Random variables

Statistical Computing and Empirical Methods
Unit EMATM0061, Data Science MSc

Rihuan Ke
rihuan.ke@bristol.ac.uk

Teaching Block 1, 2022

University of BRISTOL

What we will cover today

We will introduce the important concept of a random variable

We will discuss the concept of distributions, which play a key role in describing the stochastic behaviour of a random variable

We will also talk about distribution functions of random variables

Relevant concepts

A random experiment is a procedure (real or imagined) which:

1. has a well-defined set of possible outcomes;
2. could (at least in principle) be repeated arbitrarily many times.

An event is a set (i.e. a collection) of possible outcomes of an experiment

A sample space is the set of all possible outcomes of interest for a random experiment

A probability space consists of a triple $(\Omega, \mathcal{E}, \mathbb{P})$, where Ω is a sample space, $\mathcal{E}$ is a well-behaved collection of events in Ω, and $\mathbb{P} : \mathcal{E} \to \mathbb{R}$ is a probability function.

What is a random variable - example

We use a "random variable" to represent the outcomes of a random experiment

Example: Rolling a dice

Sample space $\Omega = \{\text{the } i\text{-th face lands face-up} : i = 1, \cdots, 6\}$

$X = 1$ if the 1-st face lands face-up

$X = 2$ if the 2-nd face lands face-up

$\vdots$

$X = 6$ if the 6-th face lands face-up

An event: $\{X \in \{1, 2, 3\}\}$ means one of the first three faces lands face-up

Probability: $\mathbb{P}(X \in \{2\})$ means $\mathbb{P}(\text{the 2-nd face lands face-up})$

What is a random variable - example

We use a "random variable" to represent the outcomes of a random experiment

Example: Flipping a coin

Sample space = {heads-up,tails-up}

$X = 0$ if heads-up
$X = 1$ if tails-up

An event: $\{X = 0\}$ means the events of heads-up

Probability: $\mathbb{P}(X = 0)$ means $\mathbb{P}(\{\text{heads-up}\})$

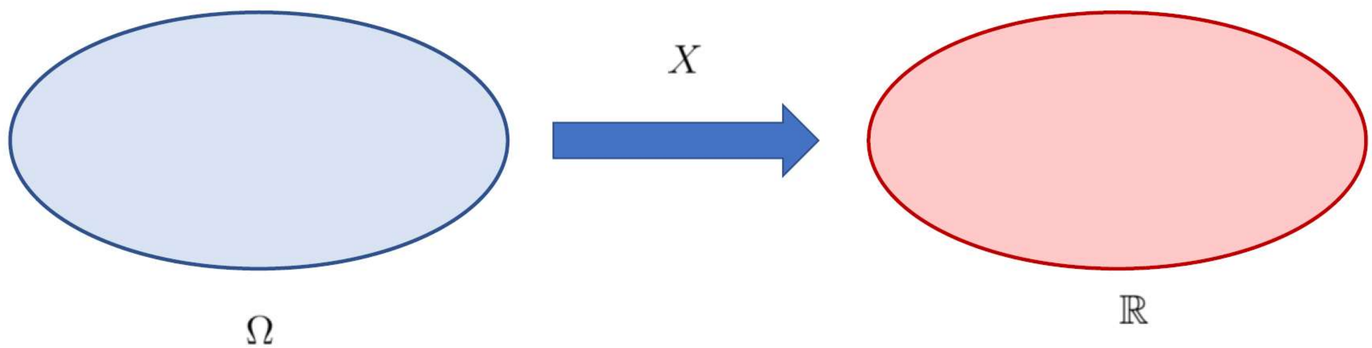
Summary:
The *random variable* X maps each outcome to a number
We can represent *events* using the random variables: $\{X = 0\}$, $\{X \in \{1, 2, 3\}\}$ etc.

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

$$\text{for every } a, b \in \mathbb{R},\ \{\omega \in \Omega : X(\omega) \in [a, b]\} \text{ is an event in } \mathcal{E}$$



$$\text{An event } (\in \mathcal{E}) \xleftarrow{\quad X^{-1} \quad} [a, b]$$

Random variables

Random variables

Suppose we have a sample space $(\Omega, \mathcal{E}, \mathbb{P})$. A random variable is a mapping $X : \Omega \to \mathbb{R}$, such that

$$\text{for every } a, b \in \mathbb{R},\ \{\omega \in \Omega : X(\omega) \in [a, b]\} \text{ is an event in } \mathcal{E}$$

Remark: The condition that "for every $a, b \in \mathbb{R}$, $\{\omega \in \Omega : X(\omega) \in [a, b]\}$ is an event in $\mathcal{E}$" is essential. With this definition, we can describe events efficiently with the values of X:

$\{\omega \in \Omega : X(\omega) \in [a, b]\}$ always represent an event.
$\mathbb{P}(\{\omega \in \Omega : X(\omega) \in [a, b]\})$ is always well defined.

Random variable: examples

Example: We roll 5 dices in a row and record each of the results.

The sample space is $\Omega = \{1, 2, \cdots, 6\}^5 = \{(x_1, \cdots, x_5) : x_i \in \{1, \cdots, 6\}\}$.

We can define a random variable as the result of the final dice roll, $X((x_1, \cdots, x_5)) = x_5$.

$\{\omega \in \Omega : X(\omega) \in [1, 1]\} = \{(x_1, \cdots, x_4, 1) : x_i \in \{1, \cdots, 6\}\}$ is an event.

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

The sample space is $\Omega = \{1, \cdots, 100\}^{10} = \{(x_1, \cdots, x_{10}) : x_i \in \{1, \cdots, 100\}\}$. The numbers $1, \cdots, 50$ represent red balls.

We can define a random variable as the number of red balls sampled $X((x_1, x_2, \cdots, x_{10})) = \sum_{i=1}^{10} \mathbb{1}_{\{1, \cdots, 50\}}(x_i)$

Notations related to random variables

Events:

For $S \subseteq \mathbb{R}$, we write $\{X \in S\}$ for an event $\{\omega \in \Omega : X(\omega) \in S\}$ which is in $\mathcal{E}$

For $a \in \mathbb{R}$, we write $\{X = a\}$ for an event $\{\omega \in \Omega : X(\omega) = a\}$ which is in $\mathcal{E}$

For $a \in \mathbb{R}$, we write $\{X \leq a\}$ for an event $\{\omega \in \Omega : X(\omega) \leq a\}$ which is in $\mathcal{E}$

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

Probability:

For $S \subseteq \mathbb{R}$, we write $\mathbb{P}(X \in S)$ for the probability $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in S\})$.

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

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

Distribution

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

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

Distribution of a random variable

The distribution of a random variable X is a function given by

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

for any $S \subseteq \mathbb{R}$ in a well-behaved collection of subsets of $\mathbb{R}$ (†).

(**Optional technical remark †**): Here the "well-behaved" collection of subsets of $\mathbb{R}$ is characterised by the Borel σ-algebra on $\mathbb{R}$, denoted by $\mathfrak{B}(\mathbb{R})$, which is the smallest σ-algebra containing all sets of the form $[a, b] \subseteq \mathbb{R}$.

Distribution defines new probability functions

The distribution P_X of a random variable defines a probability function on (well-behaved) subsets $S \subseteq \mathbb{R}$ of $\mathbb{R}$. We let $\mathfrak{B}(\mathbb{R})$ denote a collection of "well-behaved" (†) subsets $S \subseteq \mathbb{R}$.

Theorem (Distribution of random variables)

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 $P_X(S) = P(X \in S)$ for $S \in \mathfrak{B}(\mathbb{R})$ satisfies

1. For all $S \in \mathfrak{B}(\mathbb{R})$, we have $P_X(S) = P(X \in S) \geq 0$.
2. We have $P_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = 1$
3. Given a sequence of disjoint sets $A_1, A_2, \dots \in \mathfrak{B}(\mathbb{R})$, we have $P_X(\cup_j A_j) = \sum_j P_X(A_j)$.

Therefore, P_X satisfies the laws of probability, and $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P_X)$ is itself a probability space.

(Optional technical remarks †): The well-behaved subsets of $\mathbb{R}$ is a Borel σ-algebra.

Distribution functions

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

Recall: the distribution of a random variable X is given by $P_X(S) := \mathbb{P}(X \in S)$ for "well-behaved" subsets $S \subseteq \mathbb{R}$.

Distribution functions

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

$$F_X(x) = \mathbb{P}(X \leq x) \text{ for } x \in \mathbb{R}.$$

Equivalently, the distribution function is given by $F_X(x) = P_X((-\infty, x])$.

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

The distribution function F_X is a non-decreasing function on $\mathbb{R}$

Distribution, distribution function: example

Example: Rolling a fair dice

Sample space $\Omega = \{\omega_1, \cdots, \omega_6\}$ where w_i corresponds the i-th face lands face-up.

Random variable $(\Omega \to \mathbb{R})$: $Z(w_i) = i$

Distribution $(\mathfrak{B}(\mathbb{R}) \to \mathbb{R})$:
$P_Z(S) = \mathbb{P}(Z \in S) = \mathbb{P}(Z \in S \cap \{1, \cdots, 6\}) = \frac{|S \cap \{1, \cdots, 6\}|}{6} = \frac{1}{6} \sum_{x \in \{1, \cdots, 6\}} \mathbb{1}_S(x)$

Distribution function $(\mathbb{R} \to [0, 1])$: $F_Z(x) = \mathbb{P}(Z \leq x)$

$$F_Z(X) = \begin{cases} 0 & \text{if } x < 1, \\ 1/6 & \text{if } 1 \leq x < 1, \\ \vdots & \\ 5/6 & \text{if } 5 \leq x < 6, \\ 1 & \text{if } 6 \leq x. \end{cases}$$

Distribution, distribution function: example

Example: A customer in a dealership either buys a car or doesn't buy a car

Sample space $\Omega = \{\omega_0, \omega_1\}$ for outcomes ω_1 (buy a car, with probability q) and ω_0 (doesn't buy a car).

Random variable ($\Omega \to \mathbb{R}$): $X(w_i) = i$ for $i = 0, 1$.

Distribution ($\mathfrak{B}(\mathbb{R}) \to \mathbb{R}$) : $P_X(S) = \mathbb{P}(X \in S) = (1-q)\mathbb{1}_S(0) + q\mathbb{1}_S(1)$.

Distribution function ($\mathbb{R} \to [0,1]$): $F_X(x) = \mathbb{P}(X \leq x)$

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

Bernoulli distribution and Bernoulli random variable

Bernoulli distribution. A distribution P_X is called a Bernoulli distribution if there exist some $q \in [0, 1]$, such that

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

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

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

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

Example: A patient either tests positive $(X = 1)$ or negative $(X = 0)$.

Creating new random variables from old

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

Creating new random variables from old

Given random variable $X_1, \cdots, X_k : \Omega \to \mathbb{R}$ and a reasonable (†) function f: $\mathbb{R}^k \to \mathbb{R}$. We can define a random variable $Y : \Omega \to \mathbb{R}$ as a function of $X_1, \cdots, X_k$, given by

$$Y(\omega) = f(X_1(\omega), X_2(\omega), \cdots, X_k(\omega)) \text{ for } \omega \in \Omega.$$

Example:

Let Z_1, Z_2 and Z_3 be the outcomes of 3 dice rolls. Then $Y = Z_1 + Z_2 + Z_3$ defines a new variable (meaning the total accumulated score).

More precisely, we take $f : \mathbb{R}^3 \to \mathbb{R}$ by $f(x_1, x_2, x_3) = x_1 + x_2 + x_3$. So $Y(\omega) = f(Z_1(\omega), Z_2(\omega), Z_3(\omega))$ for all $\omega \in \Omega$.

Creating new random variables from old

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

Creating new random variables from old

Given random variable $X_1, \cdots, X_k : \Omega \to \mathbb{R}$ and a reasonable (†) function f: $\mathbb{R}^k \to \mathbb{R}$. We can define a random variable $Y : \Omega \to \mathbb{R}$ as a function of $X_1, \cdots, X_k$, given by

$$Y(\omega) = f(X_1(\omega), X_2(\omega), \cdots, X_k(\omega)) \text{ for } \omega \in \Omega.$$

(**Optional technical remark †**): Here "reasonable" functions can be described by the collection $\mathfrak{B}\big(\mathbb{R}^k, \mathbb{R}\big)$, which consists of all Borel-measurable functions. These are functions

$$f : \mathbb{R}^k \to \mathbb{R} \text{ such that } f^{-1}(A) \in \mathfrak{B}\left(\mathbb{R}^k\right) \text{ whenever } A \in \mathfrak{B}(\mathbb{R})$$

Here $\mathfrak{B}\big(\mathbb{R}^k\big)$ is the smallest σ-algebra containing all sets of the form $\prod_{i=1}^k [a_i, b_i]$.

What we have learned today

We introduced the important concept of a random variable.

We saw how random variables can be quantified via its distribution and distribution function.

We investigated the idea of creating new random variables by combining existing ones.

Thanks for listening!

Dr. Rihuan Ke
rihuan.ke@bristol.ac.uk

Statistical Computing and Empirical Methods
Unit EMATM0061, MSc Data Science