Probability basics

DS GA 1002 Probability and Statistics for Data Science
http://www.cims.nyu.edu/~cfgranda/pages/DSGA1002_fall17

Carlos Fernandez-Granda
Probability spaces

Conditional probability

Independence
General approach

Probabilistic modeling

1. Model phenomenon of interest as an experiment with several (possibly infinite) mutually exclusive outcomes

2. Group these outcomes in sets called events

3. Assign probabilities to the different events
A probability space is a triple \((\Omega, \mathcal{F}, P)\) consisting of

- A **sample space** \(\Omega\), which contains all possible outcomes of the experiment
- A set of events \(\mathcal{F}\), which must be a \(\sigma\)-algebra
- A **probability measure** \(P\) that assigns probabilities to the events in \(\mathcal{F}\)
Sample spaces can be

- **Discrete**: coin toss, score of a basketball game, number of people that show up at a party . . .

- **Continuous**: intervals of $\mathbb{R}$ or $\mathbb{R}^n$ used to model time, position, temperature, . . .
A $\sigma$-algebra $\mathcal{F}$ is a collection of sets in $\Omega$ such that

1. If a set $S \in \mathcal{F}$ then $S^c \in \mathcal{F}$

2. If the sets $S_1, S_2 \in \mathcal{F}$, then $S_1 \cup S_2 \in \mathcal{F}$
   Also infinite sequences; if $S_1, S_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} S_i \in \mathcal{F}$

3. $\Omega \in \mathcal{F}$
Basketball game

- Cleveland Cavaliers are playing the Golden State Warriors

- Sample space

  \[ \Omega := \{ \text{Cavs 1} - \text{Warriors 0}, \text{Cavs 0} - \text{Warriors 1}, \ldots, \text{Cavs 101} - \text{Warriors 97}, \ldots \}. \]

- Several possible \( \sigma \)-algebras

  - If we want high granularity we can choose the power set of scores
  - If we only care who wins

    \[ \mathcal{F} := \{ \text{Cavs win, Warriors win, Cavs or Warriors win, } \emptyset \}. \]
Probability measure

Function over the sets in $\mathcal{F}$ such that

1. $P(S) \geq 0$ for any event $S \in \mathcal{F}$

2. If $S_1, S_2 \in \mathcal{F}$ are disjoint then

\[ P(S_1 \cup S_2) = P(S_1) + P(S_2) \]

Also countably infinite sequences of disjoint sets: $S_1, S_2, \ldots \in \mathcal{F}$

\[ P\left( \lim_{n \to \infty} \bigcup_{i=1}^{n} S_i \right) = \lim_{n \to \infty} \sum_{i=1}^{n} P(S_i) \]

3. $P(\Omega) = 1$
Properties of a probability measure

- \( P(\emptyset) = 0 \)
- If \( A \subseteq B \) then \( P(A) \leq P(B) \)
- \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \)
- To simplify notation, we set
  \[
P(A, B, C) := P(A \cap B \cap C)
\]
Important

- Probability measure only assigns probabilities to events in the $\sigma$-algebra

- Simpler $\sigma$-algebras can make our life easy

\[
P(\text{Cavs win}) = \frac{1}{2}
\]

\[
P(\text{Warriors win}) = \frac{1}{2}
\]

\[
P(\text{Cavs or Warriors win}) = 1
\]

\[
P(\emptyset) = 0
\]
Probability spaces

Conditional probability

Independence
The conditional probability of an event $S' \in \mathcal{F}$ given $S$ is

$$P(S'|S) := \frac{P(S' \cap S)}{P(S)}$$

$P(\cdot|S)$ is a valid probability measure.

To simplify notation, we set

$$P(S|A, B, C) := P(S|A \cap B \cap C)$$
Example: Flights and rain

Probabilistic model for late arrivals at an airport

\[ \Omega = \{ \text{late and rain, late and no rain,} \]
\[ \quad \text{on time and rain, on time and no rain} \} \]

\[ \mathcal{F} = \text{power set of } \Omega, \]

\[ P(\text{late, no rain}) = \frac{2}{20}, \quad P(\text{on time, no rain}) = \frac{14}{20}, \]

\[ P(\text{late, rain}) = \frac{3}{20}, \quad P(\text{on time, rain}) = \frac{1}{20} \]
Example: Flights and rain

\[
P(\text{late}|\text{rain})
\]
Example: Flights and rain

$$P(\text{late}|\text{rain}) = \frac{P(\text{late, rain})}{P(\text{rain})}$$
Example: Flights and rain

\[
P(\text{late}|\text{rain}) = \frac{P(\text{late, rain})}{P(\text{rain})}
\]

\[
= \frac{P(\text{late, rain})}{P(\text{late, rain}) + P(\text{on time, rain})}
\]
Example: Flights and rain

\[
P(\text{late} | \text{rain}) = \frac{P(\text{late, rain})}{P(\text{rain})} = \frac{P(\text{late, rain})}{P(\text{late, rain}) + P(\text{on time, rain})} = \frac{3}{4}
\]
Chain rule

For any pair of events $A$ and $B$

$$P(A \cap B) = P(A) P(B|A) = P(B) P(A|B)$$

For any sequence of events $S_1, S_2, S_3, \ldots$

$$P(\cap_i S_i) = P(S_1) P(S_2|S_1) P(S_3|S_1 \cap S_2) \ldots$$

$$= \prod_i P(S_i|\cap_{j=1}^{i-1} S_j)$$
**Law of Total Probability**

If $A_1, A_2, \ldots \in \mathcal{F}$ is a partition of $\Omega$

- $A_i$ and $A_j$ are disjoint if $i \neq j$
- $\Omega = \bigcup_i A_i$

For any set $S \in \mathcal{F}$

$$P(S) = \sum_i P(S \cap A_i) = \sum_i P(A_i) P(S | A_i)$$
Example: Flights and rain (continued)

\[
\begin{align*}
P(\text{rain}) &= 0.2 \\
P(\text{late}|\text{rain}) &= 0.75 \\
P(\text{late}|\text{no rain}) &= 0.125 \\
P(\text{late}) &= \frac{P(\text{rain})P(\text{late}|\text{rain}) + P(\text{no rain})P(\text{late}|\text{no rain})}{P(\text{rain}) + P(\text{no rain})} \\
&= \frac{0.2 \cdot 0.75 + 0.8 \cdot 0.125}{0.2 + 0.8} \\
&= 0.25
\end{align*}
\]
Example: Flights and rain (continued)

\[
\begin{align*}
P(\text{rain}) &= 0.2 \\
P(\text{late} | \text{rain}) &= 0.75 \\
P(\text{late} | \text{no rain}) &= 0.125
\end{align*}
\]

\[
P(\text{late}) = P(\text{rain}) P(\text{late} | \text{rain}) + P(\text{no rain}) P(\text{late} | \text{no rain})
\]
Example: Flights and rain (continued)

\[
\begin{align*}
P(\text{rain}) &= 0.2 \\
P(\text{late}|\text{rain}) &= 0.75 \\
P(\text{late}|\text{no rain}) &= 0.125
\end{align*}
\]

\[
P(\text{late}) = P(\text{rain}) \cdot P(\text{late}|\text{rain}) + P(\text{no rain}) \cdot P(\text{late}|\text{no rain})
\]

\[
= 0.2 \cdot 0.75 + 0.8 \cdot 0.125 = 0.25
\]
Important!

\[ P(A \mid B) \neq P(B \mid A) \]
Bayes’ Rule

Let $A_1, A_2, \ldots \in \mathcal{F}$ be a partition of $\Omega$

For any set $S \in \mathcal{F}$

$$P(A_i|S) = \frac{P(A_i)P(S|A_i)}{\sum_j P(S|A_j)P(A_j)}$$
Example: Flights and rain (continued)

\[
\begin{align*}
\Pr(\text{rain}) &= 0.2 \\
\Pr(\text{late}|\text{rain}) &= 0.75 \\
\Pr(\text{late}|\text{no rain}) &= 0.125
\end{align*}
\]

\[
\Pr(\text{rain}|\text{late})
\]
Example: Flights and rain (continued)

\[ P(\text{rain}|\text{late}) \]
Example: Flights and rain (continued)

\[ P(\text{rain}|\text{late}) = \frac{P(\text{rain, late})}{P(\text{late})} \]
Example: Flights and rain (continued)

\[ P(\text{rain}|\text{late}) = \frac{P(\text{rain}, \text{late})}{P(\text{late})} \]

\[ = \frac{P(\text{late}|\text{rain}) P(\text{rain})}{P(\text{late}|\text{rain}) P(\text{rain}) + P(\text{late}|\text{no rain}) P(\text{no rain})} \]
Example: Flights and rain (continued)

\[
P(\text{rain}|\text{late}) = \frac{P(\text{rain}, \text{late})}{P(\text{late})} \\
= \frac{P(\text{late}|\text{rain}) P(\text{rain})}{P(\text{late}|\text{rain}) P(\text{rain}) + P(\text{late}|\text{no rain}) P(\text{no rain})} \\
= \frac{0.75 \cdot 0.2}{0.75 \cdot 0.2 + 0.125 \cdot 0.8} = 0.6
\]
Probability spaces

Conditional probability

Independence
Two sets $A, B$ are independent if

$$P(A|B) = P(A)$$

or equivalently

$$P(A \cap B) = P(A)P(B)$$
Conditional independence

$A, B$ are conditionally independent given $C$ if

$$P(A|B, C) = P(A|C)$$

where $P(A|B, C) := P(A|B \cap C)$, or equivalently

$$P(A \cap B|C) = P(A|C) P(B|C)$$
Conditional independence does not imply independence

Probabilistic model for taxi availability, flight delay and weather

\[
\begin{align*}
P(\text{rain}) &= 0.2 \\
P(\text{late} | \text{rain}) &= 0.75 \\
P(\text{late} | \text{no rain}) &= 0.125 \\
P(\text{taxi} | \text{rain}) &= 0.1 \\
P(\text{taxi} | \text{no rain}) &= 0.6
\end{align*}
\]

Given \textit{rain} and \textit{no rain}, \textit{late} and \textit{taxi} are conditionally independent

Are they also independent? \( P(\text{taxi}) = P(\text{taxi} | \text{late}) \)?
Conditional independence does not imply independence

\[ P(\text{taxi}) = P(\text{taxi} | \text{rain}) P(\text{rain}) + P(\text{taxi} | \text{no rain}) P(\text{no rain}) = 0.1 \cdot 0.2 + 0.6 \cdot 0.8 = 0.5 \]

\[ P(\text{taxi} | \text{late}) = P(\text{taxi}, \text{late}) P(\text{late}) = P(\text{taxi}, \text{late}, \text{rain}) P(\text{late}, \text{rain}) + P(\text{taxi}, \text{late}, \text{no rain}) P(\text{late}, \text{no rain}) = P(\text{taxi} | \text{late}, \text{rain}) P(\text{late} | \text{rain}) P(\text{rain}) + P(\text{taxi} | \text{late}, \text{no rain}) P(\text{late} | \text{no rain}) P(\text{no rain}) = 0.25 \]

They are not independent
Conditional independence does not imply independence

\[ P(\text{taxi}) = P(\text{taxi}|\text{rain}) P(\text{rain}) + P(\text{taxi}|\text{no rain}) P(\text{no rain}) \]
Conditional independence does not imply independence

\[ P(\text{taxi}) = P(\text{taxi}|\text{rain}) P(\text{rain}) + P(\text{taxi}|\text{no rain}) P(\text{no rain}) \]
\[ = 0.1 \cdot 0.2 + 0.6 \cdot 0.8 = 0.5 \]
Conditional independence does not imply independence

\[
P(\text{taxi}) = P(\text{taxi}|\text{rain})P(\text{rain}) + P(\text{taxi}|\text{no rain})P(\text{no rain})
\]
\[
= 0.1 \cdot 0.2 + 0.6 \cdot 0.8 = 0.5
\]

\[
P(\text{taxi}|\text{late})
\]
Conditional independence does not imply independence

\[
P(taxi) = P(taxi|rain) \cdot P(rain) + P(taxi|no \ rain) \cdot P(no \ rain)
\]
\[
= 0.1 \cdot 0.2 + 0.6 \cdot 0.8 = 0.5
\]

\[
P(taxi|late) = \frac{P(taxi, late)}{P(late)}
\]
Conditional independence does not imply independence

\[ P(\text{taxi}) = P(\text{taxi}|\text{rain}) P(\text{rain}) + P(\text{taxi}|\text{no rain}) P(\text{no rain}) \]
\[ = 0.1 \cdot 0.2 + 0.6 \cdot 0.8 = 0.5 \]

\[ P(\text{taxi}|\text{late}) = \frac{P(\text{taxi, late})}{P(\text{late})} \]
\[ = \frac{P(\text{taxi, late, rain}) + P(\text{taxi, late, no rain})}{P(\text{late})} \]
Conditional independence does not imply independence

\[ P(\text{taxi}) = P(\text{taxi} | \text{rain}) P(\text{rain}) + P(\text{taxi} | \text{no rain}) P(\text{no rain}) = 0.1 \cdot 0.2 + 0.6 \cdot 0.8 = 0.5 \]

\[ P(\text{taxi} | \text{late}) = \frac{P(\text{taxi}, \text{late})}{P(\text{late})} = \frac{P(\text{taxi}, \text{late}, \text{rain}) + P(\text{taxi}, \text{late}, \text{no rain})}{P(\text{late})} \]

\[ = \frac{P(\text{t}|l, r) P(l|r) P(r) + P(\text{t}|l, \text{no r}) P(l|\text{no r}) P(\text{no r})}{P(l)} \]
Conditional independence does not imply independence

\[ P(\text{taxi}) = P(\text{taxi}|\text{rain}) P(\text{rain}) + P(\text{taxi}|\text{no rain}) P(\text{no rain}) \]

\[ = 0.1 \cdot 0.2 + 0.6 \cdot 0.8 = 0.5 \]

\[ P(\text{taxi}|\text{late}) = \frac{P(\text{taxi}, \text{late})}{P(\text{late})} \]

\[ = \frac{P(\text{taxi}, \text{late}, \text{rain}) + P(\text{taxi}, \text{late}, \text{no rain})}{P(\text{late})} \]

\[ = \frac{P(t|l, r) P(l|r) P(r) + P(t|l, \text{no } r) P(l|\text{no } r) P(\text{no } r)}{P(l)} \]

\[ = \frac{P(\text{taxi}|r) P(\text{late}|r) P(r) + P(\text{taxi}|\text{no } r) P(\text{late}|\text{no } r) P(\text{no } r)}{P(\text{late})} \]
Conditional independence does not imply independence

\[ P(\text{taxi}) = P(\text{taxi}|\text{rain}) P(\text{rain}) + P(\text{taxi}|\text{no rain}) P(\text{no rain}) \]
\[ = 0.1 \cdot 0.2 + 0.6 \cdot 0.8 = 0.5 \]

\[ P(\text{taxi}|\text{late}) = \frac{P(\text{taxi, late})}{P(\text{late})} \]
\[ = \frac{P(\text{taxi, late, rain}) + P(\text{taxi, late, no rain})}{P(\text{late})} \]
\[ = \frac{P(\text{t}|\text{l}, r) P(\text{l}|r) P(\text{r}) + P(\text{t}|\text{l}, \text{no r}) P(\text{l}|\text{no r}) P(\text{no r})}{P(\text{l})} \]
\[ = \frac{P(\text{taxi}|\text{r}) P(\text{late}|\text{r}) P(\text{r}) + P(\text{taxi}|\text{no r}) P(\text{late}|\text{no r}) P(\text{no r})}{P(\text{late})} = \frac{0.1 \cdot 0.75 \cdot 0.2 + 0.6 \cdot 0.125 \cdot 0.8}{0.25} = 0.3 \]
Conditional independence does not imply independence

\[
P(\text{taxi}) = P(\text{taxi}|\text{rain}) P(\text{rain}) + P(\text{taxi}|\text{no rain}) P(\text{no rain})
\]
\[
= 0.1 \cdot 0.2 + 0.6 \cdot 0.8 = 0.5
\]

\[
P(\text{taxi}|\text{late}) = \frac{P(\text{taxi, late})}{P(\text{late})}
\]
\[
= \frac{P(\text{taxi, late, rain}) + P(\text{taxi, late, no rain})}{P(\text{late})}
\]
\[
= \frac{P(\text{t}|l, r) P(l|r) P(r) + P(\text{t}|l, \text{no r}) P(l|\text{no r}) P(\text{no r})}{P(l)}
\]
\[
= \frac{P(\text{taxi}|r) P(\text{late}|r) P(r) + P(\text{taxi}|\text{no r}) P(\text{late}|\text{no r}) P(\text{no r})}{P(\text{late})}
\]
\[
= \frac{0.1 \cdot 0.75 \cdot 0.2 + 0.6 \cdot 0.125 \cdot 0.8}{0.25} = 0.3
\]

They are not independent
Independence does not imply conditional independence

Probabilistic model for mechanical problems, weather and delays

\[
P(\text{rain}) = 0.2 \\
P(\text{late}|\text{rain}) = 0.75 \\
P(\text{late}|\text{no rain}) = 0.125 \\
P(\text{problem}) = 0.1 \\
P(\text{late}|\text{problem}) = 0.7 \\
P(\text{late}|\text{no problem}) = 0.2 \\
P(\text{late}|\text{no rain, problem}) = 0.5
\]

\textit{problem} and \textit{no rain} are independent

Are they also conditionally independent given \textit{late}? \\
P(\text{problem}|\text{late, no rain}) = P(\text{problem}|\text{late}) ?
Independence does not imply conditional independence

\[ P(\text{problem}|\text{late}) \]
Independence does not imply conditional independence

\[ P(\text{problem}|\text{late}) = \frac{P(\text{late, problem})}{P(\text{late})} \]
Independence does not imply conditional independence

\[
\begin{align*}
P(\text{problem}|\text{late}) &= \frac{P(\text{late}, \text{problem})}{P(\text{late})} \\
&= \frac{P(\text{late}|p)P(p)}{P(\text{late}|p)P(p) + P(\text{late}|\text{no } p)P(\text{no } p)}
\end{align*}
\]
Independence does not imply conditional independence

\[
P(\text{problem}|\text{late}) = \frac{P(\text{late, problem})}{P(\text{late})} = \frac{P(\text{late}|p)P(p)}{P(\text{late}|p)P(p) + P(\text{late}|\text{no p})P(\text{no p})} = \frac{0.7 \cdot 0.1}{0.7 \cdot 0.1 + 0.2 \cdot 0.9} = 0.28
\]
Independence does not imply conditional independence

\[
P(\text{problem}|\text{late}) = \frac{P(\text{late, problem})}{P(\text{late})} \\
= \frac{P(\text{late}|p)P(p)}{P(\text{late}|p)P(p) + P(\text{late}|\text{no p})P(\text{no p})} \\
= \frac{0.7 \cdot 0.1}{0.7 \cdot 0.1 + 0.2 \cdot 0.9} = 0.28
\]

\[
P(\text{problem}|\text{late, no rain})
\]
Independence does not imply conditional independence

\[
P(\text{problem}|\text{late}) = \frac{P(\text{late, problem})}{P(\text{late})} \\
= \frac{P(\text{late}|p)P(p)}{P(\text{late}|p)P(p) + P(\text{late}|\text{no p})P(\text{no p})} = \frac{0.7 \cdot 0.1}{0.7 \cdot 0.1 + 0.2 \cdot 0.9} = 0.28
\]

\[
P(\text{problem}|\text{late, no rain}) = \frac{P(\text{late, no rain, problem})}{P(\text{late, no rain})}
\]
Independence does not imply conditional independence

\[
P(\text{problem}|\text{late}) = \frac{P(\text{late, problem})}{P(\text{late})} = \frac{P(\text{late}|p)P(p)}{P(\text{late}|p)P(p) + P(\text{late}|\neg p)P(\neg p)} = \frac{0.7 \cdot 0.1}{0.7 \cdot 0.1 + 0.2 \cdot 0.9} = 0.28
\]

\[
P(\text{problem}|\text{late, no rain}) = \frac{P(\text{late, no rain, problem})}{P(\text{late, no rain})} = \frac{P(\text{late|no rain, p})P(\text{no rain|p})P(p)}{P(\text{late|no rain})P(\text{no rain})}
\]
Independence does not imply conditional independence

\[ P(\text{problem}|\text{late}) = \frac{P(\text{late, problem})}{P(\text{late})} = \frac{P(\text{late}|p)P(p)}{P(\text{late}|p)P(p) + P(\text{late}|\text{no } p)P(\text{no } p)} = \frac{0.7 \cdot 0.1}{0.7 \cdot 0.1 + 0.2 \cdot 0.9} = 0.28 \]

\[ P(\text{problem}|\text{late, no rain}) = \frac{P(\text{late, no rain, problem})}{P(\text{late, no rain})} = \frac{P(\text{late}|\text{no rain}, p)P(\text{no rain}|p)P(p)}{P(\text{late}|\text{no rain})P(\text{no rain})} = \frac{P(\text{late}|\text{no rain, problem})P(\text{no rain})P(\text{problem})}{P(\text{late}|\text{no rain})P(\text{no rain})} \]
Independence does not imply conditional independence

$$P(\text{problem}|\text{late}) = \frac{P(\text{late, problem})}{P(\text{late})}$$

$$= \frac{P(\text{late}|\text{p}) P(\text{p})}{P(\text{late}|\text{p}) P(\text{p}) + P(\text{late}|\text{no p}) P(\text{no p})}$$

$$= \frac{0.7 \cdot 0.1}{0.7 \cdot 0.1 + 0.2 \cdot 0.9} = 0.28$$

$$P(\text{problem}|\text{late, no rain}) = \frac{P(\text{late, no rain, problem})}{P(\text{late, no rain})}$$

$$= \frac{P(\text{late|no rain, p}) P(\text{no rain|p}) P(\text{p})}{P(\text{late|no rain}) P(\text{no rain})}$$

$$= \frac{0.5 \cdot 0.1}{0.125} = 0.4$$