UPPSALA UNIVERSITET

Bayesian Statistics and Data Analysis Lecture 9

Måns Magnusson
Department of Statistics, Uppsala University
Thanks to Aki Vehtari, Aalto University

UPPSALA UNIVERSITET

Section 1

Bayesian/Statistical Decision Theory

Introduction to Decision Theory

- Statistical Decision Theory (SDT): Formalizing decision-making under uncertainty

UPPSALA UNIVERSITET

Introduction to Decision Theory

- Statistical Decision Theory (SDT): Formalizing decision-making under uncertainty
- Early work by Condorcet (1793-1794) and Dewey (1910)
- Two types of decision theory:
  1. Normative (moral philosophy)

UPPSALA UNIVERSITET

Introduction to Decision Theory

- Statistical Decision Theory (SDT): Formalizing decision-making under uncertainty
- Early work by Condorcet (1793-1794) and Dewey (1910)
- Two types of decision theory:
  1. Normative (moral philosophy)
  2. Descriptive (empirical)

Introduction to Decision Theory

- Statistical Decision Theory (SDT): Formalizing decision-making under uncertainty
- Early work by Condorcet (1793-1794) and Dewey (1910)
- Two types of decision theory:
  1. Normative (moral philosophy)
  2. Descriptive (empirical)
- SDT: Normative theory for rational decision-making

UPPSALA UNIVERSITET

Introduction to Decision Theory

- Statistical Decision Theory (SDT): Formalizing decision-making under uncertainty
- Early work by Condorcet (1793-1794) and Dewey (1910)
- Two types of decision theory:
  1. Normative (moral philosophy)
  2. Descriptive (empirical)
- SDT: Normative theory for rational decision-making
- Although normative given a set goal, i.e. not a moral philosophy

Introduction to Decision Theory

- Statistical Decision Theory (SDT): Formalizing decision-making under uncertainty
- Early work by Condorcet (1793-1794) and Dewey (1910)
- Two types of decision theory:
  1. Normative (moral philosophy)
  2. Descriptive (empirical)
- SDT: Normative theory for rational decision-making
- Although normative given a set goal, i.e. not a moral philosophy
- Balance "values" and information/knowledge

Introduction to Decision Theory

- Statistical Decision Theory (SDT): Formalizing decision-making under uncertainty
- Early work by Condorcet (1793-1794) and Dewey (1910)
- Two types of decision theory:
  1. Normative (moral philosophy)
  2. Descriptive (empirical)
- SDT: Normative theory for rational decision-making
- Although normative given a set goal, i.e. not a moral philosophy
- Balance "values" and information/knowledge
- "Decision theory concerns goal-directed behaviour in the presence of options" (and under uncertainty) (Hansson, 1994).

Bayesian decision theory

- Potential decisions (or actions) $d \in D$
  - d may be categorical, ordinal, real, scalar, vector, etc.
    e.g. treat a person or not

Bayesian decision theory

- Potential decisions (or actions) $d \in D$
  - d may be categorical, ordinal, real, scalar, vector, etc. e.g. treat a person or not
  - fully in the control of the decision-maker

Bayesian decision theory

- Potential decisions (or actions) $d \in D$
  - d may be categorical, ordinal, real, scalar, vector, etc.
    e.g. treat a person or not
  - fully in the control of the decision-maker
- Potential consequences (or outcomes) x
  - x may be categorical, ordinal, real, scalar, vector, etc.
    e.g. a person has covid-19

Bayesian decision theory

- Potential decisions (or actions) $d \in D$
  - d may be categorical, ordinal, real, scalar, vector, etc. e.g. treat a person or not
  - fully in the control of the decision-maker
- Potential consequences (or outcomes) x
  - x may be categorical, ordinal, real, scalar, vector, etc. e.g. a person has covid-19

Bayesian decision theory

- Potential decisions (or actions) $d \in D$
  - d may be categorical, ordinal, real, scalar, vector, etc.
    e.g. treat a person or not
  - fully in the control of the decision-maker
- Potential consequences (or outcomes) x
  - x may be categorical, ordinal, real, scalar, vector, etc.
    e.g. a person has covid-19
- Probability distribution of outcome given decision $p(x|d)$
  - the decisions are controlled and thus $p(d)$ does not exist

Bayesian decision theory

- Potential decisions (or actions) $d \in D$
  - d may be categorical, ordinal, real, scalar, vector, etc. e.g. treat a person or not
  - fully in the control of the decision-maker
- Potential consequences (or outcomes) x
  - x may be categorical, ordinal, real, scalar, vector, etc. e.g. a person has covid-19
- Probability distribution of outcome given decision $p(x|d)$
  - the decisions are controlled and thus $p(d)$ does not exist

UPPSALA UNIVERSITET

Bayesian decision theory

- Potential decisions (or actions) $d \in D$
  - d may be categorical, ordinal, real, scalar, vector, etc.
    e.g. treat a person or not
  - fully in the control of the decision-maker
- Potential consequences (or outcomes) x
  - x may be categorical, ordinal, real, scalar, vector, etc.
    e.g. a person has covid-19
- Probability distribution of outcome given decision $p(x|d)$
  - the decisions are controlled and thus $p(d)$ does not exist
  - Sometimes the decision has an effect in itself, hence $p(x|d)$
    e.g. x outcome in exam, d is whether to study or not
  - If outcome doesn't depend on the decision $p(x|d) = p(x)$

Bayesian decision theory

- Utility function $U(x, d)$ maps consequences to real value
  - e.g. euro or expected lifetime
  - instead of utility, sometimes cost or loss is defined as $-U(x, d)$
  - can be multiple types of values

Bayesian decision theory

- Utility function $U(x,d)$ maps consequences to real value
  - e.g. euro or expected lifetime
  - instead of utility, sometimes cost or loss is defined as $-U(x,d)$
  - can be multiple types of values
- Expected utility for each decision decision d

$$E[U(x,d)|d] = \int U(x,d)p(x|d)dx$$

Bayesian decision theory

- Utility function $U(x, d)$ maps consequences to real value
  - e.g. euro or expected lifetime
  - instead of utility, sometimes cost or loss is defined as $-U(x, d)$
  - can be multiple types of values
- Expected utility for each decision decision d

$$E[U(x, d)|d] = \int U(x, d)p(x|d)dx$$

- Optimal decision: d^*, which maximizes the expected utility (Maximal Expected Utility, MEU)

$$d^* = \arg\max_d E[U(x, d)|d]$$

UPPSALA UNIVERSITET

Subsection 1

A Trivial Example

Example of decision making: 2 choices

- Helen is going to pick mushrooms in a forest, she notices a paw print which could made by a dog or a wolf

Example of decision making: 2 choices

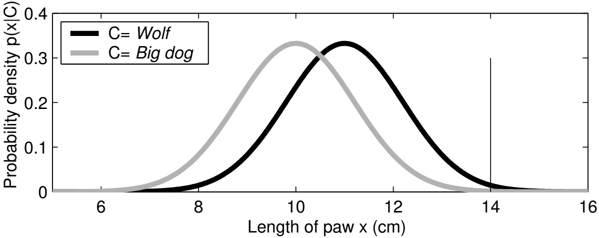
- Helen is going to pick mushrooms in a forest, she notices a paw print which could made by a dog or a wolf
- Helen measures that the length of the paw print is 14 cm
- Based on data she want to infer the probability of wolf

Example of decision making: 2 choices

UPPSALA UNIVERSITET

- Bayesian/Statistical Decision Theory
  — A Trivial Example
- Integrating Inference and Decisions
- Course Evaluation

- Helen is going to pick mushrooms in a forest, she notices a paw print which could made by a dog or a wolf
- Helen measures that the length of the paw print is 14 cm
- Based on data she want to infer the probability of wolf



observed length has been marked with a horizontal line

- Likelihood of wolf is 0.92 (alternative being dog)

Example of decision making

- Helen also assumes that in her living area there are about one hundred times more free running dogs than wolves, i.e. *a priori* probability for wolf, before observation is 1%.
- Likelihood and posterior

Animal	Likelihood	Posterior probability
Wolf	0.92	0.10
Dog	0.08	0.90

Example of decision making

- Helen also assumes that in her living area there are about one hundred times more free running dogs than wolves, i.e. *a priori* probability for wolf, before observation is 1%.
- Likelihood and posterior

Animal	Likelihood	Posterior probability
Wolf	0.92	0.10
Dog	0.08	0.90

- Posterior probability of wolf is 10%

Example of decision making

- Helen has to make decision whether to go pick mushrooms

- Utility function $U(d, x)$:

- Expected utilities

Example of decision making

- Helen has to make decision whether to go pick mushrooms
- If she doesn't go to pick mushrooms utility is zero

- Utility function $U(d, x)$:

- Expected utilities

Example of decision making

- Helen has to make decision whether to go pick mushrooms
- If she doesn't go to pick mushrooms utility is zero
- Helen assigns positive utility 1 for getting fresh mushrooms

- Utility function $U(d, x)$:

- Expected utilities

Example of decision making

- Helen has to make decision whether to go pick mushrooms
- If she doesn't go to pick mushrooms utility is zero
- Helen assigns positive utility 1 for getting fresh mushrooms
- Helen assigns negative utility -1000 for a event that she goes to the forest and wolf attacks
- Utility function $U(d, x)$:

- Expected utilities

UPPSALA UNIVERSITET

- Bayesian/Statistical Decision Theory
  - — A Trivial Example
- Integrating Inference and Decisions
- Course Evaluation

Example of decision making

- Helen has to make decision whether to go pick mushrooms
- If she doesn't go to pick mushrooms utility is zero
- Helen assigns positive utility 1 for getting fresh mushrooms
- Helen assigns negative utility -1000 for a event that she goes to the forest and wolf attacks
- Utility function $U(d, x)$:

	Animal x	
Decision d	Wolf	Dog
Stay home	0	0
Go to the forest	-1000	1

- Expected utilities

Example of decision making

- Helen has to make decision whether to go pick mushrooms
- If she doesn't go to pick mushrooms utility is zero
- Helen assigns positive utility 1 for getting fresh mushrooms
- Helen assigns negative utility -1000 for a event that she goes to the forest and wolf attacks
- Utility function $U(d, x)$:

	Animal x	
Decision d	Wolf	Dog
Stay home	0	0
Go to the forest	-1000	1

- Expected utilities

Action d	Conditional utility $E[U(x)\|d]$
Stay home	0
Go to the forest	-100+0.9 = -99.1

Example of decision making

- Maximum likelihood decision would be to assume that there is a wolf

UPPSALA UNIVERSITET

- Bayesian/Statistical Decision Theory
  - — A Trivial Example
- Integrating Inference and Decisions
- Course Evaluation

Example of decision making

- Maximum likelihood decision would be to assume that there is a wolf
- Maximum posterior decision would be to assume that there is a dog

Example of decision making

UPPSALA UNIVERSITET

- Maximum likelihood decision would be to assume that there is a wolf
- Maximum posterior decision would be to assume that there is a dog
- Maximum Expected Utility (MEU) decision is to stay home, even if it is more likely that the animal is dog

- The uncertainties (probabilities) related to all consequences need to be take into account

Section 2

Integrating Inference and Decisions

Integrating inference and decisions

- To make an optimal decision we need $p(x|d)$

UPPSALA UNIVERSITET

Integrating inference and decisions

- To make an optimal decision we need $p(x|d)$
- In many situations we can approximate $p(x|d) \approx p(x)$

Integrating inference and decisions

- To make an optimal decision we need $p(x|d)$
- In many situations we can approximate $p(x|d) \approx p(x)$
- The benefit of Bayesian inference:
  We can use $p(x|d, y)$ i.e. integrating previous data in decision making (some times referred to as Bayesianism)
- Formal Bayesian Decision making hence have two parts:
  1. Model $p(x|y, d)$ as good as possible

Integrating inference and decisions

- To make an optimal decision we need $p(x|d)$
- In many situations we can approximate $p(x|d) \approx p(x)$
- The benefit of Bayesian inference:
  We can use $p(x|d, y)$ i.e. integrating previous data in decision making (some times referred to as Bayesianism)
- Formal Bayesian Decision making hence have two parts:
  1. Model $p(x|y, d)$ as good as possible
  2. Define $U(x, d)$

Challenges with the utility function

- Utility functions are rarely linear:
  How do we set it up?

Challenges with the utility function

- Utility functions are rarely linear:
  How do we set it up?
- What is the cost of human life/illness? Can it be formulated to a utility function?

Challenges with the utility function

- Utility functions are rarely linear:
  How do we set it up?
- What is the cost of human life/illness? Can it be formulated to a utility function?
- Different parties might have different utilities (patient, physician, society)

Challenges with the utility function

- Utility functions are rarely linear:
  How do we set it up?
- What is the cost of human life/illness? Can it be formulated to a utility function?
- Different parties might have different utilities (patient, physician, society)
- Personal vs institutional decisions
  - An individual have a subjective $p(x|d)$ and a subjective $U(x, d)$. Need for formal decision-making?
  - An institution might be better suited. Decision-recommendations.
- Decision theoretical approaches is better suited when $U(x, d)$ and $p(x|d)$ is well-defined.

UPPSALA UNIVERSITET

Challenges with the probability model

- We seldom know $p(x|d)$

Challenges with the probability model

UPPSALA UNIVERSITET

- We seldom know $p(x|d)$
- We need to approximate: $\hat{p}(x|y,d) \approx p(x|d)$

Challenges with the probability model

- We seldom know $p(x|d)$
- We need to approximate: $\hat{p}(x|y, d) \approx p(x|d)$
- If our approximation is bad - we will not necessarily make the optimal decision
- And we know that all models are wrong.

Challenges with the probability model

- We seldom know $p(x|d)$
- We need to approximate: $\hat{p}(x|y,d) \approx p(x|d)$
- If our approximation is bad - we will not necessarily make the optimal decision
- And we know that all models are wrong.
- $\hat{p}(x|y,d)$ is a statistical problem. Can we trust the model?
- But it can also be extremely hard to estimate, say:

$$P(\text{a new pandemic in 2025})$$

Challenges in decision making

- Personal vs institutional decisions
  - An individual have a subjective $p(x|d)$ and a subjective $U(x, d)$. Need for formal decision-making?

Challenges in decision making

- Personal vs institutional decisions
  - An individual have a subjective $p(x|d)$ and a subjective $U(x, d)$. Need for formal decision-making?
  - An institution might be better suited. Decision-recommendations.
- Decision theoretical approaches is better suited when $U(x, d)$ and $p(x|d)$ is well-defined.

Multi-stage decision making

- Slighly more complex: Multi-stage decision-making
- We need to take all uncertainties into account
- We can also condition after the decision is made

Multi-stage decision making example

- 95-year-old has a tumor that is malignant with 90% probability
- A priori knowledge
  - expected lifetime is 34.8 months if no cancer
  - expected lifetime is 16.7 months if cancer and radiation therapy is used

Multi-stage decision making example

- 95-year-old has a tumor that is malignant with 90% probability
- A priori knowledge
  - expected lifetime is 34.8 months if no cancer
  - expected lifetime is 16.7 months if cancer and radiation therapy is used
  - expected lifetime is 20.3 months if cancer and surgery, but the probability of dying in surgery is 35% (for 95 year old)

UPPSALA UNIVERSITET

- Bayesian/Statistical Decision Theory
  - — A Trivial Example
- Integrating Inference and Decisions
- Course Evaluation

Multi-stage decision making example

- 95-year-old has a tumor that is malignant with 90% probability
- A priori knowledge
  - expected lifetime is 34.8 months if no cancer
  - expected lifetime is 16.7 months if cancer and radiation therapy is used
  - expected lifetime is 20.3 months if cancer and surgery, but the probability of dying in surgery is 35% (for 95 year old)
  - expected lifetime is 5.6 months is cancer and no treatment

Multi-stage decision making example

- 95-year-old has a tumor that is malignant with 90% probability
- A priori knowledge
  - expected lifetime is 34.8 months if no cancer
  - expected lifetime is 16.7 months if cancer and radiation therapy is used
  - expected lifetime is 20.3 months if cancer and surgery, but the probability of dying in surgery is 35% (for 95 year old)
  - expected lifetime is 5.6 months is cancer and no treatment
- Which treatment to choose?
  - quality-adjusted lifetime
  - 1 month is subtracted from the time spent in treatments

Multi-stage decision making example

- 95-year-old has a tumor that is malignant with 90% probability
- A priori knowledge
  - expected lifetime is 34.8 months if no cancer
  - expected lifetime is 16.7 months if cancer and radiation therapy is used
  - expected lifetime is 20.3 months if cancer and surgery, but the probability of dying in surgery is 35% (for 95 year old)
  - expected lifetime is 5.6 months is cancer and no treatment
- Which treatment to choose?
  - quality-adjusted lifetime
  - 1 month is subtracted from the time spent in treatments
- Quality-adjusted lifetime
  - Radiothreapy: 0.9*16.7 + 0.1*34.8 - 1 = 17.5mo

UPPSALA UNIVERSITET

- Bayesian/Statistical Decision Theory
  — A Trivial Example
- Integrating Inference and Decisions
- Course Evaluation

Multi-stage decision making example

- 95-year-old has a tumor that is malignant with 90% probability
- A priori knowledge
  - expected lifetime is 34.8 months if no cancer
  - expected lifetime is 16.7 months if cancer and radiation therapy is used
  - expected lifetime is 20.3 months if cancer and surgery, but the probability of dying in surgery is 35% (for 95 year old)
  - expected lifetime is 5.6 months is cancer and no treatment
- Which treatment to choose?
  - quality-adjusted lifetime
  - 1 month is subtracted from the time spent in treatments
- Quality-adjusted lifetime
  - Radiothreapy: 0.9*16.7 + 0.1*34.8 - 1 = 17.5mo
  - Surgery: 0.35*0 + 0.65*(0.9*20.3 + 0.1*34.8 - 1) = 13.5mo

Multi-stage decision making example

- 95-year-old has a tumor that is malignant with 90% probability
- A priori knowledge
  - expected lifetime is 34.8 months if no cancer
  - expected lifetime is 16.7 months if cancer and radiation therapy is used
  - expected lifetime is 20.3 months if cancer and surgery, but the probability of dying in surgery is 35% (for 95 year old)
  - expected lifetime is 5.6 months is cancer and no treatment
- Which treatment to choose?
  - quality-adjusted lifetime
  - 1 month is subtracted from the time spent in treatments
- Quality-adjusted lifetime
  - Radiothreapy: 0.9*16.7 + 0.1*34.8 - 1 = 17.5mo
  - Surgery: 0.35*0 + 0.65*(0.9*20.3 + 0.1*34.8 - 1) = 13.5mo
  - No treatment: 0.9*5.6 + 0.1*34.8 = 8.5mo

Multi-stage decision making example

- 95-year-old has a tumor that is malignant with 90% probability
- A priori knowledge
  - expected lifetime is 34.8 months if no cancer
  - expected lifetime is 16.7 months if cancer and radiation therapy is used
  - expected lifetime is 20.3 months if cancer and surgery, but the probability of dying in surgery is 35% (for 95 year old)
  - expected lifetime is 5.6 months is cancer and no treatment
- Which treatment to choose?
  - quality-adjusted lifetime
  - 1 month is subtracted from the time spent in treatments
- Quality-adjusted lifetime
  - Radiothreapy: 0.9*16.7 + 0.1*34.8 - 1 = 17.5mo
  - Surgery: 0.35*0 + 0.65*(0.9*20.3 + 0.1*34.8 - 1) = 13.5mo
  - No treatment: 0.9*5.6 + 0.1*34.8 = 8.5mo
- Elaborated further in Bayesian Data Analysis

Section 3

Course Evaluation

UPPSALA UNIVERSITET

Course Evaluation

- Did you get what you expected?

UPPSALA UNIVERSITET

Course Evaluation

- Did you get what you expected?
- What can be improved? What was annoying?

Course Evaluation

- Did you get what you expected?
- What can be improved? What was annoying?
- What was good? What was fun?

UPPSALA UNIVERSITET

Course Evaluation

- Did you get what you expected?
- What can be improved? What was annoying?
- What was good? What was fun?
- Anything else?