Random Variables

Mahbub Latif, PhD

January 2025

Plan

- Introduction
- Discrete random variables
- Continuous random variables
- Expectation of a random variable
- Variance of a random variable
- Jointly distributed random variables

Random variables

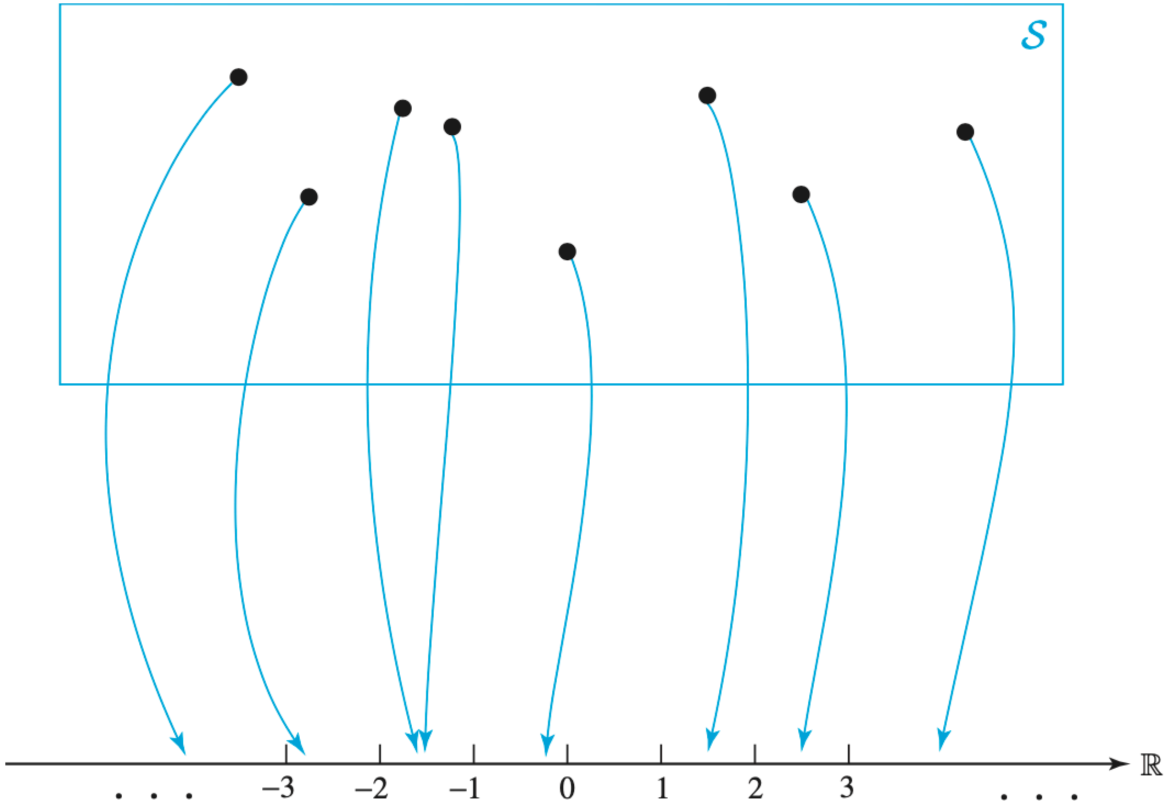
- Random variables are one of the fundamental building blocks of probability theory and statistical inference

- A random variable is formed by assigning a numerical value to each outcome in the sample space of a particular experiment

- A random variable can be thought of as being generated from a function that maps each outcome in a particular sample space onto the real number line $\mathcal{R}$

FIGURE 2.1

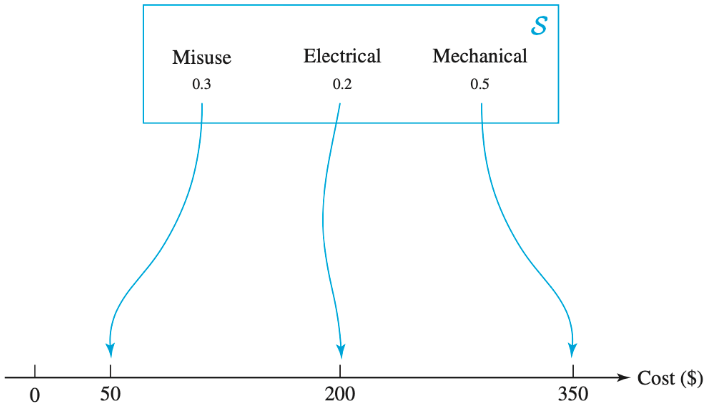
A random variable is formed by assigning a numerical value to each outcome in a sample space



- A random variable is obtained by assigning a *numerical value* to each outcome of a particular experiment.

Example 1 (Machine breakdowns)

outcome	probability	repair cost (in USD)
electrical	0.2	200
mechanical	0.5	350
misuse	0.3	50



- *Repair cost* is a random variable as it is a numeric value and it corresponds to each element of the sample space

Example 4 (power plant operation)

- $X =$ no. of power plants working and its possible values are 0, 1, 2, and 3

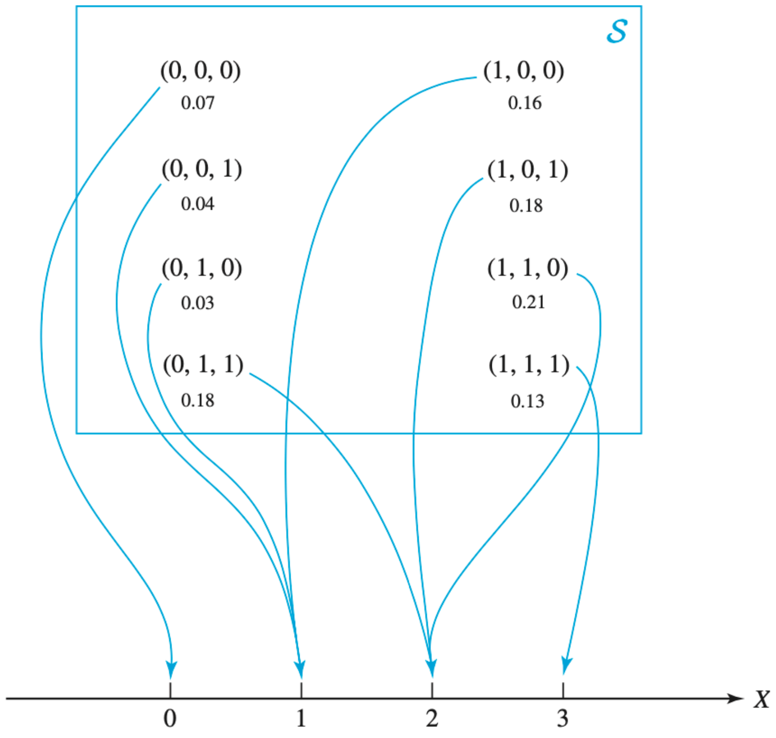
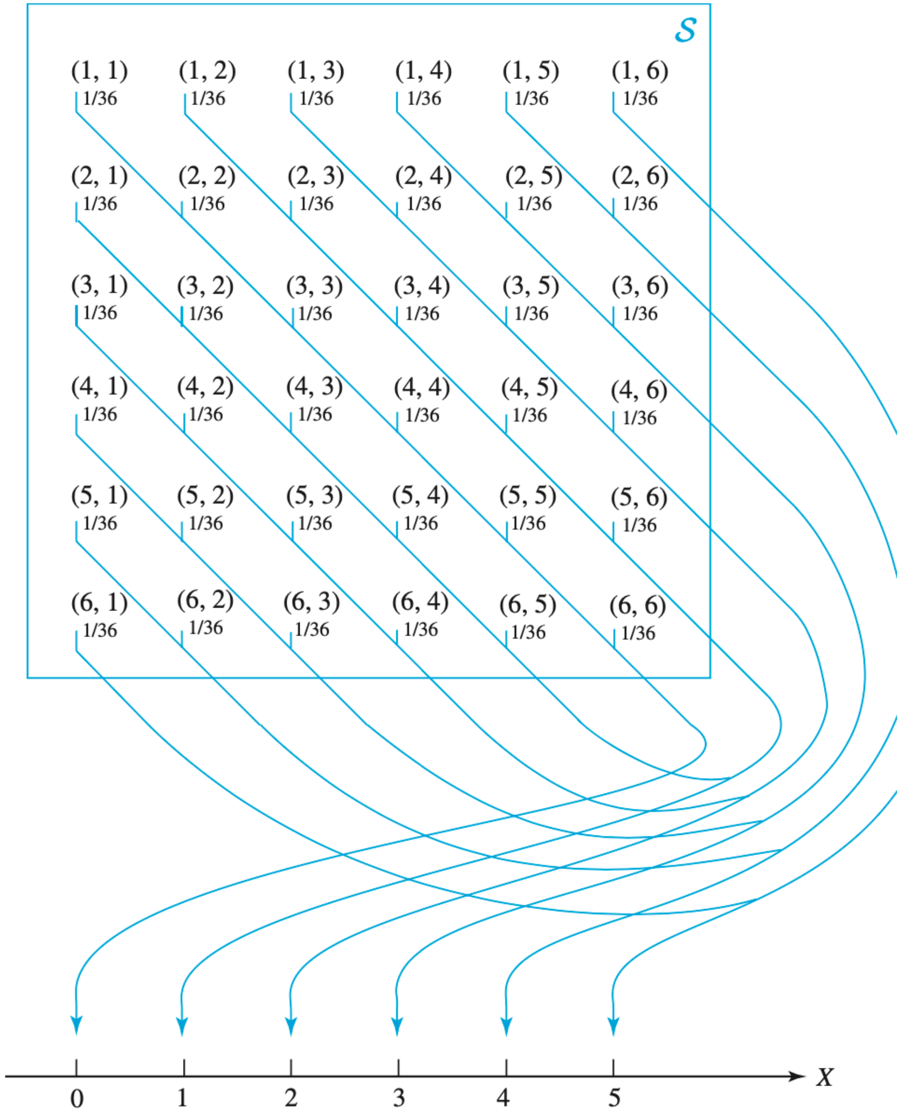


FIGURE 2.4

X = positive difference between the scores of two dice



Types of random variables

- A random variable is either discrete or continuous
  - *Number of power plants generating electricity* $\longrightarrow$ a discrete random variable
  - *Lifetime of a laptop battery* $\longrightarrow$ a continuous random variable

Types of random variables

- Random variables are generally denoted by uppercase letters, such as X, Y, Z, etc.

- Lowercase letters (e.g. x, y, etc.) are used to denote values taken by the random variable

- E.g. X denote the number of power plants generating electricity and its values are denoted by lowercase letters $x = 0$, $x = 1$, etc.

Homework 2B

2.2.1 Consider a random variable measuring the following quantities. In each case state with reasons whether you think it more appropriate to define the random variable as discrete or as continuous.

- A person’s height
- A student’s course grade (CGPA, point-grade, grade out of 100)
- The thickness of a metal plate
- A person’s age

Discrete random variables

Probability mass function

- Probability mass function (pmf) is defined for a discrete random variable and it assigns probability values to all possible values of the random variable

- The probability mass function of a discrete random variable X is a set of probability values p_i assigned to values of random variable x

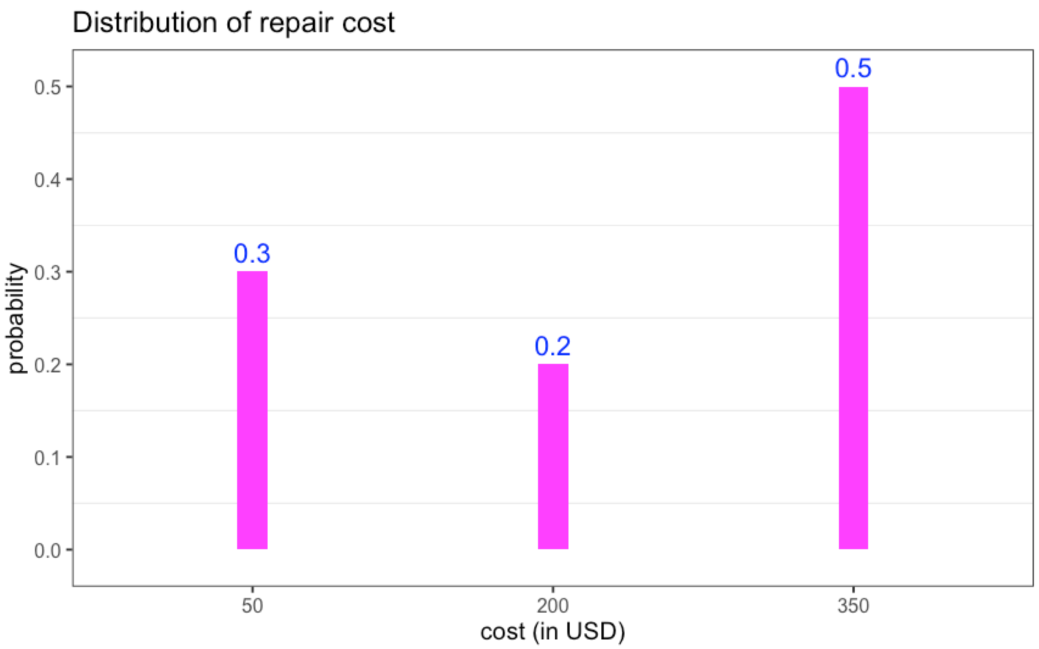
$$P(X = x) = p_x$$

- Probability values must satisfy

$$(i) \;\; 0 \leq p_x \leq 1 \;\; \text{and} \;\; (ii) \;\; \sum_x p_x = 1$$

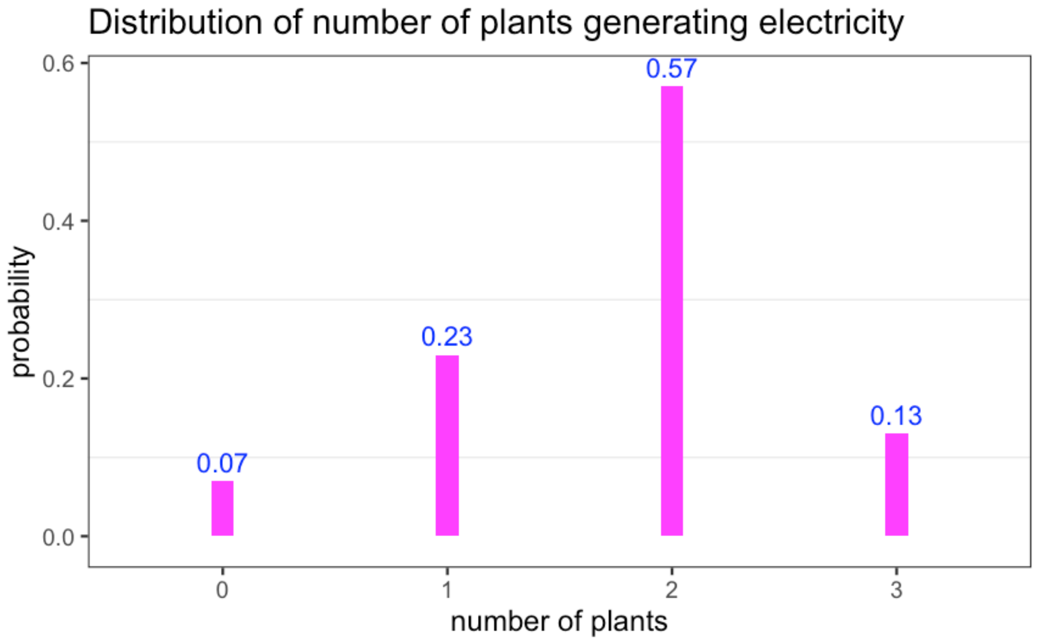
Probability distribution of repair costs of machine breakdown

outcome	repair cost, x_i	probability, p_i
electrical	200	0.2
mechanical	350	0.5
misuse	50	0.3



Probability distribution of the number of plants generating electricity

x_i	0	1	2	3
p_i	0.07	0.23	0.57	0.13



Cumulative distribution function

- The cumulative distribution function of a random variable X is defined as

$$
\begin{aligned}
F(a) &= P(X \leq a) \\
&= \sum_{y:\, y \leq a} P(X = y)
\end{aligned}
$$

- Like the probability mass function, the cumulative distribution function summarizes the probabilistic properties of a random variable.

- Knowledge of either the probability mass function or the cumulative distribution function allows the other functions to be calculated.

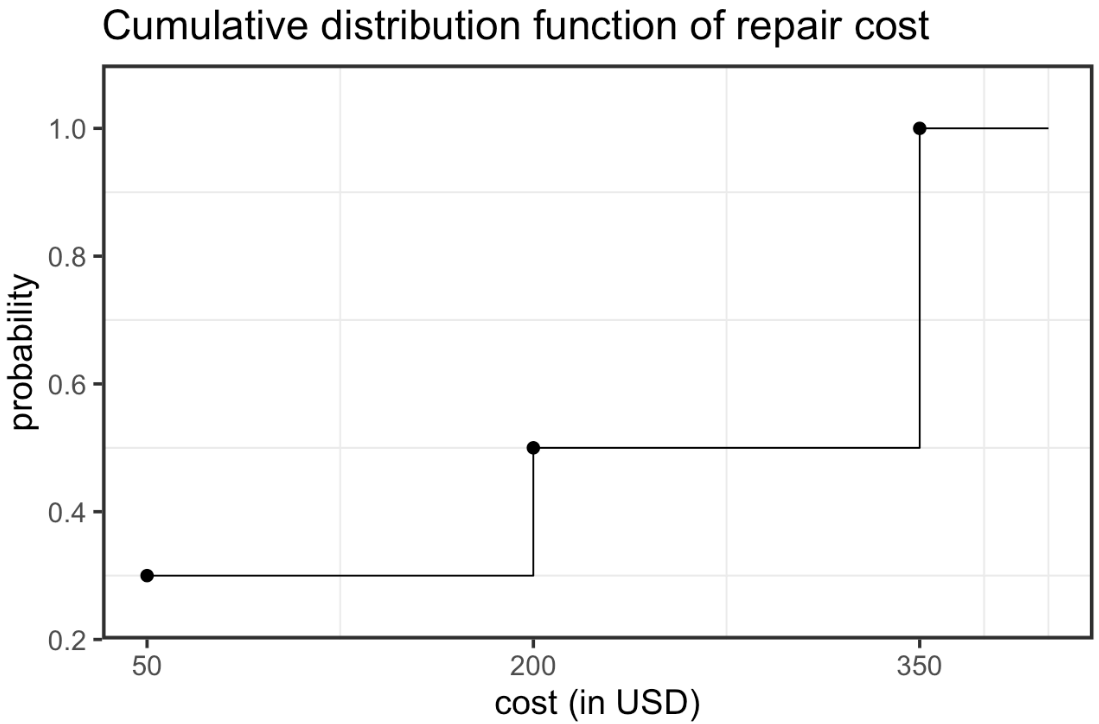
Probability distribution of repair costs of machine breakdown

outcome	repair cost, x_i	probability, p_i
electrical	200	0.2
mechanical	350	0.5
misuse	50	0.3

$$
\begin{aligned}
-\infty < x < 50 \;&\Rightarrow\; F(x) = P(cost \leqslant x) = 0 \\
50 \leq x < 200 \;&\Rightarrow\; F(x) = P(cost \leqslant x) = .30 \\
200 \leq x < 350 \;&\Rightarrow\; F(x) = P(cost \leqslant x) = .50 \\
350 \leq x < \infty \;&\Rightarrow\; F(x) = P(cost \leqslant x) = 1.0
\end{aligned}
$$

Cumulative distribution function

- For a discrete random variable, $F(x)$ is an increasing step function with steps at the values taken by the random variable



Cumulative distribution function

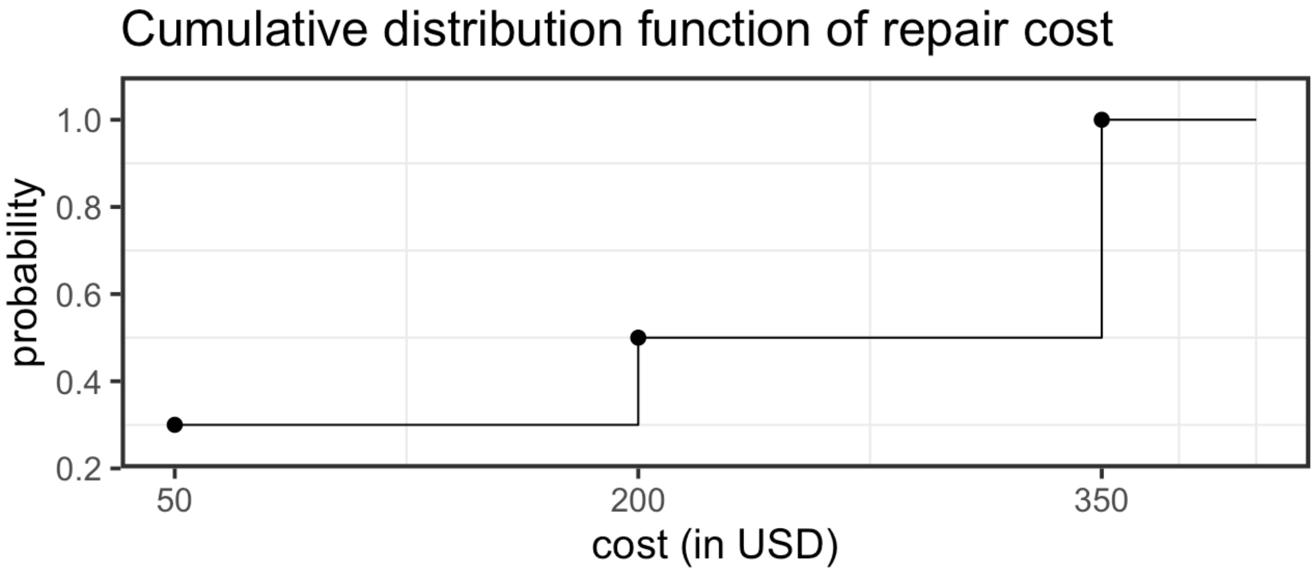
- Probability mass function (pmf) can be obtained from cumulative distribution function (cdf)

$$P(X = x) = F(x) - F(x^-)$$

- $F(x^-)$ is the limiting value from below of the cumulative distribution function

- If there is no step in the cumulative distribution function at a point x

$$F(x) = F(x^-) \; and \; P(X = x) = 0$$



$$P(X = 40) = F(40) - F(40^-) = 0$$
$$P(X = 50) = F(50) - F(50^-) = .30 - 0 = .30$$
$$P(X = 65) = F(65) - F(65^-) = .30 - .30 = 0$$

Homework 2A

2.1.1 An office has four copying machines, and the random variable X measures how many of them are in use at a particular moment in time.

- Find $P(X = 4)$ for the given probabilities

$$P(X=0)=0.08,\ P(X=1)=0.11,$$
$$P(X=2)=0.27, P(X=3)=0.33$$

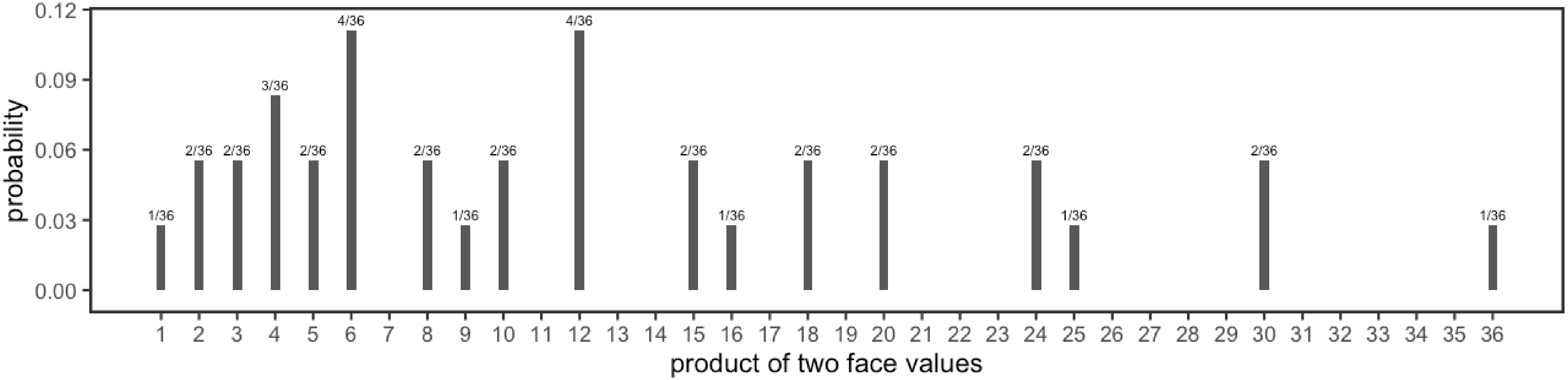
- Draw a line graph of the probability mass function

- Construct and plot the cumulative distribution function

Homework 2A

2.1.3 Suppose that two fair dice are rolled and that the two numbers recorded are multiplied to obtain a final score.

- Construct and plot the probability mass function and the cumulative distribution function of the final score



Continuous Random Variables

Example 14 (Metal Cylinder Production)

- A company manufactures metal cylinders, which are designed to have a diameter of 50 mm, but the company discovers that the cylinders can have a diameter anywhere between 49.5 and 50.5 mm

- Suppose that the random variable X is the diameter of a randomly chosen cylinder manufactured by the company

- Since this random variable can take any value between 49.5 and 50.5, it is a continuous random variable

Example 15 (Battery failure times)

- Suppose that a random variable X is the time to failure of a newly charged battery

- Failure can be defined to be the moment at which the battery can no longer supply enough energy to operate a certain appliance

- This random variable is continuous since it can hypothetically take any positive value.

- Its state space can be thought of as the interval 0 to ∞

Example 17 (Milk contents)

- A machine-filled milk container is labeled as containing 2 liters.
- However, the actual amount of milk deposited into the container by the filling machine varies between 1.95 and 2.20 liters.
- If the random variable X measures the amount of milk in a randomly chosen container, it is a continuous random variable taking any value in the interval $[1.95, 2.20]$

Probability density function

- The main distinction between discrete and continuous random variables lies in how their probabilistic properties are defined.

- The probabilistic properties of discrete random variables are defined through a *probability mass function*

- The probabilistic properties of a continuous random variable are defined through a *probability density function*

Probability density function

- The probabilistic properties of a continuous random variable are defined through a function $f(x)$

- A function $f(x)$ is said to be a density function if it satisfies the following two properties:

    - $f(x) > 0$ for all values of x

    - $\int_{-\infty}^{\infty} f(x)dx = 1, \quad -\infty < x < \infty$

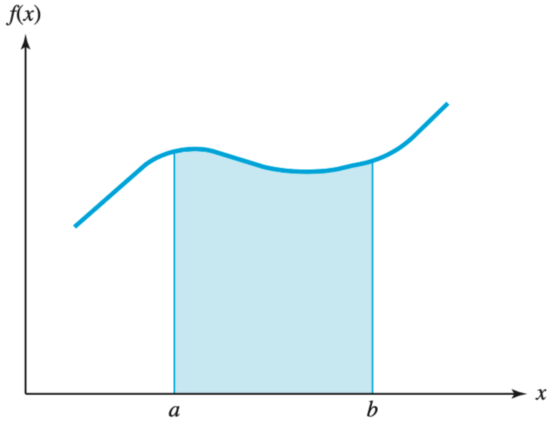
Probability density function

- The probability that a continuous random variable lies between two values a and b is obtained by integrating the probability density function between these two values

$$P(a \le X \le b) = P(a < X < b)$$
$$= \int_a^b f(x)\,dx$$

FIGURE 2.20

$P(a \le x \le b)$ is the area under the probability density function $f(x)$ between the points a and b



Probability density function

- The probability that a continuous random variable takes a specific value is zero, i.e $P(X = a) = 0$

$$P(X = a) = \int_a^a f(x)\,dx = 0$$

Example 15 (Battery failure times)

- Suppose battery failure times X (measured in hours) has a probability density function

$$f(x) = \begin{cases} \frac{2}{(1+x)^3} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Example 15 (Battery failure times)

- It can be shown that $f(x) > 0$ for all $x \geq 0$ and

$$\int_0^\infty f(x)\,dx = \int_0^\infty \frac{2}{(1+x)^3}\,dx = \frac{-1}{(1+x)^2}\Bigg|_0^\infty = 1$$

  - $f(x)$ is a valid probability density function

- What is the probability that the battery fails within first five hours?

Example 15 (Battery failure times)

- What is the probability that the battery fails within first five hours?

$$P(X \leq 5) = \int_0^5 \frac{2}{(1+x)^3}\,dx = \frac{-1}{(1+x)^2}\bigg|_0^5 = \frac{-1}{6^2} + 1 = \frac{35}{36}$$

- What is the probability that a battery lasts longer than five hours?

Example 17 (Milk container contents)

- Suppose that the probability density function of the amount of milk deposited in a milk container is

$$f(x) = 40.976 - 16x - 30e^{-x}, \quad 1.95 \leq x \leq 2.20$$

Example 17 (Milk container contents)

- It can be shown that
  - $f(1.95) = 40.976 - (16)(1.95) - (30)(e^{-1.95}) = 5.508 > 0$
  - $f(2.20) = 40.976 - (16)(2.20) - (30)(e^{-2.20}) = 2.452 > 0$ and

$$\int_{1.95}^{2.20} (40.976 - 16x - 30e^{-x})\, dx = \left(40.976x - 16(x^2/2) + 30e^{-x}\right)\Big|_{1.95}^{2.20} = 1$$

- The probability that the actual amount of milk is less than 2.0 liter

$$\int_{1.95}^{2.0} f(x)\,dx = \int_{1.95}^{2.0} (40.976 - 16x - 30e^{-x})\,dx$$

$$= \left(40.976x - 16(x^2/2) + 30e^{-x}\right)\Big|_{1.95}^{2.0}$$

$$= \left[(40.976)(2.0) - 16(2.0^2/2) + 30e^{-2.0}\right]$$

$$- \left[(40.976)(1.95) - 16(1.95^2/2) + 30e^{-1.95}\right]$$

$$= 54.012 - 53.751 = 0.261$$

  - About 26% of the milk containers are underweight.

Cumulative Distribution Function

- The cumulative distribution function of a continuous random variable X is defined in exactly the same way as for a discrete random variable

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)\, dx$$

  - $f(x) \rightarrow$ probability density function of X

Cumulative Distribution Function

- For a continuous random variable, the probability density function can also be obtained from cumulative distribution function

$$f(x) = \frac{dF(x)}{dx}$$

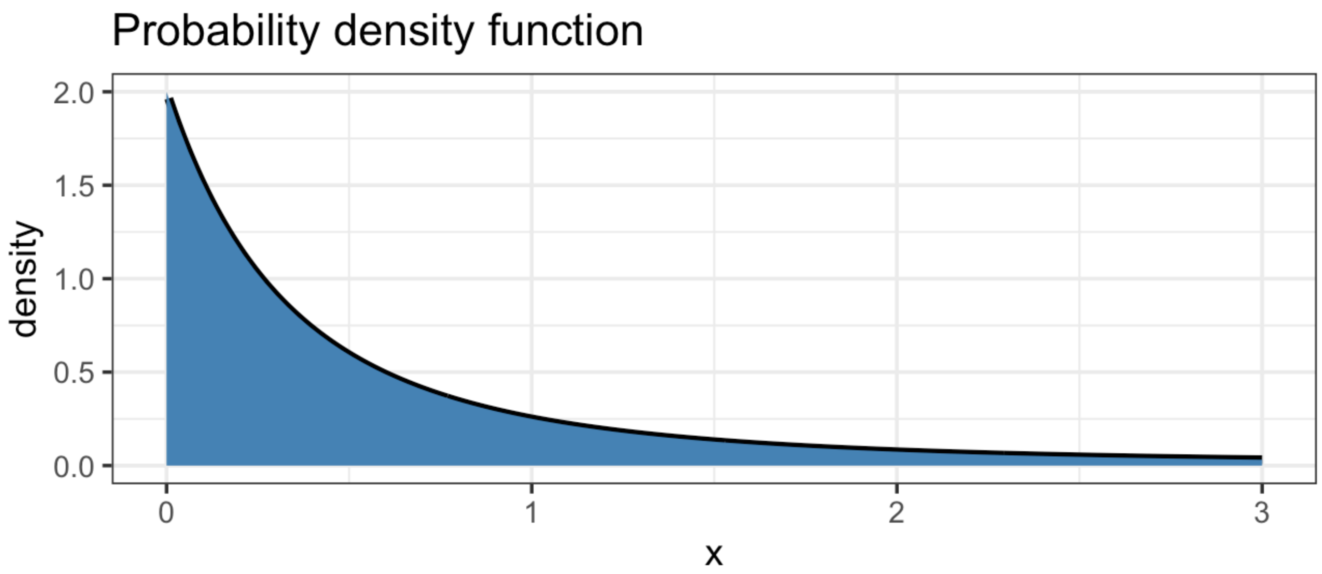
- Probability that a continuous random variable X lies between a and b can be obtained using cumulative distribution function

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

Example 15 (Battery failure times)

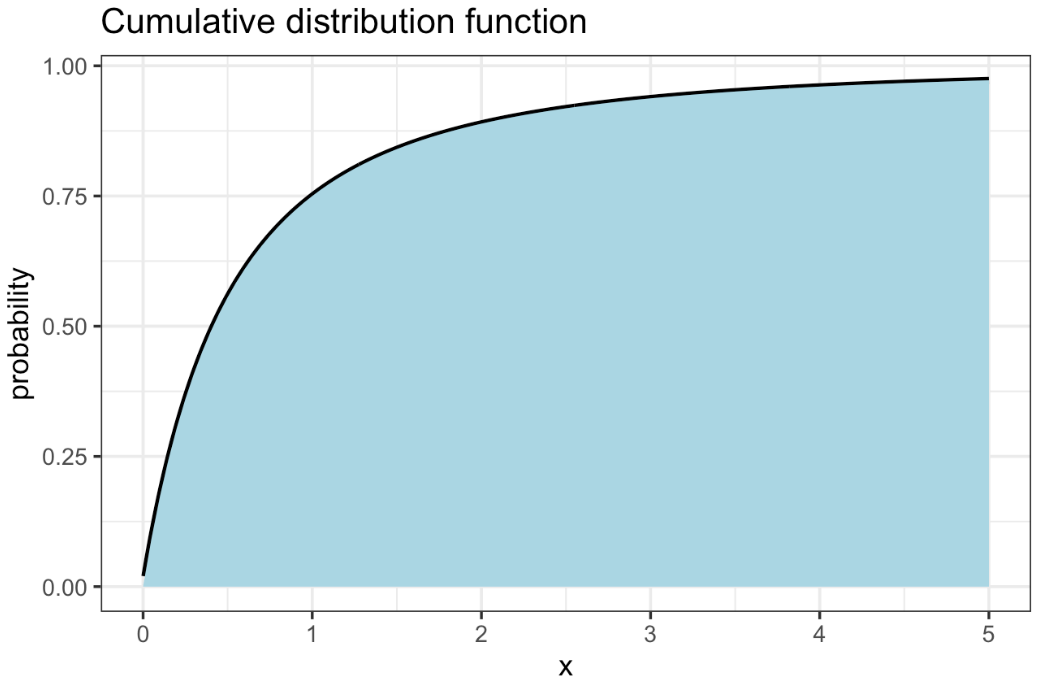
- Probability density function

$$f(x) = \frac{2}{(1+x)^3} \qquad x \geq 0$$



- Cumulative distribution function

$$
\begin{aligned}
F(x) &= \int_0^x f(y)\,dy \\
&= \int_0^x \frac{2}{(1+y)^3} dy \\
&= 1 - \frac{1}{(1+x)^2}
\end{aligned}
$$



Example 15 (Battery failure timesY)

- Probability density function

$$f(x) = \frac{2}{(1+x)^3} \qquad x \geq 0$$

- Cumulative distribution function

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

Example 15 (Battery failure timesY)

- Find the probability that a battery lasts between one and two hours

$$P(1 \leq X \leq 2) = \int_1^2 f(x)\, dx$$

$$= \int_1^2 \frac{2}{(1+x)^3}\, dx = \frac{-1}{(1+x)^2}\bigg|_1^2 = \frac{1}{4} - \frac{1}{25} = 0.21$$

$$P(1 \leq X \leq 2) = F(2) - F(1)$$

$$= \left[1 - \frac{1}{(1+2)^2}\right] - \left[1 - \frac{1}{(1+1)^2}\right] = \frac{1}{4} - \frac{1}{25} = 0.21$$

Homework 2B

2.2.2 A random variable X takes values between 4 and 6 with a probability density function

$$f(x) = \frac{1}{x \ln(1.5)}, \quad 4 \leq x \leq 6.$$

- Check that the total area under the probability density function is equal to 1.
- What is $P(4.5 \leq X \leq 5.5)$?
- Construct the cumulative distribution function.

Homework 2B

2.2.4 A random variable X takes values between 0 and 4 with a cumulative distribution function

$$F(x) = \frac{x^2}{16} \quad for\, 0 \le x \le 4$$

- What is $P(X \le 2)$?
- What is $P(1 \le X \le 3)$?
- What is the probability density function.

Homework 2B

2.2.6 A car panel is spray-painted by a machine, and the technicians are particularly interested in the thickness of the resulting paint layer.

Suppose that the random variable X measures the thickness of the paint in millimeters at a randomly chosen point on a randomly chosen car panel, and that X takes values between 0.125 and 0.5 mm with a probability density function of

$$f(x) = A[0.5 - (x - 0.5)^2], \quad 0.125 \leq x \leq 0.5$$

- Find the value of A and construct the cumulative distribution function.
- What is the probability that the paint thickness at a particular point is less than 0.2 mm?

The Expectation of a Random Variable

The Expectation of a Random Variable

- The probability mass function or the probability density function provides complete information about the probabilistic properties of a random variable

- Summary measures of a random variable would be useful, two commonly used summary measures are expectation and variance of a random variable

- Expectation represents the "average" value of the random variable, where as variance measures average distance of a variable from its mean value

- Expectation of a random variable X is denoted by $E(X)$

Expectation of a discrete random variable

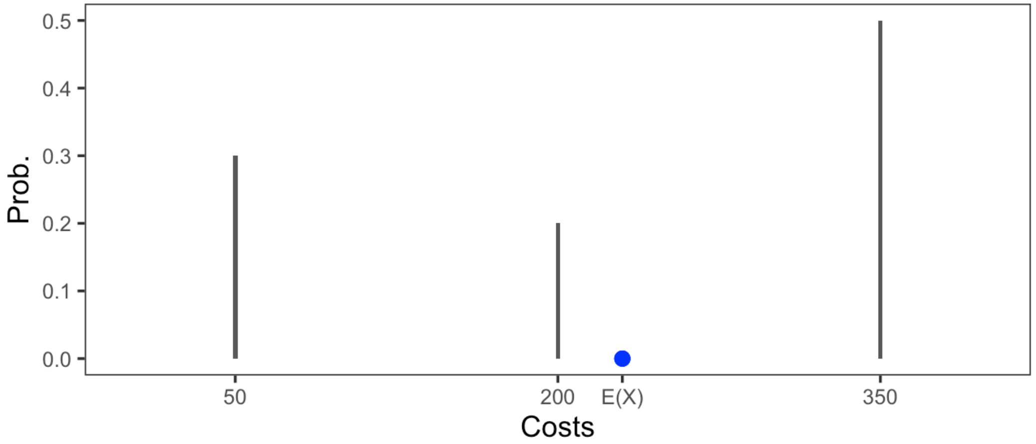
- The expected value or expectation of a discrete random variable X with a probability mass function $P(X = x_i) = p_i$ is defined as

$$\mu = E(X) = \sum_i P(X = x_i)\, x_i = \sum_i p_i x_i$$

- $E(X)$ provides a summary measure of the average value taken by the random variable and is also known as the mean of the random variable

Example 1 (Machine breakdowns)

outcome	probability	repair cost (in USD)
electrical	0.2	200
mechanical	0.5	350
misuse	0.3	50

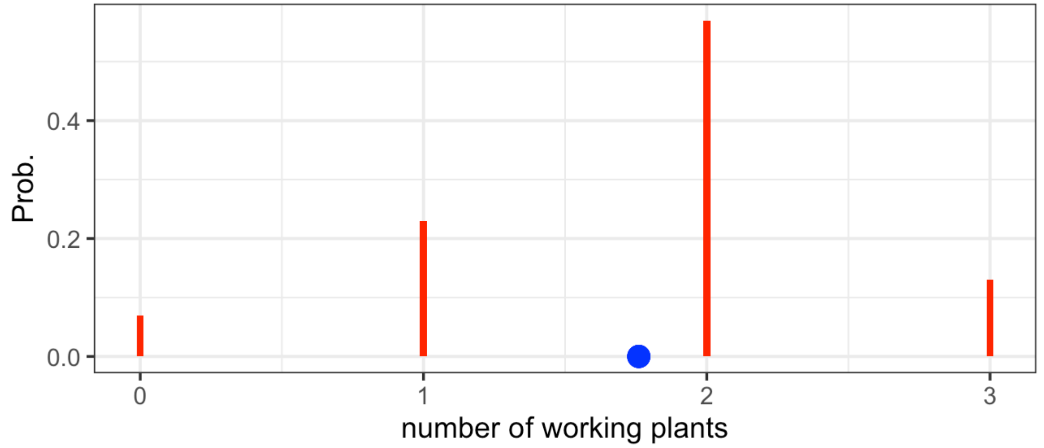


- Expected repair cost

$$E(X) = \sum_i p_i x_i = (.2)(200) + (.5)(350) + (.3)(50) = 230$$

Example 4 (power plant operation)

x_i	0	1	2	3
p_i	0.07	0.23	0.57	0.13



- Expected number of plants generating electricity

$$E(X) = \sum_i p_i x_i = (.07)(0) + (.23)(1) + (.57)(2) + (.13)(3) = 1.76$$

Expectation of Continuous Random Variables

- The expected value or expectation of a continuous random variable X with a probability density function $f(x)$ is defined as

$$E(X) = \int_x x f(x)\, dx$$

Example 15 (Battery Failure Times)

$$\begin{aligned} E(X) &= \int_0^\infty x \frac{2}{(1+x)^3}\, dx \\ &= \int_0^\infty \left[\frac{2}{(1+x)^2} - \frac{2}{(1+x)^3} \right] dx \\ &= \left[\frac{-2}{(1+x)} + \frac{1}{(1+x)^2} \right]_0^\infty \\ &= 1 \end{aligned}$$

$$\begin{aligned} E(X) &= \int_0^\infty x \frac{2}{(1+x)^3}\, dx \\ &= \int_1^\infty \left[\frac{2(y-1)}{y^3} \right] dy \\ &= \left[\frac{2}{y^2} - \frac{2}{y^3} \right]_1^\infty \\ &= 1 \end{aligned}$$

Properties of expectation

- For two constants a and b
  - $E(a) = a$
  - $E(aX) = aE(X)$
  - $E(a + bX) = a + bE(X)$

Homework 2C

2.3.1 An office has four copying machines, and the random variable X measures how many of them are in use at a particular moment in time. Suppose that

$$\begin{aligned} P(X=0) &= 0.08 \\ P(X=1) &= 0.11 \\ P(X=2) &= 0.27 \\ P(X=3) &= 0.33. \end{aligned}$$

- What is the expected number of copying machines in use at a particular moment in time?

Homework 2C

2.3.11 Consider again the random variable with a cumulative distribution function of

$$F(x) = x^2/16; \quad 0 \le x \le 4$$

- What is the expected value of this random variable?

Homework 2C

2.3.12 Consider again the car panel painting machine discussed in Problem 2.2.6, where

$$f(x) = A[0.5 - (x - 0.5)^2], \quad 0.125 \leq x \leq 0.5$$

- What is the expected paint thickness?

The Variance of a Random Variable

The Variance of a Random Variable

- Variance of a random variable measures *variability* or *spread* in the values taken by the random variable

- The variance of a random variable X is denoted by σ^2, and is defined as

$$
\begin{aligned}
\sigma^2 = Var(X) &= E\big[X - E(X)\big]^2 \\
&= E\big[X^2 - 2XE(X) + E(X)^2\big] \\
&= E(X^2) - 2E(X)^2 + E(X^2) \\
&= E(X^2) - E(X)^2 \\
&= E(X^2) - \mu^2
\end{aligned}
$$

Standard deviation

- The positive square root of the variance is known as the *standard deviation*, which is denoted by σ

$$\sigma = +\sqrt{\sigma^2} = +\sqrt{Var(X)}$$

Example 1 (Machine breakdowns)

x, repair cost	p=P(X=x)	x2	px	px2
50	0.3	2,500	15	750
200	0.2	40,000	40	8,000
350	0.5	122,500	175	61,250
Total	1.0	165,000	230	70,000

- **Expected value**

$$E(X) = \sum x_i p_i = 230$$

- **Variance**

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= \sum x_i^2 p_i - E(X)^2 \\ &= 70,000 - 230^2 = 17,100 \end{aligned}$$

- **Standard deviation**

$$\sigma = \sqrt{Var(X)} = \sqrt{17,100} = 130.767$$

Example 14 (Metal cylinder diameter)

- The probability density function of metal cylinder diameter

$$f(x) = 1.5 - 6(x - 50.0)^2, \quad 49.5 \le x \le 50.5$$

- Expected value

$$E(X) = \int_{49.5}^{50.5} x\left[1.5 - 6(x - 50.0)^2\right] dx = \int_{-.5}^{.5} (y + 50)(1.5 - 6y^2) dy$$

$$= \int_{-.5}^{.5} (1.5y - 6y^3 + 75 - 300y^2) dy$$

$$= \left[(1.5y^2/2) - (6y^4/4) + 75y - (300y^3/3)\right]_{-.5}^{.5} = 50$$

Example 14 (Metal cylinder diameter)

- The probability density function of metal cylinder diameter

$$f(x) = 1.5 - 6(x - 50.0)^2, \quad 49.5 \leq x \leq 50.5$$

- Variance

$$Var(X) = E(X^2) - E(X)^2] = 2500.05 - 50^2 = 0.05$$

$$
\begin{aligned}
E(X^2) &= \int_{-.5}^{.5} (y^2 + 100y + 2500)(1.5 - 6y^2)dy \\
&= \int_{-.5}^{.5} \left[1.5y^2 + 150y + 3750 - 6y^4 - 600y^3 - 15000y^2\right] dy \\
&= \left[.5y^3 + 75y^2 + 3750y - (6y^5/5) - (600y^4/4) - 5000y^3\right]_{-.05}^{0.5} \\
&= 1259.4 - (-1240.65) = 2500.05
\end{aligned}
$$

Properties of variance

- For two constants a and b
  - $V(a) = 0$
  - $V(aX) = a^2 V(X)$
  - $V(a + bX) = b^2 V(X)$

Quantiles

Medians of Random Variables

- The median is another summary measure of the distribution of a random variable that provides information about the "middle" value of the random variable

- The median of a continuous random variable X is x_m (say), which satisfies

$$F(x_m) = 0.5$$

Medians of Random Variables

- The cumulative distribution function of battery lifetime is

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

- The median battery lifetime

$$\begin{aligned} F(x) = 0.5 \;&\Rightarrow\; 1 - \frac{1}{(1+x)^2} = 0.5 \\ &\Rightarrow\; (1+x)^2 = 2 \\ &\Rightarrow\; x = \sqrt{2} - 1 = 0.41 \end{aligned}$$

Symmetric distribution

- A continuous random variable X is said to be symmetric about a point μ if both the median and the expectation of the random variable are equal to μ

Quantiles

- Quantiles of random variables are additional summary measures that can provide information about the spread or variability of the distribution of the random variable

- The pth quantile $(0 < p < 1)$ of a random variable X with a cumulative distribution function $F(x)$ is defined to be the value x_p for which

$$F(x_p) = P(X \leq x_p) = p$$

  - It is also referred to as the $p \times 100$th percentile of the random variable
  - There is a probability of p that the random variable takes a value less than the pth quantile x_p

Quartiles

- The $.25$ quantile $(x_{.25})$ is the first quartile Q_1
- The $.75$ quantile $(x_{.75})$ is the third quartile Q_3
- The median $(x_{.5})$ is the second quartile Q_2

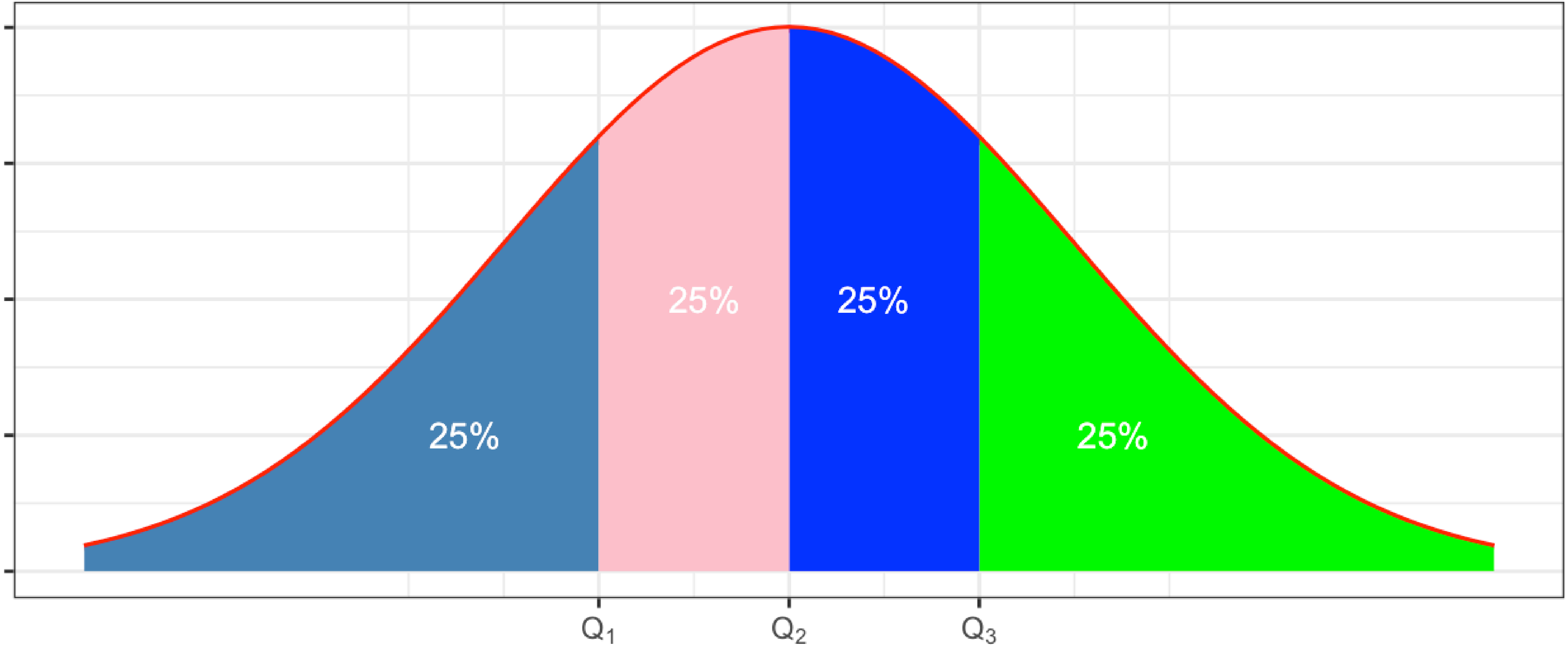
Inter-quartile range

- The **inter-quartile range** (IQR) is defined as the difference between the third and first quartile

$$IQR = x_{.75} - x_{.25} = Q_3 - Q_1$$

  - IQR is a measure of spread

Quartiles



Example 15 (Battery Failure Times)

- The cumulative distribution function of battery lifetime

$$F(x) = 1 - \frac{1}{(1+x)^2}, \quad x \geq 0$$

Example 15 (Battery Failure Times)

- Third and first quartiles

$$F(x_{.75}) = .75 \;\Rightarrow\; 1 - \frac{1}{(1 + x_{.75})^2} = .75$$

$$\Rightarrow\; \frac{1}{(1 + x_{.75})^2} = 0.25 \;\Rightarrow\; x_{.75} = 1\,\text{hr}$$

$$F(x_{.25}) = .25 \;\Rightarrow\; 1 - \frac{1}{(1 + x_{.25})^2} = .25$$

$$\Rightarrow\; \frac{1}{(1 + x_{.25})^2} = 0.75 \;\Rightarrow\; x_{.25} = .154\,\text{hr}$$

Example 15 (Battery Failure Times)

- Interquartile range (IQR)

$$IQR = x_{.75} - x_{.25} = 1 - 0.154 = 0.845$$

  - Half of the batteries will fail between 0.154 hour and 1 hour (i.e. between about 9 minutes to 60 minutes)

Homework 2D

2.4.1 Suppose that the random variable X takes the values $-2, 1, 4$, and 6 with probability values 1/3, 1/6, 1/3, and 1/6, respectively.

- Find the expectation and variance of X.

Homework 2D

2.4.5 Consider again the random variable described in Problems 2.2.2 and 2.3.10 with a probability density function of

$$f(x) = \frac{1}{x \ln(1.5)}, \quad 4 \leq x \leq 6.$$

- What is the variance of this random variable?
- What is the standard deviation of this random variable?
- Find the upper and lower quartiles of this random variable.
- What is the interquartile range?

Homework 2D

2.4.6 Find variance, standard deviation, and interquartile range of the variable X, where

$$F(x) = \frac{x^2}{6} \quad for \ \ 0 \leq x \leq 4$$

Jointly Distributed Random Variables

Jointly Distributed Random Variables

- For two discrete random variables X and Y, the joint probability mass function is defined by

$$P(X = x_i, Y = y_j) = p_{ij}$$

  - $0 < p_{ij} < 1$ for all i and j
  - $\sum_i \sum_j p_{ij} = 1$

Jointly Distributed Random Variables

- For two continuous random variables X and Y, the joint probability density function is defined by $f(x, y)$ that satisfies
  - $f(x, y) > 0$ for all x and y
  - $\int_x \int_y f(x, y)\, dy\, dx = 1$

Jointly Distributed Random Variables

- The probability that X lies between a and b, and Y lies between c and d is defined as

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)\, dy\, dx$$

- The joint cumulative distribution function is defined as

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dx\, dy$$

Example 19 (Air Conditioner Maintenance)

- A company that services air conditioner units is interested in how long a technician takes on a visit to a particular location, which depends on the number of air conditioner units at the location that need to be serviced
    - $X \in \{1, 2, 3, 4\} \to$ the service time in hours taken at a particular location
    - $Y \in \{1, 2, 3\} \to$ the number of air conditioner units at the location
- The random variables X and Y are jointly distributed

Example 19 (Air Conditioner Maintenance)

		X = service time (hrs)			
		1	2	3	4
Y = number of air conditioner units	1	0.12	0.08	0.07	0.05
	2	0.08	0.15	0.21	0.13
	3	0.01	0.01	0.02	0.07

- $P(X = 1, Y = 1) = p_{11} = 0.12$
- It is a valid joint mass function because all probabilities are non-negative and sum of all probabilities is 1, i.e. $\sum_i \sum_j p_{ij} = 1$

Example 19 (Air Conditioner Maintenance)

		X = service time (hrs)			
		1	2	3	4
Y = number of air conditioner units	1	0.12	0.08	0.07	0.05
	2	0.08	0.15	0.21	0.13
	3	0.01	0.01	0.02	0.07

$$\begin{aligned} P(X \le 1, Y \le 2) &= p_{11} + p_{12} \\ &= .12 + .08 = .20 \end{aligned}$$

Marginal probability distributions

- The marginal distribution is the individual probability distribution of the random variable, which can be obtained from a joint distribution

- For a discrete random variable

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$$

$$\text{E.g. } \; P(X = 1) = \sum_{j=1}^{3} P(X = 1, Y = y_j)$$

$$= p_{11} + p_{12} + p_{13} + p_{14} = 0.21$$

X = service time (hrs)

Y = number of air conditioner units	1	2	3	4	Marginal distribution of Y
1	0.12	0.08	0.07	0.05	0.32
2	0.08	0.15	0.21	0.13	0.57
3	0.01	0.01	0.02	0.07	0.11
Marginal distribution of X	0.21	0.24	0.30	0.25	

Marginal probability distributions

- Marginal distribution of a continuous variable

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

Marginal probability distributions

- Marginal distribution of a discrete variable

$$P(X = x) = \sum_y P(X = x, Y = y)$$

$$P(Y = y) = \sum_x P(X = x, Y = y)$$

Conditional probability distribution

- The conditional distribution of a random variable X conditional on a random variable Y taking a particular value is defined as

$$P_{x\,|\,y} = P(X = x\,|\,Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} \quad \text{(discrete)}$$

$$f_{2\,|\,2}(x|y) = \frac{f(x, y)}{f(y)} \quad \text{(continuous)}$$

Conditional probability distribution

- For example

$$P(Y = 1 \mid X = 2) = \frac{P(X = 2, Y = 1)}{P(X = 2)} = \frac{.08}{.24} = .33$$

Independent random variables

- Two random variables X and Y are defined to be independent if their joint probability mass function or joint probability density function is the product of their two marginal distributions

$$P(X = x_i, Y = Y_j) = P(X = x_i)\, P(Y = y_j)\ [discrete]$$

$$f(x, y) = f(x)\, f(y)\ [continuous]$$

Independent random variables

- Are X and Y independent for $f(x,y) = 6xy^2,\ (0 \le x \le 1,\ \ 0 \le y \le 1)$?

$$f(x) = \int_0^1 f(x,y)dy = \int_0^1 6xy^2\,dy = \left[6xy^3/3\right]_0^1 = 2x$$

$$f(y) = \int_0^1 f(x,y)dx = \int_0^1 6xy^2\,dx = \left[6x^2y^2/2\right]_0^1 = 3y^2$$

  - Since $f(x)\,f(y) = (2x)(3y^2) = f(x,y)$
  - So X and Y are independent

Homework 2E

2.5.3 Suppose that two continuous random variables X and Y have a joint probability density function

$$f(x, y) = A(x-3)y, \quad -2 \le x \le 3, \; 4 \le y \le 6$$

- What is the value of A?
- What is $P(0 \le X \le 1, 4 \le Y \le 5)$?
- Construct the marginal probability density functions of X and Y.
- Are the random variables X and Y independent?

$$\int_{-2}^{3}\int_{4}^{6} A(x-3)y\,dy\,dx = 1$$

$$A\int_{-2}^{3}(x-3)\left[(y^2/2)\Big|_4^6\right]dx = 1$$

$$10A\int_{-2}^{3}(x-3)\,dx = 1$$

$$10A(x^2/2-3x)\Big|_{-2}^{3} = 1$$

$$10A\left[(9/2)-9-2-6\right] = 1$$

$$10A(-25/2) = 1$$

$$A = -(1/125)$$