Statistics for Astronomers: Lecture 02, 2019.02.11

Prof. Sundar Srinivasan

IRyA/UNAM

Recall: 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.

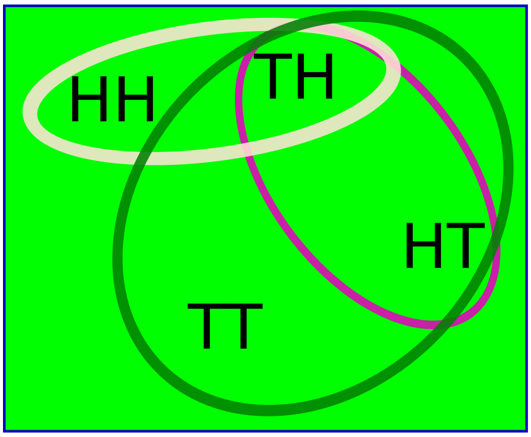
Event space: A collection of events in the sample space.
Needs to satisfy some properties ($\sigma-$algebra): must be closed over complement and closed over countable unions (see Q1 on Homework #1).
Symbol: $\mathscr{F}$.

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).

Recall: 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).

Recall: Estimating probability: an exercise

Example: In a two-coin toss, what is the probability of obtaining two tails?
Your answer was: 1/4.

Why?

Two implicit assumptions:

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

Computing probability: three ways

- **Classical interpretation**: can be computed if all outcomes are assumed equally likely.

  We just used this interpretation to compute the probability in the previous slide.

- **Frequentist interpretation**: perform an infinite sequence of experiments, find relative frequency of favourable outcomes.

  Toss the two coins N times ($N \gg 1$) and count the number of times M that we get two tails. The required probability is then $\frac{M}{N}$.

- **Bayesian interpretation**: use prior knowledge of the parameters of the problem, perform experiments, and update the priors to get posterior probabilities.

  Select your priors (*e.g.*, are the coins known to be fair from experience? "Roberto performed the experiment 500 times yesterday and only got two tails 20 times!"). Perform an experiment. Combine the resulting outcome with the prior and predict the probability of getting two tails on future trials.

Classical (naïve) interpretation of probability

"The probability of an event is the ratio of the number of cases favorable to it, to the number of all cases possible..."
– Laplace (1812).

Principle of Indifference

If N events are mutually exclusive and collectively exhaustive,

(1) they are equally likely and (2) the probability of any one occurring is $\frac{1}{N}$.

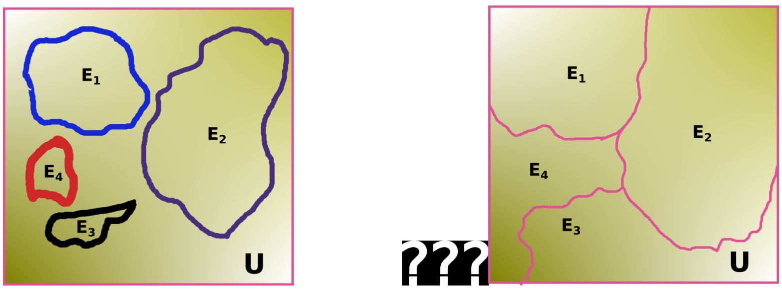


Figure: Case (1): E1, E2, E3, and E4 are mutually exclusive but not collectively exhaustive. Case (2): the events are now also collectively exhaustive.

Classical (naïve) interpretation of probability

"The probability of an event is the ratio of the number of cases favorable to it, to the number of all cases possible..."
– Laplace (1812).

Principle of Indifference

If N events are mutually exclusive and collectively exhaustive,

(1) they are equally likely and (2) the probability of any one occurring is $\frac{1}{N}$.

Two-coin toss example

4 mutually exclusive and collectively exhaustive outcomes/events:
$\omega_1 = (TT)$, $\omega_2 = (TH)$, $\omega_3 = (HT)$, and $\omega_4 = (HH)$

$\Rightarrow P(\omega_i) = 1/4$ for $i = 1, 2, 3, 4$.

Classical (naïve) interpretation of probability (contd.)

Criticism of the Principle of Indifference: not all sample spaces consist of equally likely events.

Excerpt from a Daily Show segment in which John Oliver interviews a male high-school mathematics teacher about the LHC:

John: What is the probability that the Large Hadron Collider destroys the Universe?
Teacher: 50%
John: How do you figure?
Teacher: It will either destroy the Universe, or it won't.

John: Do you know how probability works?!

Probability: formal definition

Probability P is a function on a sample space Ω such that, for any event $E \in \Omega$, $P(E) \in \mathbb{R}$.

A probability space consists of the combination $(\Omega, \mathscr{F}, P)$. (Recall: $\mathscr{F}$ is the event space, the set of all possible events in Ω).

This probability must satisfy some rules, given by the axioms of probability.

Axioms of Probability (Kolmogorov, 1933)

1. The probability that an event has occurred is always a non-negative real number.
   $\forall E \in \mathscr{F}, P(E) \in \mathbb{R}$ and $P(E) \geq 0$
   In particular, $P(\varnothing) = 0$. (At least one event in Ω occurs.)
2. **Unitarity:** The probability that at least one event in the sample space will occur is unity.
   $P(\Omega) = 1$
3. **Countable additivity:** The probability that at least one event among a set of (pairwise) disjoint events occurs is the sum of the probabilities of each of those events occurring.
   Given A_j $(j = 1, \ldots)$ such that $A_j \cap A_k = \varnothing$ for $j \neq k$,
   $$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j), \text{ if } A_j \cap A_k = \varnothing \text{ for } j \neq k$$

Some properties derived from the axioms

- $P(A^c) = 1 - P(A)$
- $A \subseteq B \Rightarrow P(A) \leq P(B)$
- $A \subseteq B \Rightarrow P(B) = P(B \cap A) + P(B \cap A^c)$
- Recall: $A \cup B = A + B - A \cap B$. This relation can be used to derive the "Inclusion-Exclusion Principle": $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Inclusion-Exclusion Principle (generalised):
$$P\left(\bigcup_{j=0}^{n} A_j\right) = \sum_{j=0}^{n} P(A_j) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \ldots$$
$$+(-1)^{n+1} P\left(\bigcap_{j=1}^{n} A_j\right)$$

The Pigeonhole Principle

"There exists no injective function whose codomain is smaller than its domain."

You cannot arrange N pigeons into $M < N$ pigeonholes without at least one pigeonhole ending up with more than one pigeon.

$N = 10$ pigeons in $M = 9$ pigeonholes.
Credit: en:User:BenFrantzDale/en:User:McKay CC BY-SA 3.0, via Wikimedia Commons.

Consequences:
In a group of 13 people, at least two must be born in the same month.
In a group of 367 people, at least two must share the same birthday.

The Birthday Problem/Paradox

Given N people, what is the probability that two people share a birthday?

Due to the Pigeonhole Principle, we know that $\mathrm{P}(N > 366) = 1$. What is N for $\mathrm{P}(N) = 0.5$?

Any two people could share a birthday in this group.
The number of ways of choosing 2 people from $N = \binom{N}{2} = N(N-1)/2$.
For $N = 20$, this is already $190 > 366/2$.

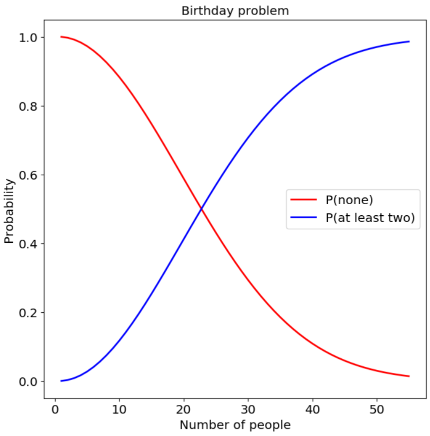
Let $A =$ "At least two people share a birthday." Assuming that there are only 365 equally likely days, first compute the probability $P(A^c)$ that no two people in the group share their birthday.

$$P(A^c) = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \ldots \frac{(365-N+1)}{365}$$

(1^{st} birthday unique and 2^{nd} birthday unique and....)

$$P(A) = 1 - P(A^c) > 0.5 \text{ for } N \gtrsim 23.$$

Code for plot available here.



Conditional probabilities and Bayes' Theorem

Conditional probability

If A and B are two events, $A \cap B$ represents the event that both A and B occur.
Examples:
1) $A =$ "Flip two coins, second coin lands heads.", $B =$ "Flip two coins, first coin lands heads."
$\Rightarrow A \cap B =$ "Flip two coins, get two heads."
2) $A =$ "Flip two coins, get two heads.", $B =$ "Flip two coins, get two tails."
$\Rightarrow A \cap B = \varnothing$.

Definition (Conditional probability)

The probability that an event A occurs, given that another event B has already occurred.
Representation: $P(A|B)$ ("probability of A given B").

Example 1: $A|B =$ "Second coin lands heads, given that first coin landed heads."
Example 2: $A|B =$ "Coins land heads, given that coins landed tails." (contradiction).

In general, the events $A \cap B$ and $A|B$ are related.
(given that B has occurred, A can only occur if $A \cap B \neq \varnothing$).

Conditional probability (contd.)

Let $P(A \cap B) = P(A|B) \times \lambda$
(λ is a normalisation constant to ensure $0 < P(A|B) < 1$).

To solve for λ, consider the special case $B \subseteq A$. Then,
(1) A always occurs if B occurs $\Rightarrow P(A|B) = 1$ in this case.
(2) $A \cap B = B \Rightarrow P(A \cap B) = P(B)$ in this case.
From (1) and (2), $\lambda = P(B)$ (since λ is a constant, valid for all cases).
$\Rightarrow P(A \cap B) = P(A|B) \times P(B)$.

Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \text{ occurs, given that } B \text{ already occurred}) = \frac{P(A \text{ and } B \text{ both occur})}{P(B \text{ occurs})}$$

Note: From the axioms of probability (unitarity),
$P(A|B) + P(A^c|B) = 1$

Independence

Two events A and B are said to be independent ("$A \perp B$") if the occurrence of one does not affect the probability of occurrence of the other.

If A and B are independent, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$.
Therefore, from the definition of conditional probability,

Definition (Independence)

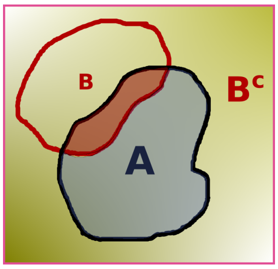
$$A \perp B \Rightarrow P(A \cap B) = P(A) \times P(B).$$

Example (Two-coin toss)

Let $A =$ "Second coin turns up heads" and $B =$ "First coin turns up heads".
If both coins are fair, then the outcome of flipping the second coin should not depend on the outcome of flipping the first one.
$\Rightarrow P(A|B) = P(A) = 1/2$.
$\Rightarrow P(\text{two heads}) = P(A|B) \times P(B) = P(A) \times P(B) = 1/4$.

Are two mutually exclusive events mutually independent?
Mutual exclusivity $\neq$ mutual independence!!

Conditionality and Marginalisation



In general, $A = (A \cap B) \cup (A \cap B^c)$. Using the Inclusion-Exclusion Principle,
$P(A) = P(A \cap B) + P(A \cap B^c) - P((A \cap B) \cap (A \cap B^c))$

Using conditional probabilities, $P(A) = P(A|B) \times P(B) + P(A|B^c) \times P(B^c)$.
$P(A)$ is obtained by "mariginalising over B".

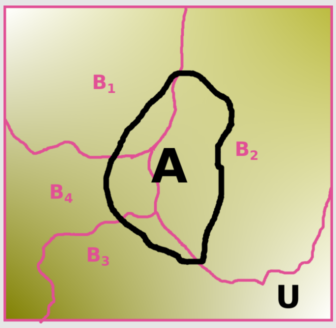
Do not confuse with the result from unitarity: $P(A|B) + P(A^c|B) = 1$.

Generalisation: Law of Total Probability

(Connects conditional probabilities to marginal probability)

Given N pairwise disjoint and collectively exhaustive events B_i ($i = 1, 2, ..., N$), the probability of occurrence of an event A is given as the weighted average of the conditional probabilities $P(A|B_i)$, with weights $P(B_i)$:

$$P(A) = \sum_{i=1}^{N} P(A \cap B_i) = \sum_{i=1}^{N} P(A|B_i) \times P(B_i)$$



$P(A)$ is then the probability of A marginalised over the events B_i.

Bayes' Theorem

Recall: $P(A \cap B) = P(A|B) \times P(B)$. Switching A and B, $P(B \cap A) = P(B|A) \times P(A)$.
But $A \cap B = B \cap A$ (order doesn't matter).$\Rightarrow P(A|B) \times P(B) = P(B|A) \times P(A)$, or

Definition (Bayes' Theorem)

$$P(A|B) = \frac{P(B|A)}{P(B)} \times P(A)$$

Under the Bayesian Interpretation of probability, this is read as
Updated deg. of belief in A = Support for A from evidence B × Original deg. of belief in A.
or

$$\text{Posterior prob. of } A \text{ given evidence } B = \frac{\text{Cond. prob. of } B \text{ given } A}{\text{Marginal prob. of } B} \times \text{Prior prob. of } A.$$

or

$$\text{Posterior prob. of } A \text{ given evidence } B = \frac{\text{Likelihood of } A \text{ given } B}{\text{Evidence } B} \times \text{Prior prob. of } A.$$

We can use the Law of Total Probability to convert the marginal probability into conditional probabilities:

$$P(A_j|B) = \frac{P(B|A_j)}{\sum_{i=1}^{N} P(B|A_i) \times P(A_i)} \times P(A_j) = \frac{P(B|A_j)}{P(B|A_j) \times P(A_j) + P(B|A_j^c) \times P(A_j^c)} \times P(A_j)$$

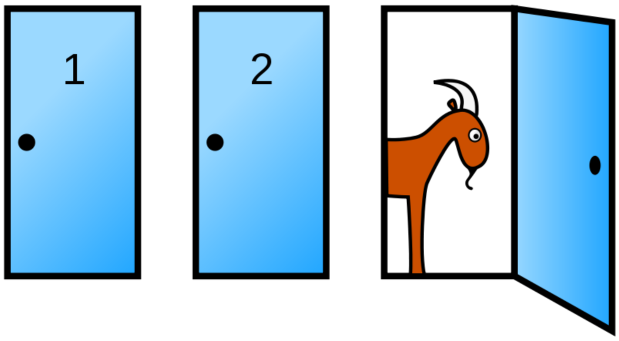
The Monty Hall Problem

"He's got diamonds on his pinky, babe,
His game's *Let's Make A Deal*.
Take it all from Monty Hall!"
–*Cocktail Desperado*, Terry Allen & The Panhandle Band

The Monty Hall Problem ("¿El Coche o la cabra?")

Named after the first host of the TV game show, *Let's Make A Deal (Trato Hecho).*

(see also Three Prisoners Problem and Bertrand's Box Problem.)



Behind one of three doors, there is a brand new car. The other two have a goat each. You choose one of the doors. The host opens **one of the other two doors**, revealing a goat. Should you switch your choice to the other closed door, or stick with your original choice?

Figure: (credit: User:Cepheus/Public Domain)

The Monty Hall Problem (contd.)

Originally solved by Selvin (1975a,b).
Also solved in *Parade* magazine column by Marilyn vos Savant (1990a,b). Cue controversy (see Wikipedia page)!

Marilyn vos Savant
(Credit: Ethan Hill/CC BY-SA 2.0).

"[E]ven Nobel physicists... give the wrong answer... [and are] ready to berate... those who propose the right answer." (vos Savant, *The Power of Logical Thinking*.)

The Monty Hall Problem – a solution via *Numberphile*

Click here for the YouTube clip.

Credit: Brady Haran, *Numberphile*.