Finite probability spaces

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 discuss finite probability spaces

We will focus on the special case of simple probability spaces

- In such cases estimating probability reduces to combinatorics.

We will consider products, permutations and combinations.

Random experiments, events and sample spaces

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

The laws of probability and their consequences

Definition: Probability

Given a sample space Ω along with a well-behaved collection of events $\mathcal{E}$, a probability $\mathbb{P}$ is a function which assigns a number $\mathbb{P}(A)$ to each event $A \in \mathcal{E}$, and satisfies rules 1, 2, and 3:

Rule 1: $\mathbb{P}(A) \geq 0$ for any event A

Rule 2: $\mathbb{P}(\Omega) = 1$ for sample space Ω

Rule 3: For pairwise disjoint events $A_1, A_2, \cdots$, we have

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

We refer to the triple $(\Omega, \mathcal{E}, \mathbb{P})$ as a probability space.

Consequence 1: $\mathbb{P}(\emptyset) = 0$

Consequence 2: If $A, B \in \mathcal{E}$ are events and $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Consequence 3: For any event $A \in \mathcal{E}$, we have $0 \leq \mathbb{P}(A) \leq 1$.

Consequence 4: For events $S_1, S_2, \cdots$, we have $\mathbb{P}(\cup_{i=1}^{\infty} S_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(S_i)$.

Finite probability spaces

Here, we consider sample spaces with finite numbers of elements (outcomes).

Finite probability spaces

A finite probability space consists of a triple $(\Omega, \mathcal{E}, \mathbb{P})$ where

1. $\Omega = \{\omega_1, \omega_2, \cdots, \omega_k\}$ is a finite sample space.
2. $\mathcal{E}$ is given by $\{A \subseteq \Omega\}$, i.e., the collection of all subsets of Ω.
3. The probability $\mathbb{P}$ on Ω is constructed in the following way.

Specify a vector $\mathbf{p} = (p_1, p_2, \cdots, p_k)$ that satisfying

$$(1).\ p_i \geq 0 \text{ for } i = 1, 2, \cdots, k. \quad \text{and} \quad (2).\ \sum_{i=1}^{k} p_i = 1$$

Define a probability $\mathbb{P}$ based on $\mathbf{p}$ by

$$\mathbb{P}(A) := \sum_{i=1}^{k} p_i \cdot \mathbb{1}_A(\omega_i) \text{ for } A \subseteq \Omega$$

By definition, we have $\mathbb{P}(\omega_i) = p_i$ for any $\omega_i \in \Omega$.

The vector $\mathbf{p} = (p_1, \cdots, p_k)$ is called a probability vector.

Finite probability spaces and three key rules

Recall that a finite probability space is a triple $(\Omega, \mathcal{E}, \mathbb{P})$ where Ω is finite, $\mathcal{E} = \{A \subseteq \Omega\}$, and $\mathbb{P}(A) := \sum_{i=1}^{k} p_i \cdot \mathbb{1}_A(\omega_i)$ for $A \subseteq \Omega$.

To make sure $\mathbb{P}$ is a probability, we must show that the fundamental rules of probability are satisfied.

Recall that the three key rules of probability are: Rule 1: $\mathbb{P}(A) \geq 0$ for any event A; Rule 2: $\mathbb{P}(\Omega) = 1$; Rule 3: $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ for disjoint $\{A_i\}$.

Rule 1: for any $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$ ✓

Proof: Since $p_i \geq 0$ for all $i = 1, 2, \cdots, k$, we have

$$\mathbb{P}(A) := \sum_{i=1}^{k} p_i \cdot \mathbb{1}_A(\omega_i) \geq 0.$$

Finite probability spaces and three key rules

Recall that a finite probability space is a triple $(\Omega, \mathcal{E}, \mathbb{P})$ where Ω is finite, $\mathcal{E} = \{A \subseteq \Omega\}$, and $\mathbb{P}(A) := \sum_{i=1}^{k} p_i \cdot \mathbb{1}_A(\omega_i)$ for $A \subseteq \Omega$.

Recall that the three key rules of probability are: Rule 1: $\mathbb{P}(A) \geq 0$ for any event A; Rule 2: $\mathbb{P}(\Omega) = 1$; Rule 3: $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ for disjoint $\{A_i\}$.

Rule 2: the sample space Ω has probability $\mathbb{P}(\Omega) = 1$ ✓

Proof: For $i = 1, 2, \cdots, k$, $\mathbb{1}_\Omega(\omega_i) = 1$, because $\omega_i \in \Omega$. Hence

$$\begin{aligned}
\mathbb{P}(\Omega) &:= \sum_{i=1}^{k} p_i \cdot \mathbb{1}_\Omega(\omega_i) \qquad \text{by the definition of } \mathbb{P} \\
&= \sum_{i=1}^{k} p_i \qquad \text{since } \omega_i \in \Omega \text{ for all } i \\
&= 1 \qquad \text{by the definition of } p_i.
\end{aligned}$$

So we have $\mathbb{P}(\Omega) = 1$.

Finite probability spaces and three key rules

Recall that a finite probability space is a triple $(\Omega, \mathcal{E}, \mathbb{P})$ where Ω is finite, $\mathcal{E} = \{A \subseteq \Omega\}$, and $\mathbb{P}(A) := \sum_{i=1}^{k} p_i \cdot \mathbb{1}_A(\omega_i)$ for $A \subseteq \Omega$.

Recall that the three key rules of probability are: Rule 1: $\mathbb{P}(A) \geq 0$ for any event A; Rule 2: $\mathbb{P}(\Omega) = 1$; Rule 3: $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ for disjoint $\{A_i\}$.

Rule 3: For a countable sequence of pairwise disjoint events $A_1, A_2, \cdots$, we have

$$\mathbb{P}\big(\cup_{j=1}^{\infty} A_j\big) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).$$

Proof: Let $S = \cup_{j=1}^{\infty} A_j$. For any $\omega \in \Omega$:

(1). If $\mathbb{1}_S(\omega) = 1$, then $\omega \in S$, hence ω is in exactly one of the (pairwise disjoint) sets $A_1, A_2, \cdots$, and hence $\sum_{j=1}^{\infty} \mathbb{1}_{A_j}(\omega) = 1$.

(2). If $\mathbb{1}_S(\omega) = 0$, then $\omega \notin S$, hence $\omega \notin A_j$ for all j, and hence $\sum_{j=1}^{\infty} \mathbb{1}_{A_j}(\omega) = 0$.

Therefore, for any $i \in \{1, \cdots, k\}$, we have $\mathbb{1}_S(\omega_i) = \sum_{j=1}^{\infty} \mathbb{1}_{A_j}(\omega_i)$, and

$$\mathbb{P}(S) = \sum_{i=1}^{k} p_i \mathbb{1}_S(\omega_i) = \sum_{i=1}^{k} p_i \left(\sum_{j=1}^{\infty} \mathbb{1}_{A_j}(\omega_i) \right) = \sum_{j=1}^{\infty} \left(\sum_{i=1}^{k} p_i \mathbb{1}_{A_j}(\omega_i) \right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).$$

Finite probability space examples

Recall that a finite probability space is a triple $(\Omega, \mathcal{E}, \mathbb{P})$ where Ω is finite, $\mathcal{E} = \{A \subseteq \Omega\}$, and $\mathbb{P}(A) := \sum_{i=1}^{k} p_i \cdot \mathbb{1}_A(\omega_i)$ for $A \subseteq \Omega$.

Example 1: Rolling a fair dice

Recall that in the previous example of rolling a dice, we have

Sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$
Set of events $\mathcal{E} = \{A \subseteq \Omega\}$
Probability $\mathbb{P}(A) = \frac{|A|}{6}$ for any $A \in \mathcal{E}$

This is a finite probability space. The probability $\mathbb{P}$ has a corresponding probability vector: $\mathbf{p} = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$, and

$$\mathbb{P}(A) = \sum_{i=1}^{k} p_i \cdot \mathbb{1}_A(\omega_i)$$

Therefore $\mathbb{P}$ is a probability

Finite probability space examples

Recall that a finite probability space is a triple $(\Omega, \mathcal{E}, \mathbb{P})$ where Ω is finite, $\mathcal{E} = \{A \subseteq \Omega\}$, and $\mathbb{P}(A) := \sum_{i=1}^{k} p_i \cdot \mathbb{1}_A(\omega_i)$ for $A \subseteq \Omega$.

Example 2: A customer in the dealership either buys a car (1) or doesn't buy a car (0)

Recall that in the last lecture we have

> Sample space $\Omega = \{0, 1\}$
> Set of events $\mathcal{E} = \{A \subseteq \Omega\} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$
> Probability $\mathbb{P}(\emptyset) = 0,\ \mathbb{P}(\{0\}) = 1 - q,\ \mathbb{P}(\{1\}) = q,\ \mathbb{P}(\{0, 1\}) = 1$ (where $0 \leq q \leq 1$)

This is a finite probability space. The probability $\mathbb{P}$ has a corresponding probability vector: $\mathbf{p} = (1 - q, q)$

Therefore $\mathbb{P}$ is a probability

Finite probability space examples

Recall that a finite probability space is a triple $(\Omega, \mathcal{E}, \mathbb{P})$ where Ω is finite, $\mathcal{E} = \{A \subseteq \Omega\}$, and $\mathbb{P}(A) := \sum_{i=1}^{k} p_i \cdot \mathbb{1}_A(\omega_i)$ for $A \subseteq \Omega$.

Example 3: A patient either tests positive (1) or negative (0) for a virus.

Sample space $\Omega = \{0, 1\}$
Set of events $\mathcal{E} = \{A \subseteq \Omega\} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$
Probability $\mathbb{P}(\emptyset) = 0,\ \mathbb{P}(\{0\}) = 1 - q,\ \mathbb{P}(\{1\}) = q,\ \mathbb{P}(\{0, 1\}) = 1$ (where $0 \leq q \leq 1$)

This is a finite probability space. The probability $\mathbb{P}$ has a corresponding probability vector: $\mathbf{p} = (1 - q, q)$

Therefore $\mathbb{P}$ is a probability

Bernoulli distribution

The Bernoulli distribution refers to the probability distribution on a sample space $\{0, 1\}$, in which the probability of outcome "1" is q, and the probability of outcome "0" is $1 - q$ for some $q \in [0, 1]$.

These are all examples of Bernoulli distributions:

> Example 2: A customer in the dealership either buys a car (1) or doesn't buy a car (0)
>
> Example 3: A patient either tests positive (1) or negative (0) for a virus.

Bernoulli distributions are typical examples of probabilities in finite sample spaces.

Simple probability spaces

Simple probability spaces are special types of finite probability spaces.

Simple probability spaces

A simple probability space is a finite probability space $(\Omega, \mathcal{E}, \mathbb{P})$ with the probability vector given by $\mathbf{p} = (\frac{1}{k}, \frac{1}{k}, \cdots, \frac{1}{k})$ where $k = |\Omega|$.

Consequently, the probability of an event $A \in \mathcal{E}$ is given by

$$\mathbb{P}(A) = \sum_{i=1}^{k} p_i \cdot \mathbb{1}_A(\omega_i) = \frac{|A|}{|\Omega|}.$$

Example: rolling a fair dice

$\mathbf{p} = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$

$\mathbb{P}(\{2, 4, 6\}) = \frac{|\{2,4,6\}|}{6} = 1/2$

Simple product probability spaces

Suppose that we have a set $\mathcal{X}$ and another set $\mathcal{Y}$.

Cartesian product

The cartesian product $\mathcal{X} \times \mathcal{Y} := \{(x, y) : x \in \mathcal{X}, \text{ and } y \in \mathcal{Y}\}$.

For example, if $\mathcal{X} = \{1, 2\}, \mathcal{Y} = \{1, 2, 3\}$, then
$\mathcal{X} \times \mathcal{Y} = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3)\}$.

If $\mathcal{X}$ and $\mathcal{Y}$ have cardinalities $|\mathcal{X}|$ and $|\mathcal{Y}|$ respectively, then $|\mathcal{X} \times \mathcal{Y}| = |\mathcal{X}| \cdot |\mathcal{Y}|$.

A simple product probability space

A simple product probability space is a simple probability space $(\Omega, \mathcal{E}, \mathbb{P})$ in which $\Omega = \mathcal{X} \times \mathcal{Y}$.

Simple product probability spaces satisfy the rules of probability.

Simple product probability spaces

Recall that a simple product probability space is a simple probability space $(\Omega, \mathcal{E}, \mathbb{P})$ in which $\Omega = \mathcal{X} \times \mathcal{Y}$.

Example 4: Suppose that I flip a coin and record "heads" (1) and "tails" (0) ($\mathcal{X} = \{0, 1\}$), and I also roll a dice and record which face up ($\mathcal{Y} = \{1, 2, 3, 4, 5, 6\}$).

So an outcome can be of the form $(1, 6)$, where 1 means "heads" up and 6 means the face of number 6.

Take a sample space $\Omega = \mathcal{X} \times \mathcal{Y}$ and for $A \subseteq \Omega$, we have

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

e.g., $\mathbb{P}(\{(1, 2), (0, 5), (1, 3)\}) = \frac{|A|}{|\Omega|} = 3/12 = 1/4$.

This a simple product probability space

Simple K-fold product probability spaces

Suppose that we have a sequence of K set $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_K$.

The (K-fold) cartesian product

The (K-fold) cartesian product is

$$\prod_{i=1}^{K} \mathcal{X}_i = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_K = \{(x_1, x_2, \cdots, x_K) : x_i \in \mathcal{X}_i \text{ for } i = 1, 2, \cdots, K\}$$

When $\mathcal{X}_1 = \mathcal{X}_2 = \cdots = \mathcal{X}_K = \mathcal{Z}$, we write $\mathcal{Z}^K = \prod_{i=1}^{K} \mathcal{X}_i$.

The Cartesian product $\prod_{i=1}^{K} \mathcal{X}_i$ has cardinality $|\prod_{i=1}^{K} \mathcal{X}_i| = |\mathcal{X}_1| \cdot |\mathcal{X}_2| \cdots |\mathcal{X}_K|$.

Example: Suppose that we roll a dice K times.
Then $\mathcal{X}_i = \{1, 2, 3, 4, 5, 6\}$ and $\Omega = \prod_{i=1}^{K} \mathcal{X}_i$. The probability of rolling all ones is $\mathbb{P}(\{1, 1, \cdots, 1\}) = \frac{1}{|\Omega|} = \frac{1}{|\prod_{i=1}^{K} \mathcal{X}_i|} = \frac{1}{6^K}$.

Permutations and Combinations for computing cardinality

Permutations

A permutation is a set of a particular choice of ordering.

Example: if we have a set $\{1, 2, 3\}$, then the different orderings are $(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)$.

There are $k! := k \cdot (k-1) \cdots 2 \cdot 1$ different permutations of a sequence of k objects.

Combinations

Given $k \leq n$, we write $\binom{n}{k}$ for the number of subsets of size k chosen from a set of n.

For example, if we have a set of $\{1, 2, 3\}$, then $\binom{3}{2}$ means the number of subsets of size 2: $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$.

The number $\binom{n}{k}$ is computed as $\binom{n}{k} = \frac{n!}{(n-k)!k!}$.

Permutations and Combinations for computing cardinality

Suppose we have a collection of n balls (with numbers $1, 2, 3, \cdots, n$) in a bag:

Example 5. Sampling with replacement

We draw a ball randomly from the bag, record the number, and then return it to bag. We repeat this $k \leq n$ times. This is called **sampling with replacement**.

Question. Let A be the event that the set of k balls drawn from the bag is exactly $\{1, 2, \cdots, k\}$. We are interested in the probability of A.

Answer.

- Since A consists of permutations of $\{1, \cdots, k\}$, we have $|A| = k!$.
- $\mathcal{X}_i = \{1, 2, \cdots, n\}$ and the sample space has cardinality $|\Omega| = |\prod_{i=1}^k \mathcal{X}_i| = n^k$
- This is a simple probability space, hence the probability of A is $\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{k!}{n^k}$.

Permutations and Combinations for computing cardinality

Suppose we have a collection of n balls (with numbers $1, 2, 3, \cdots, n$) in a bag:

Example 6. Sampling without replacement

We draw a ball randomly from the bag, record the number, and then leave the ball outside the bag. We repeat this $k \leq n$ times. This is called **sampling without replacement**.

Question. Let A be the event that the set of k balls drawn from the bag is exactly $\{1, 2, \cdots, k\}$. We are interested in the probability of A.

Answer.

- The event A consists of permutations of $\{1, \cdots, k\}$, we have $|A| = k!$.
- The sample space Ω consists of all sequences length k such that the entries of the sequence are taken from $\{1, 2, \cdots, n\}$ without repeating.
- We have $\binom{n}{k}$ choices of the subset of size k from the set $\{1, 2, \cdots, n\}$. Each choice is associated with $k!$ different sequences.
- So $|\Omega| = \binom{n}{k} \cdot k!$, and $\mathbb{P}(A) = 1/\binom{n}{k} = \frac{k!(n-k)!}{n!}$.

Permutations and Combinations for computing cardinality

Suppose we have a collection of n balls (with numbers $1, 2, 3, \cdots, n$) in a bag. Among these balls, r balls are in red (with numbers $1, 2, \cdots, k$), and $n - r$ balls are in blue (with numbers $k + 1, \cdots, n$):

Example 7. Combinations for sampling replacement

We draw a ball randomly from the bag, record the colour (but not the numbers), and then return it to the bag. We repeat this $k \leq n$ times (this is sampling with replacement).

Question. Given a number $q \leq k$, we are interested in the probability of q of these k balls being red.

Answer.

The sample space is $\Omega = \{1, 2, \cdots, n\}^k$, which consists of outcomes of the form $(a_1, a_2, \cdots, a_k)$ for $a_i \in \{1, \cdots, n\}$.

Let $A_{q,k}$ denote the event that q of these k balls are red, i.e., with numbers in $(1, 2, \cdots, r)$.

We want to compute $\mathbb{P}(A_{q,k}) = \frac{|A_{q,k}|}{|\Omega|}$. Here $|\Omega| = n^k$.

Permutations and Combinations for computing cardinality

Question. Given a number $q \leq k$, we are interested in the probability of q of these k balls being red.

Answer.

Recall: The sample space is $\Omega = \{1, 2, \cdots, n\}^k$; $A_{q,k}$ denotes the event that q of these k balls are red

Suppose that we have k boxes, and given an outcome of $(a_1, a_2, \cdots, a_k)$, we put the i^{th} ball (with number of a_i) into the i^{th} box. Then each element in $A_{q,k}$ corresponds to one way of having q red balls in the boxes.

To make sure there are q red balls in the boxes, we proceed as follows:

1. We choose q of the k boxes to put in the red balls, therefore having $\binom{k}{q}$ choices.
2. Then we put red balls in the q boxes. Each box can have one of the red balls (one red ball can be in more than 1 box). So we have r^q different ways of doing so.
3. Next we put blue balls in the remaining $k-q$ boxes. Each box can have one of the $n-r$ blue balls, so we have $(n-r)^{(k-q)}$ different ways of doing so.

In total, we have $\binom{k}{q} \cdot r^q \cdot (n-r)^{k-q}$ different ways of having q red balls in the boxes.

Permutations and Combinations for computing cardinality

We draw a ball randomly from the bag, record the colour (but not the numbers), and then return it to the bag. We repeat this $k \leq n$ times (this is sampling with replacement).

Question. Given a number $q \leq k$, we are interested in the probability of q of these k balls being red.

Answer.

In total, we have $\binom{k}{q} \cdot r^q \cdot (n-r)^{k-q}$ different ways of having q red balls in the boxes.

So we have $|A_{q,k}| = \binom{k}{q} \cdot r^q \cdot (n-r)^{k-q}$.

Finally, we have $|\Omega| = n^k$. So $\mathbb{P}(A_{q,k}) = \frac{|A_{q,k}|}{|\Omega|} = \binom{k}{q} \cdot (\frac{r}{n})^q \cdot (\frac{n-r}{n})^{k-q}$.

Countable probability spaces

Similarly to finite probability spaces, we can construct a probability for sample spaces with countably infinite numbers of elements

Countable probability spaces

A countable probability space consists of a triple $(\Omega, \mathcal{E}, \mathbb{P})$ where

1. $\Omega = \{\omega_1, \omega_2, \cdots\}$ is a countably infinite sample space.
2. $\mathcal{E}$ is given by $\{A \subseteq \Omega\}$, i.e., the collection of all subsets of Ω.
3. The probability $\mathbb{P}$ on Ω is constructed in the following way.

Specify a vector $\mathbf{p} = (p_1, p_2, \cdots, p_k, \cdots)$ that satisfying

$$(1).\ p_i \geq 0 \text{ for } i = 1, 2, \cdots. \quad \text{and} \quad (2).\ \sum_{i=1}^{\infty} p_i = 1$$

Define a probability $\mathbb{P}$ based on $\mathbf{p}$ by

$$\mathbb{P}(A) := \sum_{i=1}^{\infty} p_i \cdot \mathbb{1}_A(\omega_i) \text{ for } A \subseteq \Omega$$

What have we covered?

We introduced the concept of finite probability spaces and showed that the associated probability satisfies the laws of probability

We discussed simple probability spaces, which are a special case of finite probability spaces

We learnt how to estimate probability in simple probability space, using combinatorics.

We considered products, permutations and combinations.

Thanks for listening!

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

Statistical Computing and Empirical Methods
Unit EMATM0061, MSc Data Science