University of BRISTOL

Discrete random variables

Henry W J Reeve

henry.reeve@bristol.ac.uk

Statistical Computing & Empirical Methods (EMATM0061)

MSc in Data Science, Teaching block 1, 2021.

What will we cover today?

- We will introduce the important concept of a random variable and its distribution.
- We will focus on discrete random variables and discuss the probability mass function.
- We will also consider several important examples including Bernoulli and Binomial random variables.
- We will study several important quantities: expectation, variance, covariance, correlation.
- We will generalize our understanding of independence to the random variable setting.

Random variables

We often encounter numerical quantities which are stochastic or random in nature.

Examples

The weight of penguin selected at random from a population on an island;

The number of dots on the top face of dice which has just been thrown;

The number of red balls drawn from a bag.

We model these stochastic numerical quantities as random variables.

Random variables

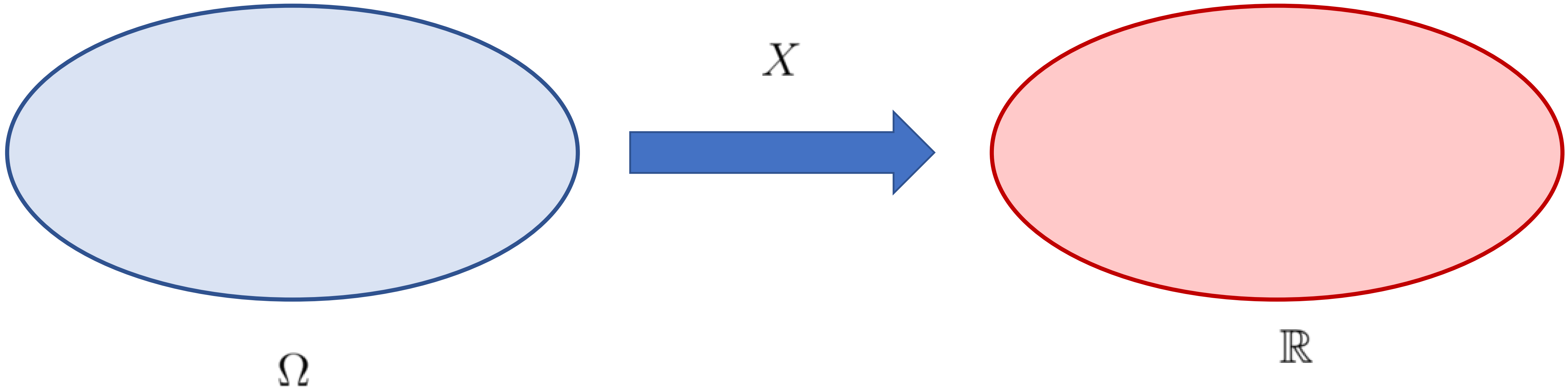
Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

A **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

Random variables

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

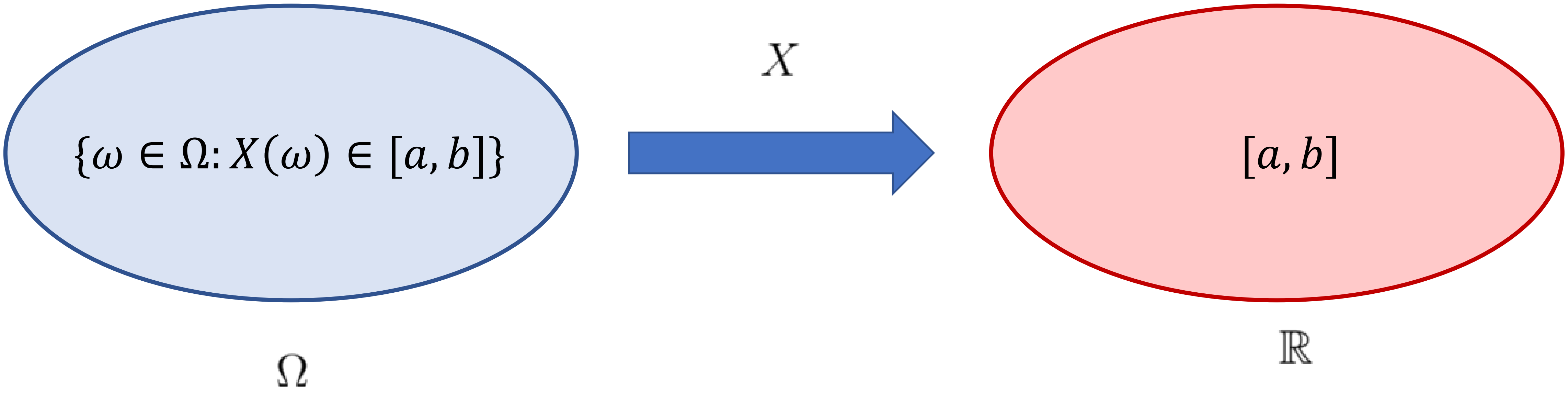
A **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.



Random variables

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

A **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.



Random variables

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

A **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

Example 1

We select 3 at random from a population and record their height and weight.

The sample space is $\Omega = \mathbb{R}^{3\times 2} = \left\{\left((x_1^h, x_1^w), (x_2^h, x_2^w), (x_3^h, x_3^w)\right)\right\}$.

Our random variable is the weight of the third penguin

$$X\left(\left((x_1^h, x_1^w), (x_2^h, x_2^w), (x_3^h, x_3^w)\right)\right) = x_3^w.$$

Random variables

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

A **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

Example 2

We roll 5 dice in a row and record each of the results.

The sample space is $\Omega = \{1, \ldots, 6\}^5 = \{(x_1, x_2, x_3, x_4, x_5)\}$.

Our random variable is the result of the final dice roll,

$$X\left((x_1, x_2, x_3, x_4, x_5)\right) = x_5.$$

Random variables

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

A **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

Example 3

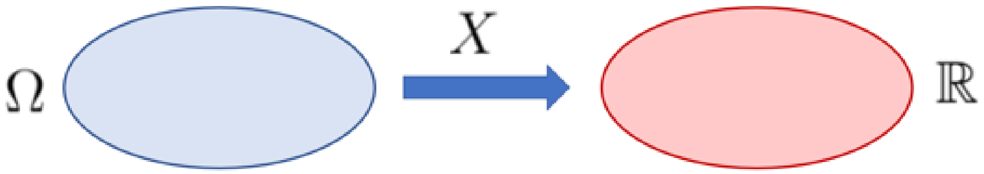
We sample 10 balls with replacement from a bag of 100 balls, 50 of which are red.

The sample space is $\Omega = \{1, \ldots, 100\}^{10} = \{(x_1, \ldots, x_{10})\}$.

Our random variable is the number of red balls sampled

$$X\left((x_1, \ldots, x_{10})\right) = \sum_{i=1}^{10} \mathbb{1}_{\{1,\ldots,50\}}(x_i).$$

Random variables

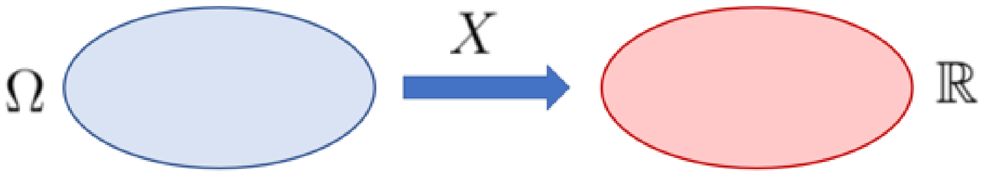


Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

A **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

We write $\{X \in S\}$ for an event $\{\omega \in \Omega \ : \ X(\omega) \in S\} \in \mathcal{E}$ with $S \subseteq \mathbb{R}$.

Random variables



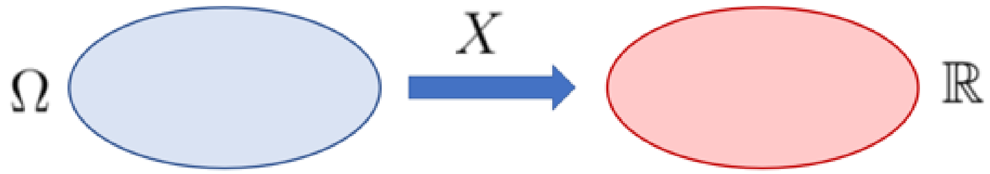
Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

A **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every $a,\ b \in \mathbb{R}$,
$\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

We write $\{X \in S\}$ for an event $\{\omega \in \Omega \ : \ X(\omega) \in S\} \in \mathcal{E}$ with $S \subseteq \mathbb{R}$.

We write $\{X = a\}$ for an event $\{\omega \in \Omega \ : \ X(\omega) = a\} \in \mathcal{E}$ with $a \in \mathbb{R}$.

Random variables



Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

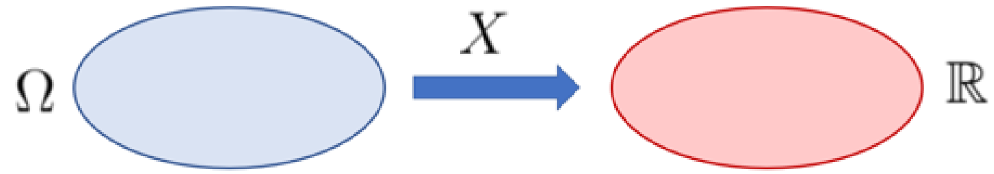
A **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \; : \; X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

We write $\{X \in S\}$ for an event $\{\omega \in \Omega \; : \; X(\omega) \in S\} \in \mathcal{E}$ with $S \subseteq \mathbb{R}$.

We write $\{X = a\}$ for an event $\{\omega \in \Omega \; : \; X(\omega) = a\} \in \mathcal{E}$ with $a \in \mathbb{R}$.

We write $\{X \leq a\}$ for an event $\{\omega \in \Omega \; : \; X(\omega) \leq a\} \in \mathcal{E}$ with $a \in \mathbb{R}$.

Random variables



Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

A **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

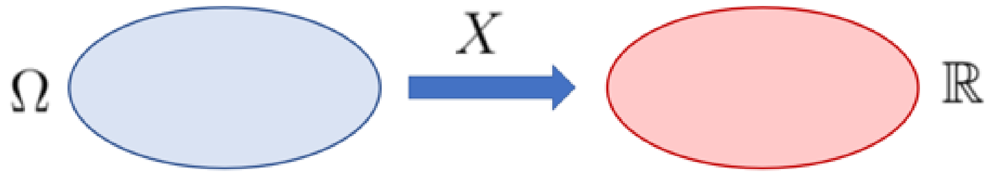
We write $\{X \in S\}$ for an event $\{\omega \in \Omega \ : \ X(\omega) \in S\} \in \mathcal{E}$ with $S \subseteq \mathbb{R}$.

We write $\{X = a\}$ for an event $\{\omega \in \Omega \ : \ X(\omega) = a\} \in \mathcal{E}$ with $a \in \mathbb{R}$.

We write $\{X \leq a\}$ for an event $\{\omega \in \Omega \ : \ X(\omega) \leq a\} \in \mathcal{E}$ with $a \in \mathbb{R}$.

In general, we write $\{F(X)\}$ for the event $\{\omega \in \Omega \ : \ F(X(\omega))\} \in \mathcal{E}$, for any predicate "$F$".

Random variables



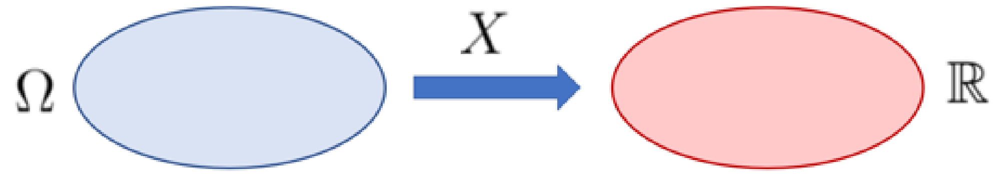
Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

A **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \; : \; X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

We write $\{X \in S\}$ for an event $\{\omega \in \Omega \; : \; X(\omega) \in S\} \in \mathcal{E}$ with $S \subseteq \mathbb{R}$.

Similarly, we write $\mathbb{P}(X \in S)$ for the probability $\mathbb{P}\left(\{\omega \in \Omega \; : \; X(\omega) \in S\right)$.

Random variables



Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

A **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

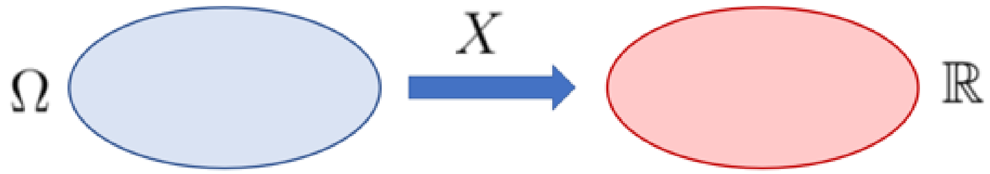
We write $\{X \in S\}$ for an event $\{\omega \in \Omega \ : \ X(\omega) \in S\} \in \mathcal{E}$ with $S \subseteq \mathbb{R}$.

Similarly, we write $\mathbb{P}(X \in S)$ for the probability $\mathbb{P}\left(\{\omega \in \Omega \ : \ X(\omega) \in S\}\right)$.

Typically, we ignore the sample space Ω, which may include extraneous information.

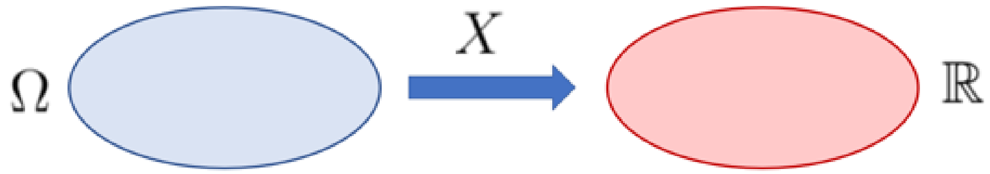
Instead, we focus on random variables and interactions between random variables.

Distributions



Given a probability space $(\Omega, \mathcal{E}, \mathbb{P})$, a **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \;:\; X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

Distributions



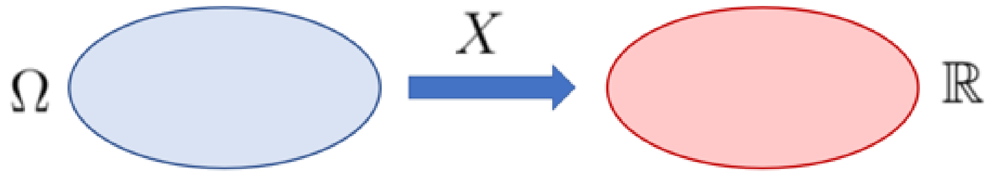
Given a probability space $(\Omega, \mathcal{E}, \mathbb{P})$, a **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

The **distribution** of a random variable X is the map

$$S \mapsto P_X(S) = \mathbb{P}(X \in S) = \mathbb{P}\left(\{\omega \in \Omega \ : \ X(\omega) \in S\}\right),$$

for all subsets $S \subseteq \mathbb{R}$ such that $\{X \in S\} = \{\omega \in \Omega \ : \ X(\omega) \in S\} \in \mathcal{E}$.

Distributions



Given a probability space $(\Omega, \mathcal{E}, \mathbb{P})$, a **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

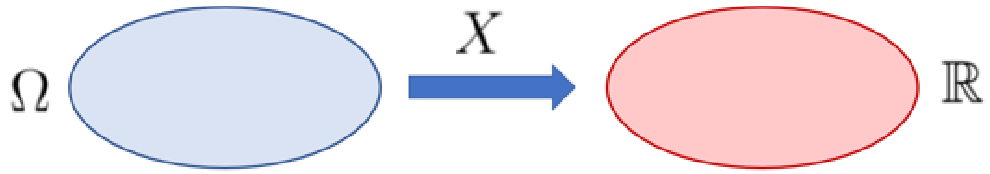
The **distribution** of a random variable X is the map

$$S \mapsto P_X(S) = \mathbb{P}(X \in S) = \mathbb{P}\left(\{\omega \in \Omega \ : \ X(\omega) \in S\}\right),$$

for all subsets $S \subseteq \mathbb{R}$ such that $\{X \in S\} = \{\omega \in \Omega \ : \ X(\omega) \in S\} \in \mathcal{E}$.

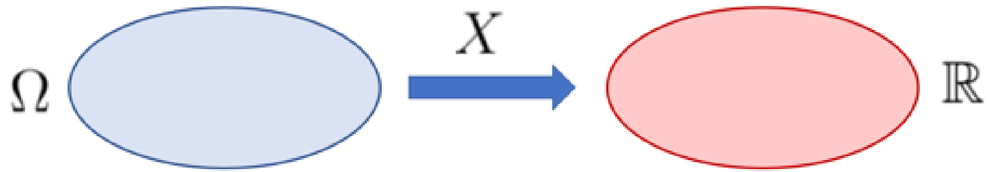
Remark: The distribution P_X of a random variable defines a probability function on (well-behaved) subsets $S \subseteq \mathbb{R}$.

Distributions



We let $\mathfrak{B}(\mathbb{R})$ denote a collection of "well-behaved" (†) subsets $S \subseteq \mathbb{R}$.

Distributions



We let $\mathfrak{B}(\mathbb{R})$ denote a collection of "well-behaved" (†) subsets $S \subseteq \mathbb{R}$.

Theorem . *Suppose we have a probability space* $(\Omega, \mathcal{E}, \mathbb{P})$ *along with a random variable* $X : \Omega \to \mathbb{R}$. *The distribution* P_X *defined by* $P_X(S) = \mathbb{P}(X \in S)$ *for* $S \in \mathfrak{B}(\mathbb{R})$ *satisfies:*

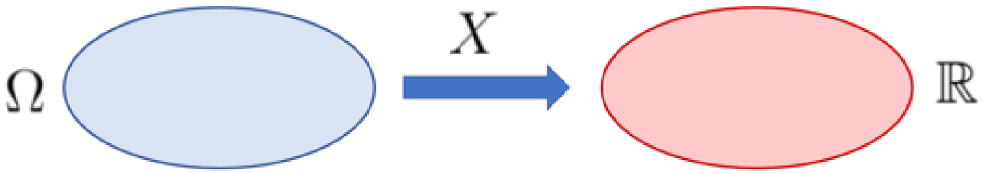
1. *For all* $S \in \mathfrak{B}(\mathbb{R})$ *we have* $P_X(S) = \mathbb{P}(X \in S) \geq 0$;

2. *We have* $P_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = 1$;

3. *Given a sequence of disjoint sets* $A_1, A_2, \ldots \in \mathfrak{B}(\mathbb{R})$ *we have* $P_X\left(\bigcup_j A_j\right) = \sum_j P_X(A_j)$.

Hence, $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P_X)$ *is itself a probability space.*

Distributions



We let $\mathfrak{B}(\mathbb{R})$ denote a collection of "well-behaved" (†) subsets $S \subseteq \mathbb{R}$.

Theorem . *Suppose we have a probability space* $(\Omega, \mathcal{E}, \mathbb{P})$ *along with a random variable* $X : \Omega \to \mathbb{R}$*. The distribution* P_X *defined by* $P_X(S) = \mathbb{P}(X \in S)$ *for* $S \in \mathfrak{B}(\mathbb{R})$ *satisfies:*

1. *For all* $S \in \mathfrak{B}(\mathbb{R})$ *we have* $P_X(S) = \mathbb{P}(X \in S) \geq 0$*;*

2. *We have* $P_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = 1$*;*

3. *Given a sequence of sets* $A_1, A_2, \ldots \in \mathfrak{B}(\mathbb{R})$ *we have* $P_X\left(\bigcup_j A_j\right) = \sum_j P_X(A_j)$*.*

Hence, $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P_X)$ *is itself a probability space.*

(†) (Optional technical remark): The collection $\mathfrak{B}(\mathbb{R})$ is the Borel σ-algebra. This the smallest σ-algebra which contains all intervals of the form $[a, b] \subseteq \mathbb{R}$.

Distributions and distribution functions

Given a probability space $(\Omega, \mathcal{E}, \mathbb{P})$, a **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

The **distribution** of a random variable X is the map

$$S \mapsto P_X(S) = \mathbb{P}(X \in S) = \mathbb{P}\left(\{\omega \in \Omega \ : \ X(\omega) \in S\}\right),$$

for all subsets $S \subseteq \mathbb{R}$ such that $\{X \in S\} = \{\omega \in \Omega \ : \ X(\omega) \in S\} \in \mathcal{E}$.

Distributions and distribution functions

Given a probability space $(\Omega, \mathcal{E}, \mathbb{P})$, a **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

The **distribution** of a random variable X is the map

$$S \mapsto P_X(S) = \mathbb{P}(X \in S) = \mathbb{P}\left(\{\omega \in \Omega \ : \ X(\omega) \in S\}\right),$$

for all subsets $S \subseteq \mathbb{R}$ such that $\{X \in S\} = \{\omega \in \Omega \ : \ X(\omega) \in S\} \in \mathcal{E}$.

The **distribution function** of a random variable X is the map $F_X : \mathbb{R} \to [0, 1]$ by

$$F_X(x) := P_X\left\{(-\infty, x]\right\} = \mathbb{P}\left(X \leq x\right) = \mathbb{P}\left(\{\omega \in \Omega \ : \ X(\omega) \leq x\}\right).$$

Distributions and distribution functions

Given a probability space $(\Omega, \mathcal{E}, \mathbb{P})$, a **random variable** is a mapping $X : \Omega \to \mathbb{R}$ such that for every a, $b \in \mathbb{R}$, $\{\omega \in \Omega \ : \ X(\omega) \in [a, b]\} \in \mathcal{E}$ is an event.

The **distribution** of a random variable X is the map

$$S \mapsto P_X(S) = \mathbb{P}(X \in S) = \mathbb{P}\left(\{\omega \in \Omega \ : \ X(\omega) \in S\}\right),$$

for all subsets $S \subseteq \mathbb{R}$ such that $\{X \in S\} = \{\omega \in \Omega \ : \ X(\omega) \in S\} \in \mathcal{E}$.

The **distribution function** of a random variable X is the map $F_X : \mathbb{R} \to [0, 1]$ by

$$F_X(x) := P_X\{(-\infty, x]\} = \mathbb{P}\left(X \leq x\right) = \mathbb{P}\left(\{\omega \in \Omega \ : \ X(\omega) \leq x\}\right).$$

The **distribution function** F_X is also referred to as the **probability distribution function** or the **cumulative distribution function**.

Example: Bernoulli distributions

We say that a random variable $X : \Omega \to \mathbb{R}$ is **Bernoulli** if P_X is a Bernoulli distribution:

There exists some $q \in [0, 1]$ such that for any $S \subseteq \mathbb{R}$ we have

$$P_X(S) = \mathbb{P}(X \in S) = \begin{cases} 0 & \text{if } S \cap \{0,1\} = \emptyset \\ 1 - q & \text{if } S \cap \{0,1\} = \{0\} \\ q & \text{if } S \cap \{0,1\} = \{1\} \\ 1 & \text{if } S \cap \{0,1\} = \{0,1\}. \end{cases}$$

Example: Bernoulli distributions

We say that a random variable $X : \Omega \to \mathbb{R}$ is **Bernoulli** if P_X is a Bernoulli distribution:

There exists some $q \in [0,1]$ such that for any $S \subseteq \mathbb{R}$ we have

$$P_X(S) = \mathbb{P}(X \in S) = (1-q) \cdot \mathbb{1}_S(0) + q \cdot \mathbb{1}_S(1).$$

Example: Bernoulli distributions

We say that a random variable $X : \Omega \to \mathbb{R}$ is **Bernoulli** if P_X is a Bernoulli distribution:

There exists some $q \in [0,1]$ such that for any $S \subseteq \mathbb{R}$ we have

$$P_X(S) = \mathbb{P}(X \in S) = (1-q) \cdot \mathbb{1}_S(0) + q \cdot \mathbb{1}_S(1).$$

Example 1

A customer in a dealership either buys a car $(X = 1)$ or doesn't buy a car $(X = 0)$;

Example: Bernoulli distributions

We say that a random variable $X : \Omega \to \mathbb{R}$ is **Bernoulli** if P_X is a Bernoulli distribution:

There exists some $q \in [0,1]$ such that for any $S \subseteq \mathbb{R}$ we have

$$P_X(S) = \mathbb{P}(X \in S) = (1-q) \cdot \mathbb{1}_S(0) + q \cdot \mathbb{1}_S(1).$$

Example 1

A customer in a dealership either buys a car $(X = 1)$ or doesn't buy a car $(X = 0)$;

Example 2

A call option either expires in the money $(X = 1)$ or out of the money $(X = 0)$;

Example: Bernoulli distributions

We say that a random variable $X : \Omega \to \mathbb{R}$ is **Bernoulli** if P_X is a Bernoulli distribution:

There exists some $q \in [0,1]$ such that for any $S \subseteq \mathbb{R}$ we have

$$P_X(S) = \mathbb{P}(X \in S) = (1-q) \cdot \mathbb{1}_S(0) + q \cdot \mathbb{1}_S(1).$$

Example 1

A customer in a dealership either buys a car $(X = 1)$ or doesn't buy a car $(X = 0)$;

Example 2

A call option either expires in the money $(X = 1)$ or out of the money $(X = 0)$;

Example 3

A patient either tests positive $(X = 1)$ or negative $(X = 0)$ for a virus.

Example: Bernoulli distributions

We say that a random variable $X : \Omega \to \mathbb{R}$ is **Bernoulli** if P_X is a Bernoulli distribution:

There exists some $q \in [0, 1]$ such that for any $S \subseteq \mathbb{R}$ we have

$$P_X(S) = \mathbb{P}(X \in S) = \begin{cases} 0 & \text{if } S \cap \{0,1\} = \emptyset \\ 1-q & \text{if } S \cap \{0,1\} = \{0\} \\ q & \text{if } S \cap \{0,1\} = \{1\} \\ 1 & \text{if } S \cap \{0,1\} = \{0,1\}. \end{cases}$$

We write $X \sim \mathcal{B}(q)$ for a Bernoulli random variable X with $\mathbb{P}\left(X = 1\right) = q$.

Example: Bernoulli distributions

We say that a random variable $X : \Omega \to \mathbb{R}$ is **Bernoulli** if P_X is a Bernoulli distribution:

There exists some $q \in [0,1]$ such that for any $S \subseteq \mathbb{R}$ we have

$$P_X(S) = \mathbb{P}(X \in S) = \begin{cases} 0 & \text{if } S \cap \{0,1\} = \emptyset \\ 1-q & \text{if } S \cap \{0,1\} = \{0\} \\ q & \text{if } S \cap \{0,1\} = \{1\} \\ 1 & \text{if } S \cap \{0,1\} = \{0,1\}. \end{cases}$$

We write $X \sim \mathcal{B}(q)$ for a Bernoulli random variable X with $\mathbb{P}(X = 1) = q$.

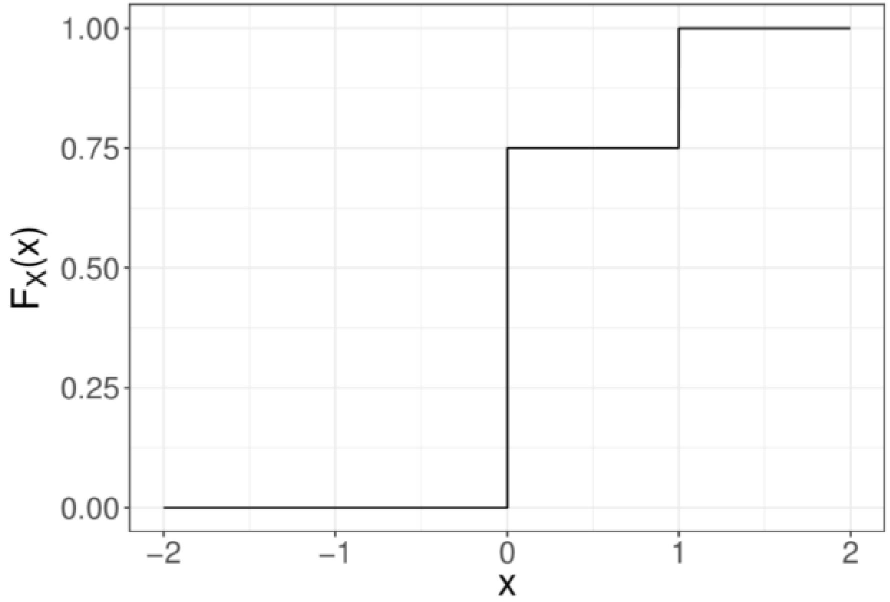
Hence, the cumulative distribution function $F_X : \mathbb{R} \to [0,1]$ satisfies,

$$F_X(x) = \mathbb{P}(X \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ 1-q & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1. \end{cases}$$

Example: Bernoulli distributions

A Bernoulli random variable $X \sim \mathcal{B}(q)$ has cumulative distribution function $F_X : \mathbb{R} \to [0, 1]$,

$$F_X(x) = \mathbb{P}(X \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - q & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1. \end{cases}$$



Example: The roll of a dice

Rolling a fair dice corresponds to a random variable $Z : \Omega \to \mathbb{R}$ with

$$P_Z(S) = \mathbb{P}(Z \in S) = \frac{|S \cap \{1, \ldots, 6\}|}{6} = \frac{1}{6} \cdot \sum_{x \in \{1,\ldots,6\}} \mathbb{1}_S(x).$$

Example: The roll of a dice

Rolling a fair dice corresponds to a random variable $Z : \Omega \to \mathbb{R}$ with

$$P_Z(S) = \mathbb{P}(Z \in S) = \frac{|S \cap \{1, \ldots, 6\}|}{6} = \frac{1}{6} \cdot \sum_{x \in \{1,\ldots,6\}} \mathbb{1}_S(x).$$

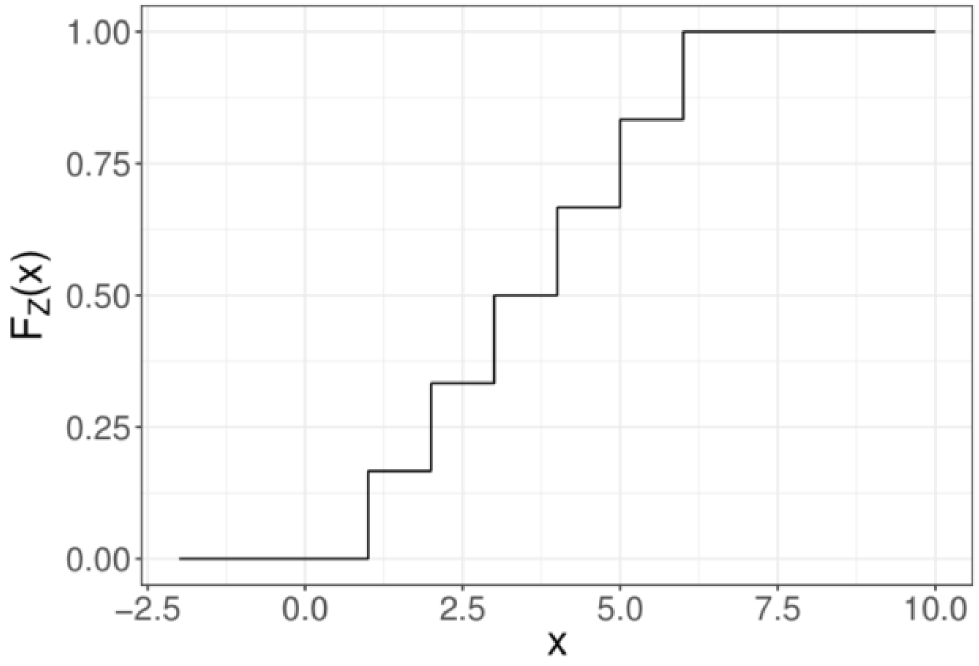
The corresponding cumulative distribution function $F_Z : \mathbb{R} \to [0, 1]$ is,

$$F_Z(x) = \mathbb{P}(Z \leq x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{1}{6} & \text{if } 1 \leq x < 2 \\ \vdots & \\ \frac{5}{6} & \text{if } 5 \leq x < 6 \\ 1 & \text{if } x \geq 6. \end{cases}$$

Example: The roll of a dice

Rolling a fair dice corresponds to a random variable $Z : \Omega \to \mathbb{R}$ with

$$P_Z(S) = \mathbb{P}(Z \in S) = \frac{|S \cap \{1, \ldots, 6\}|}{6} = \frac{1}{6} \cdot \sum_{x \in \{1,\ldots,6\}} \mathbb{1}_S(x).$$



New random variables from old

We often want to create new random variables by combining existing ones.

New random variables from old

We often want to create new random variables by combining existing ones.

Let $\mathfrak{B}(\mathbb{R}^k, \mathbb{R})$ denote the collection of "reasonable" $(\dagger)$ functions $f : \mathbb{R}^k \to \mathbb{R}$.

New random variables from old

We often want to create new random variables by combining existing ones.

Let $\mathfrak{B}(\mathbb{R}^k, \mathbb{R})$ denote the collection of "reasonable" (†) functions $f : \mathbb{R}^k \to \mathbb{R}$.

Given random variables $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ and a function $f \in \mathfrak{B}(\mathbb{R}^k, \mathbb{R})$ we let $f(X_1, \ldots, X_k)$ denote the random variable $Y : \Omega \to \mathbb{R}$ defined by $Y(\omega) := f\left(X_1(\omega), \ldots, X_k(\omega)\right)$ for $\omega \in \Omega$.

New random variables from old

We often want to create new random variables by combining existing ones.

Let $\mathfrak{B}(\mathbb{R}^k, \mathbb{R})$ denote the collection of "reasonable" (†) functions $f : \mathbb{R}^k \to \mathbb{R}$.

Given random variables $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ and a function $f \in \mathfrak{B}(\mathbb{R}^k, \mathbb{R})$ we let $f(X_1, \ldots, X_k)$ denote the random variable $Y : \Omega \to \mathbb{R}$ defined by $Y(\omega) := f\left(X_1(\omega), \ldots, X_k(\omega)\right)$ for $\omega \in \Omega$.

Example

Let Z_1, Z_2, Z_3 be the outcomes of 3 dice roles. Then $Y = Z_1 + Z_2 + Z_3$ is the total accumulated score. More precisely, we take $f : \mathbb{R}^3 \to \mathbb{R}$ by $f(z_1, z_2, z_3) := z_1 + z_2 + z_3$ so that $Y(\omega) := f\left(Z_1(\omega), Z_2(\omega), Z_3(\omega)\right)$ for all $\omega \in \Omega$.

New random variables from old

We often want to create new random variables by combining existing ones.

Let $\mathfrak{B}(\mathbb{R}^k, \mathbb{R})$ denote the collection of "reasonable" (†) functions $f : \mathbb{R}^k \to \mathbb{R}$.

Given random variables $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ and a function $f \in \mathfrak{B}(\mathbb{R}^k, \mathbb{R})$ we let $f(X_1, \ldots, X_k)$ denote the random variable $Y : \Omega \to \mathbb{R}$ defined by $Y(\omega) := f\left(X_1(\omega), \ldots, X_k(\omega)\right)$ for $\omega \in \Omega$.

(†) (Optional technical remark): The collection $\mathfrak{B}(\mathbb{R}^k, \mathbb{R})$ consists of all Borel measurable functions. These are functions $f : \mathbb{R}^k \to \mathbb{R}$ such that $f^{-1}(A) \in \mathfrak{B}(\mathbb{R}^k)$ whenever $A \in \mathfrak{B}(\mathbb{R})$. Here, $\mathfrak{B}(\mathbb{R}^k)$ is the smallest σ-algebra containing all sets of the form $\prod_{i=1}^k [a_i, b_i]$.

Now take a break!

Discrete random variables

We say that the distribution of a random variable $X : \Omega \to \mathbb{R}$ is **supported** on a set $A \subseteq \mathbb{R}$ if $P_X(A) = \mathbb{P}(X \in A) = 1$.

Discrete random variables



We say that the distribution of a random variable $X : \Omega \to \mathbb{R}$ is **supported** on a set $A \subseteq \mathbb{R}$ if $P_X(A) = \mathbb{P}(X \in A) = 1$.

A **discrete random variable** is a random variable $X : \Omega \to \mathbb{R}$ who's distribution is supported on a discrete (and hence finite or countably infinite) set $A \subseteq \mathbb{R}$.

Discrete random variables

We say that the distribution of a random variable $X : \Omega \to \mathbb{R}$ is **supported** on a set $A \subseteq \mathbb{R}$ if $P_X(A) = \mathbb{P}(X \in A) = 1$.

A **discrete random variable** is a random variable $X : \Omega \to \mathbb{R}$ who's distribution is supported on a discrete (and hence finite or countably infinite) set $A \subseteq \mathbb{R}$.

Examples

The distribution of a Bernoulli random variable $X \sim \mathcal{B}(q)$ is supported on $\{0, 1\}$.

The distribution of a random dice roll Z is supported on $\{1, 2, \ldots, 6\}$.

Probability mass functions



Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable).

The **probability mass function** of the random variable X is the function $p_X : \mathbb{R} \to [0, 1]$ defined by

$$p_X(x) := P_X(\{x\}) = \mathbb{P}\left(X = x\right).$$

Probability mass functions

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable).

The **probability mass function** of the random variable X is the function $p_X : \mathbb{R} \to [0, 1]$ defined by

$$p_X(x) := P_X(\{x\}) = \mathbb{P}\left(X = x\right).$$

Example

A Bernoulli random variable $X \sim \mathcal{B}(q)$ has probability mass function

$$p_X(x) = \begin{cases} 1 - q & \text{if } x = 0 \\ q & \text{if } x = 1 \\ 0 \text{ otherwise.} \end{cases}$$

Probability mass functions

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable).

The **probability mass function** of the random variable X is the function $p_X : \mathbb{R} \to [0, 1]$ defined by

$$p_X(x) := P_X(\{x\}) = \mathbb{P}\,(X = x)\,.$$

Example

The distribution of a random dice roll Z has probability mass function

$$p_Z(x) = \begin{cases} \frac{1}{6} & \text{if } x \in \{1, 2, \ldots, 6\} \\ 0 & \text{if } x \notin \{1, 2, \ldots, 6\}. \end{cases}$$

Probability mass functions

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable).

The **probability mass function** of the random variable X is the function $p_X : \mathbb{R} \to [0, 1]$ defined by

$$p_X(x) := P_X(\{x\}) = \mathbb{P}(X = x).$$

Key Properties

1. For all $x \in \mathbb{R}$, the probability mass function $p_X(x) \geq 0$.
2. The values of the probability mass function sum to unity $\sum_{x \in \mathbb{R}} p_X(x) = 1$.

Expectation of a discrete random variable

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}(X = x)$ be its probability mass function.

The **expectation** $\mathbb{E}(X)$ of the random variable X is defined by $\mathbb{E}(X) := \sum_{x \in \mathbb{R}} x \cdot p_X(x)$.

Expectation of a discrete random variable

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}(X = x)$ be its probability mass function.

The **expectation** $\mathbb{E}(X)$ of the random variable X is defined by $\mathbb{E}(X) := \sum_{x \in \mathbb{R}} x \cdot p_X(x)$.

Example

Given a $X \sim \mathcal{B}(q)$ be a Bernoulli random variable we have

$$\mathbb{E}(X) := \sum_{x \in \mathbb{R}} x \cdot p_X(x) = (1-q) \cdot 0 + q \cdot 1 + \overbrace{\sum_{x \in \mathbb{R} \setminus \{0,1\}} x \cdot p_X(0)}^{0} = q.$$

Expectation of a discrete random variable

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}\left(X = x\right)$ be its probability mass function.

The **expectation** $\mathbb{E}(X)$ of the random variable X is defined by $\mathbb{E}(X) := \sum_{x \in \mathbb{R}} x \cdot p_X(x)$.

Example

Given a Z be the outcome of a fair dice roll we have,

$$\mathbb{E}(Z) := \sum_{x \in \mathbb{R}} x \cdot p_Z(x) = \frac{1}{6} \cdot (1 + 2 + \ldots + 6) + \overbrace{\sum_{x \in \mathbb{R} \setminus \{0,1\}} x \cdot p_X(0)}^{0} = \frac{7}{2}.$$

Expectation of a discrete random variable

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}\left(X = x\right)$ be its probability mass function.

The **expectation** $\mathbb{E}(X)$ of the random variable X is defined by $\mathbb{E}(X) := \sum_{x \in \mathbb{R}} x \cdot p_X(x)$.

We can view the expectation of a random variable as the long run sample mean obtained by repeatedly sampling independent copies of X.

Expectation of a discrete random variable

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}(X = x)$ be its probability mass function.

The **expectation** $\mathbb{E}(X)$ of the random variable X is defined by $\mathbb{E}(X) := \sum_{x \in \mathbb{R}} x \cdot p_X(x)$.

We can view the expectation of a random variable as the long run sample mean obtained by repeatedly sampling independent copies of X.

The expectation is often referred to the population average or population mean.

Expectation of a discrete random variable

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}(X = x)$ be its probability mass function.

The **expectation** $\mathbb{E}(X)$ of the random variable X is defined by $\mathbb{E}(X) := \sum_{x \in \mathbb{R}} x \cdot p_X(x)$.

We can view the expectation of a random variable as the long run sample average obtained by repeatedly sampling independent copies of X.

Expectation is *linear* in the following sense:

Lemma *Given random variables* $X_1, \ldots, X_K : \Omega \to \mathbb{R}$ *and numbers* $\alpha_1, \ldots, \alpha_K \in \mathbb{R}$ *we have*

$$\mathbb{E}\left(\sum_{i=1}^{K} \alpha_i X_i\right) = \sum_{i=1}^{K} \alpha_i \mathbb{E}(X_i).$$

Variance of a discrete random variable

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}(X = x)$ be its probability mass function.

The **variance** $\mathrm{Var}(X)$ of the random variable X is defined by $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right]$.

Variance of a discrete random variable

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}(X = x)$ be its probability mass function.

The **variance** $\mathrm{Var}(X)$ of the random variable X is defined by $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right]$.

Example

Given a $X \sim \mathcal{B}(q)$ be a Bernoulli random variable we have

$$\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right] = \sum_{x\in\mathbb{R}} p_X(x) \cdot (x-q)^2 = (1-q) \cdot q^2 + q \cdot (1-q)^2 = q(1-q).$$

Variance of a discrete random variable

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}\left(X = x\right)$ be its probability mass function.

The **variance** $\mathrm{Var}(X)$ of the random variable X is defined by $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}\left(X\right)\}^2\right]$.

Example

Letting Z be the outcome of a fair dice roll we have,

$$\mathrm{Var}(Z) := \mathbb{E}\left[\{Z - \mathbb{E}\left(Z\right)\}^2\right] = \frac{1}{6}\sum_{x=1}^{6}\left(x - \frac{7}{2}\right)^2 = \frac{35}{12}.$$

Variance of a discrete random variable

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}\left(X = x\right)$ be its probability mass function.

The **variance** $\mathrm{Var}(X)$ of the random variable X is defined by $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}\left(X\right)\}^2\right]$.

We can view the variance of a random variable as measuring how much it typically fluctuates around its expectation $\mathbb{E}(X)$ upon repeatedly sampling independent copies of X.

Variance of a discrete random variable

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}(X = x)$ be its probability mass function.

The **variance** $\mathrm{Var}(X)$ of the random variable X is defined by $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right]$.

We can view the variance of a random variable as measuring how much it typically fluctuates around its expectation $\mathbb{E}(X)$ upon repeatedly sampling independent copies of X.

The variance of a random variable is often referred to as the population variance.

The population variance and sample variance are closely connected, as we shall see.

Variance of a discrete random variable


Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}\left(X = x\right)$ be its probability mass function.

The **variance** $\mathrm{Var}(X)$ of the random variable X is defined by $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}\left(X\right)\}^2\right]$.

We can view the variance of a random variable as measuring how much it typically fluctuates around its expectation $\mathbb{E}(X)$ upon repeatedly sampling independent copies of X.

Given a random variable $X : \Omega \to \mathbb{R}$ and $\alpha \in \mathbb{R}$ we have $\mathrm{Var}\left(\alpha \cdot X\right) = \alpha^2 \cdot \mathrm{Var}(X)$.

This motivates the (population) standard deviation $\sigma(X) := \sqrt{\mathrm{Var}(X)}$.

Given a random variable $X : \Omega \to \mathbb{R}$ and $\alpha \in \mathbb{R}$ we have $\sigma\left(\alpha \cdot X\right) = |\alpha| \cdot \sigma(X)$

Variance of a discrete random variable


Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}(X = x)$ be its probability mass function.

The **variance** $\mathrm{Var}(X)$ of the random variable X is defined by $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right]$.

We can view the variance of a random variable as measuring how much it typically fluctuates around its expectation $\mathbb{E}(X)$ upon repeatedly sampling independent copies of X.

What is the variance of a linear combination of random variables $\left(\sum_{i=1}^{K} \alpha_i X_i\right)$?

Independent random variables

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ defined by $F_X(x) := \mathbb{P}(X \leq x)$ for all $x \in \mathbb{R}$.

Independent random variables

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ defined by $F_X(x) := \mathbb{P}(X \leq x)$ for all $x \in \mathbb{R}$.

We define the joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$ by

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = \mathbb{P}\left(\{X_1 \leq x_1\} \cap \ldots \cap \{X_k \leq x_k\}\right) \text{ for all } (x_1, \ldots, x_k) \in \mathbb{R}^k.$$

Independent random variables

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ defined by $F_X(x) := \mathbb{P}(X \leq x)$ for all $x \in \mathbb{R}$.

We define the joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$ by

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = \mathbb{P}\left(\{X_1 \leq x_1\} \cap \ldots \cap \{X_k \leq x_k\}\right) \text{ for all } (x_1, \ldots, x_k) \in \mathbb{R}^k.$$

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

Independent random variables

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ defined by $F_X(x) := \mathbb{P}(X \leq x)$ for all $x \in \mathbb{R}$.

We define the joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$ by

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = \mathbb{P}(\{X_1 \leq x_1\} \cap \ldots \cap \{X_k \leq x_k\}) \text{ for all } (x_1, \ldots, x_k) \in \mathbb{R}^k.$$

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

Equivalently, $X_1, \ldots, X_k$ are independent if for all $x_1, \ldots, x_k \in \mathbb{R}$ the sequence of events $\{X_1 \leq x_1\}, \ldots, \{X_k \leq x_k\}$ are (mutually) independent.

Independent random variables

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ and joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$.

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

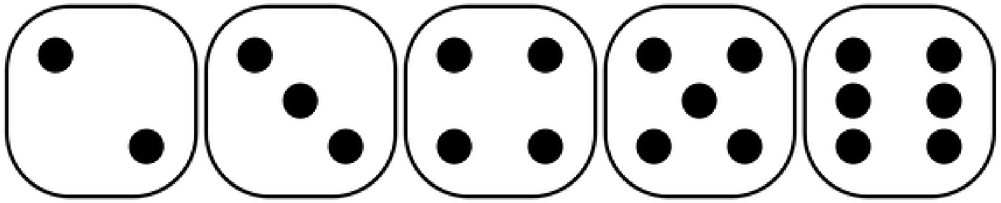
Example

Suppose that I flip k coins and let X_i correspond to the result of the i-th coin flip.

A natural assumption here is that the different coin flips have no interaction with one another.

Hence, we can model $X_1, \ldots, X_k$ as a sequence of independent random variables.

Independent random variables



Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ and joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$.

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

Example

Suppose that I roll k dice and let X_i correspond to the result of the i-th dice roll.

A natural assumption here is that the different dice rolls have no interaction with one another.

Hence, we can model $X_1, \ldots, X_k$ as a sequence of independent random variables.

Independent random variables

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ and joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$.

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

Example

Suppose that k patients are treated for a medical condition in a clinical trial.

Let X_i correspond to the recovery time of the i-th patient in the trial.

We may assume that the level of interaction between the patients is very weak.

Hence, we can model $X_1, \ldots, X_k$ as a sequence of independent random variables.

Dependent random variables

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ and joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$.

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

We say that $X_1, \ldots, X_k$ are **dependent** if they are not independent.

Dependent random variables

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ and joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$.

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

We say that $X_1, \ldots, X_k$ are **dependent** if they are not independent.

Example

Suppose that we flip k coins and let Z_j be one if the j-th coin was a head and 0 otherwise.

For each $i = 1, \ldots, k$ let $X_i = Z_1 + \ldots Z_i$, the accumulated total.

The sequence $X_1, \ldots, X_k$ is a dependent sequence of random variables.

Dependent random variables

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ and joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$.

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

We say that $X_1, \ldots, X_k$ are **dependent** if they are not independent.

Example

Let X_i be the AUDUSD exchange rate on the i-th hour of the week.

Exhange rates on consecutive time windows are closely connected, so $X_1, \ldots, X_k$ are dependent.

Dependent random variables

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ and joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$.

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

We say that $X_1, \ldots, X_k$ are **dependent** if they are not independent.

Example

Let X_i be the average temperature on the i-th day of the year.

Weather conditions on consecutive days are closely connected, so $X_1, \ldots, X_k$ are dependent.

Dependent random variables

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ and joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$.

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

We say that $X_1, \ldots, X_k$ are **dependent** if they are not independent.

Example

Let X_i be the number of tickets sold in the i-th minute for a music festival.

The total number of tickets sold is limited, so $X_1, \ldots, X_k$ are dependent.

Quantifying dependence through covariation

Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are random random variables.

The **covariance** between X and Y is defined by

$$\mathrm{Cov}(X, Y) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}\{Y - \mathbb{E}(Y)\}\right] = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Quantifying dependence through covariation

Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are random random variables.

The **covariance** between X and Y is defined by

$$\mathrm{Cov}(X,Y) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}\{Y - \mathbb{E}(Y)\}\right] = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Recall that the **variance** of a random variable X is $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right]$.
Hence, $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$.

Quantifying dependence through covariation

Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are random random variables.

The **covariance** between X and Y is defined by

$$\mathrm{Cov}(X, Y) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}\{Y - \mathbb{E}(Y)\}\right] = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Recall that the **variance** of a random variable X is $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right]$.
Hence, $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.

The covariance between random variables is a population analogue of the sample covariance.

Quantifying dependence through covariation

Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are random random variables.

The **covariance** between X and Y is defined by

$$\mathrm{Cov}(X, Y) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}\{Y - \mathbb{E}(Y)\}\right] = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Recall that the **variance** of a random variable X is $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right]$.
Hence, $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.

The covariance between random variables is a population analogue of the sample covariance.

We also define the (population) **correlation** in terms of the (population) covariance by

$$\mathrm{Corr}(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

Quantifying dependence through covariation

Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are random random variables.

The **covariance** between X and Y is defined by

$$\mathrm{Cov}(X,Y) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}\{Y - \mathbb{E}(Y)\}\right] = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Recall that the **variance** of a random variable X is $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right]$.
Hence, $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$.

The covariance between random variables is a population analogue of the sample covariance.

We also define the (population) **correlation** in terms of the (population) covariance by

$$\mathrm{Corr}(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

The correlation gives a scale-invariant quantification of the linear relationship between X and Y.

Quantifying dependence through covariation

Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are random random variables.

The **covariance** between X and Y is defined by

$$\mathrm{Cov}(X, Y) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}\{Y - \mathbb{E}(Y)\}\right] = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Recall that the **variance** of a random variable X is $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right]$.
Hence, $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.

The covariance between random variables is a population analogue of the sample covariance.

Key fact:

If X and Y are independent random variables then $\mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y) = 0$.

Quantifying dependence through covariation

Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are random random variables.

The **covariance** between X and Y is defined by

$$\mathrm{Cov}(X, Y) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}\{Y - \mathbb{E}(Y)\}\right] = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Recall that the **variance** of a random variable X is $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right]$.
Hence, $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.

The covariance between random variables is a population analogue of the sample covariance.

Key fact:

If X and Y are independent random variables then $\mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y) = 0$.

However, $\mathrm{Cov}(X, Y) = 0$ doesn't necessarily mean that X and Y are independent.

An alternative perspective on independence

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ and joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$.

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

An alternative perspective on independence

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ and joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$.

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

Recall that $\mathfrak{B}(\mathbb{R}, \mathbb{R})$ denotes the collection of reasonable (i.e. Borel) functions $f : \mathbb{R} \to \mathbb{R}$.

An alternative perspective on independence

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ and joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$.

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

Recall that $\mathfrak{B}(\mathbb{R}, \mathbb{R})$ denotes the collection of reasonable (i.e. Borel) functions $f : \mathbb{R} \to \mathbb{R}$.

Theorem *Let $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ be a sequence of random variables. Then $X_1, \ldots, X_k$ are independent if and only if the following relationship holds for every sequence of functions $f_1, \ldots, f_k \in \mathfrak{B}(\mathbb{R}, \mathbb{R})$,*

$$\mathbb{E}[f_1(X_1) \cdot \ldots \cdot f_k(X_k)] = \mathbb{E}[f_1(X_1)] \cdot \ldots \cdot \mathbb{E}[f_k(X_k)].$$

An alternative perspective on independence

Suppose that $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ are random random variables, with distribution functions $F_{X_1}, \ldots, F_{X_k} : \mathbb{R} \to [0,1]$ and joint cumulative distribution function $F_{X_1,\ldots,X_k} : \mathbb{R}^k \to \mathbb{R}$.

We say that $X_1, \ldots, X_k$ are (mutually) **independent** if for all $x_1, \ldots, x_k \in \mathbb{R}$ we have

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = F_{X_1}(x) \times \ldots \times F_{X_k}(x_k).$$

Recall that $\mathfrak{B}(\mathbb{R}, \mathbb{R})$ denotes the collection of reasonable (i.e. Borel) functions $f : \mathbb{R} \to \mathbb{R}$.

Theorem *Let $X_1, \ldots, X_k : \Omega \to \mathbb{R}$ be a sequence of random variables. Then $X_1, \ldots, X_k$ are independent if and only if the following relationship holds for every sequence of functions $f_1, \ldots, f_k \in \mathfrak{B}(\mathbb{R}, \mathbb{R})$,*

$$\mathbb{E}\left[f_1(X_1) \cdot \ldots \cdot f_k(X_k)\right] = \mathbb{E}\left[f_1(X_1)\right] \cdot \ldots \cdot \mathbb{E}\left[f_k(X_k)\right].$$

In particular, if X and Y are independent random variable then $\mathrm{Cov}(X, Y) = 0$.

The variance of a linear combination of random variables

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}(X = x)$ be its probability mass function.

The **variance** $\mathrm{Var}(X)$ of the random variable X is defined by $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right]$.

What is the variance of a linear combination of random variables $\left(\sum_{i=1}^{K} \alpha_i X_i\right)$?

The variance of a linear combination of random variables

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}(X = x)$ be its probability mass function.

The **variance** $\mathrm{Var}(X)$ of the random variable X is defined by $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right]$.

What is the variance of a linear combination of random variables $\left(\sum_{i=1}^K \alpha_i X_i\right)$?

Lemma *Given random variables* $X_1, \ldots, X_K : \Omega \to \mathbb{R}$ *and* $\alpha_1, \ldots, \alpha_K \in \mathbb{R}$ *we have*

$$\mathrm{Var}\left(\sum_{i=1}^K \alpha_i X_i\right) = \sum_{i=1}^K \alpha_i^2 \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le k} \alpha_i \cdot \alpha_j \mathrm{Cov}(X_i, X_j).$$

The variance of a linear combination of random variables

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable) and let $p_X(x) := \mathbb{P}(X = x)$ be its probability mass function.

The **variance** $\mathrm{Var}(X)$ of the random variable X is defined by $\mathrm{Var}(X) := \mathbb{E}\left[\{X - \mathbb{E}(X)\}^2\right]$.

What is the variance of a linear combination of random variables $\left(\sum_{i=1}^K \alpha_i X_i\right)$?

Lemma *Given random variables* $X_1, \dots, X_K : \Omega \to \mathbb{R}$ *and* $\alpha_1, \dots, \alpha_K \in \mathbb{R}$ *we have*

$$\mathrm{Var}\left(\sum_{i=1}^K \alpha_i X_i\right) = \sum_{i=1}^K \alpha_i^2 \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le k} \alpha_i \cdot \alpha_j \mathrm{Cov}(X_i, X_j).$$

In particular, if $X_1, \dots, X_k$ *are independent then* $\mathrm{Var}\left(\sum_{i=1}^K \alpha_i X_i\right) = \sum_{i=1}^K \alpha_i^2 \mathrm{Var}(X_i)$.

Now take a break!

Binomial distributions

We often want to model the number of successes in a sequence of approximately independent trials.

Binomial distributions

We often want to model the number of successes in a sequence of approximately independent trials.

Examples

1. The number of red balls drawn from a bag whilst sampling with replacement.

Binomial distributions

We often want to model the number of successes in a sequence of approximately independent trials.

Examples

1. The number of red balls drawn from a bag whilst sampling with replacement.

2. The number of patients who recover following treatment in a clinical trial.

Binomial distributions

We often want to model the number of successes in a sequence of approximately independent trials.

Examples

1. The number of red balls drawn from a bag whilst sampling with replacement.

2. The number of patients who recover following treatment in a clinical trial.

3. The number of customers who decide to buy a car following a test drive.

Binomial distributions

The **Binomial distribution** allows us to model the number of successes out of n independent trials, where each trial has a success probability p.

Binomial distributions

The **Binomial distribution** allows us to model the number of successes out of n independent trials, where each trial has a success probability p.

Suppose that $X_1, \ldots, X_n$ are independent random variables where each $X_i \sim \mathcal{B}(p)$ has Bernoulli distribution with $\mathbb{E}[X_i] = p$.

Binomial distributions

The **Binomial distribution** allows us to model the number of successes out of n independent trials, where each trial has a success probability p.

Suppose that $X_1, \ldots, X_n$ are independent random variables where each $X_i \sim \mathcal{B}(p)$ has Bernoulli distribution with $\mathbb{E}[X_i] = p$.

Then the sum $Z = X_1 + \ldots + X_n$ is a **Binomial** random variable with parameters n and p.

Binomial distributions

The **Binomial distribution** allows us to model the number of successes out of n independent trials, where each trial has a success probability p.

Suppose that $X_1, \ldots, X_n$ are independent random variables where each $X_i \sim \mathcal{B}(p)$ has Bernoulli distribution with $\mathbb{E}[X_i] = p$.

Then the sum $Z = X_1 + \ldots + X_n$ is a **Binomial** random variable with parameters n and p.

The probability mass function $p_Z : \mathbb{R} \to [0,1]$ is given by

$$p_Z(r) := \mathbb{P}\left(Z = r\right) = \binom{n}{r} \cdot p^r \cdot (1-p)^{n-r},$$

for $r \in \{0, 1, \ldots, n\}$ and $p_Z(r) = 0$ for $r \notin \{0, 1, \ldots, n\}$.

Binomial distributions

Suppose that Z has **Binomial distribution** with parameters n and p.

Hence, $Z = X_1 + \ldots + X_n$ where $X_1, \ldots, X_n$ are independent random variables where each $X_i \sim \mathcal{B}(p)$ has Bernoulli distribution with $\mathbb{E}[X_i] = p$.

Binomial distributions

Suppose that Z has **Binomial distribution** with parameters n and p.

Hence, $Z = X_1 + \ldots + X_n$ where $X_1, \ldots, X_n$ are independent random variables where each $X_i \sim \mathcal{B}(p)$ has Bernoulli distribution with $\mathbb{E}[X_i] = p$.

For each sequence $(s_1, \ldots, s_n) \in \{0, 1\}$ with $\sum_{i=1}^n s_i = r$ we have

$$\mathbb{P}\left(X_1 = s_1, \ldots, X_n = s_n\right) = \mathbb{P}\left(X_1 = s_1\right) \cdot \ldots \mathbb{P}\left(X_n = s_n\right) = \left(\prod_{i:s_i=1} p\right) \cdot \left(\prod_{i:s_i=0} (1-p)\right) = p^r \cdot (1-p)^{n-r}.$$

Binomial distributions

Suppose that Z has **Binomial distribution** with parameters n and p.

Hence, $Z = X_1 + \ldots + X_n$ where $X_1, \ldots, X_n$ are independent random variables where each $X_i \sim \mathcal{B}(p)$ has Bernoulli distribution with $\mathbb{E}[X_i] = p$.

For each sequence $(s_1, \ldots, s_n) \in \{0, 1\}$ with $\sum_{i=1}^n s_i = r$ we have

$$\mathbb{P}\left(X_1 = s_1, \ldots, X_n = s_n\right) = \mathbb{P}\left(X_1 = s_1\right) \cdot \ldots \mathbb{P}\left(X_n = s_n\right) = \left(\prod_{i:s_i=1} p\right) \cdot \left(\prod_{i:s_i=0} (1-p)\right) = p^r \cdot (1-p)^{n-r}.$$

There are $\binom{n}{r}$ sequences of the form $(s_1, \ldots, s_n) \in \{0, 1\}$ with $\sum_{i=1}^n s_i = r$.

Binomial distributions

Suppose that Z has **Binomial distribution** with parameters n and p.

Hence, $Z = X_1 + \ldots + X_n$ where $X_1, \ldots, X_n$ are independent random variables where each $X_i \sim \mathcal{B}(p)$ has Bernoulli distribution with $\mathbb{E}[X_i] = p$.

For each sequence $(s_1, \ldots, s_n) \in \{0,1\}$ with $\sum_{i=1}^n s_i = r$ we have

$$\mathbb{P}\left(X_1 = s_1, \ldots, X_n = s_n\right) = \mathbb{P}\left(X_1 = s_1\right) \cdot \ldots \mathbb{P}\left(X_n = s_n\right) = \left(\prod_{i:s_i=1} p\right) \cdot \left(\prod_{i:s_i=0} (1-p)\right) = p^r \cdot (1-p)^{n-r}.$$

There are $\binom{n}{r}$ sequences of the form $(s_1, \ldots, s_n) \in \{0,1\}$ with $\sum_{i=1}^n s_i = r$.

Hence, we have

$$p_Z(r) := \mathbb{P}\left(Z = r\right) = \sum_{(s_1,\ldots,s_n) \in \{0,1\} \;:\; \sum_{i=1}^n s_i = r} \mathbb{P}\left(X_1 = s_1, \ldots, X_n = s_n\right) = \binom{n}{r} \cdot p^r \cdot (1-p)^{n-r},$$

for $r \in \{0, 1, \ldots, n\}$ and $p_Z(r) = 0$ for $r \notin \{0, 1, \ldots, n\}$.

Binomial distributions

The **Binomial distribution** allows us to model the number of successes out of n independent trials, where each trial has a success probability p.

Suppose that $X_1, \ldots, X_n$ are independent random variables where each $X_i \sim \mathcal{B}(p)$ has Bernoulli distribution with $\mathbb{E}[X_i] = p$.

Then the sum $Z = X_1 + \ldots + X_n$ is a **Binomial** random variable with parameters n and p.

The probability mass function $p_Z : \mathbb{R} \to [0,1]$ is given by

$$p_Z(r) := \mathbb{P}\left(Z = r\right) = \binom{n}{r} \cdot p^r \cdot (1-p)^{n-r},$$

for $r \in \{0, 1, \ldots, n\}$ and $p_Z(r) = 0$ for $r \notin \{0, 1, \ldots, n\}$.

Binomial distributions

The **Binomial distribution** allows us to model the number of successes out of n independent trials, where each trial has a success probability p.

Suppose that $X_1, \ldots, X_n$ are independent random variables where each $X_i \sim \mathcal{B}(p)$ has Bernoulli distribution with $\mathbb{E}[X_i] = p$.

Then the sum $Z = X_1 + \ldots + X_n$ is a **Binomial** random variable with parameters n and p.

For each $i = 1, \ldots, n$ we have $\mathbb{E}[X_i] = p$ and $\mathrm{Var}(X_i) = p(1-p)$.

Since $X_1, \ldots, X_n$ are independent we have

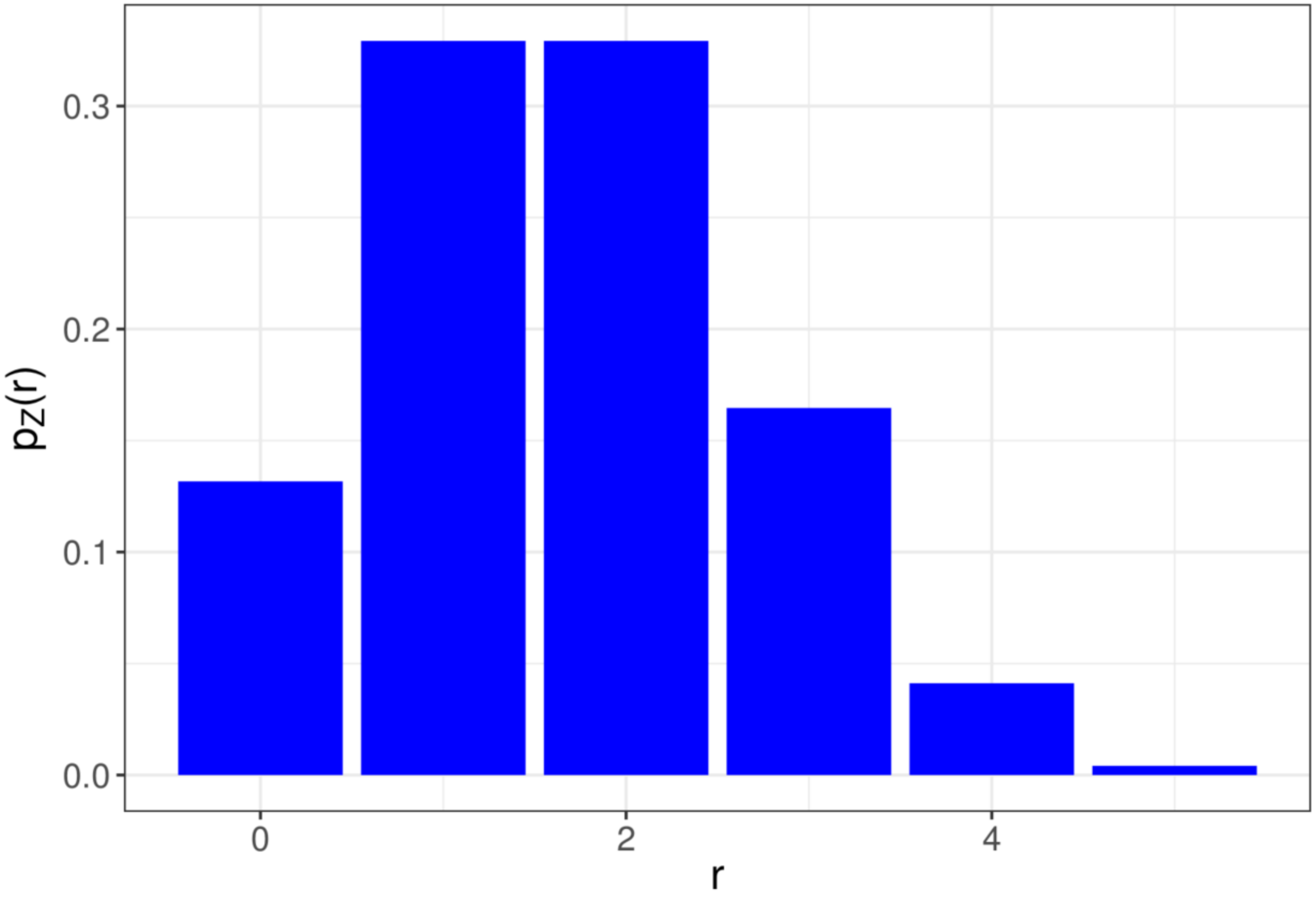
$$\mathbb{E}\left(Z\right) = \mathbb{E}\left(X_1 + \ldots + X_n\right) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n) = n \cdot p.$$
$$\mathrm{Var}\left(Z\right) = \mathrm{Var}\left(X_1 + \ldots + X_n\right) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n) = n \cdot p \cdot (1-p).$$

Binomial distributions

Let's explore the probability mass function p_Z with $p = \frac{1}{3}$ and different values of n.

Binomial distributions

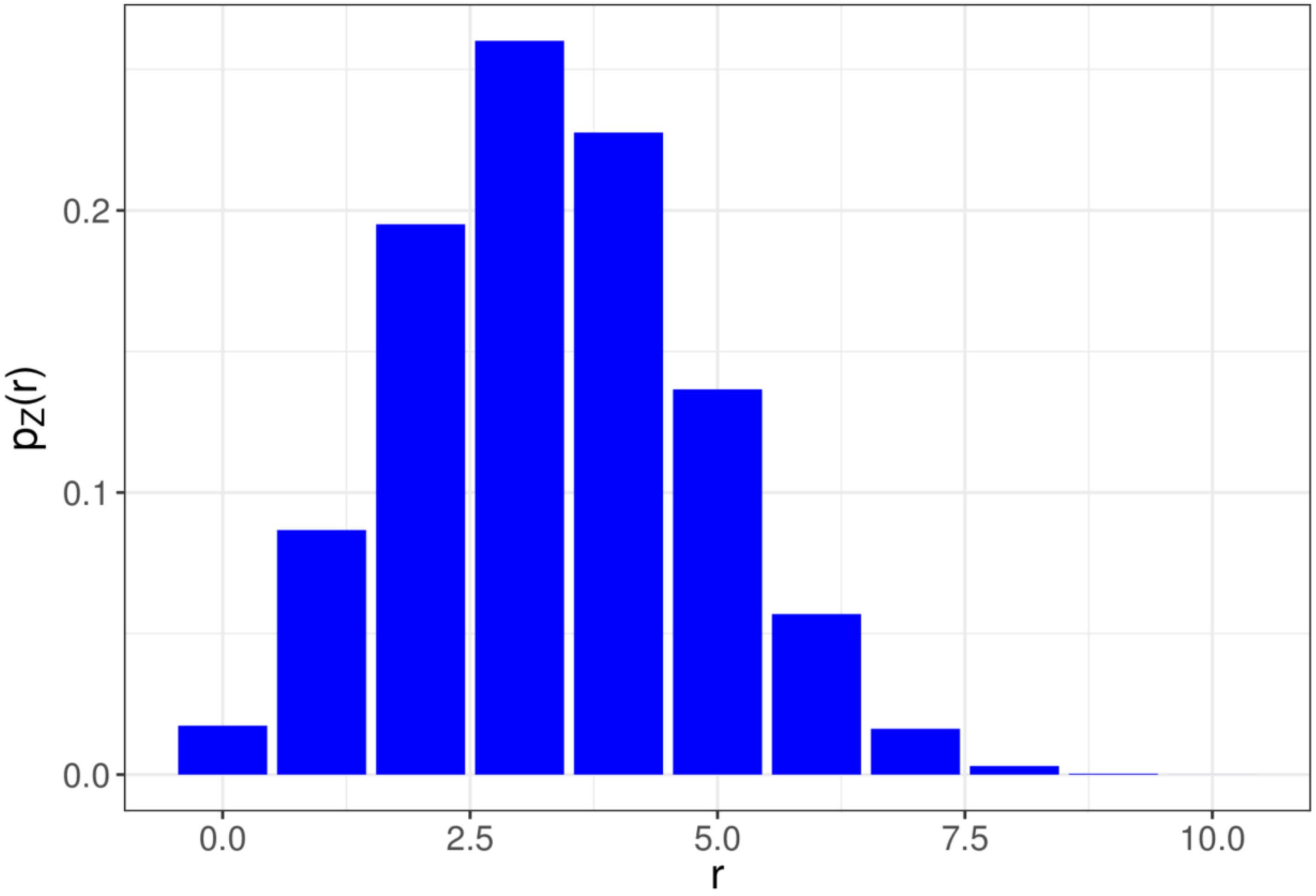
Let's explore the probability mass function p_Z with $p = \frac{1}{3}$ and different values of n.



$$n = 5$$

Binomial distributions

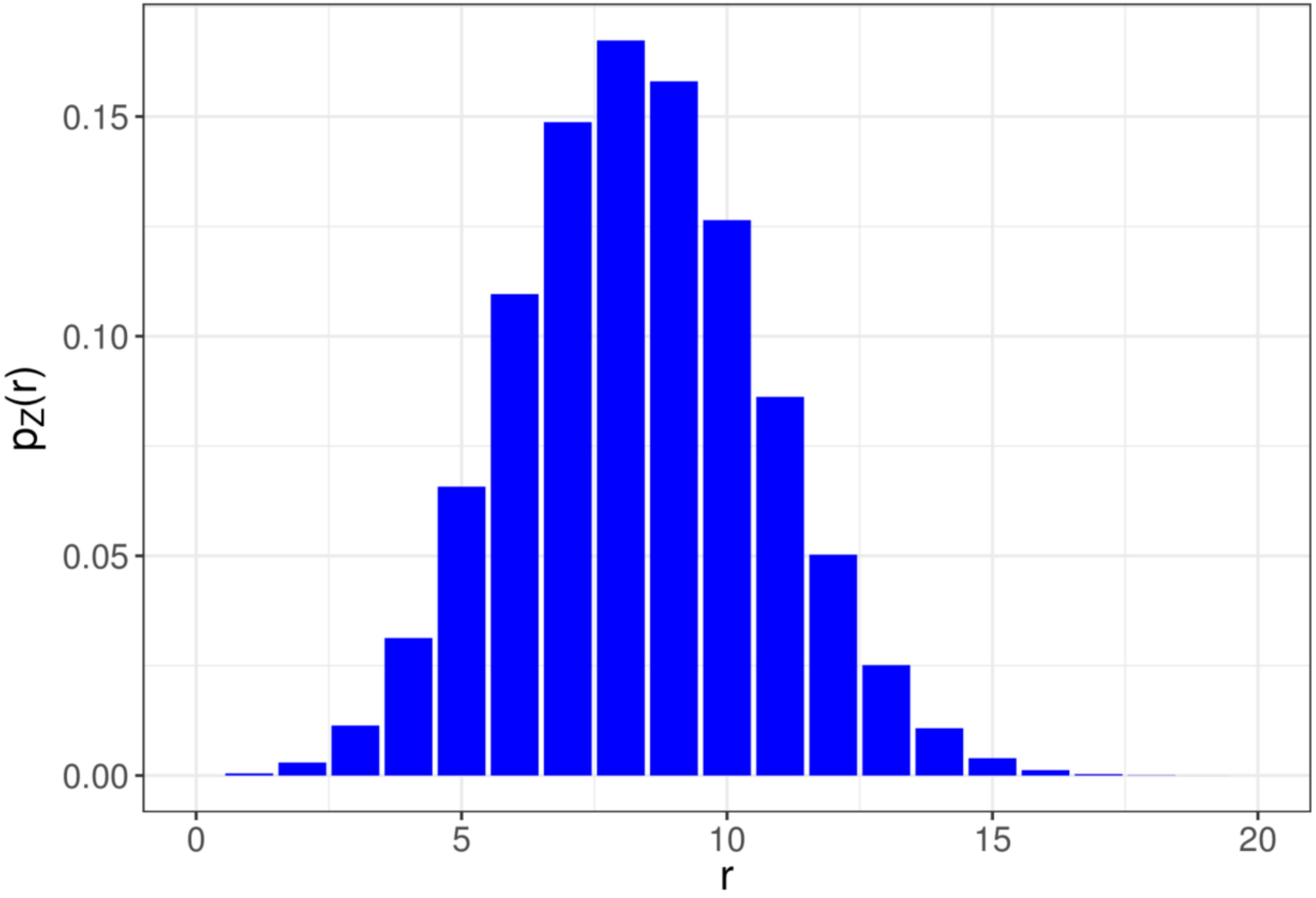
Let's explore the probability mass function p_Z with $p = \frac{1}{3}$ and different values of n.



$$n = 10$$

Binomial distributions

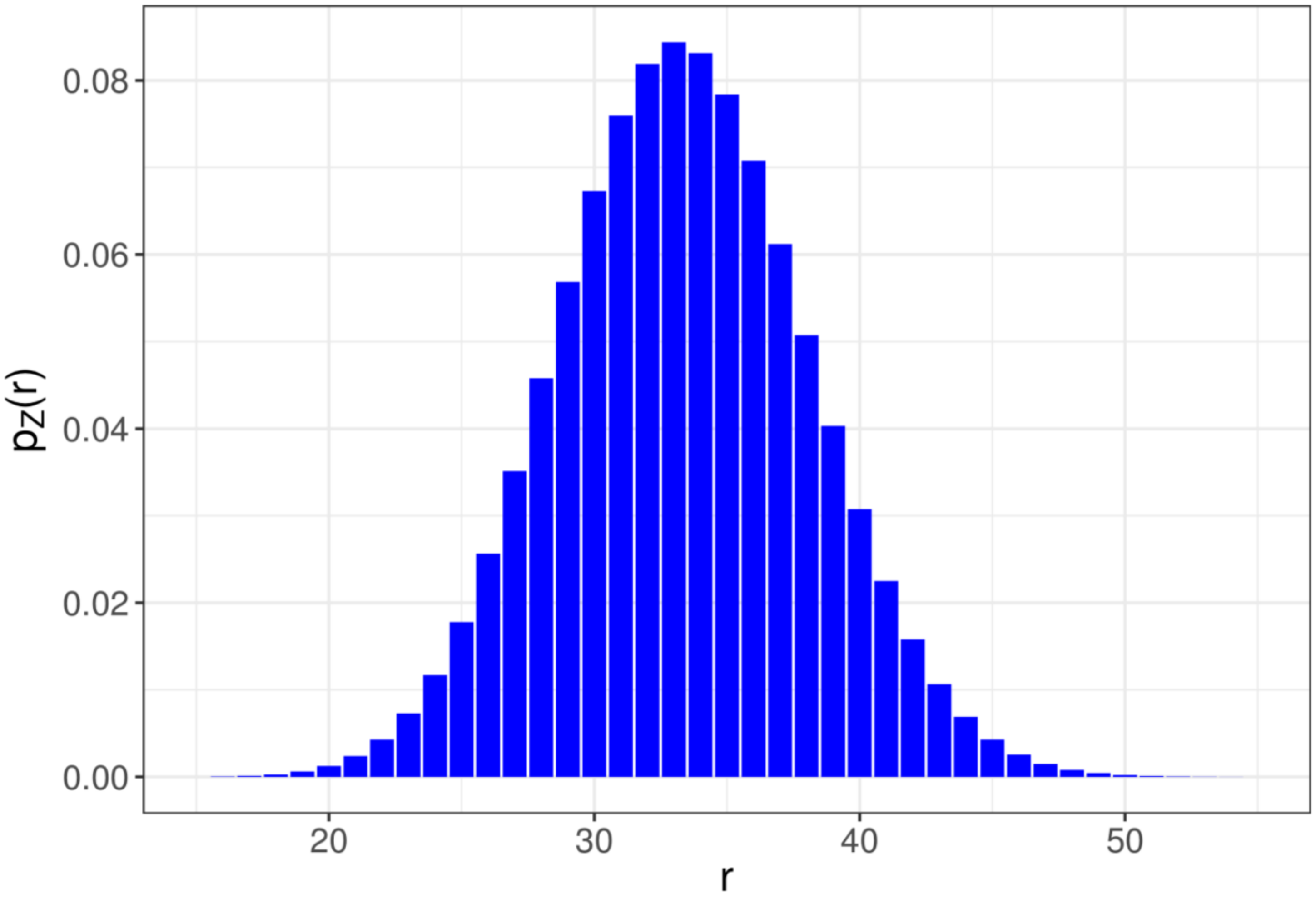
Let's explore the probability mass function p_Z with $p = \frac{1}{3}$ and different values of n.



$n = 25$

Binomial distributions

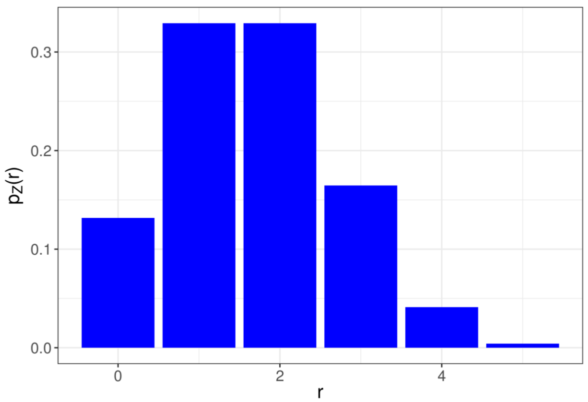
Let's explore the probability mass function p_Z with $p = \frac{1}{3}$ and different values of n.



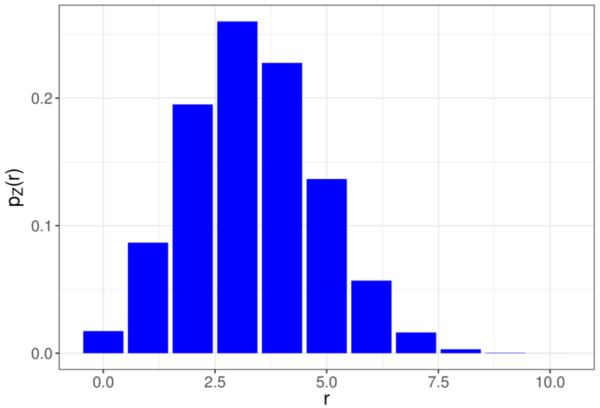
$n = 100$

Binomial distributions

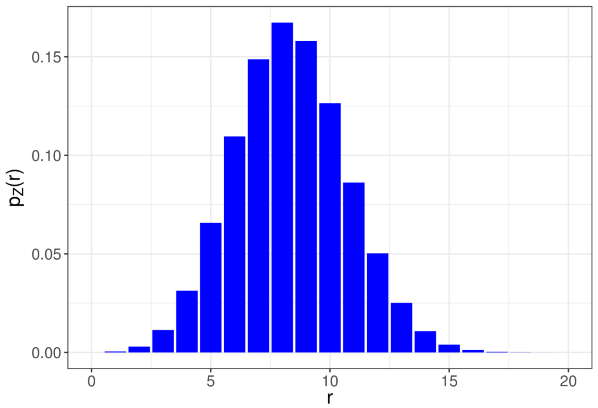
Let's explore the probability mass function p_Z with $p = \frac{1}{3}$ and different values of n.



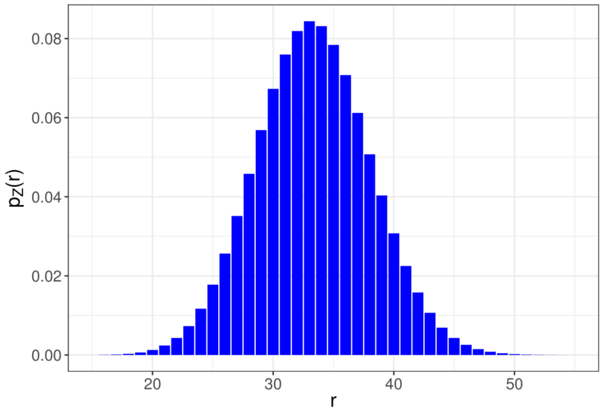
$n = 5$



$n = 10$



$n = 25$



$n = 100$

What have we covered?

- We introduced the important concept of a random variable.
- We saw how random variables can be quantified via its distribution function.
- We introduced the concept of a discrete random variable and discussed the probability mass function.
- We discussed several important examples including Bernoulli and Binomial random variables.
- We also defined the expectation, variance, covariance and correlation of random variables.
- In addition, we generalized our understanding of independence from sequences of events to sequences of random variables.

Thanks for listening!

Henry W J Reeve

henry.reeve@bristol.ac.uk

Include EMATM0061 in the subject of your email.