Basics of Probability

Janyl Jumadinova

September 17, 2018
Probability theory yields mathematical tools to deal with uncertain events.
Probability theory yields mathematical tools to deal with uncertain events.

**Used everywhere nowadays and its importance is growing.**
Applications include:
Applications include:

- **Population genetics**: tree-valued stochastic processes,
Applications include:

- Population genetics: tree-valued stochastic processes,
- **Web search engine**: Markov chain theory,
Applications include:

- Population genetics: tree-valued stochastic processes,
- Web search engine: Markov chain theory,
- Data mining, Machine learning: Stochastic gradient, Markov chain Monte Carlo,
Applications include:

- Population genetics: tree-valued stochastic processes,
- Web search engine: Markov chain theory,
- Data mining, Machine learning: Stochastic gradient, Markov chain Monte Carlo,
- **Image processing**: Markov random fields,
Probability Theory

Applications include:

- Population genetics: tree-valued stochastic processes,
- Web search engine: Markov chain theory,
- Data mining, Machine learning: Stochastic gradient, Markov chain Monte Carlo,
- Image processing: Markov random fields,
- **Design of wireless communication systems: random matrix theory**,
Applications include:

- Population genetics: tree-valued stochastic processes,
- Web search engine: Markov chain theory,
- Data mining, Machine learning: Stochastic gradient, Markov chain Monte Carlo,
- Image processing: Markov random fields,
- Design of wireless communication systems: random matrix theory,
- Optimization of engineering processes: simulated annealing, genetic algorithms,
Applications include:

- Population genetics: tree-valued stochastic processes,
- Web search engine: Markov chain theory,
- Data mining, Machine learning: Stochastic gradient, Markov chain Monte Carlo,
- Image processing: Markov random fields,
- Design of wireless communication systems: random matrix theory,
- Optimization of engineering processes: simulated annealing, genetic algorithms,
- **Computer-aided design of polymers: Markov chain Monte Carlo.**
Applications include:

- Population genetics: tree-valued stochastic processes,
- Web search engine: Markov chain theory,
- Data mining, Machine learning: Stochastic gradient, Markov chain Monte Carlo,
- Image processing: Markov random fields,
- Design of wireless communication systems: random matrix theory,
- Optimization of engineering processes: simulated annealing, genetic algorithms,
- Computer-aided design of polymers: Markov chain Monte Carlo.
- Finance (option pricing, volatility models): Monte Carlo, dynamic models,
Probability and Statistics

- Probability $\neq$ Statistics

- Probability: Known distributions $\Rightarrow$ what are the outcomes?

- Statistics: Known outcomes $\Rightarrow$ what are the distributions?
Many basic probability problems are counting problems.
Many **basic** probability problems are counting problems.

*Example:* Assume there are 1 man and 2 women in a room. You pick a person randomly. What is the probability $P_1$ that this is a man?

If you pick two persons randomly, what is the probability $P_2$ that these are a man and woman?

**Answer:** You have the possible outcomes: (M), (W1), (W2) so $P_1 = \frac{\text{# "successful" events}}{\text{# events}} = \frac{\text{# boys}}{\text{# boys} + \text{# girls}} = \frac{1}{3}$.

To compute $P_2$, you can think of all the possible events: (M,W1), (M,W2), (W1,W2) so $P_2 = \frac{\text{# "successful" events}}{\text{# events}} = \frac{2}{3}$. 

Many **basic** probability problems are counting problems.

*Example*: Assume there are 1 man and 2 women in a room. You pick a person randomly. What is the probability $P_1$ that this is a man?

If you pick two persons randomly, what is the probability $P_2$ that these are a man and woman?

**Answer**: You have the possible outcomes: (M), (W1), (W2) so $P_1 = \frac{\# \text{ "successful" events}}{\# \text{ events}} = \frac{\# \text{ boys}}{\# \text{ boys} + \# \text{ girls}} = \frac{1}{3}$.

To compute $P_2$, you can think of all the possible events: (M, W1), (M, W2), (W1, W2) so $P_2 = \frac{\# \text{ "successful" events}}{\# \text{ events}} = \frac{2}{3}$.
Many **basic** probability problems are counting problems.

**Example:** Assume there are 1 man and 2 women in a room. You pick a person randomly. What is the probability $P_1$ that this is a man? If you pick two persons randomly, what is the probability $P_2$ that these are a man and woman?

**Answer:** You have the possible outcomes: (M), (W1), (W2) so

$$P_1 = \frac{\# \text{ “successful” events}}{\# \text{ events}} = \frac{\# \text{ boys}}{\# \text{ boys} + \# \text{ girls}} = \frac{1}{3}.$$

To compute $P_2$, you can think of all the possible events: (M,W1), (M,W2), (W1,W2) so

$$P_2 = \frac{\# \text{ “successful” events}}{\# \text{ events}} = \frac{2}{3}.$$
Sample Space

Definition

The *sample space* \( S \) of an experiment (whose outcome is uncertain) is the set of all possible outcomes of the experiment.
Sample Space

- Example (child): Determining the sex of a newborn child in which case $S = \{\text{boy, girl}\}$.
Sample Space

- **Example (child):** Determining the sex of a newborn child in which case $S = \{\text{boy, girl}\}$.

- **Example (horse race):** Assume you have an horse race with 12 horses. If the experiment is the order of finish in a race, then

$$S = \{\text{all 12! permutations of } (1, 2, 3, \ldots, 11, 12)\}.$$
Example (child): Determining the sex of a newborn child in which case \( S = \{ \text{boy}, \text{girl} \} \).

Example (horse race): Assume you have an horse race with 12 horses. If the experiment is the order of finish in a race, then

\[
S = \{ \text{all 12! permutations of } (1, 2, 3, \ldots, 11, 12) \}.
\]

Example (coins): If the experiment consists of flipping two coins, then the sample space is

\[
S = \{(H, H), (H, T), (T, H), (T, T)\}.
\]
Sample Space

- **Example (child):** Determining the sex of a newborn child in which case $S = \{\text{boy, girl}\}$.

- **Example (horse race):** Assume you have an horse race with 12 horses. If the experiment is the order of finish in a race, then

  $$S = \{\text{all 12! permutations of (1, 2, 3, ..., 11, 12)}\}.$$ 

- **Example (coins):** If the experiment consists of flipping two coins, then the sample space is

  $$S = \{(H, H), (H, T), (T, H), (T, T)\}.$$ 

- **Example (lifetime):** If the experiment consists of measuring the lifetime (in years) of your pet then the sample space consists of all nonnegative real numbers: $S = \{x; 0 \leq x < \infty\}$.
Events

Any subset $E$ of the sample space $S$ is known as an event; i.e. an event is a set consisting of possible outcomes of the experiment.
Events

- Any *subset* $E$ of the sample space $S$ is known as an *event*; i.e. an event is a set consisting of possible outcomes of the experiment.
- If the outcome of the experiment is in $E$, then we say that $E$ has occurred.
Events

- *Example (child):* The event \( E = \{ \text{boy} \} \) is the event that the child is a boy.
Events

- *Example (child)*: The event $E = \{ \text{boy} \}$ is the event that the child is a boy.

- *Example (horse race)*: The event $E = \{ \text{all outcomes in } S \text{ starting with a 7} \}$ is the event that the race was won by horse 7.
Events

- *Example (child)*: The event $E = \{ \text{boy} \}$ is the event that the child is a boy.

- *Example (horse race)*: The event $E = \{ \text{all outcomes in } S \text{ starting with a 7} \}$ is the event that the race was won by horse 7.

- *Example (coins)*: The event $E = \{(H, T), (T, T)\}$ is the event that a tail appears on the second coin.
Events

- Example (child): The event $E = \{ \text{boy} \}$ is the event that the child is a boy.
- Example (horse race): The event $E = \{ \text{all outcomes in } S \text{ starting with a } 7 \}$ is the event that the race was won by horse 7.
- Example (coins): The event $E = \{(H, T), (T, T)\}$ is the event that a tail appears on the second coin.
- Example (lifetime): The event $E = \{x : 3 \leq x \leq 15\}$ is the event that your pet will live more than 3 years but won’t live more than 15 years.
Union of Events

Given events $E$ and $F$, $E \cup F$ is the set of all outcomes either in $E$ or $F$ or in both $E$ and $F$.

$E \cup F$ occurs if either $E$ or $F$ occurs.

$E \cup F$ is the union of events $E$ and $F$.
Union of Events

- *Example (coins)*: If we have $E = \{(H, T)\}$ and $F = \{(T, H)\}$ then $E \cup F = \{(H, T), (T, H)\}$ is the event that one coin is head and the other is tail.
Union of Events

- **Example (coins):** If we have $E = \{(H, T)\}$ and $F = \{(T, H)\}$ then $E \cup F = \{(H, T), (T, H)\}$ is the event that one coin is head and the other is tail.

- **Example (horse race):** If we have $E = \{\text{all outcomes in } S \text{ starting with a 7}\}$ and $F = \{\text{all outcomes in } S \text{ finishing with a 3}\}$ then $E \cup F$ is the event that the race was won by horse 7 and/or the last horse was horse 3.
Example (coins): If we have $E = \{(H, T)\}$ and $F = \{(T, H)\}$ then $E \cup F = \{(H, T), (T, H)\}$ is the event that one coin is head and the other is tail.

Example (horse race): If we have $E = \{\text{all outcomes in } S \text{ starting with a 7}\}$ and $F = \{\text{all outcomes in } S \text{ finishing with a 3}\}$ then $E \cup F$ is the event that the race was won by horse 7 and/or the last horse was horse 3.

Example (lifetime): If $E = \{x : 0 \leq x \leq 10\}$ and $F = \{x : 15 \leq x < \infty\}$ then $E \cup F$ is the event that your pet will die before 10 or will die after 15.
Intersection of Events

Given events $E$ and $F$, $E \cap F$ is the set of all outcomes which are both in $E$ and $F$.

$E \cap F$ is also denoted as $EF$. 
Intersection of Events

- **Example (coins):** If we have \( E = \{(H, H), (H, T), (T, H)\} \) (event that one H at least occurs) and \( F = \{(H, T), (T, H), (T, T)\} \) (event that one T at least occurs) then \( E \cap F = \{(H, T), (T, H)\} \) is the event that one H and one T occur.
Intersection of Events

- **Example (coins):** If we have $E = \{(H, H), (H, T), (T, H)\}$ (event that one H at least occurs) and $F = \{(H, T), (T, H), (T, T)\}$ (even that one T at least occurs) then $E \cap F = \{(H, T), (T, H)\}$ is the event that one H and one T occur.

- **Example (horse race):** If we have $E = \{\text{all outcomes in } S \text{ starting with a 7}\}$ and $F = \{\text{all outcomes in } S \text{ starting with a 8}\}$ then $E \cap F$ does not contain any outcome and is denoted by $\emptyset$. 
Intersection of Events

- **Example (coins):** If we have $E = \{(H, H), (H, T), (T, H)\}$ (event that one H at least occurs) and $F = \{(H, T), (T, H), (T, T)\}$ (even that one T at least occurs) then $E \cap F = \{(H, T), (T, H)\}$ is the event that one H and one T occur.

- **Example (horse race):** If we have 
  
  $E = \{\text{all outcomes in } S \text{ starting with a 7}\}$ and  
  $F = \{\text{all outcomes in } S \text{ starting with a 8}\}$ then $E \cap F$ does not contain any outcome and is denoted by $\emptyset$.

- **Example (lifetime):** If we have $E = \{x : 0 \leq x \leq 5\}$ and $F = \{x : 10 \leq x < 15\}$ then $E \cap F = \{x : 3 \leq x \leq 5\}$ is the event that your pet will die between 10 and 15.
Notations and Properties

For any event $E$, $E^c$ denote the \textit{complement} set of all outcomes in $S$ which are not in $E$. Hence we have $E \cup E^c = S$ and $E \cap E^c = \emptyset$. 
For any event $E$, $E^c$ denote the complement set of all outcomes in $S$ which are not in $E$.
Hence we have $E \cup E^c = S$ and $E \cap E^c = \emptyset$.

For any two events $E$ and $F$, we write $E \subset F$ is all the outcomes of $E$ are in $F$. 
Axioms of Probability

- Consider an experiment with sample space $S$. For each event $E$, we assume that a number $P(E)$, the *probability* of the event $E$, is defined and satisfies the following 3 axioms.

**Axiom 1**

$0 \leq P(E) \leq 1$

**Axiom 2**

$P(S) = 1$

**Axiom 3**

For any sequence of mutually exclusive events \( \{E_i\}_{i \geq 1} \), i.e. $E_i \cap E_j = \emptyset$ when $i \neq j$, then

$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$$
Axioms of Probability

- Consider an experiment with sample space $S$. For each event $E$, we assume that a number $P(E)$, the probability of the event $E$, is defined and satisfies the following 3 axioms.

- **Axiom 1**
  
  $$0 \leq P(E) \leq 1$$
Axioms of Probability

Consider an experiment with sample space $S$. For each event $E$, we assume that a number $P(E)$, the *probability* of the event $E$, is defined and satisfies the following 3 axioms.

- **Axiom 1**
  
  \[ 0 \leq P(E) \leq 1 \]

- **Axiom 2**
  
  \[ P(S) = 1 \]
Axioms of Probability

Consider an experiment with sample space $S$. For each event $E$, we assume that a number $P(E)$, the probability of the event $E$, is defined and satisfies the following 3 axioms.

- **Axiom 1**
  \[ 0 \leq P(E) \leq 1 \]

- **Axiom 2**
  \[ P(S) = 1 \]

- **Axiom 3.** For any sequence of mutually exclusive events $\{E_i\}_{i \geq 1}$, i.e. $E_i \cap E_j = \emptyset$ when $i \neq j$, then
  \[ P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i) \]
Proposition: \( P(E^c) = 1 - P(E) \).
Properties

- **Proposition:** $P(E^c) = 1 - P(E)$.
- **Proposition:** If $E \subset F$ then $P(E) \leq P(F)$. 
Properties

- **Proposition**: \( P(E^c) = 1 - P(E) \).
- **Proposition**: If \( E \subset F \) then \( P(E) \leq P(F) \).
- **Proposition**: We have \( P(E \cup F) = P(E) + P(F) - P(E \cap F) \).
Conditional Probabilities

- **Conditional Probability.** Consider an experiment with sample space $S$. Let $E$ and $F$ be two events, then the conditional probability of $E$ given $F$ is denoted by $P(E|F)$ and satisfies if $P(F) > 0$

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Intuition: If $F$ has occurred, then, in order for $E$ to occur, it is necessary that the occurrence be both in $E$ and $F$, hence it must be in $E \cap F$. Once $F$ has occurred, $F$ is the new sample space.
Conditional Probabilities

- **Conditional Probability.** Consider an experiment with sample space $S$. Let $E$ and $F$ be two events, then the conditional probability of $E$ given $F$ is denoted by $P(E|F)$ and satisfies if $P(F) > 0$

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

- **Intuition:** If $F$ has occurred, then, in order for $E$ to occur, it is necessary that the occurrence be both in $E$ and $F$, hence it must be in $E \cap F$. Once $F$ has occurred, $F$ is the new sample space.
Equally likely outcomes. In this case, we have

\[
P(E|F) = \frac{\text{\# outcomes in } E \cap F}{\text{\# outcomes in } F} = \frac{\text{\# outcomes in } E \cap F}{\text{\# outcomes in } S} \cdot \frac{\text{\# outcomes in } F}{\text{\# outcomes in } S} = \frac{P(E \cap F)}{P(F)}.
\]
Independence

- Events $A$ and $B$ are independent iff $P(A \cap B) = P(A)P(B)$
Independence

- Events $A$ and $B$ are independent iff $P(A \cap B) = P(A)P(B)$
- Equivalent to $P(A|B) = P(A)$
Independence

- Events $A$ and $B$ are independent iff $P(A \cap B) = P(A)P(B)$
- Equivalent to $P(A|B) = P(A)$
- One event occurring does not effect the probability of another occurring
The Multiplication Rule

Let $E_1, E_2, \ldots, E_n$ be a sequence of events, then we have

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1) P(E_2 | E_1) \times \cdots \times P(E_n | E_1 \cap \cdots \cap E_{n-1})$$
Example: You have a box with 3 blue marbles, 2 red marbles, and 4 yellow marbles. You are going to pull out one marble, record its color, put it back in the box and draw another marble. What is the probability of pulling out a red followed by a blue?
Example

Example: You have a box with 3 blue marbles, 2 red marbles, and 4 yellow marbles. You are going to pull out one marble, record its color, put it back in the box and draw another marble. What is the probability of pulling out a red followed by a blue?

Example: Consider the same box of marbles. However, we are going to pull out the first marble, leave it out and then pull out the second marble. What is the probability of pulling out a red marble followed by a blue marble?
Random Variables

- A random variable is a function $R : S \rightarrow R$
Random Variables

- A random variable is a function $R : S \rightarrow R$
- Domain of $R$ is the sample space $S$
Random Variables

- A random variable is a function $R : S \rightarrow R$
- Domain of $R$ is the sample space $S$
- Range of $R$ is the real line
Random Variables

Example: **Discrete Random Variable**

*Experiment*: flip 10 coins

*Desired outcome*: the number of heads

*We care about*: the number of heads that appear among 10 tosses (not the probability of getting a particular sequence of heads and tails)
Random Variables

Example: **Discrete Random Variable**

*Experiment:* flip 10 coins

*Desired outcome:* the number of heads

*We care about:* the number of heads that appear among 10 tosses (not the probability of getting a particular sequence of heads and tails)

Probability of a random variable $R$ taking on some specific value $k$ is:

$$P(R = k) = P(\{s : R(s) = k\})$$

, with $R(s)$ - number of heads occurring after $s$ tosses
Example: **Continuous Random Variable**

\( R(s) \) - random variable indicating the amount of time it takes for a fast food burger to decay
Random Variables

Example: **Continuous Random Variable**

\( R(s) \) - random variable indicating the amount of time it takes for a fast food burger to decay

Probability that \( R \) takes on a value between two real constants \( a \) and \( b \) is:

\[
P(a \leq R \leq b) = P(\{s : a \leq R(s) \leq b\})
\]
Bayesian approach provides mathematical rule explaining how you should change your existing beliefs in the light of new evidence.
Bayes’s theorem

\[ \text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}} \]
Bayes’s theorem

- posterior = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}}

- \( P(R = r | e) = \frac{P(e|R=r)P(R=r)}{P(e)} \)

- \( P(R = r | e) \): probability that random variable \( R \) has value \( r \) given evidence \( e \)
Bayes’s theorem

- **posterior** = \( \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}} \)

- \( P(R = r|e) = \frac{P(e|R=r)P(R=r)}{P(e)} \)

- \( P(R = r|e) \): probability that random variable \( R \) has value \( r \) given evidence \( e \)

- The denominator is just a normalizing constant (called **marginal likelihood**) that ensures the posterior adds up to 1; it can be computed by summing up the numerator over all possible values of \( R \), i.e.,
  \[
P(e) = P(R = 0, e) + P(R = 1, e) + \ldots = \sum_r P(e|R = r)P(R = r)
  \]