UPPSALA
UNIVERSITET

- Bayesian decision theory
- Examples
- Course Evaluation

# Bayesian Statistics and Data Analysis
# Lecture 9

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

# Section 1

## Bayesian decision theory

# Bayesian decision theory

- Potential decisions (or actions) $d \in D$ ($a$)

# Bayesian decision theory

- Potential decisions (or actions) $d \in D$ ($a$)

UPPSALA UNIVERSITET



# Bayesian decision theory

- Potential decisions (or actions) $d \in D$ ($a$)
- Potential consequences (or outcomes) $x$
  - $x$ may be categorical, ordinal, real, scalar, vector, etc.

UPPSALA
UNIVERSITET



# Bayesian decision theory

- Potential decisions (or actions) $d \in D$ ($a$)
- Potential consequences (or outcomes) $x$
    - $x$ may be categorical, ordinal, real, scalar, vector, etc.

# Bayesian decision theory

UPPSALA UNIVERSITET

- Bayesian decision theory
- Examples
- Course Evaluation

- Potential decisions (or actions) $d \in D$ ($a$)
- Potential consequences (or outcomes) $x$
  - $x$ may be categorical, ordinal, real, scalar, vector, etc.
- Probability distribution of outcome given decision $p(x|d)$
  - in decision making the decisions are controlled and thus $p(d)$ does not exist/are fixed

# Bayesian decision theory

- Potential decisions (or actions) $d \in D$ ($a$)
- Potential consequences (or outcomes) $x$
  - $x$ may be categorical, ordinal, real, scalar, vector, etc.
- Probability distribution of outcome given decision $p(x|d)$
  - in decision making the decisions are controlled and thus $p(d)$ does not exist/are fixed

# Bayesian decision theory

- Potential decisions (or actions) $d \in D$ ($a$)
- Potential consequences (or outcomes) $x$
  - $x$ may be categorical, ordinal, real, scalar, vector, etc.
- Probability distribution of outcome given decision $p(x|d)$
  - in decision making the decisions are controlled and thus $p(d)$ does not exist/are fixed
- Utility function $U(x)$ maps consequences to real value
  - e.g. euro or expected lifetime
  - instead of utility sometimes cost or loss is defined

# Bayesian decision theory

- Potential decisions (or actions) $d \in D$ ($a$)
- Potential consequences (or outcomes) $x$
  - $x$ may be categorical, ordinal, real, scalar, vector, etc.
- Probability distribution of outcome given decision $p(x|d)$
  - in decision making the decisions are controlled and thus $p(d)$ does not exist/are fixed
- Utility function $U(x)$ maps consequences to real value
  - e.g. euro or expected lifetime
  - instead of utility sometimes cost or loss is defined

# Bayesian decision theory

- Potential decisions (or actions) $d \in D$ ($a$)
- Potential consequences (or outcomes) $x$
  - $x$ may be categorical, ordinal, real, scalar, vector, etc.
- Probability distribution of outcome given decision $p(x|d)$
  - in decision making the decisions are controlled and thus $p(d)$ does not exist/are fixed
- Utility function $U(x)$ maps consequences to real value
  - e.g. euro or expected lifetime
  - instead of utility sometimes cost or loss is defined
- Expected utility for a decision $d$

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

# Bayesian decision theory

- Potential decisions (or actions) $d \in D$ ($a$)
- Potential consequences (or outcomes) $x$
  - $x$ may be categorical, ordinal, real, scalar, vector, etc.
- Probability distribution of outcome given decision $p(x|d)$
  - in decision making the decisions are controlled and thus $p(d)$ does not exist/are fixed
- Utility function $U(x)$ maps consequences to real value
  - e.g. euro or expected lifetime
  - instead of utility sometimes cost or loss is defined
- Expected utility for a decision $d$

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

# Bayesian decision theory

- Potential decisions (or actions) $d \in D$ ($a$)
- Potential consequences (or outcomes) $x$
  - $x$ may be categorical, ordinal, real, scalar, vector, etc.
- Probability distribution of outcome given decision $p(x|d)$
  - in decision making the decisions are controlled and thus $p(d)$ does not exist/are fixed
- Utility function $U(x)$ maps consequences to real value
  - e.g. euro or expected lifetime
  - instead of utility sometimes cost or loss is defined
- Expected utility for a decision $d$

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

- Optimal decision: $d^*$, which maximizes the expected utility

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

# Integrating inference and decisions

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

# Integrating inference and decisions

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

# Integrating inference and decisions

UPPSALA UNIVERSITET

- Bayesian decision theory
- Examples
- Course Evaluation

- 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 $p(x|d) \approx p(x)$

# 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 data in decision making

UPPSALA UNIVERSITET



# Challenges in decision making

- Actual utility functions are rarely linear

# Challenges in decision making



- Actual utility functions are rarely linear
- What is the cost of human life?

# Challenges in decision making



- Actual utility functions are rarely linear
- What is the cost of human life?
- Multipel parties having different utilities

UPPSALA UNIVERSITET



# Section 2

## Examples

UPPSALA UNIVERSITET

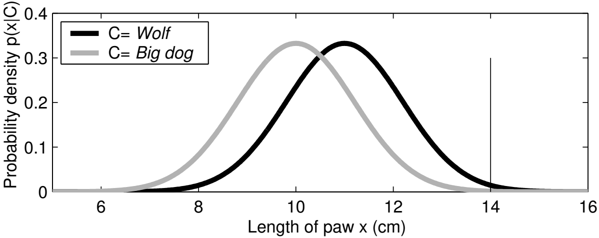
- Bayesian decision theory
- Examples
- Course Evaluation

# Example of decision making: 2 choices

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

# Example of decision making: 2 choices

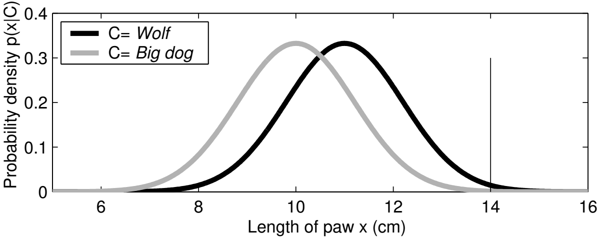
- Helen is going to pick mushrooms in a forest, while she notices a paw print which could made by a dog or a wolf
- Helen measures that the length of the paw print is 14 cm and goes home to Google how big paws dogs and wolves have, and tries then to infer which animal has made the paw print




observed length has been marked with a horizontal line

# Example of decision making: 2 choices

- Helen is going to pick mushrooms in a forest, while 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 and goes home to Google how big paws dogs and wolves have, and tries then to infer which animal has made the paw print




observed length has been marked with a horizontal line

- Likelihood of wolf is 0.92 (alternative being dog)

# Example of decision making

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

# Example of decision making

- Helen assumes also that in her living area there are about one hundred times more free running dogs than wolves, that is *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 assumes also that in her living area there are about one hundred times more free running dogs than wolves, that is *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

# 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

UPPSALA UNIVERSITET

- Bayesian decision theory
- Examples
- 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

UPPSALA UNIVERSITET

- Bayesian decision theory
- Examples
- 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 (for some reason Helen assumes that wolf will always attack)

# 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 (for some reason Helen assumes that wolf will always attack)

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

Utility matrix $U(x)$

# Example of decision making

UPPSALA UNIVERSITET

- Bayesian decision theory
- Examples
- Course Evaluation

- 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 (for some reason Helen assumes that wolf will always attack)

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

Utility matrix $U(x)$

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

Utilities for different actions

# Example of decision making

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

# Example of decision making

UPPSALA UNIVERSITET

- Bayesian decision theory
- Examples
- Course Evaluation

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

UPPSALA UNIVERSITET



# 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



- 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

- 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 utility decision is to stay home, even if it is more likely that the animal is dog

# 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
- Maximum utility decision is to stay home, even if it is more likely that the animal is dog

# 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
- Maximum utility decision is to stay home, even if it is more likely that the animal is dog
- Example illustrates that the uncertainties (probabilities) related to all consequences need to be carried on until final decision making

# Multi-stage decision making (Section 9.3)

- 95 year old has a tumor that is malignant with 90% probability

# Multi-stage decision making (Section 9.3)

- 95 year old has a tumor that is malignant with 90% probability
- Based on statistics
  - expected lifetime is 34.8 months if no cancer

# Multi-stage decision making (Section 9.3)

- 95 year old has a tumor that is malignant with 90% probability
- Based on statistics
  - 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 (Section 9.3)

- 95 year old has a tumor that is malignant with 90% probability
- Based on statistics
  - 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)

# Multi-stage decision making (Section 9.3)

- 95 year old has a tumor that is malignant with 90% probability
- Based on statistics
  - 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 (Section 9.3)

- 95 year old has a tumor that is malignant with 90% probability
- Based on statistics
  - 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 life time
  - 1 month is subtracted for the time spent in treatments

# Multi-stage decision making (Section 9.3)

- 95 year old has a tumor that is malignant with 90% probability
- Based on statistics
  - 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 life time
  - 1 month is subtracted for the time spent in treatments
- Quality adjusted life time
  - Radiothreapy: $0.9*16.7 + 0.1*34.8 - 1 = 17.5$mo

# Multi-stage decision making (Section 9.3)

- 95 year old has a tumor that is malignant with 90% probability
- Based on statistics
  - 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 life time
  - 1 month is subtracted for the time spent in treatments
- Quality adjusted life time
  - Radiothreapy: $0.9*16.7 + 0.1*34.8 - 1 = 17.5$mo
  - Surgery: $0.35*0 + 0.65*(0.9*20.3 + 0.1*34.8 - 1) =$ 13.5mo

# Multi-stage decision making (Section 9.3)

- 95 year old has a tumor that is malignant with 90% probability
- Based on statistics
  - 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 life time
  - 1 month is subtracted for the time spent in treatments
- Quality adjusted life time
  - 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

UPPSALA UNIVERSITET

- Bayesian decision theory
- Examples
- Course Evaluation

# Multi-stage decision making (Section 9.3)

- 95 year old has a tumor that is malignant with 90% probability
- Based on statistics
  - 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 life time
  - 1 month is subtracted for the time spent in treatments
- Quality adjusted life time
  - 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
- See the book for continuation of the example with additional test for cancer

UPPSALA UNIVERSITET

- Bayesian decision theory
- Examples
- Course Evaluation

# Design of experiment

- Which experiment would give most additional information
  - decide values $x_{n+1}$ for the next experiment
  - which values of $x_{n+1}$ would reduce the posterior uncertainty most
- Example
  - Imagine that in bioassay the posterior uncertainty of LD50 is too large
  - which dose should be used in the next experiment to reduce the variance of LD50 as much as possible ?
    - this way less experiments need to be made (and less animals need to be killed)

# Model selection as decision problem

UPPSALA UNIVERSITET



- Evaluate how model $M$ generalizes to unseen data $\tilde{y}$ (the *expected log predictive density*):

$$\text{elpd}_M = \int \log p_M(\tilde{y}|y) p_{\text{true}}(\tilde{y}) d\tilde{y}\,,$$

  where $\tilde{y}$ is an unseen observation generated from the true data generating process $p_{\text{true}}(\tilde{y})$, and $y$ are observed data.
- $\log p_M(\tilde{y}|y)$ is the log score (the utility of the model)

# Model selection as decision problem

- Evaluate how model $M$ generalizes to unseen data $\tilde{y}$ (the *expected log predictive density*):

$$\mathrm{elpd}_M = \int \log p_M(\tilde{y}|y) p_{\mathrm{true}}(\tilde{y}) d\tilde{y}\,,$$

  where $\tilde{y}$ is an unseen observation generated from the true data generating process $p_{\mathrm{true}}(\tilde{y})$, and $y$ are observed data.
- $\log p_M(\tilde{y}|y)$ is the log score (the utility of the model)
- Formally, our decision is to choose the model $M^\star$.

UPPSALA UNIVERSITET

- Bayesian decision theory
- Examples
- Course Evaluation

# Automated decision-making and Discrimination

Work of Holli Sargeant:

- Automated decision-making in an ML-context

$$d^\star = \sum_d u(y, d)\hat{p}(y|d, x)\,,$$

where $y$ is indicates if a person defaults, $d$ is the decision on weather granting a loan or not, $\hat{p}$ is a prediction model, and $u(y, d)$ is the utility of a bank.

# Automated decision-making and Discrimination

Work of Holli Sargeant:

- Automated decision-making in an ML-context

$$d^\star = \sum_d u(y,d)\hat{p}(y|d,x)\,,$$

  where $y$ is indicates if a person defaults, $d$ is the decision on weather granting a loan or not, $\hat{p}$ is a prediction model, and $u(y,d)$ is the utility of a bank.

- Conditional statistical parity

$$\mathbb{E}_x\left[\hat{p}(y|x) \mid x_l, x_p\right] = \mathbb{E}_x\left[\hat{p}(y|x) \mid x_l\right]\,,$$

  where $x_l$ is legitimate factors and $x_p$ are protected attributes.

UPPSALA UNIVERSITET



# Section 3

# Course Evaluation

# Course Evaluation



- What was good? What was fun?

# Course Evaluation

- What was good? What was fun?
- What can be improved? What was annoying?

# Course Evaluation

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

UPPSALA UNIVERSITET

- Bayesian decision theory
- Examples
- Course Evaluation

# Course Evaluation

- What was good? What was fun?
- What can be improved? What was annoying?
- Did you get what you expected?
- How can I get you to speak more during class?