$ becomes $\mathrm{P}(X > a) = 1 - \mathrm{P}(X \leq a) = 1 - F_X(a)$
- Probability of X falling on the interval $(a, b]$: $\mathrm{P}(a < X \leq b) = F_X(b) - F_X(a)$
- Conditional probability: $\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$ becomes

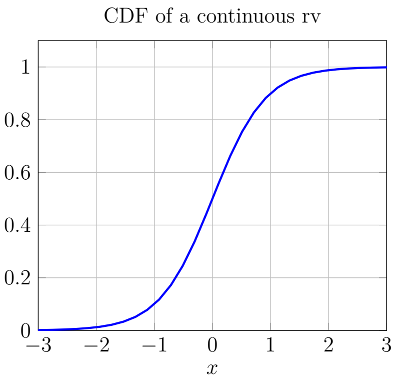
$$\begin{aligned} \mathrm{P}(a < X \leq b \mid X \leq c) &= \frac{\mathrm{P}(a < X \leq b \ \cap \ X \leq c)}{\mathrm{P}(X \leq c)} \\ &= \frac{\mathrm{P}(a < X \leq c)}{\mathrm{P}(X \leq c)} \qquad \text{(assuming } a \leq c \leq b\text{)} \\ &= \frac{F_X(c) - F_X(a)}{F_X(c)} \end{aligned}$$



Fact 2.1 Let F_X be the cumulative distribution function of a random variable X. Then,

1. F_X is a non-decreasing function.
2. $\mathrm{F}_X(\infty) = 1$.
3. $\mathrm{F}_X(-\infty) = 0$.
4. F_X is right continuous, i.e., the function is equal to its right hand limit.





2.2 Quantiles (aka percentiles)

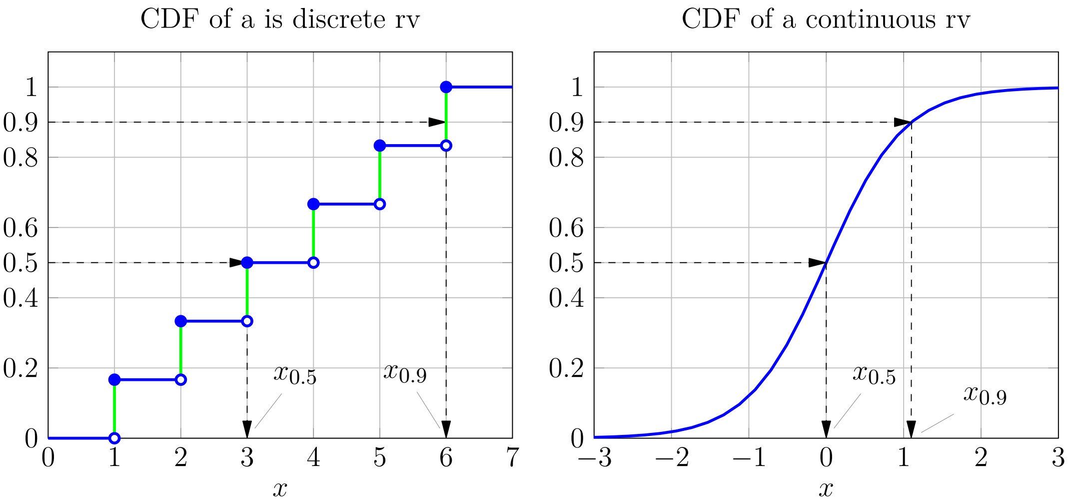
The **quantile** x_α, $0 \le \alpha \le 1$, for a random variable X is given by:

$$\mathrm{P}(X \le x_\alpha) = \alpha, \quad \rightarrow \quad x_\alpha = F_X^{-1}(\alpha).$$

The inverse function $F_X^{-1}(\alpha)$ is called the "quantile function".

Important quantiles are:

- median (50th-percentile): $x_{0.5}$
- upper quartile (75th-percentile): $x_{0.75}$
- lower quartile (25th-percentile): $x_{0.25}$
- quintiles: $x_{0.2}$, $x_{0.4}$, $x_{0.6}$, $x_{0.8}$
- deciles: $x_{0.1}$, $x_{0.2}$, $\ldots$


CDF of a is discrete rv
CDF of a continuous rv
$x_{0.5}$
$x_{0.9}$
x
0
0.2
0.4
0.5
0.6
0.8
0.9
1
0 1 2 3 4 5 6 7
−3 −2 −1 0 1 2 3

Example 1. Derive the quantile function for the continuous CDF in the above figure, where:

$$F_X(x) = \frac{1}{e^{-2x}+1}$$

Solution: Solving for x_α in $\alpha = \frac{1}{e^{-2x_\alpha}+1}$ gives

$$x_\alpha = \frac{1}{2}\log\left(\frac{\alpha}{1-\alpha}\right)$$

□

2.3 Discrete Random Variables

Probability mass function (PMF). Gives the probability of a discrete random variable X having value x is called the probability mass function of X. It is denoted as

$$p_X(x) = \mathrm{P}(X = x)$$

Observe that

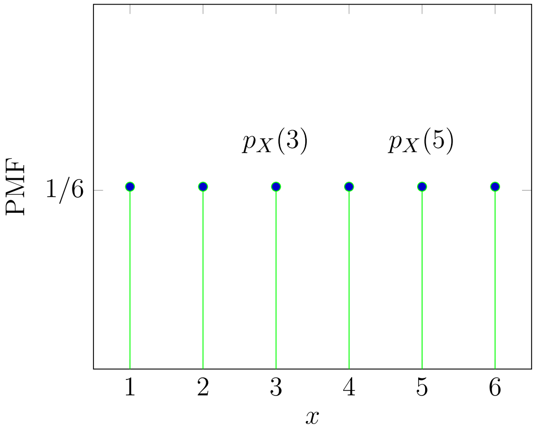
$$F_X(a) = \sum_{x \leq a} p_X(x)$$

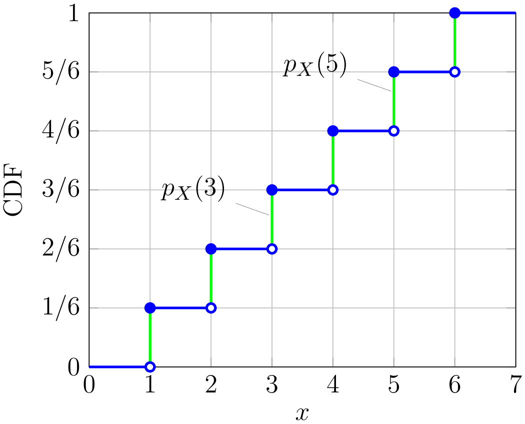
and

$$\sum_{x \in S} p_X(x) = 1$$

This is because the events $X = x$ are disjoint and hence form a *division* of the sample space S.

> **Example 2.** Let X be the result of a dice roll.
> a) Plot the CDF and PMF of X.
> b) Determine $\mathrm{P}(X > 3)$
> c) Determine $\mathrm{P}(2 < X \leq 5)$
> d) Determine $\mathrm{P}(2 < X \leq 5 \mid X \leq 3)$
> e) Determine the median, upper quartile and lower quartile





STOP! This is called the discrete uniform distribution in (1, 6).

a) Plot the CDF and PMF of X.
b) $\mathrm{P}(X > 3) = 1/2$
c) $\mathrm{P}(2 < X \leq 5) = 1/2$
d) $\mathrm{P}(2 < X \leq 5 \mid X \leq 3) = 1/3$
e) the median, upper quartile and lower quartile: 3, 5 and 2

Example 3. Consider the experiment of tossing **three fair coins**. Let X be the random variable that denotes the **number of heads** that result.

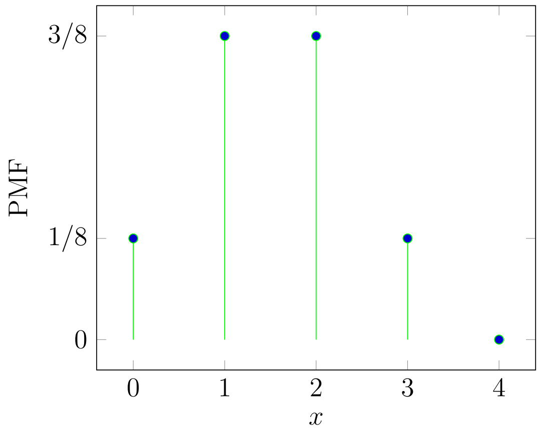
a) Plot the CDF and PMF of X.
b) Determine $\mathrm{P}(X > 2)$
c) Determine $\mathrm{P}(0 < X \leq 3)$
d) Determine $\mathrm{P}(0 < X \leq 3 \mid X \leq 2)$
e) Determine the median, upper quartile and lower quartile

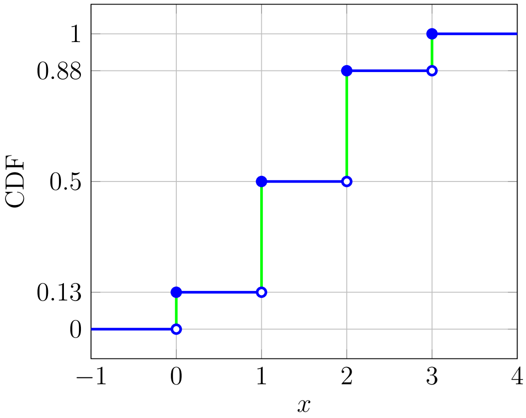
Solution: The sample space is given by

$$
\begin{array}{ccc}
H & H & H \\
H & H & T \\
H & T & H \\
H & T & T \\
T & H & H \\
T & H & T \\
T & T & H \\
T & T & T
\end{array}
$$

Therefore, the PMF of X is given by

$$
p_X(x) = \begin{cases} 1/8 & \text{if } x = 0 \text{ or } x = 3 \\ 3/8 & \text{otherwise} \end{cases}
$$





a) Plot the CDF and PMF of X.
b) $P(X > 2) = 0.125$
c) $P(0 < X \leq 3) = 0.875$
d) $P(0 < X \leq 3 \mid X \leq 2) = 0.857$
e) the median, upper quartile and lower quartile=1, 2, 1

$\square$

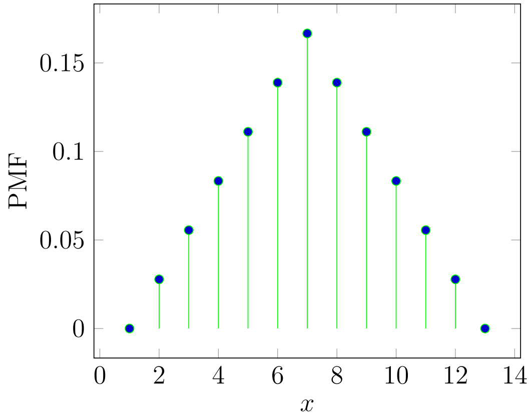
> **Example 4. — Sum of two dice*** Let X be the sum of rolling 2 dice.
> a) Plot the CDF and PMF of X.
> b) Determine $\mathrm{P}(X \geq 3)$
> c) Determine $\mathrm{P}(5 < X \leq 8)$
> d) Determine $\mathrm{P}(5 < X \leq 8 \mid X \leq 10)$
> e) Determine the median, upper quartile and lower quartile

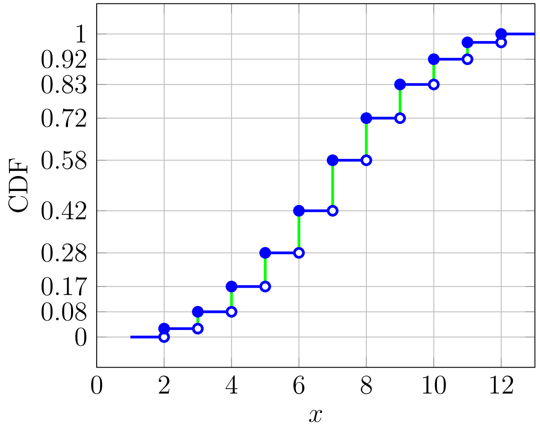
Solution The PMF for X is given by

$$p_X(x) = \begin{cases} 1/36, & x = 2, 12 \\ 2/36, & x = 3, 11 \\ 3/36, & x = 4, 10 \\ 4/36, & x = 5, 9 \\ 5/36, & x = 6, 8 \\ 6/36, & x = 7 \end{cases}$$

Die 1/Die 2	⚀	⚁	⚂	⚃	⚄	⚅
⚀	2	3	4	5	6	7
⚁	3	4	5	6	7	8
⚂	4	5	6	7	8	9
⚃	5	6	7	8	9	10
⚄	6	7	8	9	10	11
⚅	7	8	9	10	11	12

Sum of 2 dice





STOP! This is called the discrete triangular distribution

a) Plot the CDF and PMF of X.

b) $\mathrm{P}(X \geq 3) = 0.972$

c) $\mathrm{P}(5 < X \leq 8) = 0.44$

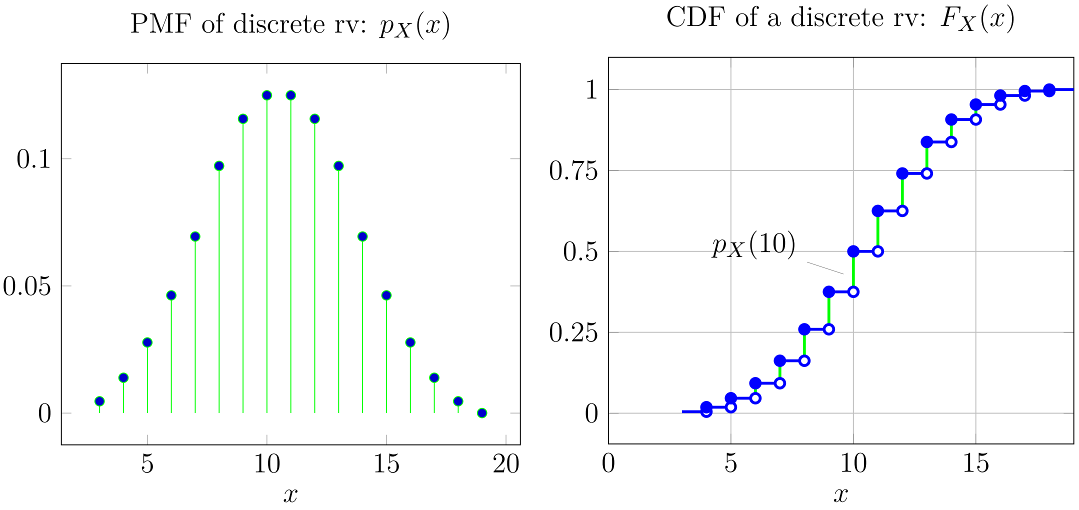
d) $\mathrm{P}(5 < X \leq 8 \mid X \leq 10) = .485$

e) the median, upper quartile and lower quartile=7, 9, 5

Example 5. — Sum of three dice Let X be the sum of rolling 3 dice.

a) Plot the CDF and PMF of X.
b) Determine $\mathrm{P}(X > 5)$
c) Determine $\mathrm{P}(5 \leq X \leq 15)$
d) Determine $\mathrm{P}(5 \leq X \leq 15 \mid X \leq 10)$
e) Determine the median, upper quartile and lower quartile=7, 9 and 5

Solution The PMF for X will be derived in future chapters. In the meantime, observe the bell shape that arises (thanks to the *Central Limit theorem*).

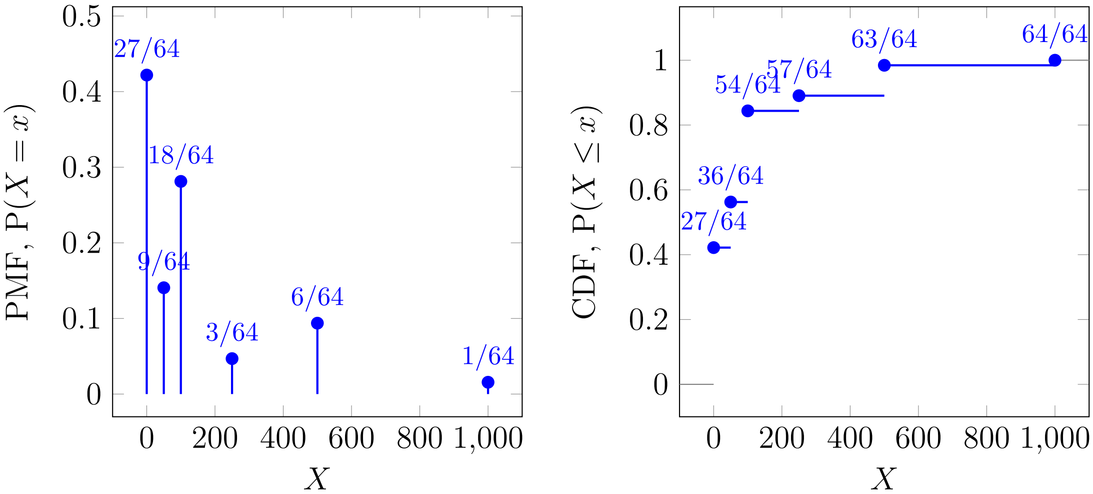


STOP! This PMF is very close to the *normal distribution*; this resemblance is a statement of the **Central Limit theorem**.

a) Plot the CDF and PMF of X.

b) $\mathrm{P}(X > 5) = 0.953$
c) $\mathrm{P}(5 \leq X \leq 15) = 0.935$
d) $\mathrm{P}(5 \leq X \leq 15 \mid X \leq 10) = 0.963$
e) the median, upper quartile and lower quartile=10, 13, 8

STOP! Notice that the sample points of a discrete random variable are not necessarily evenly spaced:



Example 6. You have a coin with probability p of getting 'Heads'. You flip this coin twice. For each flip, if the result is 'Heads', you win \$30. If the result is 'Tails', you lose \$20. Let X be your profit in the game.

a) What is the sample space?

b) Describe the probability mass function of X.

c) Describe the cumulative distribution function of X.

Solution:

a)

$$\begin{aligned} S &= \{(30+30),(30-20),(-20-20)\} \\ &= \{60,10,-40\} \end{aligned}$$

b)

$$\mathrm{P}(X=x)=\begin{cases} p^2 & ; \quad x=60 \\ 2p(1-p) & ; \quad x=10 \\ (1-p)^2 & ; \quad x=-40 \end{cases}$$

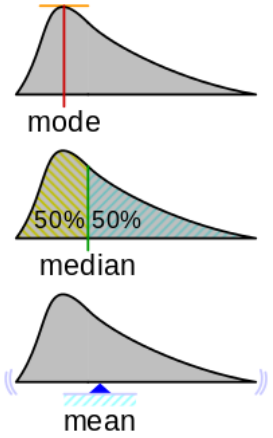
c)

$$\mathrm{F}_X(X) = \begin{cases} 0 & ; \quad x < -40 \\ (1-p)^2 & ; \quad -40 \leq x < 10 \\ (1-p)^2 + 2p(1-p) & ; \quad 10 \leq x < 60 \\ (1-p)^2 + 2p(1-p) + p^2 & ; \quad x \geq 60 \end{cases}$$

$\square$

2.4 Expectation

The expectation (also known as **mean**) is probably the most important measure of central tendency of a rv. The others are mode and median.



Expectation of a discrete rv Let X be a discrete rv with a set of possible values S and PMF $p_X(x)$. The **expected** or **mean** value of X is

$$\mathrm{E}(X) = \mu_X = \sum_{x \in S} x \cdot p_X(x).$$

The symbol μ_X will be used interchangeably with $\mathrm{E}(X)$.

In example 3, the expectation of the number of heads is given by

$$\mathrm{E}(X) = 0 \times \frac{1}{8} + 3 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} = \frac{3}{2}$$

STOP! As seen from the example, the expectation of a random variable may not be not be on its sample space.

Example 7. Find $\mathrm{E}(X)$ where X is the outcome of rolling a fair dice.

Solution: Let X be the random variable that denotes the result of a single roll of dice. The PMF for X is given by

$$p_X(x) = \frac{1}{6}, \qquad x = 1, 2, 3, 4, 5, 6.$$

The expectation of X is given by

$$\mathrm{E}(X) = \sum_{x=1}^{6} p_X(x) \cdot x = \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

$\square$

Example 8. When we roll two dice what is the expected value of their sum?

Solution Let X be the random variable denoting the sum. From example 4 above, we know that the PMF for X is given by

$$p_X(x) = \begin{cases} 1/36, & x = 2, 12 \\ 2/36, & x = 3, 11 \\ 3/36, & x = 4, 10 \\ 4/36, & x = 5, 9 \\ 5/36, & x = 6, 8 \\ 6/36, & x = 7 \end{cases}$$

The expectation of X is given by

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{x=2}^{12} p_X(x) \cdot x \\
&= \frac{1}{36} \times 2 + \frac{2}{36} \times 3 + \frac{3}{36} \times 4 + \frac{4}{36} \times 5 + \frac{5}{36} \times 6 + \frac{6}{36} \times 7 + \\
&\quad \frac{5}{36} \times 8 + \frac{4}{36} \times 9 + \frac{3}{36} \times 10 + \frac{2}{36} \times 11 + \frac{1}{36} \times 12 \\
&= \frac{252}{36} = 7
\end{aligned}$$

Example 9. A class of 120 students is driven in 3 buses to a jazz concert. There are 36 students in the first bus, 40 in the second, and 44 in the third bus. When the buses arrive, a student is randomly chosen. Let X denote the number of students on the bus of the chosen student. Find $\mathrm{E}(X)$.

Solution:

$$\mathrm{P}(X = 36) = \frac{36}{120}$$
$$\mathrm{P}(X = 40) = \frac{40}{120}$$
$$\mathrm{P}(X = 44) = \frac{44}{120}$$

Therefore,

$$\begin{aligned}
\mathrm{E}\,(X) &= \sum_x x \cdot p_X\,(x) \\
&= 36\left(\frac{36}{120}\right) + 40\left(\frac{40}{120}\right) + 44\left(\frac{44}{120}\right) \\
&= 40.2667
\end{aligned}$$

□

Example 10. — surge suppressor The owner of a small firm has just purchased a personal computer, which she expects will serve her for the next two years. The owner has been told that she "must" buy a surge suppressor to provide protection for her new hardware against possible surges or variations in the electrical current, which have the capacity to damage the computer. The amount of damage to the computer depends on the strength of the surge. It has been estimated that there is a 2% chance of incurring 350 dollar damage, 4% chance of incurring 300 dollar damage, and 11% chance of incurring 100 dollar damage from a surge within the next two years. An inexpensive suppressor, which would provide protection for only one surge, can be purchased.

How much should the owner be willing to pay if she makes decisions on the basis of expected value?

Solution: Let the random variable D be the amount of damage to the computer (in dollars) caused by a surge within the next two years. Then

$$\begin{aligned} \mathrm{E}(D) &= 350\mathrm{P}(D=350)+300\mathrm{P}(D=300)+100\mathrm{P}(D=100) \\ &= 350\cdot 0.02+300\cdot 0.04+100\cdot 0.11 = 30 \end{aligned}$$

Hence the owner expects her computer to incur \$30 of damage from a surge in the next two years and so should be willing to pay \$30 for a surge protector. □

2.4.1 Expectation of a function of X

$$Y = g(X)$$

If all we want is the expected value $\mathrm{E}(Y)$, there are **two options:**

1. compute the PMF of $Y, p_Y(y)$ (chapter 5), the sample space of Y, S_Y and use the definition of expectation
$$\mathrm{E}(Y) = \mu_Y = \sum_{Y \in S_Y} y \cdot p_Y(y).$$
2. use Fact 2.2 below:

Fact 2.2 — Expectation of a function of X. For any function $g(X)$,

$$\mathrm{E}(g(x)) = \sum_{x \in S} g(x) \cdot p_X(x)$$

Example 11. X has the following distribution.

$$\begin{aligned} P(X = -1) &= 0.2 \\ P(X = 0) &= 0.5 \\ P(X = 1) &= 0.3 \end{aligned}$$

Find $E\left[X^2\right]$.

Solution:[1] Using

$$E\left(g(x)\right) = \sum_{x \in S} g(x) \cdot p_X(x)$$

we have:

$$\begin{aligned} E\left(X^2\right) &= (-1)^2(0.2) + (0^2)(0.5) + (1^2)(0.3) \\ &= 0.5 \end{aligned}$$

□

Solution:[2] Let

$$Y = X^2$$

Using option one above, we have

$$
\begin{aligned}
\mathrm{P}(Y=0) &= \mathrm{P}(X^2=0) \\
&= \mathrm{P}(X=0) \\
&= 0.5 \\
\mathrm{P}(Y=1) &= \mathrm{P}(X^2=1) \\
&= \mathrm{P}(X=-1)+\mathrm{P}(X=1) \\
&= 0.5
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathrm{E}(Y) &= \mathrm{E}\left[X^2\right] \\
&= (0)(0.5)+(1)(0.5) \\
&= 0.5
\end{aligned}
$$

□

Example 12. Let the probability distribution of X be

X	-2	-1	0	1	2
$P(X = x)$	0.25	0.1	0.2	0.2	0.25

Calculate $\mathrm{E}(|X|)$.

Solution: Using fact 2.2:

$$\begin{aligned} E(Y) = E(|X|) &= \sum_{x \in S_X} |x| \times \mathrm{P}(X = x) \\ &= 2 * 0.25 + 1 * 0.1 + 0 * 0.2 + 1 * 0.2 + 2 * 0.25 \\ &= 1.3 \end{aligned}$$

$\square$

Solution: Using option one above, define $Y = |X|$, and we have to calculate $E(Y)$.

$$S_Y = \{0, 1, 2\}$$
$$\mathrm{P}(Y = 0) = \mathrm{P}(|X| = 0) = \mathrm{P}(X = 0) = 0.2$$
$$\mathrm{P}(Y = 1) = \mathrm{P}(X = 1) + \mathrm{P}(X = -1) = 0.2 + 0.1 = 0.3$$
$$\mathrm{P}(Y = 2) = \mathrm{P}(X = 2) + \mathrm{P}(X = -2) = 0.25 + 0.25 = 0.5$$

Hence

$$E(Y) = 0 \times 0.2 + 1 \times 0.3 + 2 \times 0.5 = 1.3$$

$\square$

Example 13. Show that for any constants a, b and random variable X,

$$\mathrm{E}(a+bX) = a + b\,\mathrm{E}(X)$$

Solution:

$$\begin{aligned}
\mathrm{E}(a+bX) &= \sum_x (a+bx)\mathrm{P}(X=x) \\
&= a\sum_x \mathrm{P}(X=x) + b\sum_x x\,\mathrm{P}(X=x) \\
&= a + b\,\mathrm{E}(X)
\end{aligned}$$

$\square$

2.4.2 Variance

Variance The variance of a random variable X is defined to be

$$\mathrm{V}(X) = \mathrm{E}\left[(X - \mathrm{E}(X))^2\right]$$

The symbol σ_X^2 will be used interchangeably with $\mathrm{V}(X)$.

Shortcut formula for the Variance

$$\mathrm{V}(X) = \mathrm{E}\left(X^2\right) - \mathrm{E}(X)^2$$

Proof.

$$\begin{aligned}
\mathrm{E}\left((X - \mathrm{E}(X))^2\right) &= \mathrm{E}\left(X^2 - 2X\mathrm{E}(X) + \mathrm{E}(X)^2\right) \\
&= \mathrm{E}\left(X^2\right) - 2\mathrm{E}(X\mathrm{E}(X)) + \mathrm{E}(X)^2 \\
&= \mathrm{E}\left(X^2\right) - 2\mathrm{E}(X)^2 + \mathrm{E}(X)^2 \\
&= \mathrm{E}\left(X^2\right) - \mathrm{E}(X)^2
\end{aligned}$$

The standard deviation of a random variable X is

$$\sigma_X = \sqrt{\mathrm{V}(X)}$$

The advantage is that it has the same units as X, so it is easier to interpret compared to the variance.

Coefficient of variation The coefficient of variation of a random variable X is defined as

$$\delta_X = \frac{\sigma_X}{|\mu_X|}$$

provided $\mu_X \neq 0$. The big advantage here is that the coefficient of variation is dimensionless! As a rule of thumb, when

$$\delta_X < 0.3$$

the random variable has moderate uncertainty.

Fact 2.3

$$\mathrm{V}(aX+b)=a^2\,\mathrm{V}(X)$$

Proof.

$$\begin{aligned}
\mathrm{V}(aX+b) &= \mathrm{E}\left[(aX+b-\mathrm{E}(aX+b))^2\right] \\
&= \mathrm{E}\left[(aX+b-a\mathrm{E}(X)-b)^2\right] \\
&= \mathrm{E}\left[a^2(X-\mathrm{E}(X))^2\right] \\
&= a^2\,\mathrm{E}\left[(X-\mathrm{E}(X))^2\right] \\
&= a^2\,\mathrm{V}(X)
\end{aligned}$$

■

Example 14. Calculate $\mathrm{V}(X)$ where X represents the outcome of rolling a fair dice.

Solution:

$$
\begin{aligned}
\mathrm{E}(X) &= \sum_{x=1}^{6} \frac{1}{6} x \\
&= \frac{7}{2} \\
\mathrm{E}\left[X^2\right] &= \sum_{x=1}^{6} \frac{1}{6} x^2 \\
&= \frac{91}{6}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathrm{V}(X) &= \mathrm{E}\left[X^2\right] - \mathrm{E}(X)^2 \\
&= \frac{91}{6} - \frac{49}{4} \\
&= \frac{35}{12}
\end{aligned}
$$

Notice that the coefficient of variation of a dice is about 0.5:

$$\delta_X = \frac{\sigma_X}{|\mu_X|} = \frac{\sqrt{35/12}}{7/2} \approx 0.49$$

$\square$

Example 15. * Recall example 6. You have a coin with probability p of getting 'Heads'. You flip this coin twice. For each flip, if the result is 'Heads', you win \$30. If the result is 'Tails', you lose \$20.
Let X be your profit in the game.

a) What is the expected value of X?
b) What is the value of p upto which you would agree to participate in the game?
c) What is $\mathrm{V}(X)$?
d) What is σ_X?

Solution:

a)

$$\begin{aligned}\mathrm{E}[X] &= (-40)(1-p)^2 + (10)2p(1-p) + (60)p^2 \\ &= 100p - 40\end{aligned}$$

b) $100p - 40 > 0 \rightarrow p > 2/5$

c) Let's use $\mathrm{V}(X) = \mathrm{E}\left[X^2\right] - \mathrm{E}(X)^2$:

$$\begin{aligned}
\mathrm{E}\left[X^2\right] &= (-40)^2(1-p)^2 + (10)^2 2p(1-p) + (60)^2 p^2 \\
&= 5000p^2 - 3000p + 1600 \\
\mathrm{E}(X)^2 &= (100p-40)^2 \\
&= 10000p^2 - 8000p + 1600
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{V}(X) &= \mathrm{E}\left[X^2\right] - \mathrm{E}(X)^2 \\
&= 5000p(1-p)
\end{aligned}$$

d)

$$\begin{aligned}
\sigma_X &= \sqrt{\mathrm{V}(X)} \\
&= \sqrt{5000p(1-p)}
\end{aligned}$$

$\square$

Example 16. Consider three random variables X, Y, Z measured **in the same units**. Their probability mass distribution is as follows.

$$\mathrm{P}(X = x) = \begin{cases} 1/2, & x = -2 \\ 1/2, & x = 2 \end{cases}$$

$$\mathrm{P}(Y = y) = \begin{cases} 0.001, & y = -10 \\ 0.998, & y = 0 \\ 0.001, & y = 10 \end{cases}$$

$$\mathrm{P}(Z = z) = \begin{cases} 1/3, & z = -10 \\ 1/3, & z = 0 \\ 1/3, & z = 10 \end{cases}$$

Which of the above random variables is more "spread out"?

Solution: It is easy to see that $\mathrm{E}(X) = \mathrm{E}(Y) = \mathrm{E}(Z) = 0$.

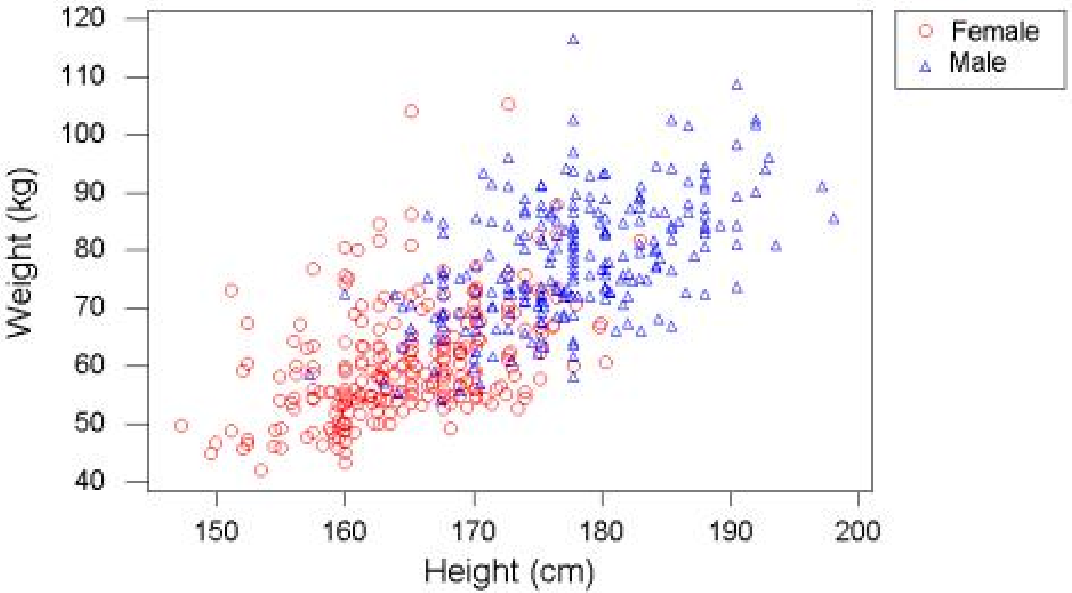
$$\begin{aligned}
\mathrm{V}(X) &= \mathrm{E}\left(X^2\right) \\
&= 0.5 \cdot (-2)^2 + 0.5 \cdot (2)^2 \\
&= 4 \\
\mathrm{V}(Y) &= \mathrm{E}\left(Y^2\right) \\
&= 0.001 \cdot (-10)^2 + 0.998 \cdot 0^2 + 0.001 \cdot (10)^2 \\
&= 0.2 \\
\mathrm{V}(Z) &= \mathrm{E}\left(Z^2\right) \\
&= (1/3) \cdot (-5)^2 + (1/3) \cdot 0^2 + (1/3) \cdot (5)^2 \\
&= 16.67
\end{aligned}$$

Thus Z is the most spread out and Y is the most concentrated. $\square$

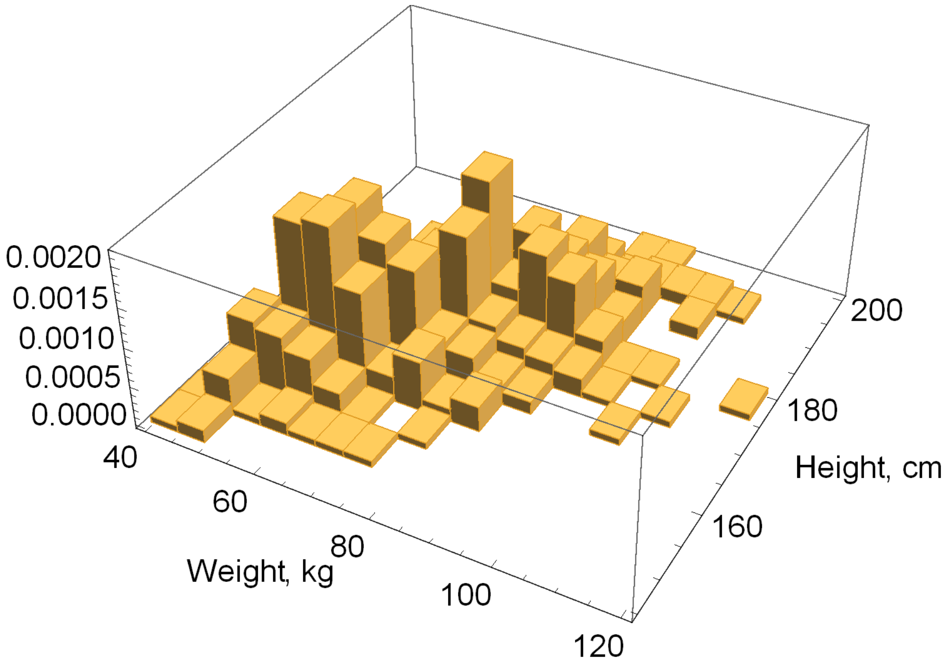
2.5 Jointly Distributed Discrete Random Variables

There are many practical situations where we need to deal with more than two measurement simultaneously. The Joint Probability Distribution of (X,Y) describes the *joint* random behavior of X and Y.

For instance we might be interested in the relationship between heights and weights of a population. Let (X_i,Y_i) denote the (weight, height) of person i, then (X_i,Y_i) are related since we can expect that if X_i is large/small then the associated Y_i tends to be large/small. Thus X_i and Y_i are **correlated** and we should describe the behavior of (X_i,Y_i) jointly (together).


Female
Male
Weight (kg)
Height (cm)
120
110
100
90
80
70
60
50
40
150
160
170
180
190
200

→ data source
Histogram of the weight/height data:

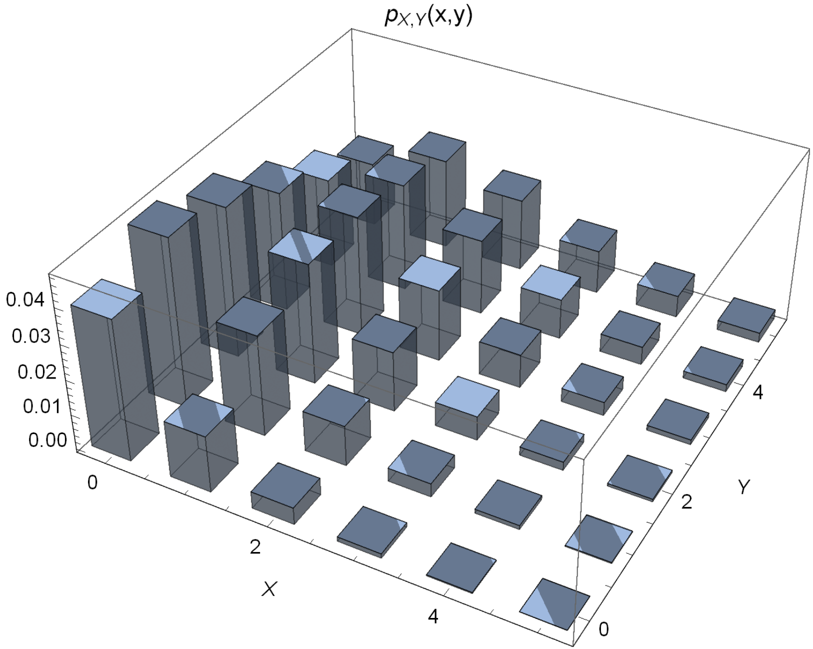

0.0020
0.0015
0.0010
0.0005
0.0000
40
60
80
100
120
Weight, kg
200
180
160
Height, cm

The data for the histogram is:

$W\backslash H$	147 - 157	157 - 167	167 - 177	177 - 187	187 - 197	197 - 207
42 - 52	3	18	6	0	0	0
52 - 62	0	21	82	21	0	0
62 - 72	0	7	57	50	6	0
72 - 82	0	2	23	61	28	1
82 - 92	0	0	9	38	35	5
92 - 102	0	0	0	10	10	5
102 - 112	0	0	1	2	2	3
112 - 122	0	0	0	1	0	0

Dividing this table entries by the sum (507 individuals) would give the joint PMF.

In general, a joint distribution looks like this:



2.5.1 Chapter 1 results in PMF notation

Recall from chapter 1 For two events A and B in sample space S:

$$\begin{aligned}
P(A|B) &= \frac{P(A\cap B)}{P(B)} && \text{(conditional probability)}\\
P(A\cap B) &= P(A|B)P(B) && \text{(multiplication rule)}\\
P(A\cap B) &= P(A)P(B) && \text{(independence)}\\
P(B) &= \sum_{i=1}^{n} P(B\cap A_i) && \text{(total probability, where } S \text{ is partitioned into } A_1,\ldots,A_n)\\
&= \sum_{i=1}^{n} P(B|A_i)P(A_i)
\end{aligned}$$

These definitions we developed for events in chapter 1 extend to random variables by defining events A and B and their intersection as

$$\boxed{A=(X=x) \qquad \text{and} \qquad B=(Y=y) \qquad \text{and} \qquad A\cap B=(X=x,Y=y)}$$

And now we express the results from chapter 1 in PMF notation, that is, in terms of probability mass functions. We start with $\mathrm{P}(A \cap B)$, which we now call the joint PMF:

Joint PMF For two random variables X and Y the joint PMF is

$$p_{X,Y}(x,y) = \mathrm{P}(X = x, Y = y)$$

Note:

$$\sum_X \sum_Y p_{X,Y}(x,y) = 1$$

The total probability rule we now call marginal PMF:

Marginal PMF The PMF of a single random variable is called marginal PMF:

$$p_X(x) = \sum_Y p_{X,Y}(x,y) \qquad \text{and} \qquad p_Y(y) = \sum_X p_{X,Y}(x,y)$$

The conditional probability rule we now call conditional PMF:

Conditional PMF Let X and Y be two discrete random variables. Then the conditional probability mass function (conditional pmf) of Y given $X = x$ is defined as,

$$p_{Y|X=x}(y) = P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)} = \frac{p_{X,Y}(x,y)}{p_X(x)}$$

for $y \in S_Y$ and $x \in S_X$. Please note that in the term $p_{Y|X=x}(y)$ we are thinking as if x is fixed and y is the variable, but obviously both can vary.

The conditional pmf of X given $Y = y$ for some $y \in S_Y$ is similarly defined as,

$$p_{X|Y=y}(x) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

STOP! The PMF version of the Multiplication Rule:

$$\begin{aligned} p_{X,Y}(x,y) &= p_Y(y)p_{X|Y}(x) \\ &= p_X(x)p_{Y|X}(y) \end{aligned}$$

can be combined with the definition of marginal PMF:

$$p_X(x) = \sum_Y p_Y(y) p_{X|Y}(x)$$

The condition for independence $\mathrm{P}(X = x, Y = y) = \mathrm{P}(X = x) \cdot \mathrm{P}(Y = y)$ in PMF notation becomes

Independence of Two Random Variables. Random variables X and Y are independent iif:

$$p_{X,Y}(x,y) = p_X(x) p_Y(y)$$

for all (x,y).

Alternative we could use $p_{X|Y}(x) = p_X(x)$ or $p_{Y|X}(y) = p_Y(y)$ to check for independence, but the above definition is the most common.

Calculating probabilities For any two random variables X and Y having joint PMF X:

$$\mathrm{P}\left((X,Y) \in A\right) = \sum_{(x,y)\in A} p_{X,Y}(x,y)$$

Finally, let's define the joint CDF although for some reason it is rarely used in this chapter.

Joint Cumulative Probability Distribution Function For any two random variables X and Y, the joint cumulative probability distribution function of X and Y is defined to be

$$\begin{aligned}\mathrm{F}_{X,Y}(a,b) &= \mathrm{P}(X \le a, Y \le b) \\ &= \sum_{x\le a, y\le b} p_{X,Y}(x,y)\end{aligned}$$

where $a \in \mathbb{R}$ and $b \in \mathbb{R}$.

Example 17. 3 balls are randomly selected from an urn containing 3 red, 4 white, and 5 blue balls.
Let X be the number of red balls chosen.
Let Y be the number of white balls chosen.
Find the joint probability mass function of X and Y.

Solution:

$$\mathrm{P}(X=x, Y=y)=\frac{\binom{3}{x}\binom{4}{y}\binom{5}{3-x-y}}{\binom{12}{3}}$$

Therefore,

$X \backslash Y$	0	1	2	3	TOT $= p_X(x)$
0	$10/220$	$40/220$	$30/220$	$4/220$	$84/220$
1	$30/220$	$60/220$	$18/220$	0	$108/220$
2	$15/220$	$12/220$	0	0	$37/220$
3	$1/220$	0	0	0	$1/220$
TOT $= p_Y(y)$	$56/220$	$112/220$	$48/220$	$4/220$	1

□

Example 18. Given the joint distribution of (X, Y):

$p_{X,Y}(x,y)$		X: 0	1	2
Y	0	$3/28$	$9/28$	$3/28$
	1	$3/14$	$3/14$	0
	2	$1/28$	0	$0/28$

are X and Y statistically independent?

Solution: Recall that random variables X and Y are independent iif:

$$p_{X,Y}(x,y) = p_X(x)p_Y(y)$$

for all (x,y), and that the marginal distributions are defined as

$$p_X(x) = \sum_Y p_{X,Y}(x,y) \qquad \text{and} \qquad p_Y(y) = \sum_X p_{X,Y}(x,y)$$

which correspond to the row totals and column totals:

		$\boxed{X}$		
$p_{X,Y}(x,y)$	0	1	2	$p_Y(y)$
0	3/28	9/28	3/28	15/28
$\boxed{Y}$ 1	3/14	3/14	0	3/7
2	1/28	0	0	1/28
$p_X(x)$	5/14	15/28	3/28	1

Since $p_{X,Y}(0,0) \neq p_X(0) \cdot p_Y(0)$, X and Y are not independent. $\square$

Example 19. In a class there are four freshman boys, six freshman girls, and six sophomore boys. How many sophomore girls must be present if gender and class are to be independent when a student is selected at random?

Solution: In class. $\square$

Example 20. * The joint distribution of X and Y is given in the following table

$p_{XY}(x,y)$	X=0	X=1	X=3	**TOT**
Y=-1	0.11	0.03	0	**0.14**
Y=2.5	0.03	0.09	0.16	**0.28**
Y=3	0.15	0.15	0.06	**0.36**
Y=4.7	0.04	0.16	0.02	**0.22**
TOT	**0.33**	**0.43**	**0.24**	1

Find a) $P(Y - X \leq 2)$, b) $P(2 \leq Y \leq 4|X = 1)$.

Solution: a) The best way to do this is to calculate the values of the required function, $Y - X$ in this case, and highlight the cells that are favorable to our condition, being less than or equal to two in this case:

Y-X	X=0	X=1	X=3
Y=-1	**-1**	**-2**	**-4**
Y=2.5	2.5	**1.5**	**-0.5**
Y=3	3	**2**	**0**
Y=4.7	4.7	3.7	**1.7**

Then, we add up all the joint probabilities corresponding to these highlighted cells:

$p_{XY}(x,y)$	X=0	X=1	X=3
Y=-1	**0.11**	**0.03**	**0**
Y=2.5	0.03	**0.09**	**0.16**
Y=3	0.15	**0.15**	**0.06**
Y=4.7	0.04	0.16	**0.02**

which gives 0.62.

b)The conditional distribution of $Y|X = 1$ is :

$p_{Y\|X=1}(y)$	X=1
Y=-1	0.03/0.43
Y=2.5	0.09/0.43
Y=3	0.15/0.43
Y=4.7	0.16/0.43
TOT	1

and the desired probability is 0.09/0.43+0.15/0.43=0.558.

□

Example 21. The joint distribution of X and Y is given in the following table

$p_{XY}(x,y)$	X=-1	X=-2	X=2	X=3	**TOT**
Y=-3	0.14	0.14	0.01	0.05	**0.34**
Y=-1	0.15	0.06	0.06	0.04	**0.31**
Y=1	0.03	0.1	0.11	0.11	**0.35**
TOT	**0.32**	**0.3**	**0.18**	**0.2**	1

Find $P(Y+X\leq 0)$, $P(-2\leq X\leq 2|Y=-1)$.

Solution: a)0.68 b)0.387 □

Example 22. Roll a balanced dice twice. Define random variables:

- X = number of 4's
- Y = number of 5's

(a) Find the joint distribution of X and Y, $p_{X,Y}(x,y)$.
(b) Find $\mathrm{P}((X,Y) \in A)$ where $A = \{2x + y < 3\}$

Solution:
Possible values of X and Y: $x = 0, 1, 2, \quad y = 0, 1, 2, \quad x + y \le 2$.
Sample space

		Roll $\boxed{2}$					
		1	2	3	4	5	6
	1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
Roll	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
$\boxed{1}$	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

(36 equally likely sample points)

		$\boxed{y}$	
$p_{X,Y}(x,y)$	0	1	2
0	16/36	8/36	1/36
$\boxed{x}$ 1	8/36	2/36	0
2	1/36	0	0

(Check if $\sum_X \sum_Y p_{X,Y}(x,y) = 1$?)

(b) Find $\mathrm{P}((X,Y) \in A)$ where $A = \{2x + y < 3\}$

$\mathrm{P}((X,Y) \in A) = p_{X,Y}(0,0) + p_{X,Y}(0,1) + p_{X,Y}(0,2) + p_{X,Y}(1,0) = 33/36$ □

Example 23. Toss a balanced coin 3 times. Define random variables

- X = number of heads
- Y = (number of heads) − (number of tails)

Find the joint distribution of X and Y.

Solution:
Possible values of X and Y:

$$\begin{aligned} x &= 0, 1, 2, 3 \\ y &= -3, -1, 1, 3 \\ 2x - y &= 3 \qquad \text{(why?)} \end{aligned}$$

Sample space

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

(8 equally likely sample points)

	$p_{X,Y}(x,y)$	y: -3	-1	1	3
x	0	1/8	0	0	0
	1	0	3/8	0	0
	2	0	0	3/8	0
	3	0	0	0	1/8

(Check if $\sum_X \sum_Y p_{X,Y}(x,y) = 1$—YES!) $\square$

Example 24. — The Titanic data :

Passenger Status	Survivors	Fatalities	TOTAL
First Class	203	122	325
Second Class	118	167	285
Third Class	178	528	706
Crew	212	673	885
Total	711	1490	2201

There are two random variables in play here,

$$\begin{aligned} X &= 0 \qquad \text{if passenger survived} \\ &= 1 \qquad \text{if passenger diced} \end{aligned}$$

and

$$
\begin{aligned}
Y &= 1 && \text{if passenger was in first class} \\
&= 2 && \text{if passenger was in second class} \\
&= 3 && \text{if passenger was in third class} \\
&= 4 && \text{if passenger was a crew member}
\end{aligned}
$$

So we approximate the joint PMF of X and Y as,

	X=0	X=1
Y=1	0.09	0.06
Y=2	0.05	0.08
Y=3	0.08	0.24
Y=4	0.10	0.30

For example $p_{X,Y}(0,1) = \frac{203}{2201} = 0.09$.
`Find` $P(X+Y \leq 2)$.

Solution:

$$\begin{aligned} P(X+Y \leq 2) &= P(X=0, Y=1) + P(X=0, Y=2) + P(X=1, Y=1) \\ &= 0.09 + 0.05 + 0.06 \\ &= 0.2 \end{aligned}$$

□

Example 25. Let's see how we can apply conditional probabilities in the titanic example,

	X=0	X=1
Y=1	0.09	0.06
Y=2	0.05	0.08
Y=3	0.08	0.24
Y=4	0.10	0.30

Determine the probability of being a survivor given the class.

Solution:

$$
\begin{aligned}
P(\text{Survivor}|\text{First Class}) &= p_{X|Y=1}(0) = \frac{p_{X,Y}(0,1)}{p_Y(1)} = \frac{0.09}{0.09+0.06} = 0.6 \\
P(\text{Survivor}|\text{Second Class}) &= p_{X|Y=2}(0) = \frac{p_{X,Y}(0,2)}{p_Y(2)} = \frac{0.05}{0.05+0.08} = 0.3846 \\
P(\text{Survivor}|\text{Third Class}) &= p_{X|Y=3}(0) = \frac{p_{X,Y}(0,3)}{p_Y(3)} = \frac{0.08}{0.08+0.24} = 0.25 \\
P(\text{Survivor}|\text{Crew}) &= p_{X|Y=4}(0) = \frac{p_{X,Y}(0,4)}{p_Y(4)} = \frac{0.1}{0.1+0.3} = 0.25
\end{aligned}
$$

So we see that probability of survival across passenger class is decreasing, although we should be careful while making remarks like this and consider other factors present. $\square$

2.5.2 Expectation with two random variables

Expected value rule for multiple random variables

$$\mathrm{E}\left(g(X,Y)\right) = \sum_X \sum_Y g(x,y) p_{X,Y}(x,y)$$

Conditional Expectation

$$\mathrm{E}\left(g(X,Y) | Y = y\right) = \sum_X g(x,y) p_{X|Y}(x)$$

Fact 2.4 — Linearity of Expectation. For any finite collection of random variables $X_1, X_2, \ldots, X_n$,

$$\mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}\left(X_i\right)$$

Proof. We will prove the statement for two random variables X and Y. The general claim can be proven using induction.

$$
\begin{aligned}
\mathrm{E}(X+Y) &= \sum_x \sum_y (x+y) p_{X,Y}(x,y) \\
&= \sum_x \sum_y (x p_{X,Y}(x,y) + y p_{X,Y}(x,y)) \\
&= \sum_x \sum_y x p_{X,Y}(x,y) + \sum_x \sum_y y p_{X,Y}(x,y) \\
&= \sum_x x \sum_y p_{X,Y}(x,y) + \sum_y y \sum_x p_{X,Y}(x,y) \\
&= \sum_x x p_X(x) + \sum_y y p_Y(y) \\
&= \mathrm{E}(X) + \mathrm{E}(Y)
\end{aligned}
$$

■

Example 26. Using linearity of expectation calculate the expected value of the sum of the numbers obtained when two dice are rolled.

Solution: Let X_1 and X_2 denote the random variables that denote the result when dice 1 and dice 2 are rolled respectively. We want to calculate $\mathrm{E}(X_1+X_2)$. By linearity of expectation

$$\begin{aligned}\mathrm{E}(X_1+X_2) &= \mathrm{E}(X_1)+\mathrm{E}(X_2)\\ &= 3.5+3.5\\ &= 7\end{aligned}$$

□

Example 27. — Indicator (aka Bernouilli) Random Variables. Let X be the random variable such that

$$X = \begin{cases} 1 & \text{if event } A \text{ occurs (with probability } p = \mathrm{P}(A)) \\ 0 & \text{otherwise} \end{cases}$$

show that

$$\begin{aligned} \mathrm{E}(X) &= p \\ \mathrm{V}(X) &= p(1-p) \end{aligned}$$

Solution:

$$\mathrm{E}(X) = 1 \times p + 0 \times (1-p) = p$$

$$\mathrm{E}\left(X^2\right) = 1^2 \times p + 0^2 \times (1-p) = p$$

$$\mathrm{V}(X) = \mathrm{E}\left(X^2\right) - \mathrm{E}(X)^2 = p - p^2 = p(1-p)$$

$\square$

Example 28. — The hat check problem. Suppose that n people leave their hats at the hat check. If the hats are randomly returned what is the expected number of people that get their own hat back?

Solution: Let X be the random variable that denotes the number of people who get their own hat back. Let $X_i, 1 \leq i \leq n$, be the random variable such that

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th person gets his/her own hat back (with probability } p = 1/n) \\ 0 & \text{otherwise} \end{cases}$$

Thus,

$$\mathrm{E}(X_i) = 1 \times \frac{1}{n} + 0 \times (1 - \frac{1}{n}) = \frac{1}{n}$$

Now clearly,

$$X = \sum_{i=1}^{n} X_i = X_1 + X_2 + X_3 + \ldots + X_n$$

By linearity of expectation we get

$$\mathrm{E}(X) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \sum_{i=1}^{n} \frac{1}{n} = n \times \frac{1}{n} = 1$$

So on average, only one person will get his/her hat back! $\square$

Example 29. Suppose we throw n balls into n bins with the probability of a ball landing in each of the n bins being equal. What is the expected number of empty bins?

Solution: Let X be the random variable denoting the number of empty bins. Let X_i be a random variable that is 1 if the i-th bin is empty and is 0 otherwise:

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th bin is empty, with probability } p = (1-1/n)^n \\ 0 & \text{otherwise} \end{cases}$$

Clearly

$$X = \sum_{i=1}^{n} X_i$$

By linearity of expectation, we have

$$\begin{aligned} \mathrm{E}(X) &= \sum_{i=1}^{n} \mathrm{E}(X_i) \\ &= \sum_{i=1}^{n} \mathrm{P}(X_i = 1) \\ &= \sum_{i=1}^{n} \left(1 - \frac{1}{n}\right)^n \\ &= n\left(1 - \frac{1}{n}\right)^n \end{aligned}$$

As $n \to \infty$, $(1 - \frac{1}{n})^n \to \frac{1}{e}$. Hence, for large enough values of n we have

$$\mathrm{E}(X) \approx n/e$$

This means that on average about one third of the bins will be empty. $\square$

Fact 2.5 If X and Y are independent, then, for any function g and h,

$$\mathrm{E}\left[g(X)h(Y)\right] = \mathrm{E}\left[g(X)\right]\mathrm{E}\left[h(Y)\right]$$

Proof.

$$\begin{aligned}
\mathrm{E}\left(g(X)h(Y)\right) &= \sum_x \sum_y g(x)h(y)p_{X,Y}(x,y) \\
&= \sum_x \sum_y g(x)h(y)p_X(x)p_Y(y) \\
&= \sum_x g(x)p_X(x) \sum_y h(y)p_Y(y) \\
&= \mathrm{E}\left[g(X)\right]\mathrm{E}\left[h(Y)\right]
\end{aligned}$$

Fact 2.6 — $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$ when X, Y are independent. This follows from the previous result with

$$g(X) = X, \qquad h(Y) = Y$$

2.6 Covariance

Covariance Let X and Y be random variables. Then

$$\operatorname{Cov}(X,Y) = \mathrm{E}\left[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right]$$

is defined to be the covariance of X and Y.

Fact 2.7 — Shortcut formula for covariance.

$$\operatorname{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$$

Proof.

$$\begin{aligned}
\operatorname{Cov}(X,Y) &= \mathrm{E}\left[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right] \\
&= \mathrm{E}\left[XY - \mathrm{E}(X)Y - X\mathrm{E}(Y) + \mathrm{E}(X)\mathrm{E}(Y)\right] \\
&= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) - \mathrm{E}(X)\mathrm{E}(Y) + \mathrm{E}(X)\mathrm{E}(Y) \\
&= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)
\end{aligned}$$

■

Fact 2.8 If X and Y are independent, then

$$\operatorname{Cov}(X,Y) = 0$$

However, the converse is not true.

Proof.

$$\operatorname{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$$

As X and Y are independent $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$, so

$$\begin{aligned}\operatorname{Cov}(X,Y) &= \mathrm{E}(X)\mathrm{E}(Y) - \mathrm{E}(X)\mathrm{E}(Y)\\ &= 0\end{aligned}$$

■

Uncorrelated random variables X and Y are said to uncorrelated if and only if

$$\operatorname{Cov}(X,Y) = 0$$

(Pearson's) correlation coefficient

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

It can be shown that

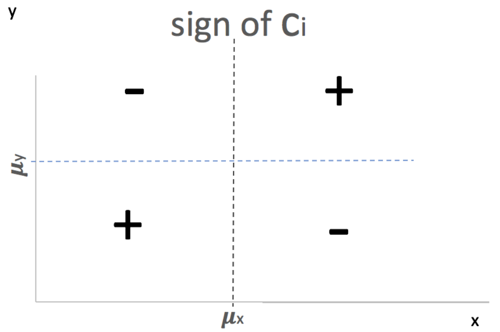
$$-1 \leq \rho_{X,Y} \leq +1$$

ρ is unitless, therefore it can be used to compare across different ramp variables.

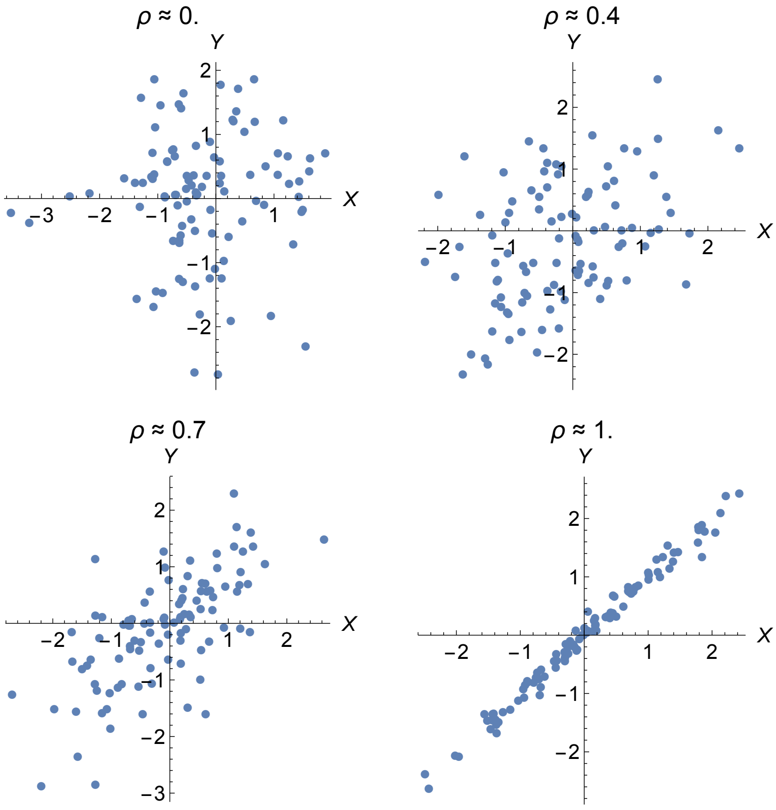
2.6.1 Interpretation of Covariance and Correlation

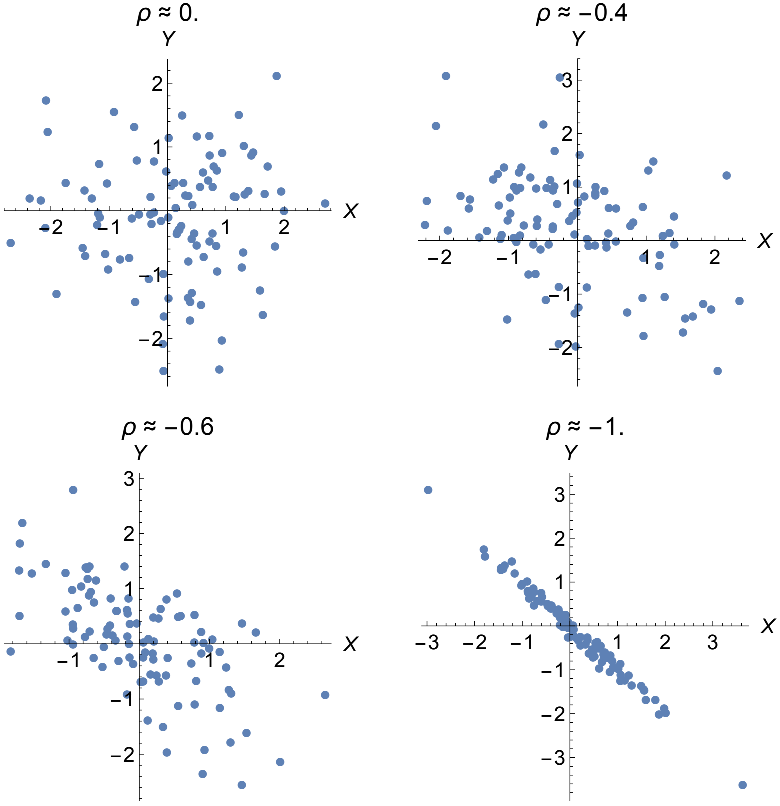
Let the following approximation of $\mathrm{Cov}(X,Y)$ based on a sample $\{(x_1,y_1),(x_2,y_2),\dots(x_n,y_n)\}$:

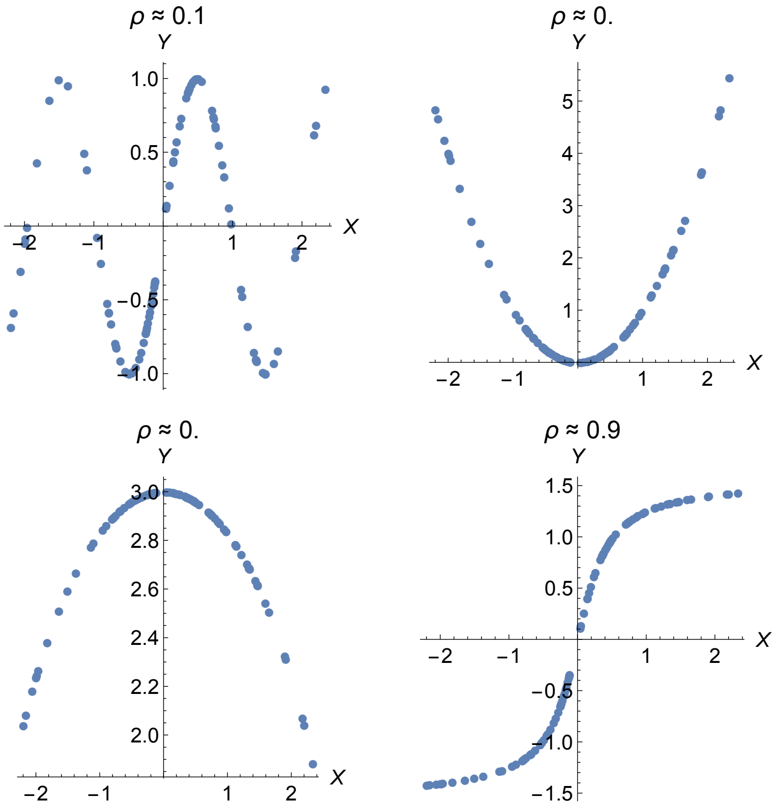
$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}\Big[\big(X-\mu_X\big)\big(Y-\mu_Y\big)\Big]\\ &= \sum_{x,y}(x-\mu_X)(y-\mu_Y)p_{X,Y}(x,y)\\ &\approx \frac{1}{n}\sum_{i=1}^{n}\underbrace{(x_i-\mu_X)(y_i-\mu_Y)}_{c_i}\end{aligned}$$



The following figures illustrate scatter plots for typical values for the correlation coefficient.


ρ ≈ 0.
Y
X
ρ ≈ 0.4
Y
X
ρ ≈ 0.7
Y
X
ρ ≈ 1.
Y
X


ρ ≈ 0.
ρ ≈ −0.4
ρ ≈ −0.6
ρ ≈ −1.
X
Y


ρ ≈ 0.1
Y
X
ρ ≈ 0.
Y
X
ρ ≈ 0.
Y
X
ρ ≈ 0.9
Y
X

These figures shows that

1. even if $\rho \approx 0$, that does not mean there is no relationship between X and Y, it just means that there is no linear relationship.
2. even if $\rho > 0.9$ the relationship may be highly nonlinear

STOP! **Correlation does not necessarily mean causation**. For example:

1. The yield of oranges and apples are highly correlated in the Monterey Valley. Therefore, to produce more apples one should produce more oranges?
2. There is a high correlation between the number of police officers and the number of crimes on a given city. Therefore, to reduce crime rates one should reduce the police force?
3. → More examples from Wikipedia...

The use of a **controlled experiment** is the most effective way of establishing causality between variables. In a controlled study, the sample or population is split in two, with both groups being comparable in almost every way. The two groups then receive different treatments, and the outcomes of each group are assessed.

2.6.2 Covariance of linear combinations

In this section we consider multiple random variables

$$\mathbf{X} = (X_1, X_2, \ldots X_n)^T$$

which we call the random vector $\mathbf{X}$, with means, variances and covariances (and correlations) given:

$$\begin{aligned} \mathrm{E}(X_i) &= \mu_i \\ \mathrm{V}(X_i) &= \sigma_i^2 \\ \mathrm{Cov}(X_i, X_j) &= \sigma_{ij} \\ &= \rho_{ij}\sigma_i\sigma_j \end{aligned}$$

Covariance matrix The covariance matrix of the random vector $\mathbf{X}$ with vector mean $\boldsymbol{\mu} = \{\mu_1, \mu_2, \ldots \mu_n\}$

is the $n \times n$ matrix whose $(i,j)^{th}$ element is $\mathrm{Cov}(X_i, X_j)$:

$$\Sigma_{\mathbf{X}} = \mathrm{E}\Big((\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T\Big) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \cdots \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Its two main properties are:

1. it is symmetric since
$$\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$$
2. its diagonal elements give the variance, since
$$\mathrm{Cov}(X_i, X_i) = \sigma_i^2$$

Fact 2.9 — Covariance of linear combinations. Let the following linear combinations:

$$U = \sum_{i=1}^{n} a_i X_i = \mathbf{a}^{\mathbf{T}} \mathbf{X}$$
$$V = \sum_{j=1}^{n} b_j X_j = \mathbf{b}^{\mathbf{T}} \mathbf{X}$$

with $\mathbf{a} = (a_1, \ldots, a_n)^T$ and $\mathbf{b} = (b_1, \ldots, b_n)^T$. Then,

$$\begin{aligned} \mathrm{Cov}(U, V) &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j \sigma_{ij} \\ &= \mathbf{a}^{\mathbf{T}} \boldsymbol{\Sigma}_{\mathbf{X}} \mathbf{b} \end{aligned}$$

Proof. From the properties of the expectation we have that:

$$\mathrm{E}(U) = \sum_{i=1}^{n} a_i \mu_i = \mathbf{a}^T \boldsymbol{\mu}, \qquad \mathrm{E}(V) = \sum_{j=1}^{n} b_j \mu_j = \mathbf{b}^T \boldsymbol{\mu}$$

therefore, by definition, the covariance of U and V is,

$$\begin{aligned}
\operatorname{Cov}(U,V) &= \mathrm{E}\Big[(U-\mathrm{E}(U))\,(V-\mathrm{E}(V))\Big] \\
&= \mathrm{E}\Big((\mathbf{a}^{\mathbf{T}}\mathbf{X}-\mathbf{a}^{\mathbf{T}}\boldsymbol{\mu})(\mathbf{b}^{\mathbf{T}}\mathbf{X}-\mathbf{b}^{\mathbf{T}}\boldsymbol{\mu})\Big) \\
&= \mathrm{E}\Big(\mathbf{a}^{\mathbf{T}}(\mathbf{X}-\boldsymbol{\mu})\mathbf{b}^{\mathbf{T}}(\mathbf{X}-\boldsymbol{\mu})\Big) \\
&= \mathrm{E}\Big(\mathbf{a}^{\mathbf{T}}(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^{T}\mathbf{b}\Big) \\
&= \mathbf{a}^{\mathbf{T}}\mathrm{E}\Big(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^{T}\Big)\,\mathbf{b} \\
&= \mathbf{a}^{\mathbf{T}}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{b}
\end{aligned}$$

Without matrix notation, the proof goes like this:

$$\begin{aligned}
\operatorname{Cov}(U,V) &= \mathrm{E}\left[(U-\mathrm{E}(U))\,(V-\mathrm{E}(V))\right] \\
&= \mathrm{E}\left[\left(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i \mu_i\right)\left(\sum_{j=1}^{m} b_i X_j - \sum_{j=1}^{m} b_j \mu_j\right)\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{n} a_i\,(X_i-\mu_i)\sum_{j=1}^{m} b_j\left(X_j-\mu_j\right)\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j\,(X_i-\mu_i)\left(X_j-\mu_j\right)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \operatorname{Cov}(X_i,X_j)
\end{aligned}$$

■

Important corollaries of theorem 2.9:

Variance of linear combination

$$\mathrm{V}\left(\mathbf{a}^{\mathbf{T}}\mathbf{X}\right) = \mathbf{a}^{\mathbf{T}}\mathbf{\Sigma}_{\mathbf{X}}\mathbf{a}$$

or in standard notation:

$$\mathrm{V}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \sigma_i^2 + \underbrace{2\sum_{i=1}^{n}\sum_{j=i+1}^{n} a_i a_j \sigma_{i,j}}_{0 \text{ if } X_i\text{'s are independent}} \tag{2.1}$$

This formula illustrates the concept of **propagation of errors**, where the errors (variances and covariances) in the X_i's produce errors in the function $U = \sum_{i=1}^{n} a_i X_i$.

Proof. Let $U = \sum_{i=1}^{n} a_i X_i$

$$
\begin{aligned}
\mathrm{V}(U) &= \mathrm{Cov}(U, U) && \text{(by definition)} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \,\mathrm{Cov}\left(X_i, X_j\right) && \text{(theorem 2.9)} \\
&= \sum_{i=1}^{n} a_i^2 \,\mathrm{V}(X_i) + \sum_{i=1}^{n} \sum_{j \neq i}^{n} a_i a_j \,\mathrm{Cov}(X_i, X_j) && (i = j \text{ terms first}) \\
&= \sum_{i=1}^{n} a_i^2 \,\mathrm{V}(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} a_i a_j \,\mathrm{Cov}(X_i, X_j) && (\text{symmetry of } \Sigma_{\mathbf{X}})
\end{aligned}
$$

■

Important results for two random variables:

$$
\begin{aligned}
\mathrm{V}(X+Y) &= \mathrm{V}(X)+\mathrm{V}(Y)+2\,\mathrm{Cov}(X,Y)\\
\mathrm{V}(X-Y) &= \mathrm{V}(X)+\mathrm{V}(Y)-2\,\mathrm{Cov}(X,Y)\\
\mathrm{Cov}(aX+b,cY+d) &= ac\ \mathrm{Cov}(X,Y)\\
\rho_{aX+b,cY+d} &= \rho_{X,Y}
\end{aligned}
$$

Example 30. The random variables X and Y have joint probability distribution specified by the following table:

	y=1	y=2	y=3
x=1	0.30	0.05	0.00
x=2	0.05	0.20	0.05
x=3	0.00	0.05	0.30

(a) Find the expectation of XY.
(b) Find the covariance $\mathrm{Cov}(X,Y)$ between X and Y.
(c) What is the correlation between X and Y?
(d) Suppose the random variables X and Y above are connected to random variables U and V by the relations

$$X = 2U + 5$$
$$Y = 4V + 5$$

What is the covariance $\mathrm{Cov}(U,V)$?
(e) What is the correlation between U and V?

Solution: **Part a)** The mass function of XY is tabulated below:

xy	1	2	3	4	6	9
$P(XY = xy)$	0.3	0.1	0.0	0.2	0.1	0.3

Therefore

$$E(XY) = 0.3 + 0.2 + 0 + 0.8 + 0.6 + 2.7 = 4.6$$

Part b)
We require $E(X), E(Y), V(X)$, and $V(Y)$; these are as follows

$$\begin{aligned}
E(X) &= 0.35 + 0.6 + 1.05 \\
&= 2.0 \\
&= E(Y), \\
E\left(X^2\right) &= 0.35 + 2^2 \times 0.3 + 3^2 \times 0.35 \\
&= 4.7 \\
&= E\left(Y^2\right), \\
\mathrm{Var}(X) &= 4.7 - 2^2 \\
&= 0.7 \\
&= \mathrm{Var}(Y).
\end{aligned}$$

Hence

$$
\begin{aligned}
Cov\left(X,Y\right) &= E\left(XY\right) - E\left(X\right)E\left(Y\right) \\
&= 4.6 - 4 \\
&= 0.6
\end{aligned}
$$

Part c) For the correlation we require

$$
\begin{aligned}
\rho\left(X,Y\right) &= \frac{Cov\left(X,Y\right)}{\sqrt{\operatorname{Var}\left(X\right)\operatorname{Var}\left(Y\right)}} \\
&= \frac{0.6}{\sqrt{0.7^2}} \\
&= \frac{6}{7}
\end{aligned}
$$

Part d)

$$
\begin{aligned}
U &= X/2 - 5/2 \\
V &= Y/4 - 5/4
\end{aligned}
$$

Since $\mathrm{Cov}(aX+b, cY+d) = ac\ \mathrm{Cov}(X,Y)$ and so $\mathrm{Cov}\left(U,V\right) = \mathrm{Cov}\left(X,Y\right)/(2\cdot 4) = 0.6/(2\cdot 4)$.
Part e) The correlation between U and V can be written

$$\begin{aligned}
\rho_{U,V} &= \frac{\mathrm{Cov}\left(U,V\right)}{\sqrt{\mathrm{V}\left(U\right)}\sqrt{\mathrm{V}\left(V\right)}} \\
&= \frac{\mathrm{Cov}\left(X,Y\right)/(2\cdot 4)}{\sqrt{(\mathrm{V}\left(X\right)/2^2)(\mathrm{V}\left(V\right)/4^2)}} \\
&= \rho_{X,Y} \\
&= \frac{6}{7}
\end{aligned}$$

$\square$

Example 31. — 2 Dice There is a blue and yellow dice. Compute the correlation between X, the number on the blue dice, and S, the total of the two dice.

Solution:
Write

$$S = X + Y$$

where Y is the number on the yellow dice. Here $n = 2$, $a_1 = 1$, $a_2 = 0$, and $b_1 = 1$, $b_2 = 1$. Therefore,

$$\begin{aligned} \mathrm{Cov}(X,S) &= \mathrm{Cov}(X, X+Y) \\ &= \mathrm{Cov}(X,X) + \mathrm{Cov}(X,Y) \\ &= \mathrm{V}(X) + 0 \end{aligned}$$

Also,

$$\mathrm{V}(S) = \mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y)$$

But $\mathrm{V}(Y) = \mathrm{V}(X)$. So the correlation between X and S is

$$\rho_{X,S} = \frac{\mathrm{V}(X)}{\sqrt{\mathrm{V}(X)}\sqrt{2\,\mathrm{V}(X)}} = 0.707$$

□

Example 32. Find the covariance matrix of $\{U,V\}$ where

$$U = X_1 + X_2$$

and

$$V = X_1 - X_2$$

Solution: Recall from Fact 2.9:

$$\begin{aligned}\operatorname{Cov}(U,V) &= \mathbf{a}^{\mathbf{T}}\boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{b} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \sigma_{ij}\end{aligned}$$

Here $n=2$, $a_1=1$, $a_2=1$ and $b_1=1$, $b_2=-1$. Therefore,

$$\begin{aligned}\operatorname{Cov}(U,V) &= \operatorname{Cov}(X_1,X_1) - \operatorname{Cov}(X_1,X_2) + \operatorname{Cov}(X_2,X_1) + \operatorname{Cov}(X_2,X_2) \\ &= \operatorname{Cov}(X_1,X_1) - \operatorname{Cov}(X_2,X_2) \\ &= \sigma_1^2 + \sigma_2^2\end{aligned}$$

Also,

$$\begin{aligned} \mathrm{V}(U) &= \sigma_1^2 + \sigma_2^2 + 2\sigma_{12} \\ \mathrm{V}(V) &= \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} \end{aligned}$$

and so the covariance matrix is:

$$\begin{pmatrix} \sigma_1^2 + \sigma_2^2 + 2\sigma_{12} & \sigma_1^2 + \sigma_2^2 \\ \sigma_1^2 + \sigma_2^2 & \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} \end{pmatrix}$$

$\square$

Example 33. *

The joint PMF of precipitation, X(in.) and runoff, Y (cfs) (discretized here for simplicity) due to storms at a given location is as follows:

	X=1	X=2	X=3
Y=10	0.0	0.25	0.10
Y=20	0.10	0.0	0.10
Y=30	0.05	0.15	0.25

(a) What is the probability that the next storm will bring a precipitation of 2 in. and a runoff of more than 20 cfs?
(b) After a storm, the rain gauge indicates a precipitation of 2 in. What is the probability that the runoff in this storm in 20 cfs or more?
(c) Are X and Y statistically independent? Substantiate your answer.
(d) Determine and plot the marginal PMF of runoff.
(e) Determine and plot the PMF of runoff for a storm whose participation is 2 in.
(f) Determine the correlation coefficient between precipitation and runoff.

Solution: Answer: (a) 0.15 (b) 0.375 (c) Not independent (f) 0.1103

(a)

$$\begin{aligned} P(X=2,Y>20) &= P(X=2,Y=30) \\ &= 0.15 \end{aligned}$$

(b)

$$\begin{aligned} P(Y \geq 20|X=2) &= \frac{P(X=2,Y=20)+P(X=2,Y=30)}{P(X=2)} \\ &= \frac{0+0.15}{0.25+0+0.15} \\ &= 0.375 \end{aligned}$$

(c)

$$\begin{aligned} P(X=1) &= 0+0.1+0.05 \\ &= 0.15 \end{aligned}$$

$$\begin{aligned} P(Y=10) &= 0+0.25+0.1 \\ &= 0.35 \end{aligned}$$

$$\begin{aligned}P(X=1,Y=10) &= 0 \\ &\neq P(X=1)\cdot P(Y=10)\end{aligned}$$

So they are not independent.
(d)

$$\begin{aligned}P(Y=10) &= 0+0.25+0.1 \\ &= 0.35\end{aligned}$$

$$\begin{aligned}P(Y=20) &= 0.1+0+0.1 \\ &= 0.2\end{aligned}$$

$$\begin{aligned}P(Y=30) &= 0.05+0.15+0.25 \\ &= 0.45\end{aligned}$$

(e)

$$\begin{aligned}P(Y=10|X=2) &= \frac{0.25}{0.25+0+0.15} \\ &= 0.625\end{aligned}$$

$$
\begin{aligned}
P(Y=20|X=2) &= \frac{0}{0.25+0+0.15} \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
P(Y=30|X=2) &= \frac{0.15}{0.25+0+0.15} \\
&= 0.375
\end{aligned}
$$

(f)

$$\begin{aligned}
E(X) &= 0.15 \times 1 + 0.4 \times 2 + 0.45 \times 3 \\
&= 2.3 \\
E(Y) &= 0.35 \times 10 + 0.2 \times 20 + 0.45 \times 30 \\
&= 21 \\
E(X^2) &= 0.15 \times 1^2 + 0.4 \times 2^2 + 0.45 \times 3^2 \\
&= 5.8 \\
E(Y^2) &= 0.35 \times 10^2 + 0.2 \times 20^2 + 0.45 \times 30^2 \\
&= 520 \\
E(XY) &= 0 \times 10 + 0.25 \times 20 + 0.10 \times 30 + 0.1 \times 20 + 0 \times 40 + 0.1 \times 60 + 0.05 \times 30 + 0.15 \times 60 + 0.25 \times 90 \\
&= 49 \\
\rho &= \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \\
&= \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E(X)^2}\sqrt{E(Y^2) - E(Y)^2}} \\
&= \frac{49 - 2.3 \times 21}{\sqrt{5.8 - 2.3^2}\sqrt{520 - 21^2}} \\
&= 0.1103
\end{aligned}$$

□

Example 34. — speeding tickets** A study investigates a group of people who have received speeding tickets in the past year. The joint PMF is summarized in the following table. Let X be the number of tickets received by a person in the past year and Y be the age of that person.

	X=1	X=2
Y=30	c	c + 1/8
Y=40	c	c - 1/8

(a) Determine the value of c. (5 points)
(b) Determine the marginal PMF of X, and the conditional PMF of X. (10 points)
(c) If someone is 30 years old, find the probability that this person gets exact one ticket. (5 points)
(d) Are X and Y statistically independent? Substantiate your answer. (5 points)
(e) Determine the correlation coefficient between X and Y. (10 points)

Solution: (a)

$$\begin{aligned} 1 &= c+(c+\frac{1}{8})+c+(c-\frac{1}{8}) \\ c &= 0.25 \end{aligned}$$

(b) Marginal:

$$\begin{aligned} p_{X=1} &= c+c \\ &= 0.5 \\ p_{X=2} &= (c+\frac{1}{8})+(c-\frac{1}{8}) \\ &= 0.5 \end{aligned}$$

Conditional:

$$
\begin{aligned}
p_{X=1|Y=30} &= \frac{c}{c+(c+\frac{1}{8})} \\
&= 0.4 \\
p_{X=2|Y=30} &= \frac{c+\frac{1}{8}}{c+(c+\frac{1}{8})} \\
&= 0.6 \\
p_{X=1|Y=40} &= \frac{c}{c+(c-\frac{1}{8})} \\
&= 0.67 \\
p_{X=2|Y=40} &= \frac{c-\frac{1}{8})}{c+(c-\frac{1}{8})} \\
&= 0.33
\end{aligned}
$$

(c)

$$
\begin{aligned}
p_{X=1|Y=30} &= \frac{c}{c+(c+\frac{1}{8})} \\
&= 0.4
\end{aligned}
$$

(d) Not independent. Because $p_{X=1|Y=30} \neq p_{X=1}$

(e)

$$
\begin{aligned}
E(X) &= 1.5 \\
E(Y) &= \frac{270}{8} \\
E(X^2) &= 2.5 \\
E(Y^2) &= 1162.5 \\
E(XY) &= c \times 1 \times 30 + c \times 1 \times 40 + (c + \frac{1}{8}) \times 2 \times 30 + (c - \frac{1}{8}) \times 2 \times 40 \\
&= 50 \\
\rho &= \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} \\
&= \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \\
&= -0.258
\end{aligned}
$$

□

Example 35. * Given the joint distribution of (X, Y):

$p_{X,Y}(x,y)$		X			
		-1	0	1	Σ
Y	-1	1/16	3/16	1/16	5/16
	0	3/16	0	3/16	3/8
	1	1/16	3/16	1/16	5/16
	Σ	5/16	3/8	5/16	1

Calculate the covariance of X and Y. Are they statistically independent?

Solution: The covariance is zero but since $p_{X,Y}(0,0) \neq p_X(0) \cdot p_Y(0)$, X and Y are not independent.
$\square$

Example 36. — Tornadoes, take 2 100 structures are located in a region where tornado wind force must be considered in its design. Suppose that from the records of tornadoes for the past 200 years, it is estimated that (i) during any given year the probability of having 0, 1 and 2 tornadoes is 0.5, 0.3 and 0.2, respectively, (ii) the number of tornadoes in different years are independent, and (iii) if a tornado occurs, a structure will be damaged with probability **p=5**%.

a) if two tornadoes occurred last year, how many structures do you expect to have been damaged?
b) what is the probability the a structure will be damaged in the next five years?
c) calculate the mean and variance of the number of structures damaged in the next five years?

d) If you're a contractor in charge of rehabilitating the structures in the region after a tornado damage, compute the mean and variance of your yearly income, U, if you charge c dollars per rehabilitation work.
e) calculate the coefficient of variation of your yearly income, and comment.

Solution: Let
$X =$ number of tornadoes on a given year, $S_X = \{0, 1, 2\}$ tornadoes.
Y_i number of times structure $i = 1, 2, \ldots 100$ is damaged due to tornadoes on a given year.
The event $(Y_i > 0|X = x)$ is similar to obtaining at least one Head out of x tosses of a coin with $\mathrm{P}(Head) = p$, therefore:

$$\mathrm{P}(Y_i > 0|X = x) = 1 - (1 - p)^x$$

a) If two tornadoes occurred last year, how many structures do you expect to have been damaged?

Let $Z =$ number of structures damaged last year. Let

$$Z_i = \begin{cases} 1 & \text{if structure } i \text{ was damaged last years} \\ 0 & \text{otherwise} \end{cases}$$

We are interested in the expected value of $Z = \sum_{i=1}^{100} Z_i$. By linearity of expectation we get

$$\mathrm{E}(Z) = \sum_{i=1}^{100} \mathrm{E}(Z_i) = \sum_{i=1}^{100} \mathrm{P}(Z_i = 1) = \sum_{i=1}^{100} \mathrm{P}(Y_i > 0|\mathbf{X{=}2}) = 100(1 - (1 - p)^2) = 9.75$$

$\rightarrow$ 10 structure.

b) what is the probability the a structure will be damaged in the next five years?

Since the number of tornadoes are independent from year to year, we can focus on a single year, calculate the probability of $D =$ **damage in one year for one structure**, and then " flip a coin" five times with

$$P(D) = P(Y_i > 0) = 1 - P(Y_i = 0)$$

Since we don't know the number of tornadoes that will occur, we use the total prob. rule:

$$\begin{aligned}
P(Y_i = 0) &= \sum_{x=0}^{2} P(Y = 0|X = x)P(X = x) \\
&= \sum_{x=0}^{2} (1-p)^x P(X = x) \\
&= (1-p)^0 P(X = 0) + (1-p)^1 P(X = 1) + (1-p)^2 P(X = 2) \\
&= ((1)(0.5)) + ((0.95)(0.3)) + \left((0.95^2)(0.2)\right) = 0.9655
\end{aligned}$$

then $P(D) = 1 - 0.9655 = 0.0345$ for one year. The desired probability is $1 - (1 - P(D))^5 = 0.16$.

c) calculate the mean and variance of the number of structures damaged in the next five years?

Let

$$Z_i = \begin{cases} 1 & \text{if structure } i \text{ is damaged in five years} \\ 0 & \text{otherwise} \end{cases}$$

Thus, $\mathrm{P}(Z_i = 1) = 0.16$ and $\mathrm{E}(Z_i) = 0.16$, $\mathrm{V}(Z_i) = 0.16(1-0.16) = 0.13$. We are interested in

$$Z = \sum_{i=1}^{100} Z_i$$

By linearity of expectation we get

$$\mathrm{E}(Z) = \sum_{i=1}^{100} \mathrm{E}(Z_i) = \sum_{i=1}^{100} 0.16 = 16$$

and by result (2.1) for the variance of a sum we have,

$$\mathrm{V}(Z) = \sum_{i=1}^{100} \mathrm{V}(Z_i) = \sum_{i=1}^{100} 0.13 = 13$$

d) If you're a contractor in charge of rehabilitating the structures in the region after a tornado damage, compute the mean and variance of your yearly income, U, if you charge c dollars per rehabilitation work.

Let Y_i be the number of times structure i is damaged due to tornadoes on a given year. We are interested in $U = \sum_{i=1}^{100} cY_i$. By linearity of expectation we get

$$\mathrm{E}(U) = \sum_{i=1}^{100} c\mathrm{E}(Y_i) = 100c\mathrm{E}(Y_i)$$

and by result (2.1) for the variance of a sum we have,

$$\mathrm{V}(U) = \sum_{i=1}^{100} c^2\mathrm{V}(Y_i) = 100c^2\mathrm{V}(Y_i)$$

To compute the mean and variance of Y_i we need its PMF. From the problem statement we have p_X, and we can determine $p_{Y_i|X}$ as we did on part a), and then we use the multiplication rule,

$$p_{X,Y_i}(x,y) = p_X(x)p_{Y_i|X}(y).$$

marginal distribution of X :

x	0	1	2
$p_X(x)$	0.5	0.3	0.2

conditional distribution of $Y_i|X : \mathrm{P}(Y_i = y|X = x) = \binom{x}{y} p^y (1-p)^{x-y}$:

y	0	1	2	Σ
$p_{Y_i\|X=0}(y)$	1	0	0	1
$p_{Y_i\|X=1}(y)$	$1-p = 0.95$	$p = 0.05$	0	1
$p_{Y_i\|X=2}(y)$	$(1-p)^2 = 0.9025$	$2p(1-p) = 0.095$	$p^2 = 0.0025$	1

and we can calculate the joint distribution of (X, Y_i):

			$\boxed{Y_i}$		
$p_{X,Y_i}(x,y)$		0	1	2	Σ
	0	0.5	0	0	0.5
$\boxed{X}$	1	0.285	0.015	0	0.3
	2	0.1805	0.019	0.0005	0.2
	Σ	0.9655	0.034	0.0005	1

and finally we have the marginal distribution of Y_i :

y	0	1	2
$p_Y(y)$	0.9655	0.034	0.0005

and we obtain $\mathrm{E}(Y_i) = 0.035, \mathrm{V}(Y_i) = 0.0347$. Therefore,

$$\mathrm{E}(U) = 3.5c$$

$$\mathrm{V}(U) = 3.47c^2$$

e) $\delta_U = \mathrm{V}(U)^{1/2} / \mathrm{E}(U) = 0.53$
Since this coefficient of variation is greater than 30 %, the yearly income has a large variability.

$\square$

Example 37. — Stock Prices, simple portfolio Model Let

X_i = rate of return for stock i

μ_i = expected rate of return (historically 10% annually)

σ_i = price volatility (standard deviation of X_i, historically 15% monthly)

The value of a portfolio of the stocks

$$\mathbf{X} = (X_1, X_2, \dots X_n)^T$$

is:

$$U = \sum_{i=1}^{n} a_i X_i = \mathbf{a}^{\mathbf{T}}\mathbf{X}$$

$$\mathrm{E}(U) = \sum_{i=1}^{n} a_i \mu_i = \mathbf{a}^T \boldsymbol{\mu}$$

where the a_i's represent the weight of each stock in the portfolio, with:

$$\begin{aligned} &0 \leq a_i \leq 1 \\ &\sum_{i=1}^{n} a_i = 1 \end{aligned}$$

The risk of the portfolio is given by its variance:

$$\begin{aligned} \mathrm{V}(U) &= \mathbf{a}^{\mathbf{T}} \boldsymbol{\Sigma}_{\mathbf{X}} \mathbf{a} &(2.2)\\ &= \sum_{i=1}^{n} a_i^2 \sigma_i^2 + \underbrace{2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} a_i a_j \sigma_{i,j}}_{0 \text{ if } X_i\text{'s are independent}} &(2.3) \end{aligned}$$

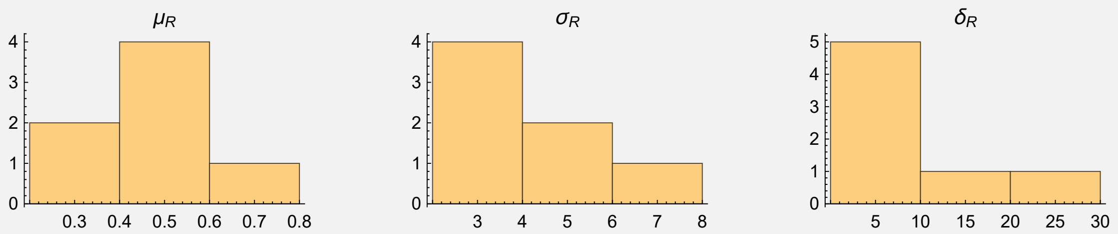
where:

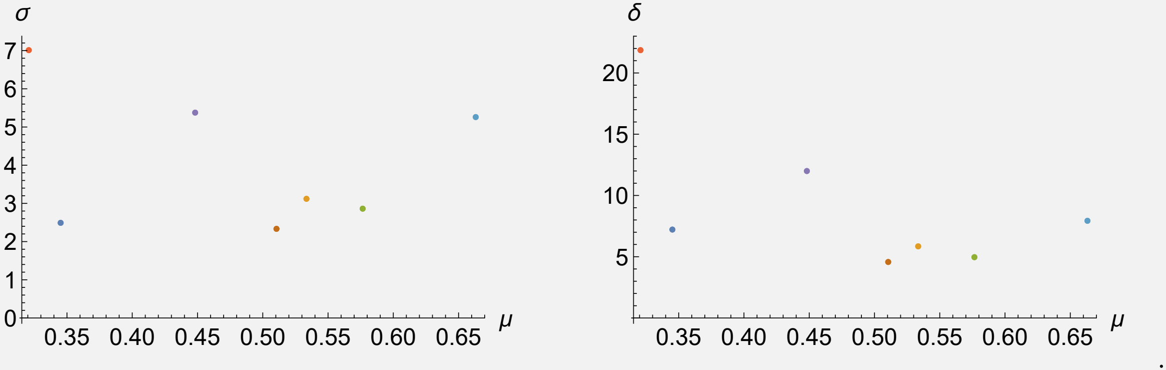
$$\Sigma_{\mathbf{X}} = \mathrm{E}\Big((\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T\Big) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \cdots \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

is the covariance matrix for the stock returns. The idea is to find that the weights a_i that minimizes the variance while maximizing the value of U.

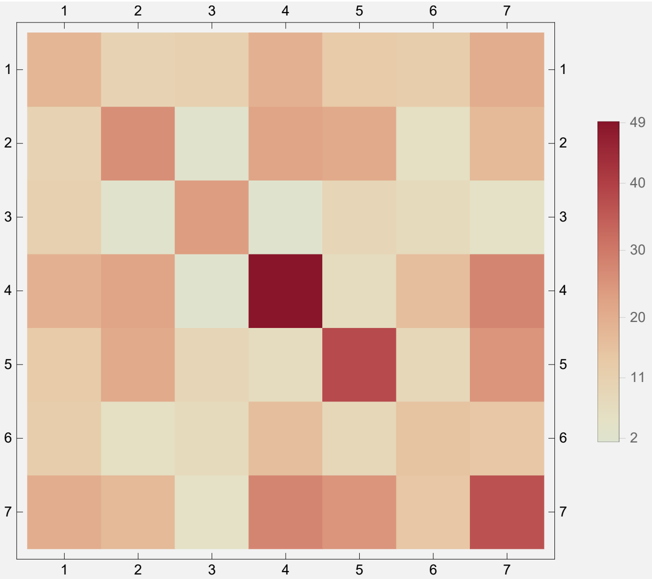
Consider the following tech stocks weekly return data for 2016-2017:

	stock	μ_i	σ_i	δ_i
1	GOOGL	0.35	2.49	7.22
2	AAPL	0.53	3.12	5.85
3	FB	0.58	2.86	4.96
4	TWTR	0.32	7.01	21.86
5	TSLA	0.45	5.38	12.
6	MSFT	0.51	2.33	4.57
7	NFLX	0.66	5.26	7.93





$\Sigma_{\mathbf{X}} =$



$$\Sigma_{\mathbf{X}} =$$

$$\begin{bmatrix} 6.2 & 4. & 4.1 & 6.5 & 4.6 & 4.6 & 6.5 \\ 4. & 9.7 & 1.4 & 6.8 & 6.6 & 3.2 & 6. \\ 4.1 & 1.4 & 8.2 & 1.2 & 4. & 3.7 & 3.2 \\ 6.5 & 6.8 & 1.2 & 49.2 & 3.7 & 5.8 & 11.6 \\ 4.6 & 6.6 & 4. & 3.7 & 28.9 & 3.9 & 9.5 \\ 4.6 & 3.2 & 3.7 & 5.8 & 3.9 & 5.4 & 5.1 \\ 6.5 & 6. & 3.2 & 11.6 & 9.5 & 5.1 & 27.7 \end{bmatrix}$$

a) if your portfolio only has stocks from Netflix and Facebook, what are the weights that minimize the variance?
b) what is the expected value, variance and coefficient of variation of the return of the portfolio in part a)?
c) repeat a) and b) for Apple and Twitter?
d) which portfolio would you recommend buying, why?

Solution:

a) if your portfolio only has stocks from Netflix and Facebook, what are the weights that minimize the variance?

$$\begin{aligned} \mathrm{V}(U) &= a_3^2 \mathrm{V}(X_3) + a_7^2 \mathrm{V}(X_7) + 2a_3 a_7 \operatorname{Cov}(X_3, Y_7) \\ &= a_3^2 \cdot 8.2 + (1 - a_3)^2 \cdot 27.7 + 2a_3(1 - a_3) \cdot 3.2 \\ &= a_3 (29.5186 a_3 - 48.9958) + 27.666 \end{aligned}$$

which is minimized at $a_3^* = 0.83$, which implies that $a_7^* = 1 - 0.83 = 0.17$.

b) what is the expected value, variance and coefficient of variation of the return of the portfolio in part a)?

Evaluating the respective formulas with weights $a_3^* = 0.83$ and $a_7^* = 0.17$ gives

$$0.59, 2.7, 4.6,$$

respectively.

c) repeat a) and b) for Apple and Twitter?

$$\mathrm{V}(U) = a_2 (45.2135 a_2 - 84.6517) + 49.1704$$

which is minimized at $a_2^* = 0.94$, which implies that $a_4^* = 1 - 0.94 = 0.06$. The expected value, variance and coefficient of variation are

$$0.52, 9.55, 5.94,$$

resp.

d) which portfolio would you recommend buying, why?

$\square$

Midterm 1 solution

Example 38. Consider the experiment of tossing **4 fair coins**. Let X be the random variable that denotes the number of heads that result. The sample space for this experiment is illustrated in the table below, which also shows the number of heads in each possible case.

coin 1	H	H	H	H	H	H	H	H	T	T	T	T	T	T	T	T
coin 2	H	H	H	H	T	T	T	T	H	H	H	H	T	T	T	T
coin 3	H	H	T	T	H	H	T	T	H	H	T	T	H	H	T	T
coin 4	H	T	H	T	H	T	H	T	H	T	H	T	H	T	H	T
ΣH	4	3	3	2	3	2	2	1	3	2	2	1	2	1	1	0

a) Find the CDF and PMF of X and draw sketches of each one.

b) Determine the median, upper quartile and lower quartile and show them graphically in one of the sketches of part a)

c) Determine $\mathrm{P}(0 < X \leq 3 \mid X \leq 2)$

d) BONUS: suppose two players play this game, and the one with the largest number of Hs wins. Let X_1 and X_2 denote their corresponding random variables, the distribution of each one corresponding to the one you calculate it in part a). Find the joint PMF and the probability that player one wins by more than one point. Hint: X_1 and X_2 are independent.

Solution:

a) Find the CDF and PMF of X and draw sketches of each one.

The PMF of X is given by

$$p_X(x) = \begin{cases} 1/16 & \text{if } x = 0 \text{ or } x = 4 \\ 1/4 & \text{if } x = 1 \text{ or } x = 3 \\ 3/8 & \text{if } x = 2 \end{cases}$$

b) Determine the median, upper quartile and lower quartile and show them graphically in one of the sketches of part a): $\{2, 3, 1\}$

c) $\mathrm{P}(0 < X \leq 3 \mid X \leq 2) = 0.91$

d) BONUS: suppose two players play this game, and the one with the largest number of Hs wins. Let X_1 and X_2 denote their corresponding random variables, the distribution of each one corresponding to the one you calculate it in part a). Find the joint PMF and the probability that player one wins by more than one point. Hint: X_1 and X_2 are independent.

$\square$

Example 39. The state of GA has license plates showing three numbers and four letters. How many different license plates are possible

a) if the numbers must come after the letters?
b) if there is no restriction on where the letters and numbers appear?
c) as in part b) but replacing all numbers by an "A" and all letters by "B"?
d) BONUS: as in part b) but replacing all numbers < 5 by an "A", all numbers ≥ 5 by "B" and all letters by "C"?

Solution:

a) $10^3 \times 26^4$
b) The three number places can be arranged in $\binom{7}{3}$ different ways within the 7 places, each arrangement having 10^3 possibilities for the digits. Similarly for the letter places which have one way to be arranged in the remaining four places. Therefore,

$$\binom{7}{3} 10^3 26^4$$

c) Letters and numbers are treated as identical objects: of the $n = 7$ objects $k = 3$ are identical of type A and the remaining $(n-k) = 4$ of type B. The number of ways one can arrange these

n objects is

$$\binom{7}{3}$$

d) BONUS: Here we don't know how many of the numbers will be <5 so we add all the possibilities:

$$\sum_{x=0}^{3} \binom{7}{x}$$

$\square$

Example 40. The annual rainfall, R, (accumulated generally during the winter and spring of each year) in Orange County, California, has the following CDF:

r	0	1	2	3	4	5	6	7	...
$F_R(r)$	0.00003	0.00048	0.0036	0.0175	0.0592	0.15	0.30	0.5	...

in inches, and assuming that R can only take on integer values. Suppose the current water policy of the county is such that if the annual rainfall is less than 6 in. for a given year, water rationing will be required during the summer and fall of that year.
Whenever the annual rainfall is less than 6 in. in any given year, the probability of damage to the agricultural crops in the county is 30%. Assuming that crop damages between dry years (i.e., with rainfall less than 6 in.) are statistically independent. Determine:

a) the probability of water rationing in Orange County in any given year? and in the next 20 years?
b) the probability of crop damage in the next 4 years?
c) If there is an economic loss of c dollars every time there is crop damage in the region, calculate the mean and variance of the economic loss in 4 years.
d) BONUS: Derive the joint PMF of the number of times there is rationing, and the number of times there is crop damage, both during a period of three years.

Solution:

a) the probability of water rationing in Orange County in any given year is $\mathrm{P}(R < 6) = 0.15$. In the next 20 years the probability that there will be rationing at least once, is similar to flipping a biased coin 20 times and obtaining the probability of no heads, where the probability of heads is $p = 0.15$:

$$1-(1-p)^{20} = 0.961$$

b) the probability of crop damage in the next 4 years can be done in two ways:

Method 1: focus on one year and then flip a coin four times.
Let

$$Y_i = \begin{cases} 1 & \text{if during the year } i \text{ there is crop damage} \\ 0 & \text{otherwise} \end{cases}$$

Thus, in this case $p = \mathrm{P}(Y_i = 1) = 0.15 \times 0.3 = 0.045$ and the desired probability is

$$1-(1-p)^4 = 0.168$$

Notice:

$$\mathrm{E}(Y_i) = 1 \times p + 0 \times (1-p) = p = 0.045$$

$$\mathrm{E}\left(Y_i^2\right) = 1^2 \times p + 0^2 \times (1-p) = p = 0.045$$

$$\mathrm{V}(Y_i) = \mathrm{E}\left(Y_i^2\right) - \mathrm{E}(Y_i)^2 = p(1-p) = 0.045 \times 0.055 = 0.0429$$

Method 2: focus on 4-year period and use total probability. Let X be the rv representing the number of drought years within the four years of interest, $S_X = \{0,1,2,3,4\}$, and Y be the number of times there was crop damage in the region during the same period. Then, with $p = 0.3$:

$$\mathrm{P}(Y = 0|X = i) = (1-p)^i$$

The distribution of X can be obtained with the help of the table in problem 38 with probability of heads of $a = 0.15$. For example, $X = 1$ appears 4 times on the table, and therefore $\mathrm{P}(X = 1) = 4a(1-a)^3 = 0.368$:

x	0	1	2	3	4
$p_X(x)$	0.522006	0.368475	0.0975375	0.011475	0.00050625

and therefore the desired probability is with $p = 0.3$:

$$\begin{aligned}
\mathrm{P}(Y > 0) &= 1 - \sum_{i=0}^{4} \mathrm{P}(Y = 0 | X = i)\mathrm{P}(X = i) \\
&= 1 - \sum_{i=0}^{4} (1-p)^i \mathrm{P}(X = i) \\
&= 0.168
\end{aligned}$$

c) If there is an economic loss of c dollars every time there is crop damage in the region, calculate the mean and variance of the economic loss in 4 years.

We are interested in $U = \sum_{i=1}^{4} cY_i$. By linearity of expectation we get

$$\mathrm{E}\left(U\right) = \sum_{i=1}^{4} c\mathrm{E}\left(Y_i\right) = 4c\mathrm{E}\left(Y_i\right) = 4C \times 0.045 = 0.18c$$

and by result (2.1) for the variance of a sum we have,

$$\mathrm{V}\left(U\right) = \sum_{i=1}^{4} c^2\mathrm{V}\left(Y_i\right) = 4c^2\mathrm{V}\left(Y_i\right) = 4c^2 \times 0.0429 = 0.1719c^2$$

To compute the mean and variance of Y_i we used part b)

d) BONUS: Derive the joint PMF of $X =$ the number of times there is rationing, and $Y =$ the number of times there is crop damage, both during a period of three years.

The distribution of X can be obtained with the help of the table in problem 3:

coin 1	H	H	H	H	T	T	T	T
coin 2	H	H	T	T	H	H	T	T
coin 3	H	T	H	T	H	T	H	T
ΣH	3	2	2	1	2	1	1	0

Here the probability of heads is $a = 0.15$. For example, $X = 3$ appears 1 times on the table, and therefore $\mathrm{P}(X = 3) = 1 \times a^3(1-a)^0 = 0.003375$:

x	0	1	2	3
$p_X(x)$	0.614125	0.325125	0.057375	0.003375

conditional distribution of $Y|X$, $(p = 0.3)$:

y	0	1	2	3	Σ
$p_{Y\|X=0}(y)$	1	0	0	0	1
$p_{Y\|X=1}(y)$	$(1-p)^1$	p^1	0	0	1
$p_{Y\|X=2}(y)$	$(1-p)^2$	$2(1-p)^1p^1$	p^2	0	1
$p_{Y\|X=3}(y)$	$(1-p)^3$	$3(1-p)^2p^1$	$3(1-p)^1p^2$	p^3	1

or:

y	0	1	2	3	Σ
$p_{Y\|X=0}(y)$	1	0	0	0	1
$p_{Y\|X=1}(y)$	0.7	0.3	0	0	1
$p_{Y\|X=2}(y)$	0.49	0.42	0.09	0	1
$p_{Y\|X=3}(y)$	0.343	0.441	0.189	0.027	1

and we can calculate the joint distribution of (X,Y):

$p_{X,Y}(x,y)$		Y: 0	1	2	3	Σ
	0	0.614	0.	0.	0.	0.614
X	1	0.228	0.098	0.	0.	0.325
	2	0.028	0.024	0.005	0.	0.057
	3	0.0012	0.0015	0.0006	0.0001	0.0034
	Σ	0.871	0.123	0.006	0.0001	1.

□

Example 41. Let

X_i = rate of return for stock i

μ_i = expected rate of return (historically 10% annually)

σ_i = price volatility (standard deviation of X_i, historically 15% monthly)

Consider the following tech stocks weekly return data for 2016-2017:

	stock	μ_i	σ_i	δ_i
1	GOOGL	0.35	2.49	7.22
2	AAPL	0.53	3.12	5.85
3	FB	0.58	2.86	4.96
4	TWTR	0.32	7.01	21.86
5	TSLA	0.45	5.38	12.
6	MSFT	0.51	2.33	4.57
7	NFLX	0.66	5.26	7.93

The covariance matrix for the stock returns:

$$\begin{pmatrix} 6.2 & 4. & 4.1 & 6.5 & 4.6 & 4.6 & 6.5 \\ 4. & 9.7 & 1.4 & 6.8 & 6.6 & 3.2 & 6. \\ 4.1 & 1.4 & 8.2 & 1.2 & 4. & 3.7 & 3.2 \\ 6.5 & 6.8 & 1.2 & 49.2 & 3.7 & 5.8 & 11.6 \\ 4.6 & 6.6 & 4. & 3.7 & 28.9 & 3.9 & 9.5 \\ 4.6 & 3.2 & 3.7 & 5.8 & 3.9 & 5.4 & 5.1 \\ 6.5 & 6. & 3.2 & 11.6 & 9.5 & 5.1 & 27.7 \end{pmatrix}$$

a) if your portfolio only has stocks from Apple and Twitter, what are the weights that minimize the variance?
b) what is the expected value, variance and coefficient of variation of the return of the portfolio in part a)?
c) repeat a) and b) for Google and Microsoft?
d) Bonus: which portfolio would you recommend buying, why?

Solution:

a) for Apple and Twitter

$$\begin{aligned}\mathrm{V}(U) &= (1-a_2)(49.1704(1-a_2)+6.84451a_2)+a_2(6.84451(1-a_2)+9.73216a_2)\\ &= a_2(45.2135a_2-84.6517)+49.1704\end{aligned}$$

which is minimized at $a_2^* = 0.94$, which implies that $a_4^* = 1-0.94 = 0.06$.

b) The expected value, variance and coefficient of variation are

$$0.52, 9.55, 5.94,$$

resp.

c) for Google and Microsoft: $a_1 = 0.349$ and the expected value, variance and coefficient of variation are

$$0.45, 5.14, 5.0$$

resp.

d) which portfolio would you recommend buying, why? no right or wrong answers in this particular case, it depends on each individual's trade-off between value and risk.

□