Foundations of Probability

Random experiments

Sample spaces

Venn diagrams of random events
Random Experiments

• An experiment is an operation or procedure, carried out under controlled conditions
  – Example: measure the metabolic flux through a reaction catalyzed by enzyme pA

• An experiment that can result in different outcomes, even if repeated in the same manner every time, is called a random experiment
  – Cell-to-cell variability due to history/genome variants
  – Noise in external parameters such as temperature, nutrients, pH, etc.

• Evolution offers ready-made random experiments
  – Genomes of different species
  – Genomes of different individuals within a species
  – Individual cancer cells
Variability/Noise Produce Output Variation

- Controlled variables
  - e.g. Temperature, Nutrients, pH

- Input
  - What I want to change in the experiment, e.g. expression level of a gene A
  - Internal state of individual cells, Signals from neighbors

- Biological system: cell/organism/population

- Noise variables

- Output
  - What I measure in the experiment, e.g. metabolic flux catalyzed by the enzyme encoded by the gene A
Sample Spaces

• Random experiments have unique outcomes.
• The set of all possible outcomes of a random experiment is called the sample space, $S$.
• $S$ is discrete if it consists of a finite or countable infinite set of outcomes.
• $S$ is continuous if it contains an interval (either a finite or infinite width) of real numbers.
Examples of a Sample Space

• Experiment measuring the abundance of mRNA expressed from a single gene
  \( S = \{x | x > 0\} \): continuous.

• Bin it into four groups
  \( S = \{\text{below 10, 10-30, 30-100, above 100}\} \): discrete.

• Is gene “on” or “off”?\n  \( S = \{\text{true, false}\} \): logical/Boolean/discrete.
Event

An event \((E)\) is a subset of the sample space of a random experiment, i.e., one or more outcomes of the sample space.

- The **union** of two events is the event that consists of all outcomes that are contained in either of the two events. We denote the union as \(E_1 \cup E_2\).

- The **intersection** of two events is the event that consists of all outcomes that are contained in both of the two events. We denote the intersection as \(E_1 \cap E_2\).

- The **complement** of an event in a sample space is the set of outcomes in the sample space that are not in the event. We denote the complement of the event \(E\) as \(E'\) (sometimes \(E^c\) or \(\bar{E}\)).
Examples

Discrete
1. Assume you toss a coin once. The sample space is $S = \{H, T\}$, where $H =$ head and $T =$ tail and the event of a head is $\{H\}$.

2. Assume you toss a coin twice. The sample space is $S = \{(H, H), (H, T), (T, H), (T, T)\}$, and the event of obtaining exactly one head is $\{(H, T), (T, H)\}$.

Continuous

Sample space for the expression level of a gene: $S = \{x|x \geq 0\}$

Two events:
- $E_1 = \{x|10 < x < 100\}$
- $E_2 = \{x|30 < x < 300\}$

- $E_1 \cap E_2 = \{x|30 < x < 100\}$
- $E_1 \cup E_2 = \{x|10 < x < 300\}$
- $E_1' = \{x|x \leq 10 \text{ or } x \geq 100\}$
Venn diagrams

John Venn (1843-1923)
British logician

John Venn (1990-)
Brooklyn hipster

Find 5 differences in beard and hairstyle
Venn diagrams

Which formula describes the blue region?
A. $A \cup B$
B. $A \cap B$
C. $A'$
D. $B'$

Get your i-clickers
Venn diagrams

Which formula describes the blue region?
A. $A \cup B$
B. $A \cap B$
C. $A'$
D. $B'$
Venn diagrams

Which formula describes the blue region?
A. \((A \cup B) \cap C\)
B. \((A \cap B) \cap C\)
C. \((A \cup B) \cup C\)
D. \((A \cap B) \cup C\)

Get your i-clickers
Which formula describes the blue region?

A. $(A \cup B) \cap C$
B. $(A \cap B) \cap C$
C. $(A \cup B) \cup C$
D. $(A \cap B) \cup C$
Venn diagrams

Which formula describes the blue region?
A. $A \cap C$
B. $A' \cup C'$
C. $(A \cap B \cap C)'$
D. $(A \cap B) \cap C$

Get your i-clickers
Venn diagrams

Which formula describes the blue region?

A. $A \cap C$
B. $A' \cup C'$
C. $(A \cap B \cap C)'$
D. $(A \cap B) \cap C$

B. $A' \cup C'$
Definitions of Probability
Two definitions of probability

• (1) **STATISTICAL PROBABILITY**: the relative frequency with which an event occurs in the long run

• (2) **INDUCTIVE PROBABILITY**: the degree of belief which it is reasonable to place in a proposition on given evidence

Bulmer, M. G.. Principles of Statistics (Dover Books on Mathematics)
A statistical probability of an event is the limiting value of the relative frequency with it occurs in a very large number of independent trials.

Empirical
Statistical Probability of a Coin Toss

N(Heads out of T tosses) - N(Tails out of T tosses)

Excess of heads among 2,000 coin tosses (Kerrich 1946)
Statistical Probability of a Coin Toss

Probability(Heads) = \frac{N(\text{Heads out of T tosses})}{T}\text{ limit for large T}

Proportion of heads among 10,000 coin tosses (Kerrich 1946)
Credit: XKCD
questions
found in google autocomplete

Why do whales jump
why are witches green
why are there mirrors above beds
why do i say uh
why is sea salt better
why are there trees in the middle of fields
why is there not a pokemon mmo
why is there laughing in tv shows
why are there doors on the freeway
why are there so many schooled e. running
why aren't there any countries in antarctica
why are there scary sounds in minecraft
why is there kinking in my stomach
why are there two slashes after http
why are there celebrities
why do snakes exist
why do oysters have pearls
why are ducks called ducks
why do they call it the clap
why are kyle and cartman friends
why is there an arrow on hang's head
why are there text messages blue
why are there mustaches on clothes
why are there mustaches on cars
why are there mustaches everywhere
why are there so many birds in ohio
why is there so much rain in ohio
why is ohio weather so weird
why are there male and female bikes
why are there tiny spiders in my house
why do spiders come inside
why are there huge spiders in my house
why are there lots of spiders in my house
why are there spiders in my room
why are there so many spiders in my room
why do spider bites itch
why is dying so scary

Why is mt. vesuvius there
why do they say t minus
why are there ocelots
why are wrestlers always wet
why are oceans becoming more acidic
why is arwen dying
why aren't my quail laying eggs
why aren't my quail eggs hatching
why aren't there any foreign military bases in america
why are there spiders on my phone
why is there a lion on my head
why is there a red line through https
why is there a red line through https on facebook
why is https important
why aren't my arms growing
why are there so many rows in rochester, mn
why is there, philip
why are there sharks of canada
why is there, philip
why do children get cancer
why is poseidon angry with odysseus
why is there ice in space
why is there an owl in my backyard
why is there an owl outside my window
why is there an owl on the dollar bill
why do owls attack people
why are ak-47's so expensive
why are there helicopters circling my house
why are there gods
why are there two spocks
why are cigarettes legal
why are there ducks in my pool
why is jesus white
why is there, liquid in my ear
why do cats feel good
why do good people die
why aren't there guns in harry potter
why are ultrasounds important
why are laser guns made of diamond
why is stealing wrong
why aren't there any foreign military bases in america
Two definitions of probability

• (1) **STATISTICAL PROBABILITY:** the relative frequency with which an event occurs in the long run

• (2) **INDUCTIVE PROBABILITY:** the degree of belief which it is reasonable to place in a proposition on given evidence

Bulmer, M. G.. Principles of Statistics (Dover Books on Mathematics)
Inductive Probability

An inductive probability of an event the degree of belief which it is rational to place in a hypothesis or proposition on given evidence.
**Principle of indifference**

- **Principle of Indifference** states that two events are equally probable if we have no reason to suppose that one of them will happen rather than the other. (Laplace, 1814)

- Unbiased coin:
  - probability Heads = probability Tails = $\frac{1}{2}$

- Symmetric die:
  - probability of each side = $\frac{1}{6}$

---

Pierre-Simon, marquis de Laplace (1749 –1827)
French mathematician, physicist, astronomer
Inductive probability can lead to trouble

• Glass contains a mixture of wine and water
• We know: proportion of water to wine can be anywhere between 1:1 and 2:1
• We can argue that the proportion of water to wine is equally likely to lie between 1 and 1.5 as between 1.5 and 2.
• Consider now the ratio of wine to water. This quantity must lie between 0.5 and 1, and we can use the same argument to show that it is equally likely to lie between 1/2 and 3/4 as it is to lie between 3/4 and 1.
• But this means that the water to wine ratio is equally likely to lie between 1 and 4/3=1.333... as it is to lie between 1.333.. and 2
• This is clearly inconsistent with the previous calculation

Bertrand’s paradox
Inductive probability relies on combinatorics or the art of counting combinations.
Counting – Multiplication Rule

• Multiplication rule:
  – Let an operation consist of k steps and
    • $n_1$ ways of completing step 1,
    • $n_2$ ways of completing step 2, ... and
    • $n_k$ ways of completing step $k$.
  – Then, the total number of ways or outcomes are:
    • $n_1 \times n_2 \times \ldots \times n_k$

• Example:
  – $S = \{A, C, G, T\}$ the set of 4 DNA bases
  – Number of k-mers is $4^k = 4 \times 4 \times 4 \times \ldots \times 4$ (k –times)
    Important example: 64 triplets in the genetic code
  – A protein-coding part of the gene is typically 1000 bases long
    There are $4^{1000} \approx 2^{2000} \approx 10^{600}$ possible sequences of
    just one gene. Or $(10^{600})^{25,000} = 10^{15,000,000}$ of 25,000 human
    genes.
    For comparison, the Universe has between $10^{78}$ and $10^{80}$ atoms
    and is $4 \times 10^{17}$ seconds old.
Counting – Permutation Rule

• A permutation is a unique sequence of distinct items.
• If $S = \{a, b, c\}$, then there are 6 permutations
  – Namely: $abc, acb, bac, bca, cab, cba$ (order matters)
• # of permutations for a set of $n$ items is $n!$
• $n!$ (factorial function) = $n*(n-1)*(n-2)*...*2*1$
• $7! = 7*6*5*4*3*2*1 = 5,040$
• By definition: $0! = 1$
Counting - Similar Item Permutations

• Used for counting the sequences when not all the items are different.

• The number of permutations of:
  – \( n = n_1 + n_2 + \ldots + n_r \) items of which
    • \( n_1 \) are identical,
    • \( n_2 \) are identical, \ldots, and
    • \( n_r \) are identical.

• Is calculated as: \( \frac{n!}{n_1!n_2!\ldots n_r!} \)