Probability with finite sample spaces

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 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 and sample spaces

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

(a) has a well-defined set of possible outcomes;

(b) could (at least in principle) be repeated arbitrarily many times.

Random experiments and sample spaces

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

(a) has a well-defined set of possible outcomes;

(b) could (at least in principle) be repeated arbitrarily many times.

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

Random experiments and sample spaces

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

(a) has a well-defined set of possible outcomes;

(b) could (at least in principle) be repeated arbitrarily many times.

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

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

The laws of probability and their consequences

Let's suppose we have a sample space Ω along with a well-behaved collection of events $\mathcal{E}$ (a σ-algebra), where every $A \in \mathcal{E}$ is a subset $A \subseteq \Omega$.

The laws of probability and their consequences

Let's suppose we have a sample space Ω along with a well-behaved collection of events $\mathcal{E}$ (a σ-algebra), where every $A \in \mathcal{E}$ is a subset $A \subseteq \Omega$.

A probability $\mathbb{P}$ is a function which assigns a number $\mathbb{P}(A)$ each event $A \in \mathcal{E}$ and satisfies the following rules:

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

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

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \ldots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

The laws of probability and their consequences

Let's suppose we have a sample space Ω along with a well-behaved collection of events $\mathcal{E}$ (a σ-algebra), where every $A \in \mathcal{E}$ is a subset $A \subseteq \Omega$.

A probability $\mathbb{P}$ is a function which assigns a number $\mathbb{P}(A)$ each event $A \in \mathcal{E}$ and satisfies the following rules:

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

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

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \dots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

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

The laws of probability and their consequences

Let's suppose we have a sample space Ω along with a well-behaved collection of events $\mathcal{E}$ (a σ-algebra), where every $A \in \mathcal{E}$ is a subset $A \subseteq \Omega$.

A probability $\mathbb{P}$ is a function which assigns a number $\mathbb{P}(A)$ each event $A \in \mathcal{E}$ and satisfies the following rules:

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

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

Rule 3: For every sequence of pairwise disjoint events $A_1, A_2, \dots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

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

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

Consequence 3: If $A \in \mathcal{E}$ is any event, then $0 \leq \mathbb{P}(A) \leq 1$.

Consequence 4: Given any sequence of sets $S_1, S_2, \dots,$ we have $\mathbb{P}(\bigcup_{i=1}^{\infty} S_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(S_i)$.

Finite probability spaces

Let's suppose we have a finite sample space $\Omega = \{\omega_1, \ldots, \omega_k\}$.

The collection of events $\mathcal{E}$ consists of all possible subsets $A \subseteq \Omega$.

Finite probability spaces

Let's suppose we have a finite sample space $\Omega = \{\omega_1, \ldots, \omega_k\}$.

The collection of events $\mathcal{E}$ consists of all possible subsets $A \subseteq \Omega$.

Then we can define a probability $\mathbb{P}$ on Ω by specifying a vector $\mathbf{p} = (p_1, \ldots, p_k)$ such that:

1. $p_i \geq 0$ for each $i = 1, \ldots, k$;
2. $\sum_{i=1}^{k} p_i = 1$.

A vector $\mathbf{p} = (p_1, \ldots, p_k)$ which satisfies (1) & (2) is called a probability vector.

Finite probability spaces

Let's suppose we have a finite sample space $\Omega = \{\omega_1, \ldots, \omega_k\}$.

The collection of events $\mathcal{E}$ consists of all possible subsets $A \subseteq \Omega$.

Then we can define a probability $\mathbb{P}$ on Ω by specifying a vector $\mathbf{p} = (p_1, \ldots, p_k)$ such that:

1. $p_i \geq 0$ for each $i = 1, \ldots, k$;
2. $\sum_{i=1}^{k} p_i = 1$.

A vector $\mathbf{p} = (p_1, \ldots, p_k)$ which satisfies (1) & (2) is called a probability vector.

We can define a probability $\mathbb{P}$ based on our probability vector $\mathbf{p}$ by setting

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

Finite probability spaces

A probability vector $\mathbf{p} = (p_1, \ldots, p_k)$ satisfies (1) $p_i \geq 0$ (2) $\sum_{i=1}^k p_i = 1$.

Given a probability vector $\mathbf{p} = (p_1, \ldots, p_k)$ we define a probability $\mathbb{P}$ by

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

We should check each of the rules are satisfied:

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

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

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

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).$$

Probability with finite sample spaces

Given a probability vector $\mathbf{p} = (p_1, \ldots, p_k)$ we define a probability $\mathbb{P}$ by

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

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

Probability with finite sample spaces

Given a probability vector $\mathbf{p} = (p_1, \ldots, p_k)$ we define a probability $\mathbb{P}$ by

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

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

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

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

as required.

Probability with finite sample spaces

Given a probability vector $\mathbf{p} = (p_1, \ldots, p_k)$ we define a probability $\mathbb{P}$ by

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

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

Probability with finite sample spaces

Given a probability vector $\mathbf{p} = (p_1, \ldots, p_k)$ we define a probability $\mathbb{P}$ by

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

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

Proof: Since $\mathbf{p}$ is a probability vector we have $\sum_{i=1}^{k} p_i = 1$ for all $i = 1, \ldots, k$, so

$$\mathbb{P}(\Omega) := \sum_{i=1}^{k} p_i \cdot \mathbb{1}_\Omega(\omega_i) = \sum_{i=1}^{k} p_i = 1,$$

as required.

Probability with finite sample spaces

Given a probability vector $\mathbf{p} = (p_1, \ldots, p_k)$ we define a probability $\mathbb{P}$ by

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

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

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).$$

Probability with finite sample spaces

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

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).$$

Proof: Write $A = \bigcup_{j=1}^{\infty} A_j$. Since $A_1, A_2, \ldots$ are pairwise disjoint, for each i with $\omega_i \in A$ we must have $\omega_i \in A_{j_i}$ for exactly one $j_i \in \mathbb{N}$. One the other hand, if $\omega_i \notin A$ then $\omega \notin A_j$ for all $j \in \mathbb{N}$.

Hence, we for all $i = 1, \ldots, k$ we have $\mathbb{1}_A(\omega_i) = \sum_{j=1}^{\infty} \mathbb{1}_{A_j}(\omega_i)$.

Probability with finite sample spaces

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

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).$$

Proof: Write $A = \bigcup_{j=1}^{\infty} A_j$. Since $A_1, A_2, \ldots$ are pairwise disjoint, for each i with $\omega_i \in A$ we must have $\omega_i \in A_{j_i}$ for exactly one $j_i \in \mathbb{N}$. One the other hand, if $\omega_i \notin A$ then $\omega \notin A_j$ for all $j \in \mathbb{N}$.

Hence, we for all $i = 1, \ldots, k$ we have $\mathbb{1}_A(\omega_i) = \sum_{j=1}^{\infty} \mathbb{1}_{A_j}(\omega_i)$, so

$$\begin{aligned}\mathbb{P}(A) &= \sum_{i=1}^{k} p_i \cdot \mathbb{1}_A(\omega_i) = \sum_{i=1}^{k} p_i \cdot \left(\sum_{j=1}^{\infty} \mathbb{1}_{A_j}(\omega_i)\right) \\ &= \sum_{j=1}^{\infty} \left(\sum_{i=1}^{k} p_i \cdot \mathbb{1}_{A_j}(\omega_i)\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).\end{aligned}$$

Probability with finite sample spaces

A probability vector $\mathbf{p} = (p_1, \ldots, p_k)$ satisfies (1) $p_i \geq 0$ (2) $\sum_{i=1}^{k} p_i = 1$.

Given a probability vector $\mathbf{p} = (p_1, \ldots, p_k)$ we define a probability $\mathbb{P}$ by

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

We have checked all of the rules are satisfied:

Rule 1: For every event $A \in \mathcal{E}$, we have $\mathbb{P}(A) \geq 0$.

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

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

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).$$

Countable probability spaces

We can construct probabilities for countably infinite spaces in the same way.

Suppose we have a countably infinite sample space $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$.

The collection of events $\mathcal{E}$ consists of all possible subsets $A \subseteq \Omega$.

Then we can define a probability $\mathbb{P}$ on Ω by specifying an infinite probability vector $\mathbf{p} = (p_i)_{i=1}^{\infty}$ such that:

1. $p_i \geq 0$ for each $i = 1, 2, 3, \ldots$;
2. $\sum_{i=1}^{\infty} p_i = 1$.

We can define a probability $\mathbb{P}$ based on our infinite probability vector $\mathbf{p}$ by setting

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

Finite probability spaces

Let's suppose we have a finite sample space $\Omega = \{\omega_1, \ldots, \omega_k\}$.

The collection of events $\mathcal{E}$ consists of all possible subsets $A \subseteq \Omega$.

Then we can define a probability $\mathbb{P}$ on Ω by specifying a vector $\mathbf{p} = (p_1, \ldots, p_k)$ such that:

1. $p_i \geq 0$ for each $i = 1, \ldots, k$;
2. $\sum_{i=1}^{k} p_i = 1$.

A vector $\mathbf{p} = (p_1, \ldots, p_k)$ which satisfies (1) & (2) is called a probability vector.

We can define a probability $\mathbb{P}$ based on our probability vector $\mathbf{p}$ by setting

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

Examples of finite probability spaces

Example 1

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

The sample space is $\Omega = \{0, 1\}$ with events $\mathcal{E} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$;

Let $q \in (0, 1)$ denote the probability that a purchase is made;

Taking probability vector $\mathbf{p}_q = (p_0, p_1) = (1 - q, q)$ defines a probability $\mathbb{P}$ on $\mathcal{E}$,

$$\mathbb{P}(A) := p_0 \cdot \mathbb{1}_A(0) + p_1 \cdot \mathbb{1}_A(1) = \begin{cases} 0 & \text{if } A = \emptyset \\ 1 - q & \text{if } A = \{0\} \\ q & \text{if } A = \{1\} \\ 1 & \text{if } A = \{0, 1\}. \end{cases}$$

Examples of finite probability spaces

Example 2

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

The sample space is $\Omega = \{0, 1\}$ with events $\mathcal{E} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$;

Let $q \in (0, 1)$ denote the probability that the call option expires in the money;

Taking probability vector $\mathbf{p}_q = (p_0, p_1) = (1 - q, q)$ defines a probability $\mathbb{P}$ on $\mathcal{E}$,

$$\mathbb{P}(A) := p_0 \cdot \mathbb{1}_A(0) + p_1 \cdot \mathbb{1}_A(1) = \begin{cases} 0 & \text{if } A = \emptyset \\ 1 - q & \text{if } A = \{0\} \\ q & \text{if } A = \{1\} \\ 1 & \text{if } A = \{0, 1\}. \end{cases}$$

Examples of finite probability spaces

Example 3

A patient either tests positive (1) or negative (0) for a virus;

The sample space is $\Omega = \{0, 1\}$ with events $\mathcal{E} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$;

Let $q \in (0, 1)$ denote the probability that the patient tests positive;

Taking probability vector $\mathbf{p}_q = (p_0, p_1) = (1 - q, q)$ defines a probability $\mathbb{P}$ on $\mathcal{E}$,

$$\mathbb{P}(A) := p_0 \cdot \mathbb{1}_A(0) + p_1 \cdot \mathbb{1}_A(1) = \begin{cases} 0 & \text{if } A = \emptyset \\ 1 - q & \text{if } A = \{0\} \\ q & \text{if } A = \{1\} \\ 1 & \text{if } A = \{0, 1\}. \end{cases}$$

Bernoulli distributions

These are all examples of **Bernoulli** distributions with a binary sample space $\Omega = \{0, 1\}$.

Example 1

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

Example 2

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

Example 3

A patient either tests positive (1) or negative (0) for a virus;

Finite probability spaces with multiple outcomes

Example 4

Bristol FC either loses (0), draws (1) or wins (2) a match;

The sample space is $\Omega = \{0, 1, 2\}$;

The set of events $\mathcal{E} = \{\emptyset, \{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}$;

Taking probability vector $\mathbf{p} = (p_0, p_1, p_2)$ defines a probability $\mathbb{P}$ on $\mathcal{E}$ by

$$\mathbb{P}(A) = \sum_{i \in \{0,1,2\}} p_i \cdot \mathbb{1}_A(i).$$

The number of events in a finite probability space

If our sample space Ω is binary then $\mathcal{E} := \{A \subseteq \Omega\}$ contains $4 = 2^2$ events;

If our sample space Ω contains three elements then $\mathcal{E} := \{A \subseteq \Omega\}$ contains $8 = 2^3$ events;

If our sample space Ω contains four elements then $\mathcal{E} := \{A \subseteq \Omega\}$ contains $16 = 2^4$ events;

General rule

If our sample space Ω contains k elements then $\mathcal{E} := \{A \subseteq \Omega\}$ contains 2^k events.

The collection of all events $\mathcal{E} := \{A \subseteq \Omega\}$ when Ω is of size k is in one-to-one correspondence with the set of binary strings of length k.

Rolling a dice

A classic illustrative example of a finite probability space is rolling a dice.

Example 5

A dice is rolled and we record which side lands "face up".

The sample space is the set of dice rolls $\Omega = \{1, 2, 3, 4, 5, 6\}$;

The set of events $\mathcal{E} = \{A \ : \ A \subseteq \Omega\}$ is of size 2^6;

The probability of an event $A \in \mathcal{E}$ is

$$\mathbb{P}(A) := \frac{|A|}{6},$$

where $|A|$ denotes the cardinality of $|A|$.

Rolling a dice

A classic illustrative example of a finite probability space is rolling a dice.

Example 5

A dice is rolled and we record which side lands "face up".

The sample space is the set of dice rolls $\Omega = \{1, 2, 3, 4, 5, 6\}$;

The set of events $\mathcal{E} = \{A \ : \ A \subseteq \Omega\}$ is of size 2^6;

The probability of an event $A \in \mathcal{E}$ is $\mathbb{P}(A) := \frac{|A|}{6}$,
where $|A|$ denotes the cardinality of $|A|$.

This corresponds to the probability vector $(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$.

Simple probability spaces

Dice rolling is an example of a simple probability space.

A **simple probability space** consists of a triple $(\Omega, \mathcal{E}, \mathbb{P})$ such that:

1. Ω is a finite set (with cardinality $|\Omega|$);
2. $\mathcal{E} := \{A \ : \ A \subseteq \Omega\}$ consists of all finite subsets;
3. Each event $A \in \mathcal{E}$ of cardinality $|A|$ has probability $\mathbb{P}(A) := \frac{|A|}{|\Omega|}$.

Example 5 (cont.)

Dice rolling is a simple probability space with $\Omega = \{1, 2, 3, 4, 5, 6\}$;

Simple probability spaces

A **simple probability space** consists of a triple $(\Omega, \mathcal{E}, \mathbb{P})$ such that:

1. Ω is a finite set (with cardinality $k = |\Omega|$);
2. $\mathcal{E} := \{A \; : \; A \subseteq \Omega\}$ consists of all finite subsets;
3. Each event $A \in \mathcal{E}$ of cardinality $|A|$ has probability $\mathbb{P}(A) := \frac{|A|}{k} = \frac{|A|}{|\Omega|}$.

Write out $\Omega = \{\omega_1, \ldots, \omega_k\}$, so Ω has cardinality $|\Omega| = k$.

This corresponds to finite probability space with probability vector $(p_1, \ldots, p_k) = \left(\frac{1}{k}, \ldots, \frac{1}{k}\right)$.

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

Simple probability spaces and combinatorics

A **simple probability space** consists of a triple $(\Omega, \mathcal{E}, \mathbb{P})$ such that:

1. Ω is a finite set (with cardinality $|\Omega|$);
2. $\mathcal{E} := \{A \ : \ A \subseteq \Omega\}$ consists of all finite subsets;
3. Each event $A \in \mathcal{E}$ of cardinality $|A|$ has probability $\mathbb{P}(A) := \frac{|A|}{|\Omega|}$.

For simple probability spaces, calculating probabilities is just a matter of counting!

Example 5 (cont.)

The probability of rolling an even number is $\mathbb{P}(\{2,4,6\}) = \frac{|\{2,4,6\}|}{6} = \frac{1}{2}$.

Simple product probability spaces

Suppose we have a pair of sets $\mathcal{X}$ and $\mathcal{Y}$.

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

If $\mathcal{X}$ has cardinality $|\mathcal{X}|$ and $\mathcal{Y}$ has cardinality $|\mathcal{Y}|$ then $\mathcal{X} \times \mathcal{Y}$ has cardinality $|\mathcal{X} \times \mathcal{Y}| = |\mathcal{X}| \cdot |\mathcal{Y}|$.

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

Simple product probability spaces

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

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

Example 6

Suppose I flip a coin and record "heads" (1) or "tails" (0) and also throw a dice and record which face lands up;

Take a sample space $\Omega = \mathcal{X} \times \mathcal{Y}$ where $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y} = \{1, 2, 3, 5, 6\}$.

The probability of the coin landing "heads up" and the dice rolling a prime number is

$$\mathbb{P}(\{1\} \times \{2, 3, 5\}) = \frac{|\{1\} \times \{2, 3, 5\}|}{2 \cdot 6} = \frac{|\{(1, 2), (1, 3), (1, 5)\}|}{2 \cdot 6} = \frac{3}{12} = \frac{1}{4}.$$

Simple k-fold product probability spaces

Suppose we have a sequence of K sets $\mathcal{X}_1, \ldots, \mathcal{X}_K$.

The (K-fold) **Cartesian product** is

$$\prod_{k=1}^{K} \mathcal{X}_k = \mathcal{X}_1 \times \ldots \times \mathcal{X}_K = \{(x_1, x_2, \ldots, x_K) \ : \ x_k \in \mathcal{X}_k \text{ for all } k = 1, \ldots, K\}\,.$$

Note that the Cartesian product $\prod_{k=1}^{K} \mathcal{X}_k$ has cardinality $|\prod_{k=1}^{K} \mathcal{X}_k| = |\mathcal{X}_1| \times \ldots \times |\mathcal{X}_K|$.

A **simple product probability space** is a simple probability space $(\Omega, \mathcal{E}, \mathbb{P})$ with $\Omega = \prod_{k=1}^{K} \mathcal{X}_k$.

Simple k-fold product probability spaces

Suppose we have a sequence of K sets $\mathcal{X}_1, \ldots, \mathcal{X}_K$.

The (K-fold) **Cartesian product** is

$$\prod_{k=1}^{K} \mathcal{X}_k = \mathcal{X}_1 \times \ldots \times \mathcal{X}_K = \{(x_1, x_2, \ldots, x_K) \ : \ x_k \in \mathcal{X}_k \text{ for all } k = 1, \ldots, K\}.$$

Note that the Cartesian product $\prod_{k=1}^{K} \mathcal{X}_k$ has cardinality $|\prod_{k=1}^{K} \mathcal{X}_k| = |\mathcal{X}_1| \times \ldots \times |\mathcal{X}_K|$.

A **simple product probability space** is a simple probability space $(\Omega, \mathcal{E}, \mathbb{P})$ with $\Omega = \prod_{k=1}^{K} \mathcal{X}_k$.

Example 7

Suppose we roll K dice. The probability of rolling all ones is $\mathbb{P}(\{(\overbrace{1, 1, \ldots, 1}^{K})\}) = \frac{1}{6^K}$.

Simple k-fold product probability spaces

Suppose we have a sequence of K sets $\mathcal{X}_1, \ldots, \mathcal{X}_K$.

The (K-fold) **Cartesian product** is

$$\prod_{k=1}^{K} \mathcal{X}_k = \mathcal{X}_1 \times \ldots \times \mathcal{X}_K = \{(x_1, x_2, \ldots, x_K) \ : \ x_k \in \mathcal{X}_k \text{ for all } k = 1, \ldots, K\}.$$

Note that the Cartesian product $\prod_{k=1}^{K} \mathcal{X}_k$ has cardinality $|\prod_{k=1}^{K} \mathcal{X}_k| = |\mathcal{X}_1| \times \ldots \times |\mathcal{X}_K|$.

A **simple product probability space** is a simple probability space $(\Omega, \mathcal{E}, \mathbb{P})$ with $\Omega = \prod_{k=1}^{K} \mathcal{X}_k$.

When $\mathcal{X}_1 = \mathcal{X}_2 = \ldots = \mathcal{X}_K = \mathcal{Z}$ are all equal we write $\mathcal{Z}^K$ for $\prod_{k=1}^{K} \mathcal{X}_k$.

Permutations

Permutations from
Yi Ching (~1000BC)

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

There are $k! = k \cdot (k-1) \cdot (k-2) \cdot \ldots \cdot 2 \cdot 1$ possible permutations of a set of size k.

Permutations

Permutations from
Yi Ching (~1000BC)

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

There are $k! = k \cdot (k-1) \cdot (k-2) \cdot \ldots \cdot 2 \cdot 1$ possible permutations of a set of size k.

Example 8

Suppose we have a collection of n balls in a bag with numbers $1, 2, \ldots, n$.

We draw a ball at random from the bag, before replacing it. We repeat this $k \leq n$ times, recording the label of the ball each time. This is called **sampling with replacement**.

Permutations

Permutations from
Yi Ching (~1000BC)

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

There are $k! = k \cdot (k-1) \cdot (k-2) \cdot \ldots \cdot 2 \cdot 1$ possible permutations of a set of size k.

Example 8

Suppose we have a collection of n balls in a bag with numbers $1, 2, \ldots, n$.

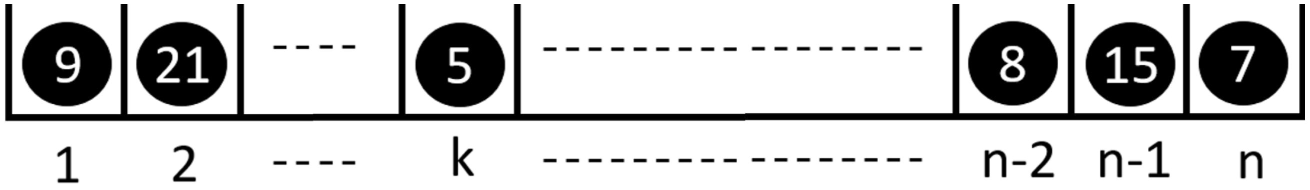
We draw a ball at random from the bag, before replacing it. We repeat this $k \leq n$ times, recording the label of the ball each time. This is called **sampling with replacement**.

Let A be the event that the **set** of k balls drawn from the bag is exactly $\{1, 2, \ldots, k\}$.

A consists of $k!$ permutations of $\{1, \ldots, k\}$, each with probability $\frac{1}{n^k}$, so $\mathbb{P}(A) = \frac{k!}{n^k}$.

Combinations

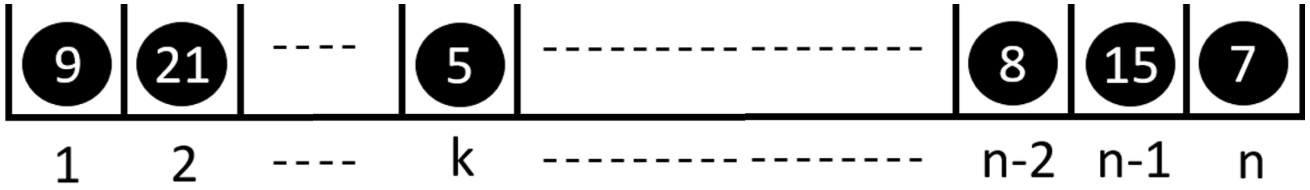
Given $k \leq n$ we write $\binom{n}{k}$ for the number of subsets of size k chosen from a set of size n.



Combinations

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

To compute $\binom{n}{k}$ we consider the n numbered balls placed in n numbered boxes. The selected subset of size k corresponds to the numbers of balls placed in the first k boxes.

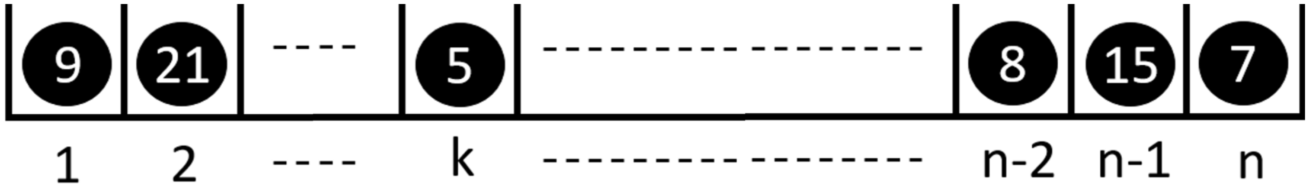


Combinations

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

To compute $\binom{n}{k}$ we consider the n numbered balls placed in n numbered boxes. The selected subset of size k corresponds to the numbers of balls placed in the first k boxes.

1. There are $n!$ permutations of a set S of size n;

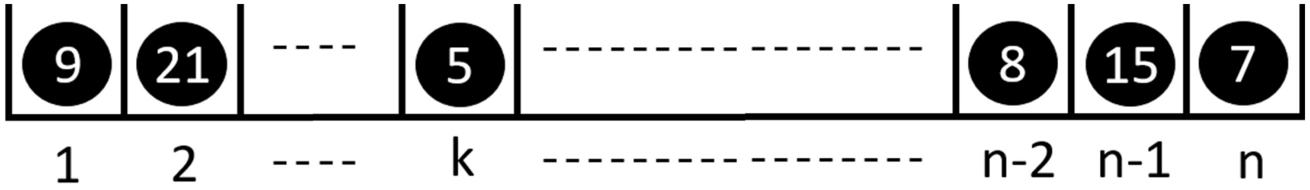


Combinations

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

To compute $\binom{n}{k}$ we consider the n numbered balls placed in n numbered boxes. The selected subset of size k corresponds to the numbers of balls placed in the first k boxes.

1. There are $n!$ permutations of a set S of size n;
2. There are $k!$ permuations of the selected set $A \subseteq S$;

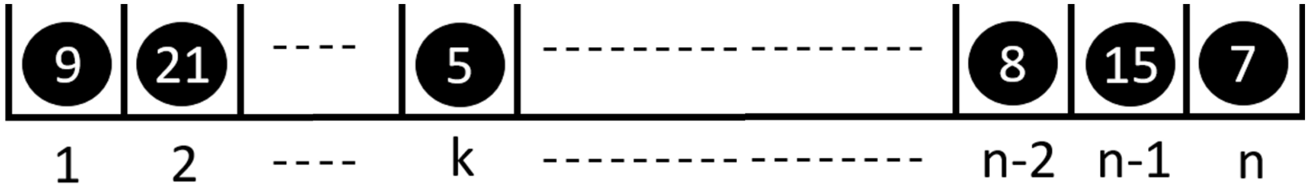


Combinations

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

To compute $\binom{n}{k}$ we consider the n numbered balls placed in n numbered boxes. The selected subset of size k corresponds to the numbers of balls placed in the first k boxes.

1. There are $n!$ permutations of a set S of size n;
2. There are $k!$ permuations of the selected set $A \subseteq S$;
3. There are $(n-k)!$ permutations of the unselected set $S \setminus A$.



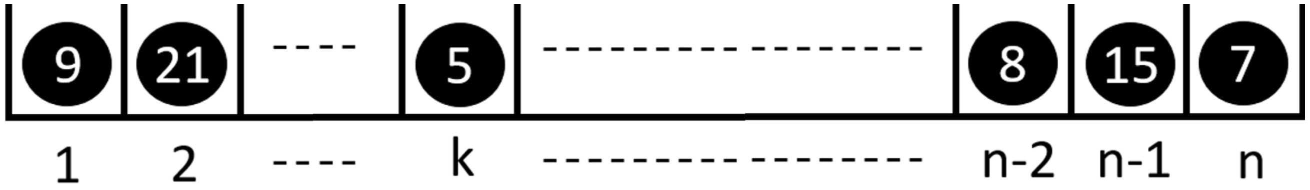
Combinations

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

To compute $\binom{n}{k}$ we consider the n numbered balls placed in n numbered boxes. The selected subset of size k corresponds to the numbers of balls placed in the first k boxes.

1. There are $n!$ permutations of a set S of size n;
2. There are $k!$ permuations of the selected set $A \subseteq S$;
3. There are $(n-k)!$ permutations of the unselected set $S \setminus A$.

Hence, there are $\binom{n}{k} = \frac{n!}{(n-k)!k!}$ subsets of size k from a set of size n.



Combinations for sampling without replacement

Given $k \leq n$ let $\binom{n}{k} = \frac{n!}{(n-k)!k!}$ be the number of subsets of size k chosen from a set of size n.

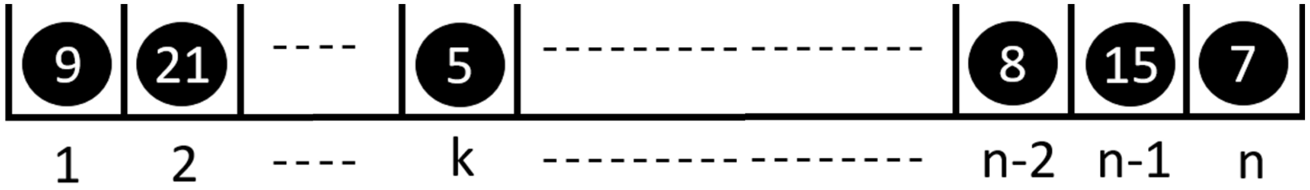
Example 9

Suppose we have a collection of n balls in a bag with numbers $1, 2, \ldots, n$.

We draw a ball at random from the bag, write down the number and leave the ball outside the bag. We repeat this $k \leq n$ times, recording the label of the ball each time. This is called **sampling without replacement**.

Letting A be the event that the **set** of k balls drawn from the bag is exactly $\{1, 2, \ldots, k\}$,

$\mathbb{P}(A) = \frac{1}{\binom{n}{k}} = \frac{k!(n-k)!}{n!}$.



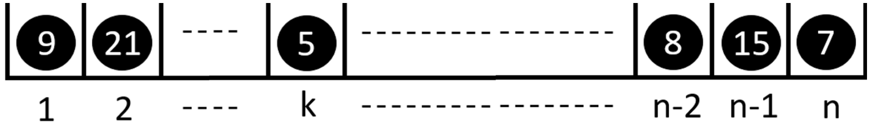
Combinations for sampling with replacement

Example 10

Suppose we have a collection of n balls in a bag. Of those, r are red and $(n-r)$ are blue.

We draw a ball at random from the bag, record its color and then replace it to the bag.
We repeat this process $k \leq n$ times (sampling with replacement each time).

Given $q \leq k$ we are interested in the probability that q of the k balls are red.



Combinations for sampling with replacement

Example 10

Suppose we have a collection of n balls in a bag. Of those, r are red and $(n-r)$ are blue.

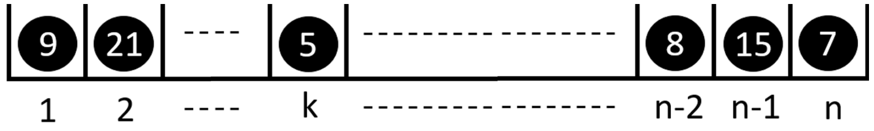
We draw a ball at random from the bag, record its color and then replace it to the bag.
We repeat this process $k \leq n$ times (sampling with replacement each time).

Given $q \leq k$ we are interested in the probability that q of the k balls are red.

Imagine the balls are numbered from $\{1, \ldots, n\}$, with red balls numbered from $\{1, \ldots, r\}$.

Consider the sample space $\Omega = \{1, \ldots, n\}^k$ corresponding to the numbers of the k selected balls.

Let $A_{q,k}$ denote the event that q of the k balls are from $\{1, \ldots, r\}$ i.e. red balls.



Combinations for sampling with replacement

Example 10

Suppose we have a collection of n balls in a bag. Of those, r are red and $(n-r)$ are blue.

We draw a ball at random from the bag, record its color and then replace it to the bag.
We repeat this process $k \leq n$ times (sampling with replacement each time).

Given $q \leq k$ we are interested in the probability that q of the k balls are red.

Imagine the balls are numbered from $\{1, \ldots, n\}$, with red balls numbered from $\{1, \ldots, r\}$.

Consider the sample space $\Omega = \{1, \ldots, n\}^k$ corresponding to the numbers of the k selected balls.

Let $A_{q,k}$ denote the event that q of the k balls are from $\{1, \ldots, r\}$ i.e. red balls.

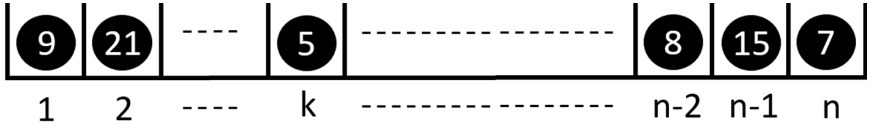
That is $A_{q,k}$ denotes the set of all length k sequences $(a_1, a_2, \ldots, a_k) \in \{1, \ldots, n\}^k$ such that $\mathcal{I} := \{i \in \{1, \ldots, k\} \ : \ a_i \in \{1, \ldots, r\}\}$ is of cardinality $|\mathcal{I}| = q$.

Combinations for sampling with replacement

Let $A_{q,k}$ denote the event that q of the k balls are from $\{1, \ldots, r\}$ i.e. red balls.

That is $A_{q,k}$ denotes the set of all length k sequences $(a_1, a_2, \ldots, a_k) \in \{1, \ldots, n\}^k$ such that $\mathcal{I} := \{i \in \{1, \ldots, k\} \ : \ a_i \in \{1, \ldots, r\}\}$ is of cardinality $|\mathcal{I}| = q$.

To count $A_{q,k}$ we proceed as follows:



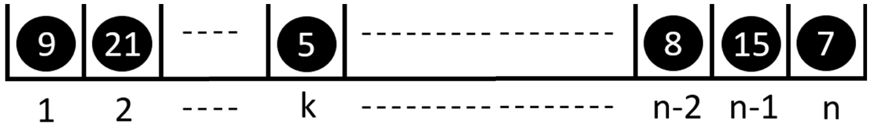
Combinations for sampling with replacement

Let $A_{q,k}$ denote the event that q of the k balls are from $\{1,\ldots,r\}$ i.e. red balls.

That is $A_{q,k}$ denotes the set of all length k sequences $(a_1,a_2,\ldots,a_k) \in \{1,\ldots,n\}^k$ such that $\mathcal{I} := \{i \in \{1,\ldots,k\} \ : \ a_i \in \{1,\ldots,r\}\}$ is of cardinality $|\mathcal{I}| = q$.

To count $A_{q,k}$ we proceed as follows:

1. There are $\binom{k}{q}$ possible subsets of indices $\mathcal{I} \subseteq \{1,\ldots,k\}$ with $|\mathcal{I}| = q$;



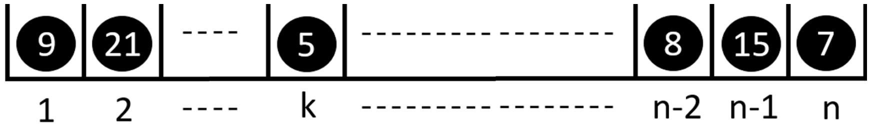
Combinations for sampling with replacement

Let $A_{q,k}$ denote the event that q of the k balls are from $\{1, \ldots, r\}$ i.e. red balls.

That is $A_{q,k}$ denotes the set of all length k sequences $(a_1, a_2, \ldots, a_k) \in \{1, \ldots, n\}^k$ such that $\mathcal{I} := \{i \in \{1, \ldots, k\} \ : \ a_i \in \{1, \ldots, r\}\}$ is of cardinality $|\mathcal{I}| = q$.

To count $A_{q,k}$ we proceed as follows:

1. There are $\binom{k}{q}$ possible subsets of indices $\mathcal{I} \subseteq \{1, \ldots, k\}$ with $|\mathcal{I}| = q$;
2. There are r^q possibilities for the values of $(a_i)_{i \in \mathcal{I}}$ (r for each $i \in \mathcal{I}$);



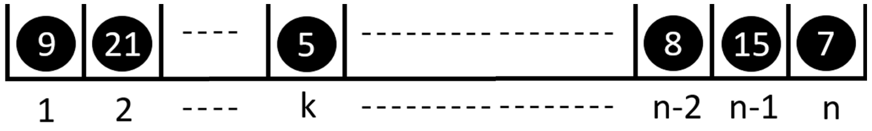
Combinations for sampling with replacement

Let $A_{q,k}$ denote the event that q of the k balls are from $\{1, \ldots, r\}$ i.e. red balls.

That is $A_{q,k}$ denotes the set of all length k sequences $(a_1, a_2, \ldots, a_k) \in \{1, \ldots, n\}^k$ such that $\mathcal{I} := \{i \in \{1, \ldots, k\} \ : \ a_i \in \{1, \ldots, r\}\}$ is of cardinality $|\mathcal{I}| = q$.

To count $A_{q,k}$ we proceed as follows:

1. There are $\binom{k}{q}$ possible subsets of indices $\mathcal{I} \subseteq \{1, \ldots, k\}$ with $|\mathcal{I}| = q$;
2. There are r^q possibilities for the values of $(a_i)_{i \in \mathcal{I}}$ (r for each $i \in \mathcal{I}$);
3. There are $(n-r)^{k-q}$ possibilities for the values of $(a_i)_{i \notin \mathcal{I}}$ ($(n-r)$ for each $i \notin \mathcal{I}$);



Combinations for sampling with replacement

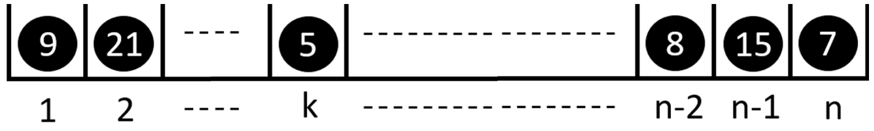
Let $A_{q,k}$ denote the event that q of the k balls are from $\{1, \ldots, r\}$ i.e. red balls.

That is $A_{q,k}$ denotes the set of all length k sequences $(a_1, a_2, \ldots, a_k) \in \{1, \ldots, n\}^k$ such that $\mathcal{I} := \{i \in \{1, \ldots, k\} \ : \ a_i \in \{1, \ldots, r\}\}$ is of cardinality $|\mathcal{I}| = q$.

To count $A_{q,k}$ we proceed as follows:

1. There are $\binom{k}{q}$ possible subsets of indices $\mathcal{I} \subseteq \{1, \ldots, k\}$ with $|\mathcal{I}| = q$;
2. There are r^q possibilities for the values of $(a_i)_{i \in \mathcal{I}}$ (r for each $i \in \mathcal{I}$);
3. There are $(n-r)^{k-q}$ possibilities for the values of $(a_i)_{i \notin \mathcal{I}}$ ($(n-r)$ for each $i \notin \mathcal{I}$);

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



Combinations for sampling with replacement

Example 10

Suppose we have a collection of n balls in a bag. Of those, r are red and $(n-r)$ are blue.

We draw a ball at random from the bag, record its color and then replace it to the bag.
We repeat this process $k \leq n$ times (sampling with replacement each time).

Given $q \leq k$ we are interested in the probability that q of the k balls are red.

Imagine the balls are numbered from $\{1, \ldots, n\}$, with red balls numbered from $\{1, \ldots, r\}$.

Consider the sample space $\Omega = \{1, \ldots, n\}^k$ corresponding to the numbers of the k selected balls.

Let $A_{q,k}$ denote the event that q of the k balls are from $\{1, \ldots, r\}$ i.e. red balls.

Since $|A_{q,k}| = \binom{k}{q} \cdot r^q \cdot (n-r)^{k-q}$ and $\mathbb{P}(A_{q,k}) = \binom{k}{q} \left(\frac{r}{n}\right)^q \cdot \left(1 - \frac{r}{n}\right)^{k-q}$.

Capture and recapture

Example 11

Suppose there are n squirrels living on an island.

A conservationist captures t squirrels at random before tagging and releasing them.

A week later, the conservationist captures k squirrels at random again, and counts how many have already been tagged.

For simplicity assume that the population of n squirrels is constant over the time period, and on both occasions the squirrels are selected purely at random.

What is the probability that $q \leq \min\{t, k\}$ of the tagged squirrels are recaptured?

Capture and recapture

Example 11

Suppose there are n squirrels living on an island.

A conservationist captures t squirrels at random before tagging and releasing them.

A week later, the conservationist captures k squirrels at random again, and counts how many have already been tagged.

For simplicity assume that the population of n squirrels is constant over the time period, and on both occasions the squirrels are selected purely at random.

What is the probability that $q \leq \min\{t, k\}$ of the tagged squirrels are recaptured?

To compute the probability let's imagine the n squirrels are assigned numbers $1, \ldots, n$, with the first $t \leq n$ squirrels corresponding to those which were tagged in the first round.

Our sample space Ω corresponds to all subsets $S \subseteq \{1, \ldots, n\}$ of size k, corresponding to the set of squirrels recaptured in the second round.

Capture and recapture

Our sample space Ω corresponds to all subsets $S \subseteq \{1, \ldots, n\}$ of size k, corresponding to the set of squirrels recaptured in the second round.

The event $A_q := \{S \in \Omega \ : \ |S \cap \{1, \ldots, t\}| = q\}$ corresponds to q of the squirrels in the second round being tagged.

Capture and recapture

Our sample space Ω corresponds to all subsets $S \subseteq \{1, \ldots, n\}$ of size k, corresponding to the set of squirrels recaptured in the second round.

The event $A_q := \{S \in \Omega \ : \ |S \cap \{1, \ldots, t\}| = q\}$ corresponds to q of the squirrels in the second round being tagged.

1. There $\binom{n}{k}$ subsets of size k, so $|\Omega| = \binom{n}{k}$;
2. There are $\binom{t}{q}$ possibilities for $S \cap \{1, \ldots, t\}$ with $|S \cap \{1, \ldots, t\}| = q$;
3. There are $\binom{n-t}{k-q}$ possibilities for $S \cap \{t+1, \ldots, n\}$ with $|S \cap \{t+1, \ldots, n\}| = k - q$.

Capture and recapture

Our sample space Ω corresponds to all subsets $S \subseteq \{1, \ldots, n\}$ of size k, corresponding to the set of squirrels recaptured in the second round.

The event $A_q := \{S \in \Omega \ : \ |S \cap \{1, \ldots, t\}| = q\}$ corresponds to q of the squirrels in the second round being tagged.

1. There $\binom{n}{k}$ subsets of size k, so $|\Omega| = \binom{n}{k}$;
2. There are $\binom{t}{q}$ possibilities for $S \cap \{1, \ldots, t\}$ with $|S \cap \{1, \ldots, t\}| = q$;
3. There are $\binom{n-t}{k-q}$ possibilities for $S \cap \{t+1, \ldots, n\}$ with $|S \cap \{t+1, \ldots, n\}| = k - q$.

Hence, $|A_q| = \binom{t}{q} \cdot \binom{n-t}{k-q}$, so $\mathbb{P}(A_q) = \frac{|A_q|}{|\Omega|} = \frac{\binom{t}{q} \cdot \binom{n-t}{k-q}}{\binom{n}{k}}$.

Capture and recapture

Suppose there are n squirrels living on an island.

A conservationist captures t squirrels at random before tagging and releasing them.

A week later, the conservationist captures k squirrels at random again, and counts how many have already been tagged.

For simplicity assume that the population of n squirrels is constant over the time period, and on both occasions the squirrels are selected purely at random.

The probability that $q \leq \min\{t, k\}$ of the tagged squirrels are recaptured is $\frac{|A_q|}{|\Omega|} = \frac{\binom{t}{q} \cdot \binom{n-t}{k-q}}{\binom{n}{k}}$.

What have we covered?

- 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.
- We will discussed several examples including the capture-recapture model.

University of BRISTOL

Thanks for listening!

Henry W J Reeve

henry.reeve@bristol.ac.uk

Include EMATM0061 in the subject of your email.