Introduction and Overview

Motonobu Kanagawa

Introduction to Statistics, EURECOM

February 19, 2024

Outline

Probability: Language for Describing Uncertainty

Uncertainty: Degree of being not certain about a certain statement.

Uncertainty arises from various sources.

- The statement may be about the future

— e.g., It will rain tonight.

- Lack of information

— e.g., Does this patient have a cancer?

— If we only have limited information about the patient, we cannot be certain about this statement (unless we conduct a more detailed investigation).

Why Uncertainty/Probability Matters?

Question. Why do we need to quantify uncertainty?

Why Uncertainty/Probability Matters?

Question. Why do we need to quantify uncertainty?

One answer. Because we need to make "decisions".

Example 1: Whether to Bring an Umbrella

- You need to decide whether to bring an umbrella for going out.

Example 1: Whether to Bring an Umbrella

- You need to decide whether to bring an umbrella for going out.

- This involves the uncertainty regarding whether it will rain today.

Example 1: Whether to Bring an Umbrella

- You need to decide whether to bring an umbrella for going out.

- This involves the uncertainty regarding whether it will rain today.

  - If you believe that it will rain, you would bring the umbrella.

Example 1: Whether to Bring an Umbrella

- You need to decide whether to bring an umbrella for going out.

- This involves the uncertainty regarding whether it will rain today.

  - If you believe that it will rain, you would bring the umbrella.
  - If the weather forecast says probability 0% of rain, you would not bring the umbrella.

Example 1: Whether to Bring an Umbrella

- You need to decide whether to bring an umbrella for going out.

- This involves the uncertainty regarding whether it will rain today.

  - If you believe that it will rain, you would bring the umbrella.
  - If the weather forecast says probability 0% of rain, you would not bring the umbrella.

- It also involves costs:

  - The cost of bringing the umbrella.

Example 1: Whether to Bring an Umbrella

- You need to decide whether to bring an umbrella for going out.

- This involves the uncertainty regarding whether it will rain today.

  - If you believe that it will rain, you would bring the umbrella.
  - If the weather forecast says probability 0% of rain, you would not bring the umbrella.

- It also involves costs:

  - The cost of bringing the umbrella.
  - The cost of being wet without bringing the umbrella.

Example 1: Whether to Bring an Umbrella

- You need to decide whether to bring an umbrella for going out.

- This involves the uncertainty regarding whether it will rain today.

  - If you believe that it will rain, you would bring the umbrella.
  - If the weather forecast says probability 0% of rain, you would not bring the umbrella.

- It also involves costs:

  - The cost of bringing the umbrella.
  - The cost of being wet without bringing the umbrella.

- If you don't care about being wet, you would not bring the umbrella anyway.

- If your clothes are fragile, you would bring the umbrella even when the probability of rain is low.

Example 2: Which Drug Should a Doctor Give to a Patient?

- Assume that you are a doctor and seeing a patient.

- You need to decide which drug you should give to the patient.

Example 2: Which Drug Should a Doctor Give to a Patient?

- Assume that you are a doctor and seeing a patient.
- You need to decide which drug you should give to the patient.
- The decision should take various uncertainties into account.

Example 2: Which Drug Should a Doctor Give to a Patient?

- Assume that you are a doctor and seeing a patient.
- You need to decide which drug you should give to the patient.
- The decision should take various uncertainties into account.
  - Which disease does the patient have (disease A, B, C, ...)?

Example 2: Which Drug Should a Doctor Give to a Patient?

- Assume that you are a doctor and seeing a patient.

- You need to decide which drug you should give to the patient.

- The decision should take various uncertainties into account.

  - Which disease does the patient have (disease A, B, C, ...)?
  - Whether the drug α (or β, γ ...) is effective to the disease A (or B, C, ...).

Example 2: Which Drug Should a Doctor Give to a Patient?

- Assume that you are a doctor and seeing a patient.

- You need to decide which drug you should give to the patient.

- The decision should take various uncertainties into account.

  - Which disease does the patient have (disease A, B, C, ...)?
  - Whether the drug α (or β, γ ...) is effective to the disease A (or B, C, ...).

- The decision should also take costs into account.

  - How expensive / risky is the drug?
  - If the drug is risky, and if the disease is not serious, you would choose not to give the drug.

Probability, Statistics, and Decision Theory

The course involves the following three disciplines.

Probability, Statistics, and Decision Theory

The course involves the following three disciplines.

Probability Theory:

- provides a way of quantitatively modeling uncertainties.

Probability, Statistics, and Decision Theory

The course involves the following three disciplines.

Probability Theory:

- provides a way of quantitatively modeling uncertainties.

Statistics:

- provides a way of estimating probabilities/uncertainties from data.

Probability, Statistics, and Decision Theory

The course involves the following three disciplines.

Probability Theory:

- provides a way of quantitatively modeling uncertainties.

Statistics:

- provides a way of estimating probabilities/uncertainties from data.

Statistical Decision Theory:

- provides a way of determining the optimal decisions/policies under uncertainty.

Outline

Schedule

Week 1: Introduction and Probability Theory.

Week 2: Mean Estimation and Introduction to Estimation Theory.

Week 3: Parametric Models and Maximum Likelihood Estimation.

Week 4: Statistical Hypothesis Testing

Week 5: Statistical Hypothesis Testing (Contd)

Week 6: Bayesian Inference I

Week 7: Bayesian Inference II: Hypothesis Testing

Outline

Grade Evaluation

- Exam (13.8 pt) + Homework (5 pt) + Class attendance (1.2 pt)

- Class attendance ($0.2 \times 6 = 1.2$ pt)

  - At the beginning of each lecture, I will check your attendance.
  - If you cannot come to the lectures for a valid reason (e.g., being sick, strike, etc), please write me an email.

Grade Evaluation

- Homework: Read **Sections 1.1 to 1.6** of the following book by Berger, and write a report. The length of the report is up to 2 pages. Please summarize and discuss what you find important and/or interesting.

- **James O Berger. Statistical Decision Theory and Bayesian Analysis. Springer 1985.**

- **Deadline:** April 16th, 2024.

- Please submit the report as a pdf to motonobu.kanagawa@eurecom.fr by an email titled "IntroStat Homework Submission".

- **Caution:** If it turns out that you essentially copied and pasted your friend's report, this friend and you will not get any credit from the course. You also cannot use ChatGPT or any other AI tools.

Downloading the book

You can download the book from the following link:

`https://link.springer.com/book/10.1007/978-1-4757-4286-2`

To download it, you first need to log in from the following link (via Shibboleth) with your EURECOM account:

`https://wayf.springernature.com/?redirect_uri=https://link.springer.com`

After that you should be able to download.

Outline

Recommended Books

- B. Efron and T. Hastie, "Computer Age Statistical Inference: Algorithms, Evidence and Data Science", Cambridge University Press, 2016.

The textbook is freely available at the authors' website: `https://web.stanford.edu/~hastie/CASI/index.html`

- J. O. Berger, "Statistical Decision Theory and Bayesian Analysis", Springer, 1985.

You can download the book at the publisher's site (see the guidline above): `https://link.springer.com/book/10.1007/978-1-4757-4286-2`