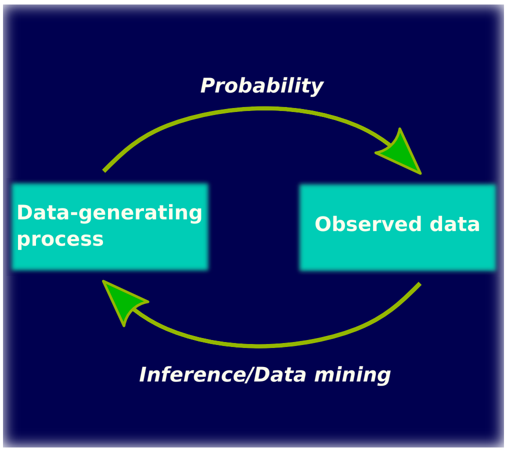
Statistics for Astronomers: Lecture 01, 2019.02.07

Prof. Sundar Srinivasan

IRyA/UNAM

Overview

Probability Theory and Statistical Inference[1]



- Given a data-generating process, what are the properties of the outcomes?
  Probability theory.
- Given the outcome of an observation/experiment, what can be said about the process(es) that generated the data? Statistical inference.

Why Must Astronomers Care About Probability?[2]

Probability quantifies uncertainty

"Uncertain knowledge + Knowledge of the amount of uncertainty in it = Usable knowledge."
– C. R. Rao (1997)

- Astronomers have little to no control over their "test subjects".
- Experiments observe events with uncertain outcomes, and are affected by random error as well as uncertainty due to small samples.
- Typically either costly to repeat observations, or there are physical/instrumental constraints (e.g., resolution, weather).
- Can provide quantitative information about whether our measured interval captures the "true" value of a quantity. Can also quantitatively compare properties of two different samples.
- Uncertainty due to ignorance of the underlying physical processes could be reduced using past experience (e.g., "updating priors").
- Predicting exact behaviour of individual members of a class difficult, but easier to describe the class as a whole with some confidence.
  e.g.: stellar populations, globular clusters, kinetic theory of gases, radioactive decay, mortality, prevalence of a disease, the science of psychohistory in the **Foundation** series by Isaac Asimov.

Why Must Astronomers Care About Statistics?[3]

A relevant definition of statistics

"[T]he theory and methods of collecting, organizing, presenting, analyzing, and interpreting data sets so as to determine their essential characteristics."

– Panik (2005)

- Statistics: what to compute.
  Informatics: how to compute.
- Data-driven vs. hypothesis-driven science.
- Most data will never be seen by humans (already the case).
- Patterns in (multidimensional) data cannot be comprehended by humans directly.
- **Why you should care:** Jobs! Jobs! Jobs! Data science! Big data! LSST! SKA! Flatiron Institute!

Review

Review: Sets

Universal set U, Ω
Null set $\varnothing$
Operations on sets: Complement, "Addition", "Difference", Union, Intersection.

Review: Sets (contd.)

For the following, let $U =$ the set of natural numbers, $A = \{1, 2, 3, 4\}$, and $B = \{2, 4, 5, 6\}$:

- Complement: A^c or $\neg A = \{x \mid x \notin A\}$
  e.g., $A^c = \{5,\ 6,\ 7,\ 8,\ ...\}$ (all natural numbers except 1, 2, 3, and 4).
- Addition ("inclusive OR"): $A + B = \{x \mid x \in A \ \vee \ x \in B\}$
  e.g., $A + B = \{1, 2, 2, 3, 4, 4, 5, 6\}$
- Union ("exclusive OR"): $A \cup B = \{x \mid x \in A \ \veebar \ x \in B\}$
  e.g., $A \cup B = \{1, 2, 3, 4, 5, 6\}$
- Difference: $A - B = \{x \mid x \in A \ \wedge \ x \notin B\}$
  e.g., $A - B = \{1, 3\}$
- Intersection: $A \cap B = \{x \mid x \in A \ \wedge \ x \in B\}$
  e.g., $A \cap B = \{2, 4\}$

Review: Sets (contd.)

The four binary relations defined above are related as follows (verify using a Venn diagram):

$$A \cup B = A + B - A \cap B \text{ (to avoid double-counting)}$$

This will be useful later when computing probabilities.

Review: Combinatorics

Combinatorics deals with the number of ways of choosing k objects from a total of n.
Notation: we represent by the symbol $\binom{n}{k}$ the quantity

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

where $k!$ is the factorial of k.

Review: Combinatorics[4] (contd.)

Choosing k objects from a total of n. The number of ways this can be done depends on whether **replacement is allowed** (*e.g.*, an object selected from the pile is replaced before the next selection), and on whether the **order in which the objects are selected matters**.

	Order doesn't matter	Order matters
Without replacement ($k \leq n$)	$\binom{n}{k}$	$\binom{n}{k} \times k!$ (First select k objects from n, then arrange the k objects.)
With replacement ($k > n$ allowed)	$\binom{n+k-1}{k}$ (Select and replace same object k times = $k-1$ replacements = $n+k-1$ "objects.")	n^k

Review: Binomial Coefficients

The quantity $\binom{n}{k}$ is called a binomial coefficient because it features in the expansion of the binomial:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

The number of ways of forming a term of the type $x^k y^{n-k}$ is exactly $\binom{n}{k}$.

E.g., number of ways of forming the term $x^1 y^3 = \binom{4}{1} = \binom{4}{3} = 4$.

Review: Binomial Coefficients (contd.)

Some properties of the binomial coefficients:

- Choosing k objects from n is the same as **rejecting** $n-k$ objects from n:

$$\binom{n}{k} = \binom{n}{n-k}$$

- The coefficients for a total of n objects sum to 2^n (the number of ways of choosing **at least one object** from n objects is $2^n - 1$):

$$(1+1)^n = \sum_{k=0}^{n} \binom{n}{k} = 2^n$$

- The term $n-k = (n-s)+(k-s)$ for any s:

$$\frac{(k+s)!}{(n+s)!} \times \binom{n+s}{k+s} = \frac{k!}{n!} \times \binom{n}{k}$$

- Simplification for $s=1$:

$$\frac{(k+1)}{(n+1)} \times \binom{n+1}{k+1} = \binom{n}{k}$$

Probability

Some definitions

Outcome: One of many possible results of an experiment/observation.
Symbol: ω.

Sample space: Set of all possible outcomes. Also called outcome space.
Symbol: Ω.

Event: A subset of the sample space Ω, consisting of zero or more outcomes ω.

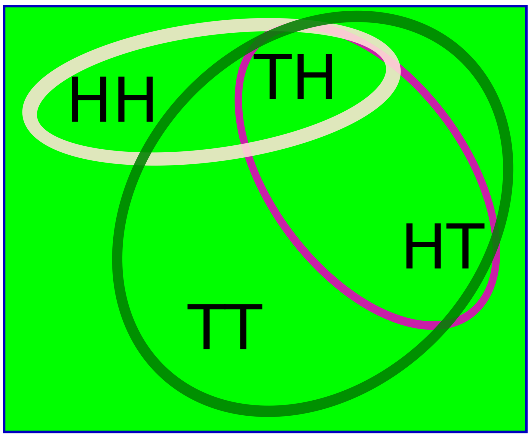
Event space: A collection of events in the sample space.
Needs to satisfy some properties ($\sigma-$algebra): must be closed over complement and closed over countable unions (see Q1 on Homework #1).
Symbol: $\mathscr{F}$.

Warning: some sources refer to Ω as the event space.

We usually talk about probabilities of events (which are a more general construct than outcomes).

Events and Event Spaces

Experiment: Two-coin toss.
Sample space: Ω ={(TT), (TH), (HT), (HH)}.



Consider the following events:

1. E_1 = At least one tail = {(TT), (TH), (HT)},
2. E_2 = Exactly one head = {(TH), (HT)}, and
3. E_3 = Second coin ends up heads = {(TH), (HH)}.

Are the collections $\mathscr{F}_1 = \{E_1, E_3\}$ and $\mathscr{F}_2 = \{E_1, E_2, E_3\}$ event spaces? See Q1 in Homework #1 (solutions here).

Estimating probability: an exercise

Example: In a two-coin toss, what is the probability of obtaining two tails?

Why?

Two assumptions:

1. All outcomes are equally likely, and
2. The sum of probabilities of all outcomes is 1.