MT4531/MT5731: (Advanced) Bayesian Inference

Introduction

Nicolò Margaritella

School of Mathematics and Statistics, University of St Andrews

–

University of St Andrews

Outline

Outline

Bayesian approach in statistics

- The Bayesian approach is a complete theory in statistics, different in many ways from the standard (classical/frequentist) approach.
- We can design experiments, construct statistical models, estimate parameters, make predictions, test hypotheses, compare models, and in general measure the uncertainty that is associated with our inferences.
- In some situations, inference outcomes (e.g. parameter estimates) from the Bayesian and the frequentist approaches can be similar, however...
- ... the methodology and interpretation under the two approaches of such outcomes is very different! There is a long-lasting, and often passionate, debate about which is the most appropriate statistical approach.

Bayesian approach in statistics

- The Bayesian approach is a complete theory in statistics, different in many ways from the standard (classical/frequentist) approach.
- We can design experiments, construct statistical models, estimate parameters, make predictions, test hypotheses, compare models, and in general measure the uncertainty that is associated with our inferences.
- In some situations, inference outcomes (e.g. parameter estimates) from the Bayesian and the frequentist approaches can be similar, however...
- ... the methodology and interpretation under the two approaches of such outcomes is very different! There is a long-lasting, and often passionate, debate about which is the most appropriate statistical approach.

Bayesian approach in statistics

- The Bayesian approach is a complete theory in statistics, different in many ways from the standard (classical/frequentist) approach.
- We can design experiments, construct statistical models, estimate parameters, make predictions, test hypotheses, compare models, and in general measure the uncertainty that is associated with our inferences.
- In some situations, inference outcomes (e.g. parameter estimates) from the Bayesian and the frequentist approaches can be similar, however...
- ... the methodology and interpretation under the two approaches of such outcomes is very different! There is a long-lasting, and often passionate, debate about which is the most appropriate statistical approach.

Bayesian approach in statistics

- The Bayesian approach is a complete theory in statistics, different in many ways from the standard (classical/frequentist) approach.
- We can design experiments, construct statistical models, estimate parameters, make predictions, test hypotheses, compare models, and in general measure the uncertainty that is associated with our inferences.
- In some situations, inference outcomes (e.g. parameter estimates) from the Bayesian and the frequentist approaches can be similar, however...
- ... the methodology and interpretation under the two approaches of such outcomes is very different! There is a long-lasting, and often passionate, debate about which is the most appropriate statistical approach.

Outline

1. Introduction
2. The basic idea of Bayesian statistics. Examples
3. Putting Bayesian statistics in perspective
4. Why using Bayesian statistics?

The basic idea of Bayesian statistics. Examples

- Ex1. I wish to estimate the proportion θ of blemished apples that my apple tree produces.(Lecture Notes, Section 0.1)
- Ex2. Suppose that you are working hard at your desk, and glance out of the window to see a large wooden-looking object with branches covered in green things... (LN Section 1.1)

The basic idea of Bayesian statistics. Examples

- Ex1. I wish to estimate the proportion θ of blemished apples that my apple tree produces.(Lecture Notes, Section 0.1)
- Ex2. Suppose that you are working hard at your desk, and glance out of the window to see a large wooden-looking object with branches covered in green things... (LN Section 1.1)

The basic idea of Bayesian statistics. Example 3

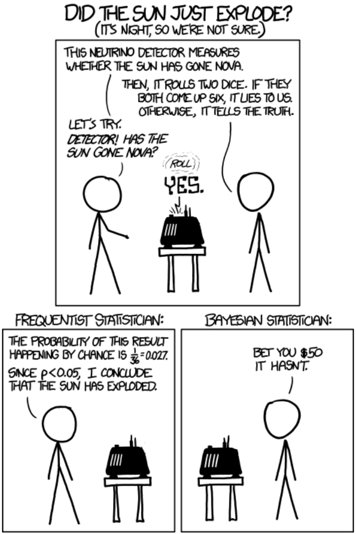

Figure: xkcd.com

- We call $\pi(\theta|data)$ the *posterior* distribution of θ as opposed to the *prior* distribution, $p(\theta)$.
- These are the probability distributions before (*a priori*) and after (*a posteriori*) observing the data.
- Bayes theorem allows us to make the transition from prior to posterior probability distributions for an unknown quantity of interest.
- Bayesian statistics provides a way of formalising the gathering of information and uncertainty before new observations are made available and then updating this information based on the new observations.
- It is a natural way of scientific investigation.

- We call $\pi(\theta|data)$ the *posterior* distribution of θ as opposed to the *prior* distribution, $p(\theta)$.
- These are the probability distributions before (*a priori*) and after (*a posteriori*) observing the data.
- Bayes theorem allows us to make the transition from prior to posterior probability distributions for an unknown quantity of interest.
- Bayesian statistics provides a way of formalising the gathering of information and uncertainty before new observations are made available and then updating this information based on the new observations.
- It is a natural way of scientific investigation.

- We call $\pi(\theta|data)$ the *posterior* distribution of θ as opposed to the *prior* distribution, $p(\theta)$.
- These are the probability distributions before (*a priori*) and after (*a posteriori*) observing the data.
- Bayes theorem allows us to make the transition from prior to posterior probability distributions for an unknown quantity of interest.
- Bayesian statistics provides a way of formalising the gathering of information and uncertainty before new observations are made available and then updating this information based on the new observations.
- It is a natural way of scientific investigation.

- We call $\pi(\theta|data)$ the *posterior* distribution of θ as opposed to the *prior* distribution, $p(\theta)$.
- These are the probability distributions before (*a priori*) and after (*a posteriori*) observing the data.
- Bayes theorem allows us to make the transition from prior to posterior probability distributions for an unknown quantity of interest.
- Bayesian statistics provides a way of formalising the gathering of information and uncertainty before new observations are made available and then updating this information based on the new observations.
- It is a natural way of scientific investigation.

- We call $\pi(\theta|data)$ the *posterior* distribution of θ as opposed to the *prior* distribution, $p(\theta)$.
- These are the probability distributions before (*a priori*) and after (*a posteriori*) observing the data.
- Bayes theorem allows us to make the transition from prior to posterior probability distributions for an unknown quantity of interest.
- Bayesian statistics provides a way of formalising the gathering of information and uncertainty before new observations are made available and then updating this information based on the new observations.
- It is a natural way of scientific investigation.

Outline

Frequentists and Bayesians

(b) Classical statistics

- The standard approach that you have encountered in previous statistics modules is typically called "classical" or "frequentist".
- It is not a unified approach, but rather a collection of different approaches.
- Some of the key classical statistical methods are based solely on the likelihood function of observed data (e.g. maximum likelihood estimates, p-values, likelihood ratio function).
- Other methods are based on satisfying certain criteria, such as the least square estimation (LSE) that minimises squared differences between observations and fitted values.
- Methods typically focus on deriving procedures that satisfy optimal long-run properties (e.g. unbiasedness, consistency, fixed error rates) if the experiment is repeated infinite times.
- A key characteristic is that they consider parameters as having a *fixed* (typically unknown) value.

Frequentists and Bayesians

(b) **Classical statistics**

- The standard approach that you have encountered in previous statistics modules is typically called "classical" or "frequentist".
- It is not a unified approach, but rather a collection of different approaches.
- Some of the key classical statistical methods are based solely on the likelihood function of observed data (e.g. maximum likelihood estimates, p-values, likelihood ratio function).
- Other methods are based on satisfying certain criteria, such as the least square estimation (LSE) that minimises squared differences between observations and fitted values.
- Methods typically focus on deriving procedures that satisfy optimal long-run properties (e.g. unbiasedness, consistency, fixed error rates) if the experiment is repeated infinite times.
- A key characteristic is that they consider parameters as having a *fixed* (typically unknown) value.

Frequentists and Bayesians

(b) **Classical statistics**

- The standard approach that you have encountered in previous statistics modules is typically called "classical" or "frequentist".
- It is not a unified approach, but rather a collection of different approaches.
- Some of the key classical statistical methods are based solely on the likelihood function of observed data (e.g. maximum likelihood estimates, p-values, likelihood ratio function).
- Other methods are based on satisfying certain criteria, such as the least square estimation (LSE) that minimises squared differences between observations and fitted values.
- Methods typically focus on deriving procedures that satisfy optimal long-run properties (e.g. unbiasedness, consistency, fixed error rates) if the experiment is repeated infinite times.
- A key characteristic is that they consider parameters as having a *fixed* (typically unknown) value.

Frequentists and Bayesians

(b) **Classical statistics**

- The standard approach that you have encountered in previous statistics modules is typically called "classical" or "frequentist".
- It is not a unified approach, but rather a collection of different approaches.
- Some of the key classical statistical methods are based solely on the likelihood function of observed data (e.g. maximum likelihood estimates, p-values, likelihood ratio function).
- Other methods are based on satisfying certain criteria, such as the least square estimation (LSE) that minimises squared differences between observations and fitted values.
- Methods typically focus on deriving procedures that satisfy optimal long-run properties (e.g. unbiasedness, consistency, fixed error rates) if the experiment is repeated infinite times.
- A key characteristic is that they consider parameters as having a *fixed* (typically unknown) value.

Frequentists and Bayesians

(b) **Classical statistics**

- The standard approach that you have encountered in previous statistics modules is typically called "classical" or "frequentist".
- It is not a unified approach, but rather a collection of different approaches.
- Some of the key classical statistical methods are based solely on the likelihood function of observed data (e.g. maximum likelihood estimates, p-values, likelihood ratio function).
- Other methods are based on satisfying certain criteria, such as the least square estimation (LSE) that minimises squared differences between observations and fitted values.
- Methods typically focus on deriving procedures that satisfy optimal long-run properties (e.g. unbiasedness, consistency, fixed error rates) if the experiment is repeated infinite times.
- A key characteristic is that they consider parameters as having a *fixed* (typically unknown) value.

Frequentists and Bayesians

(b) **Classical statistics**

- The standard approach that you have encountered in previous statistics modules is typically called "classical" or "frequentist".
- It is not a unified approach, but rather a collection of different approaches.
- Some of the key classical statistical methods are based solely on the likelihood function of observed data (e.g. maximum likelihood estimates, p-values, likelihood ratio function).
- Other methods are based on satisfying certain criteria, such as the least square estimation (LSE) that minimises squared differences between observations and fitted values.
- Methods typically focus on deriving procedures that satisfy optimal long-run properties (e.g. unbiasedness, consistency, fixed error rates) if the experiment is repeated infinite times.
- A key characteristic is that they consider parameters as having a *fixed* (typically unknown) value.

Frequentists and Bayesians

(b) **Classical statistics**

- The standard approach that you have encountered in previous statistics modules is typically called "classical" or "frequentist".
- It is not a unified approach, but rather a collection of different approaches.
- Some of the key classical statistical methods are based solely on the likelihood function of observed data (e.g. maximum likelihood estimates, p-values, likelihood ratio function).
- Other methods are based on satisfying certain criteria, such as the least square estimation (LSE) that minimises squared differences between observations and fitted values.
- Methods typically focus on deriving procedures that satisfy optimal long-run properties (e.g. unbiasedness, consistency, fixed error rates) if the experiment is repeated infinite times.
- A key characteristic is that they consider parameters as having a *fixed* (typically unknown) value.

Frequentists and Bayesians

(b) **Classical statistics**

- The common criticism for the standard approach is:
  1. the interpretation of outcomes (e.g. confidence intervals or p-values) based on their long-run properties, even in cases where the experiments, surveys, etc. are practically unrepeatable.
  2. the lack of formal methods for incorporating background information and that they often use background information implicitly (e.g. experimental design, choice of likelihood, choice of error rates in testing).
  3. the disobedience of some classical procedures (e.g. p-values) to the likelihood principle.

Frequentists and Bayesians

(b) **Classical statistics**

- The common criticism for the standard approach is:
  1. the interpretation of outcomes (e.g. confidence intervals or p-values) based on their long-run properties, even in cases where the experiments, surveys, etc. are practically unrepeatable.
  2. the lack of formal methods for incorporating background information and that they often use background information implicitly (e.g. experimental design, choice of likelihood, choice of error rates in testing).
  3. the disobedience of some classical procedures (e.g. p-values) to the likelihood principle.

Frequentists and Bayesians

(b) **Classical statistics**

- The common criticism for the standard approach is:
  1. the interpretation of outcomes (e.g. confidence intervals or p-values) based on their long-run properties, even in cases where the experiments, surveys, etc. are practically unrepeatable.
  2. the lack of formal methods for incorporating background information and that they often use background information implicitly (e.g. experimental design, choice of likelihood, choice of error rates in testing).
  3. the disobedience of some classical procedures (e.g. p-values) to the likelihood principle.

Frequentists and Bayesians

(b) **Classical statistics**

- The common criticism for the standard approach is:
  1. the interpretation of outcomes (e.g. confidence intervals or p-values) based on their long-run properties, even in cases where the experiments, surveys, etc. are practically unrepeatable.
  2. the lack of formal methods for incorporating background information and that they often use background information implicitly (e.g. experimental design, choice of likelihood, choice of error rates in testing).
  3. the disobedience of some classical procedures (e.g. p-values) to the likelihood principle.

Frequentists and Bayesians

(b) **Classical statistics**

- The common criticism for the standard approach is:
  1. the interpretation of outcomes (e.g. confidence intervals or p-values) based on their long-run properties, even in cases where the experiments, surveys, etc. are practically unrepeatable.
  2. the lack of formal methods for incorporating background information and that they often use background information implicitly (e.g. experimental design, choice of likelihood, choice of error rates in testing).
  3. the disobedience of some classical procedures (e.g. p-values) to the likelihood principle.

Frequentists and Bayesians

(c) **Bayesian statistics:**

- In Bayesian statistics all unknown quantities, including parameters, are random and therefore they have probability distributions.
- These probability distributions are constructed using the rules of probability theory that ensure that they are well-defined.
- All probability distributions describe the "degree of belief" or uncertainty of one person about a quantity of interest.
- Probability distributions are always conditional to the information used to construct them. Therefore the use of background information, or indeed any form of information, and the subjective method used to construct them is explicitly open to scrutiny.

Frequentists and Bayesians

(c) **Bayesian statistics:**

- In Bayesian statistics all unknown quantities, including parameters, are random and therefore they have probability distributions.
- These probability distributions are constructed using the rules of probability theory that ensure that they are well-defined.
- All probability distributions describe the "degree of belief" or uncertainty of one person about a quantity of interest.
- Probability distributions are always conditional to the information used to construct them. Therefore the use of background information, or indeed any form of information, and the subjective method used to construct them is explicitly open to scrutiny.

Frequentists and Bayesians

(c) **Bayesian statistics:**

- In Bayesian statistics all unknown quantities, including parameters, are random and therefore they have probability distributions.
- These probability distributions are constructed using the rules of probability theory that ensure that they are well-defined.
- All probability distributions describe the "degree of belief" or uncertainty of one person about a quantity of interest.
- Probability distributions are always conditional to the information used to construct them. Therefore the use of background information, or indeed any form of information, and the subjective method used to construct them is explicitly open to scrutiny.

Frequentists and Bayesians

(c) **Bayesian statistics:**

- In Bayesian statistics all unknown quantities, including parameters, are random and therefore they have probability distributions.
- These probability distributions are constructed using the rules of probability theory that ensure that they are well-defined.
- All probability distributions describe the "degree of belief" or uncertainty of one person about a quantity of interest.
- Probability distributions are always conditional to the information used to construct them. Therefore the use of background information, or indeed any form of information, and the subjective method used to construct them is explicitly open to scrutiny.

Frequentists and Bayesians

(c) **Bayesian statistics:**

- In Bayesian statistics all unknown quantities, including parameters, are random and therefore they have probability distributions.
- These probability distributions are constructed using the rules of probability theory that ensure that they are well-defined.
- All probability distributions describe the "degree of belief" or uncertainty of one person about a quantity of interest.
- Probability distributions are always conditional to the information used to construct them. Therefore the use of background information, or indeed any form of information, and the subjective method used to construct them is explicitly open to scrutiny.

Frequentists and Bayesians

(c) **Bayesian statistics:**

- The openness in the use of background information is one of the most common arguments in support of the Bayesian approach.
- It is particularly useful in situations where the data are either very limited or non-existent (e.g. when studying rare events).
- However, it is also one of its common criticism.
- The argument against Bayesian Statistics is that it is subjective, and represents personal judgements. However...
- Non-informative priors, as well as prior sensitivity analysis (comparing inferences that result from different priors), can be used to alleviate this criticism.
- Bayesian methods can also be more computationally intensive, although complex models are often easier to fit within the Bayesian framework.

Frequentists and Bayesians

(c) **Bayesian statistics:**

- The openness in the use of background information is one of the most common arguments in support of the Bayesian approach.
- It is particularly useful in situations where the data are either very limited or non-existent (e.g. when studying rare events).
- However, it is also one of its common criticism.
- The argument against Bayesian Statistics is that it is subjective, and represents personal judgements. However...
- Non-informative priors, as well as prior sensitivity analysis (comparing inferences that result from different priors), can be used to alleviate this criticism.
- Bayesian methods can also be more computationally intensive, although complex models are often easier to fit within the Bayesian framework.

Frequentists and Bayesians

(c) **Bayesian statistics:**

- The openness in the use of background information is one of the most common arguments in support of the Bayesian approach.
- It is particularly useful in situations where the data are either very limited or non-existent (e.g. when studying rare events).
- However, it is also one of its common criticism.
- The argument against Bayesian Statistics is that it is subjective, and represents personal judgements. However...
- Non-informative priors, as well as prior sensitivity analysis (comparing inferences that result from different priors), can be used to alleviate this criticism.
- Bayesian methods can also be more computationally intensive, although complex models are often easier to fit within the Bayesian framework.

Frequentists and Bayesians

(c) **Bayesian statistics:**

- The openness in the use of background information is one of the most common arguments in support of the Bayesian approach.
- It is particularly useful in situations where the data are either very limited or non-existent (e.g. when studying rare events).
- However, it is also one of its common criticism.
- The argument against Bayesian Statistics is that it is subjective, and represents personal judgements. However...
- Non-informative priors, as well as prior sensitivity analysis (comparing inferences that result from different priors), can be used to alleviate this criticism.
- Bayesian methods can also be more computationally intensive, although complex models are often easier to fit within the Bayesian framework.

Frequentists and Bayesians

(c) **Bayesian statistics:**

- The openness in the use of background information is one of the most common arguments in support of the Bayesian approach.
- It is particularly useful in situations where the data are either very limited or non-existent (e.g. when studying rare events).
- However, it is also one of its common criticism.
- The argument against Bayesian Statistics is that it is subjective, and represents personal judgements. However...
- Non-informative priors, as well as prior sensitivity analysis (comparing inferences that result from different priors), can be used to alleviate this criticism.
- Bayesian methods can also be more computationally intensive, although complex models are often easier to fit within the Bayesian framework.

Frequentists and Bayesians

(c) **Bayesian statistics:**

- The openness in the use of background information is one of the most common arguments in support of the Bayesian approach.
- It is particularly useful in situations where the data are either very limited or non-existent (e.g. when studying rare events).
- However, it is also one of its common criticism.
- The argument against Bayesian Statistics is that it is subjective, and represents personal judgements. However...
- Non-informative priors, as well as prior sensitivity analysis (comparing inferences that result from different priors), can be used to alleviate this criticism.
- Bayesian methods can also be more computationally intensive, although complex models are often easier to fit within the Bayesian framework.

Frequentists and Bayesians

(c) **Bayesian statistics:**

- The openness in the use of background information is one of the most common arguments in support of the Bayesian approach.
- It is particularly useful in situations where the data are either very limited or non-existent (e.g. when studying rare events).
- However, it is also one of its common criticism.
- The argument against Bayesian Statistics is that it is subjective, and represents personal judgements. However...
- Non-informative priors, as well as prior sensitivity analysis (comparing inferences that result from different priors), can be used to alleviate this criticism.
- Bayesian methods can also be more computationally intensive, although complex models are often easier to fit within the Bayesian framework.

Outline

Why using Bayesian statistics?

- It is a natural and coherent way of thinking about science and learning based solely on standard probability theory.
- All unknowns are random variables. Random variables is what humans have agreed to use to express their uncertainty for something that is not known!
- Probability distributions are updated once more information becomes available, according to the laws of probability. Different procedures (e.g. LSE, MLE) are not required.
- It is becoming more acceptable to use prior background information in many areas of application.

Why using Bayesian statistics?

- It is a natural and coherent way of thinking about science and learning based solely on standard probability theory.
- All unknowns are random variables. Random variables is what humans have agreed to use to express their uncertainty for something that is not known!
- Probability distributions are updated once more information becomes available, according to the laws of probability. Different procedures (e.g. LSE, MLE) are not required.
- It is becoming more acceptable to use prior background information in many areas of application.

Why using Bayesian statistics?

- It is a natural and coherent way of thinking about science and learning based solely on standard probability theory.
- All unknowns are random variables. Random variables is what humans have agreed to use to express their uncertainty for something that is not known!
- Probability distributions are updated once more information becomes available, according to the laws of probability. Different procedures (e.g. LSE, MLE) are not required.
- It is becoming more acceptable to use prior background information in many areas of application.

Why using Bayesian statistics?

- It is a natural and coherent way of thinking about science and learning based solely on standard probability theory.
- All unknowns are random variables. Random variables is what humans have agreed to use to express their uncertainty for something that is not known!
- Probability distributions are updated once more information becomes available, according to the laws of probability. Different procedures (e.g. LSE, MLE) are not required.
- It is becoming more acceptable to use prior background information in many areas of application.

Why using Bayesian statistics?

- Non-informative/vague priors that express huge uncertainty can be used, also to examine the degree to which informative priors have affected the analysis.
- The audience to which the analyses are directed can then assess the validity of the results and decide if the inferences are justified. Sometimes, when data are not easily obtained, using prior information is the only option!
- Posterior distributions and their summaries are easier to interpret compared to, say, p-values and confidence intervals.
- It is often computationally easier and also more informative to do the analysis with the Bayesian approach (this is particularly true for complex analyses).

Why using Bayesian statistics?

- Non-informative/vague priors that express huge uncertainty can be used, also to examine the degree to which informative priors have affected the analysis.
- The audience to which the analyses are directed can then assess the validity of the results and decide if the inferences are justified. Sometimes, when data are not easily obtained, using prior information is the only option!
- Posterior distributions and their summaries are easier to interpret compared to, say, p-values and confidence intervals.
- It is often computationally easier and also more informative to do the analysis with the Bayesian approach (this is particularly true for complex analyses).

Why using Bayesian statistics?

- Non-informative/vague priors that express huge uncertainty can be used, also to examine the degree to which informative priors have affected the analysis.
- The audience to which the analyses are directed can then assess the validity of the results and decide if the inferences are justified. Sometimes, when data are not easily obtained, using prior information is the only option!
- Posterior distributions and their summaries are easier to interpret compared to, say, p-values and confidence intervals.
- It is often computationally easier and also more informative to do the analysis with the Bayesian approach (this is particularly true for complex analyses).

Why using Bayesian statistics?

- Non-informative/vague priors that express huge uncertainty can be used, also to examine the degree to which informative priors have affected the analysis.
- The audience to which the analyses are directed can then assess the validity of the results and decide if the inferences are justified. Sometimes, when data are not easily obtained, using prior information is the only option!
- Posterior distributions and their summaries are easier to interpret compared to, say, p-values and confidence intervals.
- It is often computationally easier and also more informative to do the analysis with the Bayesian approach (this is particularly true for complex analyses).

Further Readings (optional)

- Jordi Vallverdú. Bayesians Versus Frequentists. A Philosophical Debate on Statistical Reasoning. Springer (2016) [Available at https://www.st-andrews.ac.uk/library/]