TODAY’S LECTURE

Data collection → Data processing → Exploratory analysis & Data viz → Analysis, hypothesis testing, & ML → Insight & Policy Decision

BIG THANKS: Zico Kolter (CMU) & Amol Deshpande (UMD)
Statistical inference is the discipline that concerns itself with the development of procedures, methods, and theorems that allow us to extract meaning and information from data that has been generated by stochastic (random) processes.

- Process of going from the world to the data, and then back to the world
- Often the goal is to develop a statistical model of the world from observed data

**Conclusion is typically:**

- an estimate;
- or a confidence interval;
- or rejection of a hypothesis
- or clustering or classification of data points into groups
Probability is concerned with the outcome of a trial (also called experiment or observation)

Sample Space: Set of all possible outcomes of a trial
- Probability of Sample Space = 1

Event is the specification of the outcome of a trial
- For example: Trial = Tossing a coin; Sample Space = {Heads, Tails}; Event = Heads

If two events E and F are independent, then: Probability of E does not change if F has already happened = \( P(E) \), i.e., \( P(E \mid F) = P(E) \)

Also: \( P(E \text{ AND } F) = P(E) \times P(F) \)

If two events E and F are mutually exclusive, then: \( P(E \text{ UNION } F) = P(E) + P(F) \)
BASIC PROBABILITY II

Bayes Theorem \( P(A \mid B) = P(B \mid A) \times P(A) / P(B) \)
Simple equation, but fundamental to Bayesian inference
Conditional Independence: A and B are conditionally independent given C if: \( \Pr(A \text{ AND } B \mid C) = \Pr(A \mid C) \times \Pr(B \mid C) \)
Powerful in reducing the computational efforts in storing and manipulating large joint probability distributions
Entropy: A measure of the uncertainty in a probability distribution

Wikipedia Article
Informal definition of probability

- Probability that event happens:
Informal definition of probability

• Probability that event happens:
  \[
  \frac{\text{# of possibilities that an event happens}}{\text{# of all possibilities}}
  \]

• This definition is owed to Andrey Kolmogorov, and assumes that all possibilities are equally likely!
First examples

• Experiment #1: Tossing the same coin 3 times.
First examples

- Experiment #1: Tossing the same coin 3 times.
  - What is the probability that I don’t get any heads?
First examples

• Experiment #1: Tossing the same coin 3 times.
  • What is the probability that I don’t get any heads?
  • Why?
    • Set of different events?
      • (8 of them)
    • Set of events with no heads:
      • (1 of them)
    • Hence the answer:
First examples

• Experiment #1: Tossing the same coin 3 times.
  • What is the probability that I don’t get any heads?
  • Why?
    • Set of different events?
      • (8 of them)
    • Set of events with no heads:
      • (1 of them)
  • Hence the answer: 1/8

Implicit assumption: all individual outcomes (HHH, HHT, HTH, ....) are considered equally likely (probability 1/8)
Practice

• Experiment #2: I roll two dice.
  • Probability that I hit seven = ?
Practice

• Experiment #2: I roll two dice.
  • Probability that I hit seven = ?
• Why?
  • Set of different events?
    • (36 of them)
  • Set of events where we hit 7.
    • (6 of them)
  • Hence the answer:
Practice

• Experiment #2: I roll two dice.
  • Probability that I hit seven = ?
  • Why?
    • Set of different events?
      • (36 of them)
    • Set of events where we hit 7.
      • (6 of them)
    • Hence the answer: 1/6
  • Probability that I hit two = ?
Practice

• Experiment #2: I roll two dice.
  • Probability that I hit seven = ?
  • Why?
    • Set of different events?
      • (36 of them)
    • Set of events where we hit 7.
      • (6 of them)
    • Hence the answer:
  • Probability that I hit two = \(\frac{1}{36}\)
    • Same procedure
Joint probability ("AND" of two events)

• The probability that two events A and B occur simultaneously is known as the joint probability of A and B and is denoted in number of ways:
  • Set theoretic perspective, \( P(A \cap B) \)
  • More commonly used in Physics, \( P(A, B) \)
  • More common, \( P(AB) \)
Calculating joints

• Probability that the first coin toss is heads and the second coin toss is tails
Calculating joints

• Probability that the first coin toss is heads and the second coin toss is tails \( \frac{1}{2} \times \frac{1}{2} \)
Calculating joints

- Probability that the first coin toss is heads and the second coin toss is tails
  \[ \frac{1}{2} \times \frac{1}{2} \]
- Probability that the first die is at most a 2 and the second one is 5 or 6
Calculating joints

- Probability that the first coin toss is heads and the second coin toss is tails
  \[ \frac{1}{2} \times \frac{1}{2} \]

- Probability that the first die is at most a 2 and the second one is 5 or 6
  - # outcomes of die roll is
  - # outcomes where first die is at most 2 is 2
    - Hence, probability of first die roll being at most 2 is
Calculating joints

• Probability that the first coin toss is heads and the second coin toss is tails \(\frac{1}{2} \times \frac{1}{2}\)

• Probability that the first die is at most a 2 and the second one is 5 or 6
  • # outcomes of die roll is
  • # outcomes where first die is at most 2 is 2
    • Hence, probability of first die roll being at most 2 is \(\frac{1}{3}\)
  • Similarly, probability of second die roll being 5 or 6 is \(\frac{1}{3}\)
  • Hence, probability that both events happen (joint probability) is \(\frac{1}{9}\)
Calculating joints

• Flip a coin and then pick a card from a 52-card deck.
  • Probability that the coin is heads and the card has rank 8?
The law of joint probability

\[ P(A \cap B) = P(A) \cdot P(B) \]

\[ P(A_1 \cap A_2 \cap \ldots \cap A_n) = \prod_{i=1}^{n} P(A_i) \]
The law of joint probability

\[ P(A \cap B) = P(A) \cdot P(B) \]

\[ P(A_1 \cap A_2 \cap \ldots \cap A_n) = \prod_{i=1}^{n} P(A_i) \]

• Unfortunately, this “law” is not always applicable!
• It is applicable only when all the different events are independent (sometimes called marginally independent) of each other.
• Let’s look at an example.
Calculating joints of dependent events

- Probability that a die is even and that it is 2.
Calculating joints of dependent events

• Probability that a die is even and that it is 2.
  • Probability that the die is even = $\frac{1}{2}$
Calculating joints of dependent events

• Probability that a die is even and that it is 2.
  • Probability that the die is even = \( \frac{1}{2} \)
  • Probability that the die is two = \( \frac{1}{6} \)
Calculating joints of dependent events

- Probability that a die is even and that it is 2.
  - Probability that the die is even = $\frac{1}{2}$
  - Probability that the die is two = $\frac{1}{6}$
  - Probability the die is even and the die is two = $\frac{1}{12}$
Calculating joints

• A university offers a Course where the possible grades are A through G. (No + or -)
• What is the probability that a student gets both an A and a G in that course?
Calculating joints

• A university offers a Course where the possible grades are A through G. (No + or -)

• What is the probability that a student gets both an A and a G in that course?
  • Clearly, it can’t be

\[
\text{Probability of an A} \times \text{Probability of a G} = \frac{1}{7} \times \frac{1}{7} = \frac{1}{49}
\]
Calculating joints

- The University of Southern North Dakota offers a Discrete Mathematics Course where the possible grades are A through G. (No + or -)

- What is the probability that Jason gets both an A and a G in that course?
  - Clearly, it can’t be

\[
\text{Probability of an A} \times \text{Probability of a G} = \frac{1}{7} \times \frac{1}{7} = \frac{1}{49}
\]
Calculating joints

- The University of Southern North Dakota offers a Discrete Mathematics Course where the possible grades are A through G. (No + or -)
- What is the probability that Jason gets both an A and a G in that course?
  - Clearly, it can’t be
    - Probability of an A $\times$ Probability of a G $= \frac{1}{7} \times \frac{1}{7} = \frac{1}{49}$
  - It is 0. Those two events cannot happen jointly!
Calculating joints

- The University of Southern North Dakota offers a Discrete Mathematics Course where the possible grades are A through G. (No + or -)
- What is the probability that Jason gets both an A and a G in that course?
  - Clearly, it can’t be
    - Probability of an A x Probability of a G = \( \frac{1}{7} \times \frac{1}{7} = \frac{1}{49} \)
  - It is 0. Those two events cannot happen **jointly**!
  - Events such as these are called **disjoint or mutually disjoint**.
Recap: “Disjoint” vs “independent”

<table>
<thead>
<tr>
<th>Disjoint</th>
<th>Independent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has a set-theoretic interpretation!</td>
<td>Has a causality interpretation!</td>
</tr>
<tr>
<td>Means that $P(A \cap B) = 0$</td>
<td>Means that $P(A \cap B) = P(A) \cdot P(B)$</td>
</tr>
<tr>
<td>Means that $P(A \cup B) = P(A) + P(B)$</td>
<td>Means that $P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B)$</td>
</tr>
</tbody>
</table>
Conditional Probability

• If A occurs, then is B
  a) More likely?
  b) Equally likely?
  c) Less likely?
Conditional Probability

• If A occurs, then is B
  a) More likely?
  b) Equally likely?
  c) Less likely?

• Any of these could happen, it depends on the relationship between A and B.
Conditional Probability

• If A occurs, then is B
  a) More likely?
  b) Equally likely?
  c) Less likely?

• Any of these could happen, it depends on the relationship between A and B.
Examples

• We roll two dice
  • Event A = “Sum of the dice is a multiple of 4”
    • Note that, since we have nine rolls of the dice that sum to a multiple of 4:
      (1, 3), (2, 2), (3, 1), (2, 6), (3, 5), (4, 4), (5, 3), (6, 2), (6, 6)
  • Event B = “The first die comes up 3”
    • Note that \( P(B) = \frac{6}{36} = \frac{1}{6} \)
Examples

• We roll two dice
  • Event A = “Sum of the dice
    • Note that, since we have nine rolls of the dice that sum to a multiple of 4:
      (1, 3), (2, 2), (3, 1), (2, 6), (3, 5), (4, 4), (5, 3), (6, 2), (6, 6)
  • Event B = “The first die comes up 3”
    • Note that $P(B) = \frac{6}{36} = \frac{1}{6}$

• What is the probability of A given B?
Examples

• What is the probability of A given B?
Examples

• What is the probability of A given B?
  • Outcomes of A are (1, 3), (2, 2), (3, 1), (2, 6), (3, 5), (4, 4), (5, 3), (6, 2), (6, 6)
  • Outcomes of B are (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)
  • Outcomes of rolling two dice: (1, 1), (1, 2), ...., (6, 5), (6, 6)
Examples

• What is the probability of $A$ given $B$?
  • Outcomes of $A$ are $(1, 3), (2, 2), (3, 1), (2, 6), (3, 5), (4, 4), (5, 3), (6, 2), (6, 6)$
  • Outcomes of $B$ are $(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)$
  • Outcomes of rolling two dice: $(1, 1), (1, 2), \ldots, (6, 5), (6, 6)$
Examples

• What is the probability of A given B?
  • Outcomes of A are (1, 3), (2, 2), (3, 1), (2, 6), (3, 5), (4, 4), (5, 3), (6, 2), (6, 6)
  • Outcomes of B are (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)
  • Outcomes of rolling two dice: (1, 1), (1, 2), …. (6, 5), (6, 6)

• As discussed,

• However, once B occurs, instead of 36 outcomes, we now have...
Examples

• What is the probability of A given B?
  • Outcomes of A are (1, 3), (2, 2), (3, 1), (2, 6), (3, 5), (4, 4),
    (5, 3), (6, 2), (6, 6)
    • Outcomes of B are (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)
  • Outcomes of rolling two dice: (1, 1), (1, 2), …. , (6, 5), (6, 6)

• As discussed,
  • However, once B occurs, instead of 36 outcomes, we now have... 6 outcomes.
Examples

• What is the probability of A given B?
  • Outcomes of A are (1, 3), (2, 2), (3, 1), (2, 6), (3, 5), (4, 4), (5, 3), (6, 2), (6, 6)
  • Outcomes of B are (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)
  • Outcomes of rolling two dice: (1, 1), (1, 2), ...., (6, 5), (6, 6)

• As discussed,
  • However, once B occurs, instead of 36 outcomes, we now have... 6 outcomes.
    • Only 2 of them are outcomes that correspond to A.
Examples

• What is the probability of **A given B**?
  • Outcomes of A are (1, 3), (2, 2), (3, 1), (2, 6), (3, 5), (4, 4), (5, 3), (6, 2), (6, 6)
  • Outcomes of B are (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)
  • Outcomes of rolling two dice: (1, 1), (1, 2), ...., (6, 5), (6, 6)

• As discussed,
  • However, once B occurs, instead of 36 outcomes, we now have... 6 outcomes.
    • Only 2 of them are outcomes that correspond to A.
    • Therefore, the probability of **A given B** is \( \frac{2}{6} = \frac{1}{3} \)
Examples

• We once again roll two dice
  • Event A = “Sum of the dice is $\geq 8$”
  • Event B = “First die is 4”
Examples

• We once again roll two dice
  • Event A = “Sum of the dice is ≥ 8”
  • Event B = “First die is 4”

• If B happens, what is your intuition about the probability of A?
Examples

• We once again two roll dice
  • Event A = “Sum of the dice is $\geq 8$”
  • Event B = “First die is 4”

• If B happens, what is your intuition about the probability of A?
Examples

• We once again two roll dice
  • Event A = “Sum of the dice is $\geq 8”$
  • Event B = “First die is 4”

• If B happens, what is your intuition about the probability of A?

Let’s see if your intuition was correct!
Examples

• We once again two roll dice
  • Event A = “Sum of the dice is $\geq 8$” (work on it)
  • Event B = “First die is 4”
Examples

• We once again two roll dice
  • Event A = “Sum of the dice is $\geq 8$” $P(A) = \frac{15}{36} = \frac{5}{12}$
  • Event B = “First die is a 4”
Examples

• We once again two roll dice
  • Event A = “Sum of the dice is $\geq 8$”
  • Event B = “First die is a 4”

\[ P(A) = \frac{15}{36} = \frac{5}{12} \]

\[ P(B) = \frac{1}{6} \]
Examples

• We once again two roll dice
  • Event A = “Sum of the dice is $\geq 8$”
  • Event B = “First die is a 4” $P(B) = \frac{1}{6}$

• Prob of A given B = Prob second dice is 4, 5, or 6

\[
P(A) = \frac{15}{36} = \frac{5}{12}
\]

$\frac{3}{6} = \frac{1}{2} > \frac{5}{12}$
Conditional probability

• Let $A$ and $B$ be two events. The conditional probability of $A$ given $B$, denoted $P(A|B)$, is defined as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
Re-thinking independent events

• **Alternative definition of independent events**: Two events $A$ and $B$ will be called marginally independent, or just independent for short, if and only if

\[ P(A | B) = P(A) \]
Re-thinking independent events

• **Alternative definition of independent events**: Two events $A$ and $B$ will be called marginally independent, or just independent for short, if and only if

\[
P(A|B) = P(A)
\]

• Applying the definition of $P(A|B)$ we have:

\[
\frac{P(A \cap B)}{P(B)} = P(A) \implies P(A \cap B) = P(A) \cdot P(B),
\]

which is a relationship we had reached earlier when discussing the joint probability.
RECALL: NORMAL DISTRIBUTION

99.7% values will fall within 3 standard deviations (around the mean)
- 95% for 2 standard deviations;
  68% for 1

Central Limit Theorem: As sample size approaches infinity, distribution of sample means will follow a normal distribution irrespective of the original distribution.
HYPOTHESIS TESTING

Accepting or rejecting a statistical hypothesis about a population

H_0: null hypothesis, and H_1: the alternative hypothesis
• Mutually exclusive and exhaustive
• H_0 can never be proven to be true, but can be rejected
• Sometimes don’t have H_1 at all (Fisher’s test)

Statistical significance: probability that the result is not due to chance

Example: Deciding if a coin is fair
• http://20bits.com/article/hypothesis-testing-the-basics
HYPOTHESIS TESTING

$H_0$: *null hypothesis*, and $H_1$: the *alternative hypothesis*

- Mutually exclusive and exhaustive
- $H_0$ can never be proven to be true

**Statistical significance:** probability that the result is not due to chance

**Process:**
- Decide on $H_0$ and $H_1$
- Decide which test statistic is appropriate
  - Roughly, how well does my sample agree with the null hypothesis?
  - Key question: what is the distribution of the test statistic over samples?
- Select a significance level (sigma), a probability threshold below which the null hypothesis will be rejected -- typically 5% or 1%.
- Compute the observed value of the test statistic $t_{obs}$ from the sample
- Compute p-value: the probability that the test statistic took that value by chance
  - Use the distribution above to compute the *p-value*
- Reject the null hypothesis if the *p-value* $< \sigma$
Hypothesis testing allows us to formulate beliefs about investment attributes and subject those beliefs to rigorous testing following the scientific method.

- For parametric hypothesis testing, we formulate our beliefs (hypotheses), collect data, and calculate a value of the investment attribute in which we are interested (the test statistic) for that set of data (the sample), and then we compare that with a value determined under assumptions that describe the underlying population (the critical value). We can then assess the likelihood that our beliefs are true given the relationship between the test statistic and the critical value.

Commonly tested beliefs associated with the expected return and variance of returns for a given investment or investments can be formulated in this way.

Outline

Motivation

Background: sample statistics and central limit theorem

Basic hypothesis testing

Experimental design
Motivating setting

For a data science course, there has been very little “science” thus far...

“Science” as I’m using it roughly refers to “determining truth about the real world”

Sad truth: Most “mad scientists” are actually just mad engineers
Asking scientific questions

Suppose you work for a company that is considering a redesign of their website; does their new design (design B) offer any statistical advantage to their current design (design A)?

In linear regression, does a certain variable impact the response? (E.g., does energy consumption depend on whether or not a day is a weekday or weekend?)

In both settings, we are concerned with making actual statements about the nature of the world
Outline

Motivation

Background: sample statistics and central limit theorem

Basic hypothesis testing

Experimental design
SAMPLE STATISTICS

To be more consistent with standard statistics notation, we’ll introduce the notion of a population and a sample.

Mean

\[ \mu = E[X] \]

Variance

\[ \sigma = E[(X - \mu)^2] \]

\[ \bar{x} = \frac{1}{m} \sum_{i=1}^{m} x^{(i)} \]

\[ s^2 = \frac{1}{m-1} \sum_{i=1}^{m} (x^{(i)} - \bar{x})^2 \]
Sample mean as random variable

The same mean is an empirical average over $m$ independent samples from the distribution; it can also be considered as a random variable.

This new random variable has the mean and variance:

$$ E[\bar{x}] = E\left[\frac{1}{m} \sum_{i=1}^{m} x^{(i)} \right] = \frac{1}{m} \sum_{i=1}^{m} E[X] = E[X] = \mu $$

$$ \text{Var}[\bar{x}] = \text{Var}\left[\frac{1}{m} \sum_{i=1}^{m} x^{(i)} \right] = \frac{1}{m^2} \sum_{i=1}^{m} \text{Var}[X] = \frac{\sigma^2}{m} $$

where we used the fact that for independent random variables $X_1, X_2$

$$ \text{Var}[X_1 + X_2] = \text{Var}[X_1] + \text{Var}[X_2] $$

When estimating variance of sample, we use $s^2/m$ (the square root of this term is called the standard error)