Probability I

Ekrem T. Başer

University of Illinois

baser2@illinois.edu

Elevator pitch for learning probability

- Uncertainty is everywhere!
  - # heads if I flip a fair coin 10 times? Weather tomorrow? # of Trump tweets today? # of protests next month in the US? Turnout in the next elections?
- Probability is a formal model of uncertainty.
- Statistics is the act of trying to get at the underlying probability distribution of the data we have

I benefited quite a bit from Prof. Anil Bera's Econ 532 notes for these slides. Take his class.
I also benefited from Harvard Political Science math camp lecture notes.

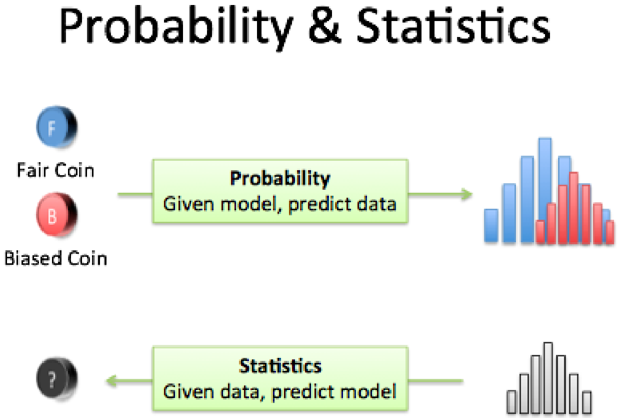


Figure: BetterExplained.com

Basic set theory

Definition 1 (Set)

A set is any (well defined) collection of objects.

Example

- $C = \{2, 4, 6, ...\}$ set of all positive even integers.
- $A = \{$All grad students in U of I$\}$
- $P = \{$Students in the new Poli Sci PhD cohort$\}$

An object in a set is an element, e.g., James is an element of the set "P". So we have: James $\in$ P, and we also have $P \subset A$.

Definition 2 (Subset)

Set B is a subset of set A, denoted by $B \subset A$ if $x \in B$ implies $x \in A$.

What does "implies" mean?? A slight detour to logic..

Logic: Necessary and Sufficient Conditions

Necessity *(I don't shower) Only if it rains do I get wet*

- X is a necessary condition for Y
- Y is a sufficient condition for X
- X is implied by Y (Only if X then Y)
- $X \Leftarrow Y$

Sufficiency *(I have no umbrella) If it rains then I get wet*

- X is a sufficient condition for Y
- Y is a necessary condition for X
- X implies Y (or If X then Y)
- $X \Rightarrow Y$

Necessity and Sufficiency *(I don't shower and I don't have an umbrella) I get wet if and only if it rains*

- X is a necessary and sufficient condition for Y
- Y is a necessary and sufficient condition for X
- X if and only if Y (or X iff Y)
- $X \iff Y$

Back to biz..

Definition 3 (Subset)

Set B is a subset of set A, denoted by $B \subset A$ if $x \in B$ implies $x \in A$.

In other words, we know that if James is in the set of our new cohort, then he also must be in the set of all grad students in Illinois

Set Operations: Union

Definition 4 (Union, "$\cup$")

Union of two sets A and B is C, defined by $C = A \cup B$, if $C = \{x \mid x \in A$ or $x \in B.\}$

(In math "or" means "and/or." Also $\mid$ means "given" or "such that")
By union of two sets we mean collection of elements which belong to at least one of the two sets.
The operation can be defined for more than two sets. Suppose we have n sets $A_1, A_2, .., A_n$. Then $A_1 \cup A_2 ... \cup A_n$ denoted by $\cup_{i=1}^{n} A_i$ is defined as:

Definition 5 (Union, "$\cup$")

$$\bigcup_{i=1}^{n} A_i = \{x \mid x \in \text{at least one } A_i, i = 1, 2, ..., n\}$$

(We can have $n = \infty$, but should be countable)

Union!!

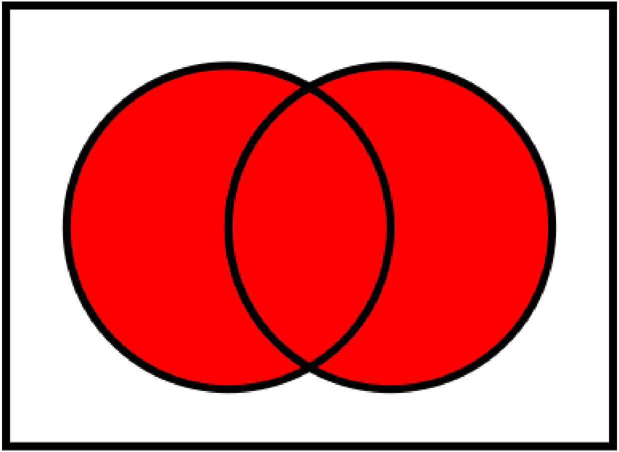


Figure: Wikipedia

Set Operations: Intersection

Definition 6 (Intersection, "$\cap$")

Intersection of two sets A and B is C, denoted by $C = A \cap B$,if $C = \{x \mid x \in A \text{ and } x \in B\}$.

James, being an element of both your cohort and all grad students in Illinois, is also an element of the intersection of these two sets.

Definition 7 (Intersection, "$\cap$")

$$\bigcap_{i=1}^{n} A_i = \{x \mid x \in A_i, \forall i = 1, 2, ..., n\}$$

(We can have $n = \infty$, but should be countable)

Intersection!!

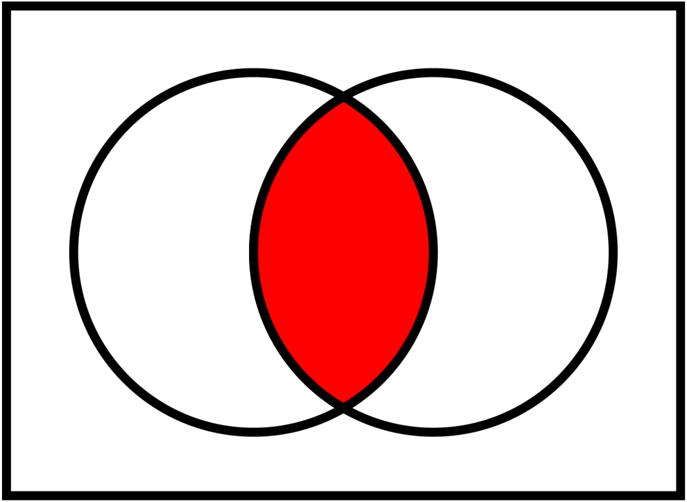


Figure: Wikipedia

Set Operations: Difference

Definition 8 (Difference, "\\")

The difference between two sets A and B, denoted by $A \setminus B$, is defined as $C = A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}$.

Definition 9 (Complement)

Complement of a set A with respect to a Space S, denoted by $A^{\complement} = \{x \in S \mid x \notin A\}$.

In most cases, the reference set S will be obvious from the context.

Difference and Complement!!

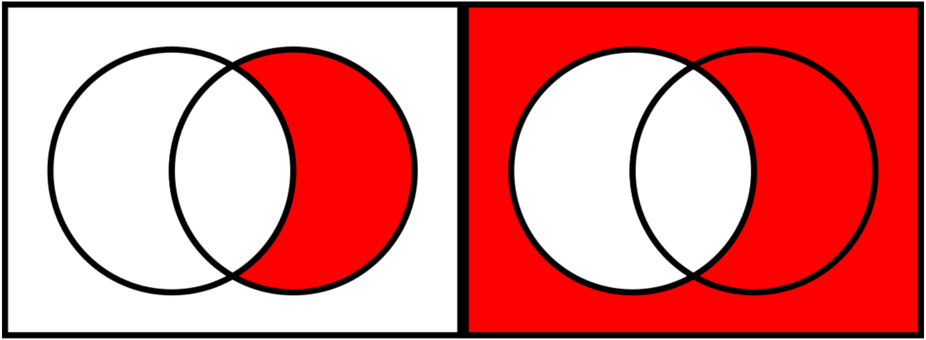


Figure: $B \setminus A$ and A^{C}

Properties of Set Operations

1. Commutative: $A \cup B = B \cup A; A \cap B = B \cap A$
2. Associative: $A \cup (B \cup C) = (A \cup B) \cup C; A \cap (B \cap C) = (A \cap B) \cap C$
3. Distributive:
   $A \cap (B \cup C) = (A \cap B) \cup (A \cap C); A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
4. de Morgan's laws: $(A \cup B)^{\complement} = A^{\complement} \cap B^{\complement}; (A \cap B)^{\complement} = A^{\complement} \cup B^{\complement}$

And a few other things:
Disjoint: Sets are disjoint when they do not intersect; $A \cap B = \{\varnothing\}$.
Pairwise disjoint: A collection of sets is pairwise disjoint if, for all $i \neq j$, $A_i \cap A_j = \{\varnothing\}$.
Partition: A collection of sets form a partition of set S if they are pairwise disjoint and they cover set S, such that $\cup_{i=1}^{n} A_i = S$.

Let's connect set theory with "event" and "probability."

- Suppose we throw one coin twice. The coin has two sides, head (H) and tail (T). What are the possible outcomes?
  - Both tail (T T)
  - Both head (H H)
  - Tail head (T H)
  - Head tail (H T)
- Collect these together in a set $\Omega = \{(TT), (HH), (TH), (HT)\}$, this is the collection of all possible outcomes. In probability theory, the collection of all possible outcomes is known as **sample space**.
- We may be interested in the following special outcomes:
  - $A_l = \{\text{outcomes with first head}\} = \{\text{(HH),(HT)}\}$.
  - $A_2 = \{\text{outcomes with at least one head}\} = \{\text{(HH),(HT),(TH)}\}$.
  - $A_3 = \{\text{outcomes with no tail}\} = \{\text{(HH)}\}$.

Example

1. Toss a coin once, the sample space is $\Omega = \{H, T\}$
2. Toss a coin twice, the sample space is $\Omega = \{(TT), (HH), (TH), (HT)\}$

Let's connect set theory with "event" and "probability."

- A_1, A_2, A_3...are all **events**, and note $A_1, A_2, A_3 \subset \Omega$
- We can think of a collection of subsets of Ω and a particular event will be an element of that collection.
  - Under this framework, we can define the probabilities of different events.
- So far we have considered sets that are collections of single elements (e.g. $F = \{1, 2, 3\}$), now let's think of sets of sets.
  - **Sigma Field ($\mathcal{A}$):** A collection of sets $A_1, A_2, A_3, ...$ satisfying the following properties:
    - $A_1, A_2, .. \in \mathcal{A} \Rightarrow \cup_{i=1}^{\infty} \in \mathcal{A}$
    - $A \in \mathcal{A} \Rightarrow A^{\complement} \in \mathcal{A}$

Example

Let $\Omega = \{1, 2, 3, 4\}$
A σ-field on Ω can be written as $\mathcal{A} = \{\varnothing, (1, 2), (3, 4), (1, 2, 3, 4)\}$

Probability and Random Variable

- As you can guess, the word "random" is associated with some sort of uncertainty
- If we toss a coin, we know the possibilities: head (H) or tail (T); but we are uncertain about exactly which one will appear.
  - Therefore, "tossing a coin" can be regarded as a random experiment where the possibilities are known but not the exact outcome
  - Instead of assigning symbols, we can give these outcomes, some numbers (real numbers).
  - For this case, we can define:

$$X = \begin{cases} 0 & \text{If the outcome is T} \\ 1 & \text{If the outcome is H} \end{cases}$$

- Let's formally define **probability**: For this example, we have the **sample space** $\Omega = \{H, T\}$. A σ-field defined on Ω is $\mathcal{A} = \{\varnothing, \Omega, (H), (T)\}$. Elements of $\mathcal{A}$ are called the **events**. **"Probability"** is nothing but assigning real numbers (satisfying some conditions) to each of these events.

Probability

Definition 10 (Probability)

Probability denoted as P is a function from $\mathcal{A}$ to $[0,1]$, that is $P : \mathcal{A} \to [0,1]$, satisfying the following conditions:

1. $P(\Omega) = 1$
2. For any event A, $P(A) \geq 0$
3. If $A_1, A_2, ... \in \mathcal{A}$ are disjoint(mutually exclusive) then

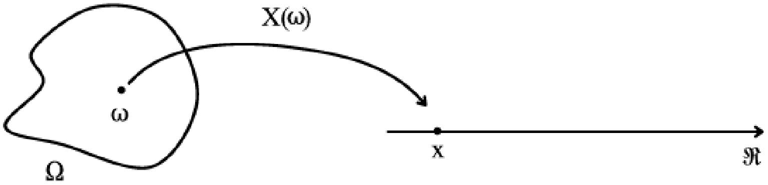
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Example (Roll a fair dice once)

$$\Omega = \{H, T\},\ \mathcal{A} = \{\varnothing, \Omega, (H), (T)\}$$

$$P(\varnothing) = 0,\ P(\Omega) = 1,\ P(H) = \frac{1}{2},\ P(T) = \frac{1}{2}$$

A **random variable** is a *measurable function* from the **sample space** to the **real line**. "Measurability" is defined by requiring that the inverse image of X is an element of the σ-field, i.e., an event.



Recall that, probability is defined only for events. By requiring that X is measurable, in a sense, we are assuring its probability distribution.

Probability and Random Variable

All that mumbo jumbo was to say:

- Probabilities are defined only for **events**.
- Random variables take **outcomes** and assign them a (real) number.
- Random variables are measurable functions. This means we ensure that every value generated by a random variable can be traced back to an **event**, which then can be asssigned a **probability**. This means we have a **probability distribution** for every random variable
- This event-outcome distinction all seems like semantics when we are talking about coin tosses and dice rolls. Not so much when the sample space of a distribution is the whole real line, like in normal distribution.
  - If you like to assign a probability to every outcome in the real line, be my guest..

Back to Probability

Basic Rules of Probability: All rules below can be deduced from the three axioms mentioned earlier

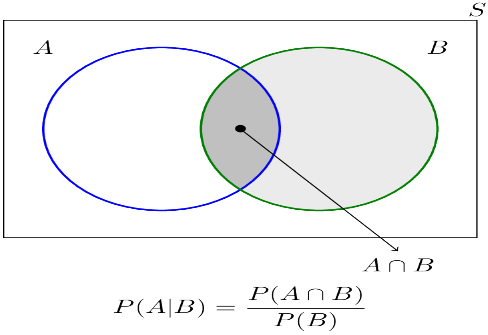
1. $P(\varnothing) = 0$
2. For any event A, $0 \leq P(A) \leq 1$
3. $P(A^{\complement}) = 1 - P(A)$
4. If $A \subset B$ then $P(A) \leq P(B)$
5. For *any* two events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
6. For any sequence of n events $A_1, A_2, ..., A_n$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

Conditional Probability

The conditional probability $P(A \mid B)$ of an event A is the probability of A, given that another event B has occurred. Conditional probability allows for the inclusion of other information into the calculation of the probability of an event. It is calculated as

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



Conditional Probability

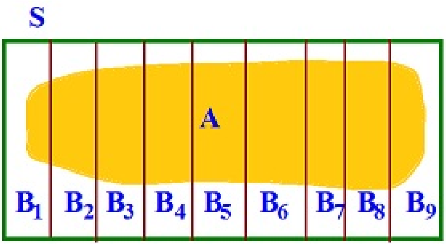
Example (Conditional Probability)

A six-sided die is rolled. What is the probability of a 1, given the outcome is an odd number?

Multiplicative Law of Probability: The probability of the intersection of two events A and B is $P(A \cap B) = P(A)P(B \mid A) = Pr(B)P(A \mid B)$ which follows directly from the definition of conditional probability.

Law of Total Probability: Let S be the sample space of some experiment and let the disjoint k events $B_1, ..., B_k$ partition S, such that $P(B_1 \cup ... \cup B_k) = P(S) = 1$. If A is some other event in S, then the events $A \cap B_1, A \cap B_2, ..., A \cap B_k$ will form a partition of A and we can write A as $A = (A \cap B_1) \cup ... \cup (A \cap B_k)$.

Conditional Probability



Since the k events are disjoint,

$$P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(B)P(A \mid B_i)$$

Sometimes it is easier to calculate these conditional probabilities and sum them than it is to calculate P(A) directly.

Bayes' Rule

Assume that events $Innocent(I)$ and $Guilty(G)$ form a partition of the sample space S regarding a suspect, and $Evidence(E)$ represents a set of evidence we gathered. Then by the Law of Total Probability:

$$P(I \mid E) = \frac{P(E \cap I)}{P(E)} = \frac{P(I)P(E \mid I)}{P(I)P(E \mid I) + P(G)P(E \mid G)}$$

Prior and Posterior Probabilities: Above, $P(I)$ is called the prior probability, since it's the probability of the suspect being innocent before seeing the evidence. $P(I \mid E)$ is called the posterior probability, since it's the probability of the suspect being innocent after the evidence is taken into account.

Bayes Rule

Example

A test for cancer correctly detects it 90% of the time, but incorrectly identifies a person as having cancer 10% of the time. If 10% of all people have cancer at any given time, what is the probability that a person who tests positive actually has cancer?

The probability that the test detects cancer correctly is $P(T \mid C) = 0.9$
The test giving a false positive has probability $P(T \mid C') = 0.1$.
10% of the population have cancer, so $P(C) = 0.1$
We are interested in a person who tests positive, so $P(C \mid T) = ?$

$$P(C \mid T) = \frac{P(C \cap T)}{P(T)} = \frac{P(C)P(T \mid C)}{P(C)P(T \mid C) + P(C')P(T \mid C')}$$

$$= \frac{(0.1 * 0.9)}{(0.1 * 0.9) + (0.9 * 0.1)} = 0.5$$

Independence

Independence: If the occurrence or non-occurrence of either events A and B have no effect on the occurrence or non-occurrence of the other, then A and B are independent. If A and B are independent, then:

1. $P(A \mid B) = P(A)$
2. $P(B \mid A) = P(B)$
3. $P(A \cap B) = P(A)P(B)$

Mutually exclusive events are NOT independent of each other. If A and B are mutually exclusive, then they cannot happen simultaneously. If we know that A occurred, then we know that B couldn't have occurred. Because of this, A and B aren't independent.