

Introduction to Bayesian inference and its application in medical biology

P. Tsiamyrtzis^{1,2}

¹Dept. of Mechanical Engineering, Politecnico di Milano, Italy
panagiotis.tsiamyrtzis@polimi.it

²Dept. of Statistics, Athens University of Economics and Business, Greece
pt@aueb.gr

CNBH - Paris 2024



COLLEGE NATIONAL DE BIOCHIMIE DES HÔPITAUX

Organisme de formation continue n°82 07 00551 07

32^{èmes} Journées Nationales

Jeudi 25 et vendredi 26 janvier 2024

hôtel Ibis Paris 17 Clichy-Batignolles

**DECLARATION OF INTEREST
AS PART OF TRAINING MISSIONS
MADE FOR THE CNBH**

Professor Panagiotis Tsiamyrtzis

Working at Politecnico di Milano & Athens University of Economics and Business
(Establishment – service)

solemnly declare

I do not have any interest, direct or indirect (financial), with pharmaceutical, diagnostic or software publishing companies likely to modify my judgment or my comments, concerning the subject and the DMDIVs presented.

The Bayesian versus the Frequentist Approach

Data and Parameter(s) in Statistics

- In statistics a data set is becoming available via a random mechanism.

Data and Parameter(s) in Statistics

- In statistics a data set is becoming available via a random mechanism.
- A distribution $f(X|\theta)$ is used to describe the data generation procedure.

Data and Parameter(s) in Statistics

- In statistics a data set is becoming available via a random mechanism.
- A distribution $f(X|\theta)$ is used to describe the data generation procedure.
- For a given setup, the distribution is either:

Data and Parameter(s) in Statistics

- In statistics a data set is becoming available via a random mechanism.
- A distribution $f(X|\theta)$ is used to describe the data generation procedure.
- For a given setup, the distribution is either:
 - given by sampling design, e.g. a Binomial experiment with known number of trials, where $X|\theta \sim \text{Bin}(n, \theta)$, or

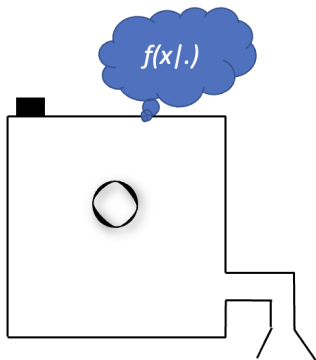
Data and Parameter(s) in Statistics

- In statistics a data set is becoming available via a random mechanism.
- A distribution $f(X|\theta)$ is used to describe the data generation procedure.
- For a given setup, the distribution is either:
 - given by sampling design, e.g. a Binomial experiment with known number of trials, where $X|\theta \sim \text{Bin}(n, \theta)$, or
 - we need to elicit it from the data, i.e. collect the data $\{x_1, \dots, x_n\}$ and derive the distribution, e.g. $X|(\theta_1, \theta_2) \sim N(\theta_1, \theta_2)$.

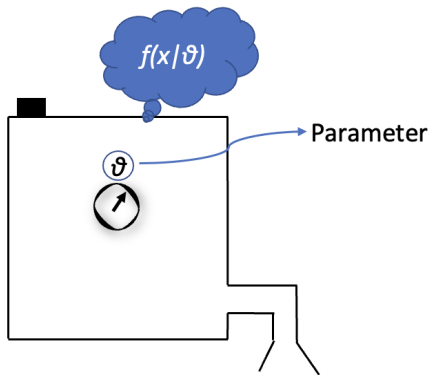
Data and Parameter(s) in Statistics

- In statistics a data set is becoming available via a random mechanism.
- A distribution $f(X|\theta)$ is used to describe the data generation procedure.
- For a given setup, the distribution is either:
 - given by sampling design, e.g. a Binomial experiment with known number of trials, where $X|\theta \sim \text{Bin}(n, \theta)$, or
 - we need to elicit it from the data, i.e. collect the data $\{x_1, \dots, x_n\}$ and derive the distribution, e.g. $X|(\theta_1, \theta_2) \sim N(\theta_1, \theta_2)$.
- Goodness of fit tests and plots should be used to examine how well an assumed theoretical distribution fits the observed data.

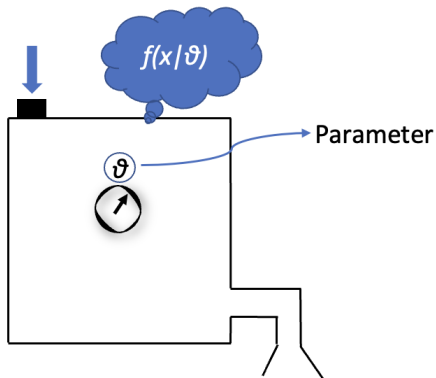
Distribution: random number generation engine



Distribution: random number generation engine

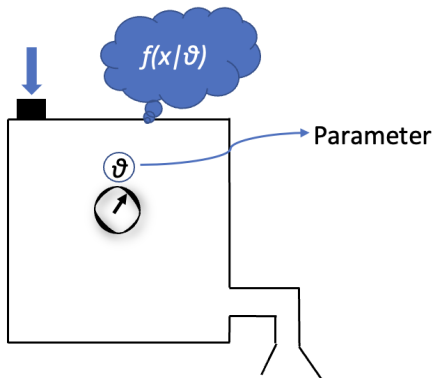


Distribution: random number generation engine



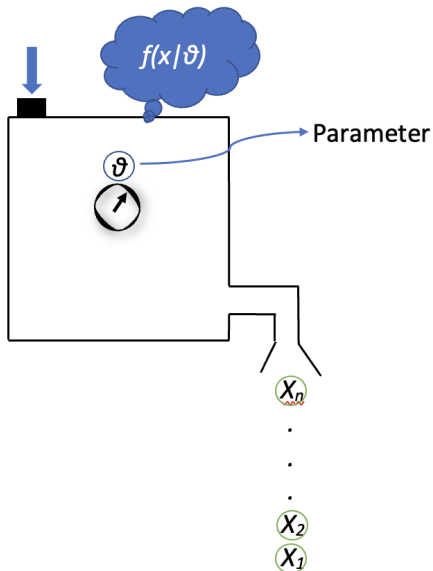
X_1

Distribution: random number generation engine

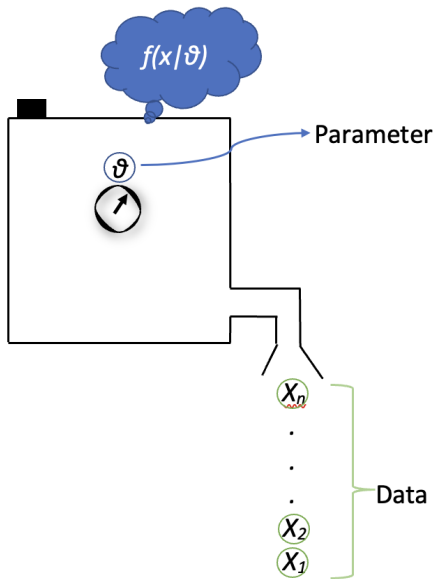


X_2
 X_1

Distribution: random number generation engine



Distribution: random number generation engine



Treating the unknown parameter θ

There are two (mainstream) schools of thought in statistics regarding the way we deal with the unknown parameter θ :

Treating the unknown parameter θ

There are two (mainstream) schools of thought in statistics regarding the way we deal with the unknown parameter θ :

(1) Frequentist (Neyman - Pearson - Wald)

Treating the unknown parameter θ

There are two (mainstream) schools of thought in statistics regarding the way we deal with the unknown parameter θ :

- (1) **Frequentist** (Neyman - Pearson - Wald)
- (2) **Bayesian** (Bayes)

Treating the unknown parameter θ

There are two (mainstream) schools of thought in statistics regarding the way we deal with the unknown parameter θ :

(1) **Frequentist** (Neyman - Pearson - Wald)

(2) **Bayesian** (Bayes)

- Both schools share the idea of the likelihood function, $f(\mathbf{x}|\theta)$, that is the joint distribution of the data $\mathbf{x} = \{x_1, \dots, x_n\}$, where the likelihood is considered to capture all the information about θ , that is available in the observed data $\mathbf{x} = \{x_1, \dots, x_n\}$ (i.e., the likelihood links the observed data with the unknown parameter via the distribution function)

Treating the unknown parameter θ

There are two (mainstream) schools of thought in statistics regarding the way we deal with the unknown parameter θ :

(1) **Frequentist** (Neyman - Pearson - Wald)

(2) **Bayesian** (Bayes)

- Both schools share the idea of the likelihood function, $f(\mathbf{x}|\theta)$, that is the joint distribution of the data $\mathbf{x} = \{x_1, \dots, x_n\}$, where the likelihood is considered to capture all the information about θ , that is available in the observed data $\mathbf{x} = \{x_1, \dots, x_n\}$ (i.e., the likelihood links the observed data with the unknown parameter via the distribution function)
- The two schools differ drastically on the way they handle the unknown parameter θ .

Frequentist School

- Within this school we adopt the frequency based interpretation of probability.

Frequentist School

- Within this school we adopt the frequency based interpretation of probability.
- **Frequentist probability** of an event is defined as the limit of its relative frequency in a large number of trials.

Frequentist School

- Within this school we adopt the frequency based interpretation of probability.
- **Frequentist probability** of an event is defined as the limit of its relative frequency in a large number of trials.

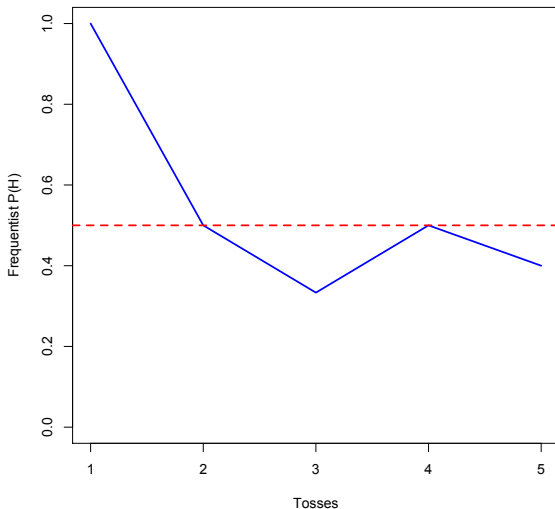
- **Example:**

In the case of a fair coin toss (la pièce lancée en l'air), the frequentist probability $P(H)$ is $1/2$, not because there are two equally likely outcomes (classical interpretation of probability) but because in repeated trials the empirical frequency converges to the limit $1/2$ as the number of trials goes to infinity, i.e.

$$P(H) = \frac{\# \text{ of heads}}{\# \text{ of trials}} = \frac{n_H}{n} \xrightarrow{n \rightarrow \infty} \frac{1}{2}$$

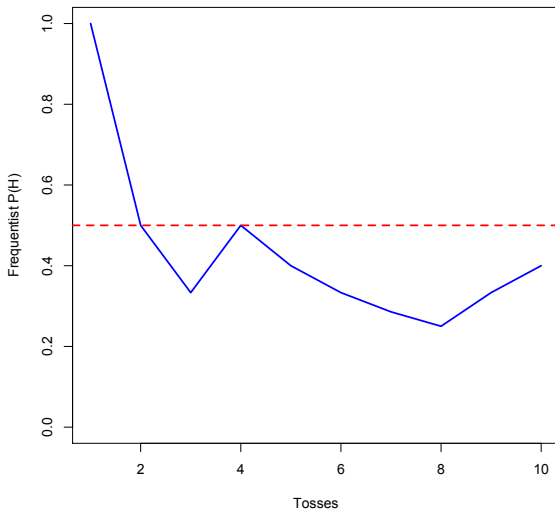
Frequentist School: probability interpretation

The proportion of Heads in 5 tosses of a fair coin



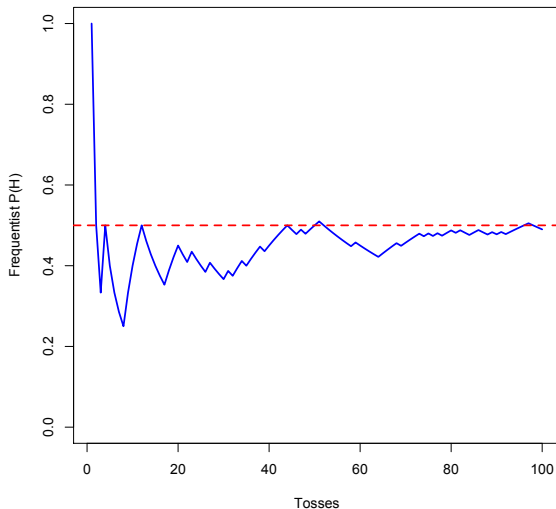
Frequentist School: probability interpretation

The proportion of Heads in 10 tosses of a fair coin



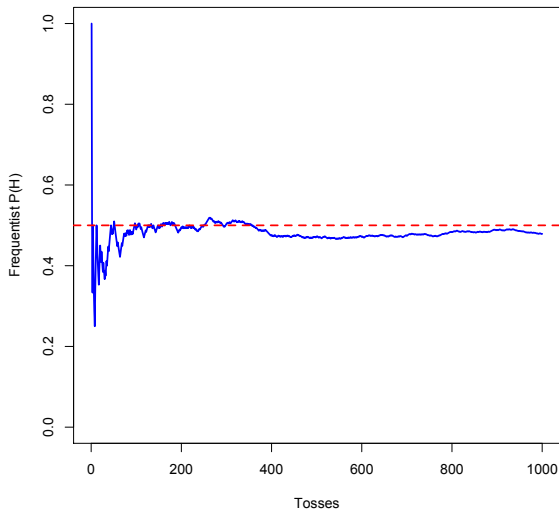
Frequentist School: probability interpretation

The proportion of Heads in 100 tosses of a fair coin



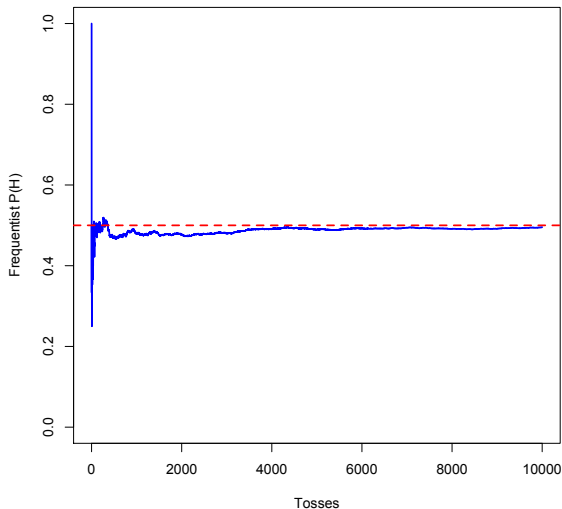
Frequentist School: probability interpretation

The proportion of Heads in 1000 tosses of a fair coin



Frequentist School: probability interpretation

The proportion of Heads in 10000 tosses of a fair coin



Data and Parameter(s) in Statistics

- The distribution that we chose for our data, comes along with a (univariate or multivariate) set of parameters that fully describe the random mechanism which produces the data. For example:
 - $X|\theta \sim \text{Bin}(n, \theta)$

Data and Parameter(s) in Statistics

- The distribution that we chose for our data, comes along with a (univariate or multivariate) set of parameters that fully describe the random mechanism which produces the data. For example:
 - $X|\theta \sim \text{Bin}(n, \theta)$
 - $X|(\theta_1, \theta_2) \sim N(\theta_1, \theta_2)$

Data and Parameter(s) in Statistics

- The distribution that we chose for our data, comes along with a (univariate or multivariate) set of parameters that fully describe the random mechanism which produces the data. For example:
 - $X|\theta \sim \text{Bin}(n, \theta)$
 - $X|(\theta_1, \theta_2) \sim N(\theta_1, \theta_2)$
- Usually we are interested in drawing either:

Data and Parameter(s) in Statistics

- The distribution that we chose for our data, comes along with a (univariate or multivariate) set of parameters that fully describe the random mechanism which produces the data. For example:
 - $X|\theta \sim \text{Bin}(n, \theta)$
 - $X|(\theta_1, \theta_2) \sim N(\theta_1, \theta_2)$
- Usually we are interested in drawing either:
 - **inference** for the unknown parameter(s) θ , i.e., derive point/interval estimates, hypothesis testing and/or

Data and Parameter(s) in Statistics

- The distribution that we chose for our data, comes along with a (univariate or multivariate) set of parameters that fully describe the random mechanism which produces the data. For example:
 - $X|\theta \sim \text{Bin}(n, \theta)$
 - $X|(\theta_1, \theta_2) \sim N(\theta_1, \theta_2)$
- Usually we are interested in drawing either:
 - **inference** for the unknown parameter(s) θ , i.e., derive point/interval estimates, hypothesis testing and/or
 - **predictions** for future observable(s).

Frequentist School

- The frequentist approach attempts to be “**objective**” in setting the probabilities. However, it relies heavily on the assumption that we are capable to repeat an experiment, infinite number of times, under “identical” conditions. The latter is restrictive in several real life applications making it really difficult to provide a frequentist based evaluation. Here are some examples:

Frequentist School

- The frequentist approach attempts to be “**objective**” in setting the probabilities. However, it relies heavily on the assumption that we are capable to repeat an experiment, infinite number of times, under “identical” conditions. The latter is restrictive in several real life applications making it really difficult to provide a frequentist based evaluation. Here are some examples:
- **Example 1:**
What is the probability that France will win the next World Cup in Rugby?

Frequentist School

- The frequentist approach attempts to be “**objective**” in setting the probabilities. However, it relies heavily on the assumption that we are capable to repeat an experiment, infinite number of times, under “identical” conditions. The latter is restrictive in several real life applications making it really difficult to provide a frequentist based evaluation. Here are some examples:
- **Example 1:**
What is the probability that France will win the next World Cup in Rugby?
- **Example 2:**
What is the probability that tomorrow noon will rain at Paris?

Frequentist School

- The frequentist approach attempts to be “**objective**” in setting the probabilities. However, it relies heavily on the assumption that we are capable to repeat an experiment, infinite number of times, under “identical” conditions. The latter is restrictive in several real life applications making it really difficult to provide a frequentist based evaluation. Here are some examples:
- **Example 1:**
What is the probability that France will win the next World Cup in Rugby?
- **Example 2:**
What is the probability that tomorrow noon will rain at Paris?
- The lack of repeatability in the previous examples lead to the embarrassing answer that this probability can either be 0 (if the event will not occur) or 1 (if it occurs).

- **Example 3:**

What is the probability that the first lot of the raw material arriving tomorrow at the lab will be of acceptable quality?

- **Example 3:**

What is the probability that the first lot of the raw material arriving tomorrow at the lab will be of acceptable quality?

- From a frequentist perspective, once we will examine the lot, the above probability is either 0 or 1. However, we tend to think that there exists an infinite population of cases exactly like the one above, from which we can randomly sample.

- **Example 3:**

What is the probability that the first lot of the raw material arriving tomorrow at the lab will be of acceptable quality?

- From a frequentist perspective, once we will examine the lot, the above probability is either 0 or 1. However, we tend to think that there exists an infinite population of cases exactly like the one above, from which we can randomly sample.

In all the above examples we have a Bernoulli distribution, $B(\theta)$, where θ is the probability of success, and we observe only one data point x , and so:

- **Example 3:**

What is the probability that the first lot of the raw material arriving tomorrow at the lab will be of acceptable quality?

- From a frequentist perspective, once we will examine the lot, the above probability is either 0 or 1. However, we tend to think that there exists an infinite population of cases exactly like the one above, from which we can randomly sample.

In all the above examples we have a Bernoulli distribution, $B(\theta)$, where θ is the probability of success, and we observe only one data point x , and so:

- $x = 0 \Rightarrow \hat{\theta} = 0$ or

- **Example 3:**

What is the probability that the first lot of the raw material arriving tomorrow at the lab will be of acceptable quality?

- From a frequentist perspective, once we will examine the lot, the above probability is either 0 or 1. However, we tend to think that there exists an infinite population of cases exactly like the one above, from which we can randomly sample.

In all the above examples we have a Bernoulli distribution, $B(\theta)$, where θ is the probability of success, and we observe only one data point x , and so:

- $x = 0 \Rightarrow \hat{\theta} = 0$ or
- $x = 1 \Rightarrow \hat{\theta} = 1$

- Based on the frequentist's approach and moving to inference, the parameter θ in the likelihood is considered to be a **fixed** unknown constant.

- Based on the frequentist's approach and moving to inference, the parameter θ in the likelihood is considered to be a **fixed** unknown constant.
- Thus inference regarding θ becomes available thanks to long term frequency properties. Precisely, we consider infinite repeated sampling, for **fixed** value of θ .

- Based on the frequentist's approach and moving to inference, the parameter θ in the likelihood is considered to be a **fixed** unknown constant.
- Thus inference regarding θ becomes available thanks to long term frequency properties. Precisely, we consider infinite repeated sampling, for **fixed** value of θ .
- While point estimation seems to be well aligned in this school, the assumption of a **fixed** parameter value can cause great difficulty in the interpretation of interval estimates (confidence intervals) and/or hypotheses testing.

Bayesian School: Subjective probability

- Within this school we have the **subjective** Bayesian interpretation of probabilities.

Bayesian School: Subjective probability

- Within this school we have the **subjective** Bayesian interpretation of probabilities.

Subjective probability:

A probability derived from an individual's personal belief, regarding the occurrence of a specific event. Subjective probabilities contain no formal calculations and reflect strictly the subject's personal information.

Bayesian School: Subjective probability

- Within this school we have the **subjective** Bayesian interpretation of probabilities.

Subjective probability:

A probability derived from an individual's personal belief, regarding the occurrence of a specific event. Subjective probabilities contain no formal calculations and reflect strictly the subject's personal information.

- **Example:**

In the case of tossing a coin, the subjective probability $P(H) = 0.4$ and $P(T) = 0.6$ can express one's personal opinion for a specific coin. A different person can come with a different pair of values based on his/her personal information.

Bayesian School: Coherence

- In contrast to the frequentist “objective” conjecture in the Bayesian school the probabilities are set **subjectively**, depending on **what bet a person is willing to take** regarding the occurrence of the specific outcome.

Bayesian School: Coherence

- In contrast to the frequentist “objective” conjecture in the Bayesian school the probabilities are set **subjectively**, depending on **what bet a person is willing to take** regarding the occurrence of the specific outcome.
- The assignment of the Bayesian probabilities needs to be done in a coherent way (De Finetti).

Bayesian School: Coherence

- In contrast to the frequentist “objective” conjecture in the Bayesian school the probabilities are set **subjectively**, depending on **what bet a person is willing to take** regarding the occurrence of the specific outcome.
- The assignment of the Bayesian probabilities needs to be done in a coherent way (De Finetti).

Coherence:

A system of bets (probabilities) is coherent when it does not allow a participant to be a sure loser, independently of the outcome of the experiment (i.e. we do not allow a Dutch book).

Bayesian School: Coherence

- In contrast to the frequentist “objective” conjecture in the Bayesian school the probabilities are set **subjectively**, depending on **what bet a person is willing to take** regarding the occurrence of the specific outcome.
- The assignment of the Bayesian probabilities needs to be done in a coherent way (De Finetti).

Coherence:

A system of bets (probabilities) is coherent when it does not allow a participant to be a sure loser, independently of the outcome of the experiment (i.e. we do not allow a Dutch book).

- Conformity to the probability calculus is necessary and sufficient for coherence.

- According to the Bayesian interpretation we can now easily provide (**subjective**) probabilities for the non-repeatable event examples (that the frequentist's probability failed):

- According to the Bayesian interpretation we can now easily provide (**subjective**) probabilities for the non-repeatable event examples (that the frequentist's probability failed):
- **Example 1:**
What is the probability that France will win the next World Cup in Rugby?

- According to the Bayesian interpretation we can now easily provide (**subjective**) probabilities for the non-repeatable event examples (that the frequentist's probability failed):
- **Example 1:**
What is the probability that France will win the next World Cup in Rugby?
- **Example 2:**
What is the probability that tomorrow noon will rain at Paris?

- According to the Bayesian interpretation we can now easily provide (**subjective**) probabilities for the non-repeatable event examples (that the frequentist's probability failed):
- **Example 1:**
What is the probability that France will win the next World Cup in Rugby?
- **Example 2:**
What is the probability that tomorrow noon will rain at Paris?
- **Example 3:**
What is the probability that the first lot of the raw material arriving tomorrow in the lab will be of acceptable quality?

- The data \mathbf{x} become available via the random mechanism $f(\mathbf{x}|\theta)$, where θ is an unknown constant.

Bayesian School

- The data \mathbf{x} become available via the random mechanism $f(\mathbf{x}|\theta)$, where θ is an unknown constant.
- When something is unknown to you, i.e. θ , a reasonable thing to do is to use probability theory to quantify your uncertainty.

Bayesian School

- The data \mathbf{x} become available via the random mechanism $f(\mathbf{x}|\theta)$, where θ is an unknown constant.
- When something is unknown to you, i.e. θ , a reasonable thing to do is to use probability theory to quantify your uncertainty.
- So in the Bayesian school the parameter θ is considered to be a random variable.

Bayesian School

- The data \mathbf{x} become available via the random mechanism $f(\mathbf{x}|\theta)$, where θ is an unknown constant.
- When something is unknown to you, i.e. θ , a reasonable thing to do is to use probability theory to quantify your uncertainty.
- So in the Bayesian school the parameter θ is considered to be a random variable.
- Its distribution, will quantify our (**subjective**) opinion regarding θ (before looking the data) with a prior distribution: $\pi(\theta)$.

Bayesian School

- The data \mathbf{x} become available via the random mechanism $f(\mathbf{x}|\theta)$, where θ is an unknown constant.
- When something is unknown to you, i.e. θ , a reasonable thing to do is to use probability theory to quantify your uncertainty.
- So in the Bayesian school the parameter θ is considered to be a random variable.
- Its distribution, will quantify our (**subjective**) opinion regarding θ (before looking the data) with a prior distribution: $\pi(\theta)$.
- Then Bayes theorem will do the magic **updating** the prior distribution to posterior, in the light of the data.

Bayes theorem

- The Bayes theorem for distributions is given by:

$$p(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int f(\mathbf{x}|\theta)\pi(\theta)d\theta}$$

where:

Bayes theorem

- The Bayes theorem for distributions is given by:

$$p(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int f(\mathbf{x}|\theta)\pi(\theta)d\theta}$$

where:

- $\mathbf{x} = \{x_1, \dots, x_n\}$ are the observed data.
- $f(\mathbf{x}|\theta)$ refers to the likelihood of the data
- $\pi(\theta)$ is the prior distribution of the parameter θ ,
- $p(\theta|\mathbf{x})$ is the posterior distribution of the parameter θ , given the observed data $\mathbf{x} = \{x_1, \dots, x_n\}$.

Bayes theorem

- The Bayes theorem for distributions is given by:

$$p(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int f(\mathbf{x}|\theta)\pi(\theta)d\theta}$$

where:

- $\mathbf{x} = \{x_1, \dots, x_n\}$ are the observed data.
 - $f(\mathbf{x}|\theta)$ refers to the likelihood of the data
 - $\pi(\theta)$ is the prior distribution of the parameter θ ,
 - $p(\theta|\mathbf{x})$ is the posterior distribution of the parameter θ , given the observed data $\mathbf{x} = \{x_1, \dots, x_n\}$.
- So the Bayes theorem is nothing more than an **updating mechanism**, where the prior is updated to posterior in the light of evidence coming from the available data.

Bayesian School

The Bayesian approach consists of the following steps:

- (a) Define the likelihood: $f(\mathbf{x}|\theta)$.
- (b) Define the prior distribution: $\pi(\theta)$.
- (c) Compute the posterior distribution: $p(\theta|\mathbf{x})$.
- (d) Decision Making: Draw inference regarding θ (point/interval estimates and hypothesis testing).
- (e) Derive the predictive distribution $f(x_{n+1}|x_1, x_2, \dots, x_n)$ of a future observable and make predictions.

Prior Distribution

- This is the key element of the Bayesian approach.

- This is the key element of the Bayesian approach.
- **Subjective Bayesian approach:**
The parameter of interest takes eventually a single number, which is used in the likelihood to provide the data. Since we do not know this value, we use a random mechanism (the prior $\pi(\theta)$) to describe the uncertainty about this parameter value. Thus, we simply use probability theory to model the uncertainty.

- This is the key element of the Bayesian approach.
- **Subjective Bayesian approach:**
The parameter of interest takes eventually a single number, which is used in the likelihood to provide the data. Since we do not know this value, we use a random mechanism (the prior $\pi(\theta)$) to describe the uncertainty about this parameter value. Thus, we simply use probability theory to model the uncertainty.
- The prior should reflect our personal (**subjective**) opinion (information) regarding the parameter, **before** we look at the data. The only thing we need to be careful about, is to be coherent, which will happen if we will obey the probability laws.

Prior Distribution

Main issues regarding prior distributions:

Prior Distribution

Main issues regarding prior distributions:

- Posterior lives in the range defined by the prior.

Prior Distribution

Main issues regarding prior distributions:

- Posterior lives in the range defined by the prior.
- The more data we get the less the effect of the prior in determining the posterior distribution (unless extreme choices, like point mass priors are made).

Prior Distribution

Main issues regarding prior distributions:

- Posterior lives in the range defined by the prior.
- The more data we get the less the effect of the prior in determining the posterior distribution (unless extreme choices, like point mass priors are made).
- Different priors applied on the same data will lead to different posteriors.

Prior Distribution

Main issues regarding prior distributions:

- Posterior lives in the range defined by the prior.
- The more data we get the less the effect of the prior in determining the posterior distribution (unless extreme choices, like point mass priors are made).
- Different priors applied on the same data will lead to different posteriors.
- The last bullet, raised (and keeps raising) the major criticism from non-Bayesians (see for example Efron (1986), “Why isn’t everyone a Bayesian”). However, Bayesians love the opportunity to be subjective. Lets see an example:

Prior Distribution: example

We have two different binomial experiments.

Prior Distribution: example

We have two different binomial experiments.

- **Setup 1:** We ask from a sommelier (wine expert) to taste 10 glasses of wine and decide whether each glass is Merlot or Pinot Noir.

Prior Distribution: example

We have two different binomial experiments.

- **Setup 1:** We ask from a sommelier (wine expert) to taste 10 glasses of wine and decide whether each glass is Merlot or Pinot Noir.
- **Setup 2:** We ask from a drunk man to guess the sequence of H and T in 10 tosses of a fair coin.

Prior Distribution: example

We have two different binomial experiments.

- **Setup 1:** We ask from a sommelier (wine expert) to taste 10 glasses of wine and decide whether each glass is Merlot or Pinot Noir.
- **Setup 2:** We ask from a drunk man to guess the sequence of H and T in 10 tosses of a fair coin.

In both cases we have a Binomial $B(10, \theta)$, with unknown the probability of success (θ) .

Prior Distribution: example

We have two different binomial experiments.

- **Setup 1:** We ask from a sommelier (wine expert) to taste 10 glasses of wine and decide whether each glass is Merlot or Pinot Noir.
- **Setup 2:** We ask from a drunk man to guess the sequence of H and T in 10 tosses of a fair coin.

In both cases we have a Binomial $B(10, \theta)$, with unknown the probability of success (θ).

The data become available and let's assume that we have 10 successes in both setups (i.e. 10 successes in 10 trials).

Prior Distribution: example

We have two different binomial experiments.

- **Setup 1:** We ask from a sommelier (wine expert) to taste 10 glasses of wine and decide whether each glass is Merlot or Pinot Noir.
- **Setup 2:** We ask from a drunk man to guess the sequence of H and T in 10 tosses of a fair coin.

In both cases we have a Binomial $B(10, \theta)$, with unknown the probability of success (θ) .

The data become available and let's assume that we have 10 successes in both setups (i.e. 10 successes in 10 trials).

Please write down your best guess the value of (θ) in each setup.

Prior Distribution: example

We have two different binomial experiments.

- **Setup 1:** We ask from a sommelier (wine expert) to taste 10 glasses of wine and decide whether each glass is Merlot or Pinot Noir.
- **Setup 2:** We ask from a drunk man to guess the sequence of H and T in 10 tosses of a fair coin.

In both cases we have a Binomial $B(10, \theta)$, with unknown the probability of success (θ).

The data become available and let's assume that we have 10 successes in both setups (i.e. 10 successes in 10 trials).

Please write down your best guess the value of (θ) in each setup.

Based on the frequentist approach the Maximum Likelihood Estimate of the unknown parameter is $\hat{\theta} = \bar{x} = 1$ in both cases.

Prior Distribution: example

But is this really what we believe?

Prior Distribution: example

But is this really what we believe?

Before looking in the data, if you were to bet money in which setup has higher success probability θ , would you put your money to setup 1 or 2? or did you think that the probabilities of success were equal in both setups?

Prior Distribution: example

But is this really what we believe?

Before looking in the data, if you were to bet money in which setup has higher success probability θ , would you put your money to setup 1 or 2? or did you think that the probabilities of success were equal in both setups?

For the sommelier we expect to have the probability of success close to 1, while for the drunk man we would expect his success rate to be close to $1/2$.

Prior Distribution: example

But is this really what we believe?

Before looking in the data, if you were to bet money in which setup has higher success probability θ , would you put your money to setup 1 or 2? or did you think that the probabilities of success were equal in both setups?

For the sommelier we expect to have the probability of success close to 1, while for the drunk man we would expect his success rate to be close to $1/2$.

Adopting the appropriate prior distribution for each setup would lead to different posteriors, in contrast to the frequentist based methods that will force to have identical results.

Prior Distribution: example

But is this really what we believe?

Before looking in the data, if you were to bet money in which setup has higher success probability θ , would you put your money to setup 1 or 2? or did you think that the probabilities of success were equal in both setups?

For the sommelier we expect to have the probability of success close to 1, while for the drunk man we would expect his success rate to be close to $1/2$.

Adopting the appropriate prior distribution for each setup would lead to different posteriors, in contrast to the frequentist based methods that will force to have identical results.

If the two probability numbers that you wrote down about θ in the previous setups were not the same number, then you are thinking and acting as **Bayesian** (even though you might not know about it!).

Computing the posterior: example

- Assume that we have a random sample of size n : $\mathbf{x} = (x_1, x_2, \dots, x_n)$ from the Normal (Gaussian) distribution, i.e.:

$$X_i | \theta \sim N(\theta, \sigma^2)$$

with σ^2 being known.

- If the prior distribution of the unknown parameter θ is Normal, i.e. :

$$\pi(\theta) \sim N(\mu, \tau^2)$$

with μ and τ^2 being known.

- Then the posterior distribution of $\theta | \mathbf{x}$ will be given by:

$$\theta | \mathbf{x} \sim N \left(\frac{\tau^2 \bar{X} + \frac{\sigma^2}{n} \mu}{\frac{\sigma^2}{n} + \tau^2}, \frac{\frac{\sigma^2}{n} \tau^2}{\frac{\sigma^2}{n} + \tau^2} \right)$$

Sensitivity Analysis

We will use the posterior distribution of $\theta|\mathbf{x}$ derived in the last example:

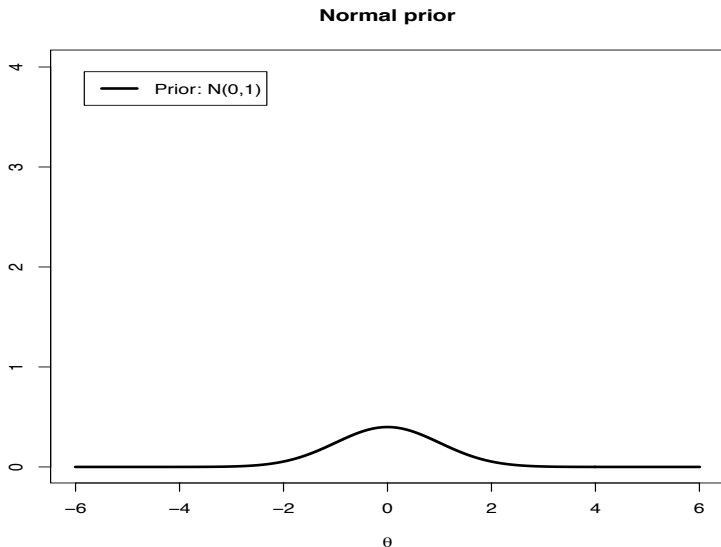
$$\theta|\mathbf{x} \sim N \left(\frac{\tau^2 \bar{x} + \frac{\sigma^2}{n} \mu}{\frac{\sigma^2}{n} + \tau^2}, \frac{\frac{\sigma^2}{n} \tau^2}{\frac{\sigma^2}{n} + \tau^2} \right)$$

Let's look on some graphical illustrations regarding the effect of the sample size n and the variance of the prior distribution, τ^2 . Specifically, let's assume that $\bar{x} = 4$ and:

- $n = 1, 10, 100$ with $\pi(\theta) \sim N(0, 1)$
- $n = 1$ with $\pi(\theta) \sim N(0, 10^2)$
- $n = 1, 10, 100$ with $\pi(\theta) \sim N(0, 0.1^2)$

Sensitivity Analysis

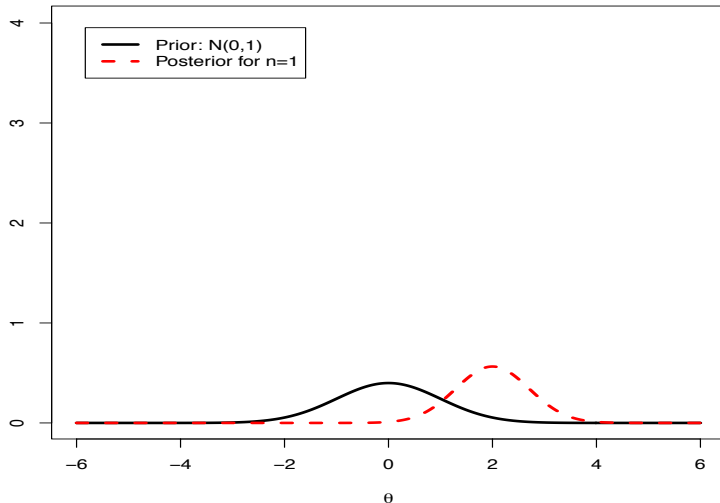
Plot of the $N(0, 1)$ prior distribution:



Sensitivity Analysis

Plot of $p(\theta|x)$, when $n = 1$ with $\pi(\theta) \sim N(0,1)$:

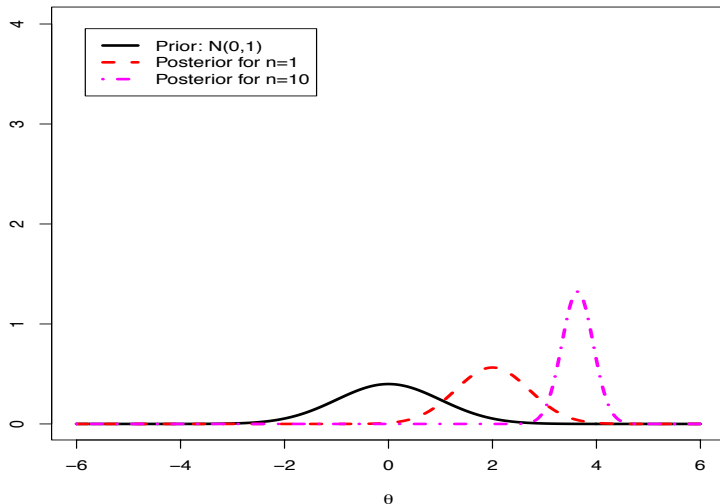
Normal prior and likelihood with various sample sizes n and $\bar{x}=4$



Sensitivity Analysis

Plot of $p(\theta|x)$, when $n = 1, 10$ with $\pi(\theta) \sim N(0, 1)$:

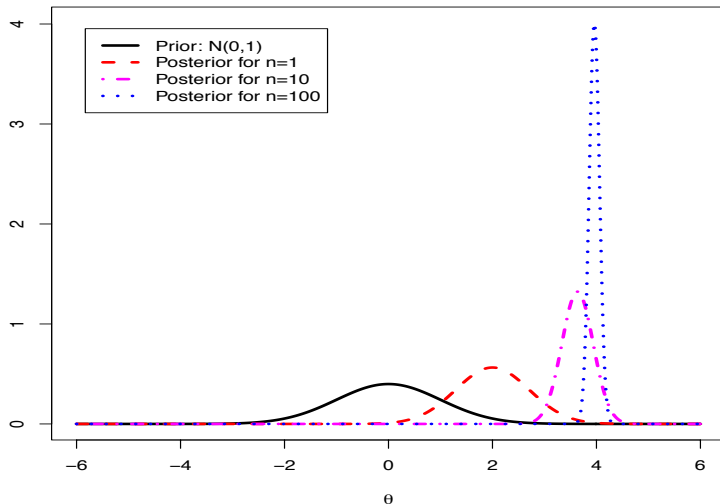
Normal prior and likelihood with various sample sizes n and $\bar{x}=4$



Sensitivity Analysis

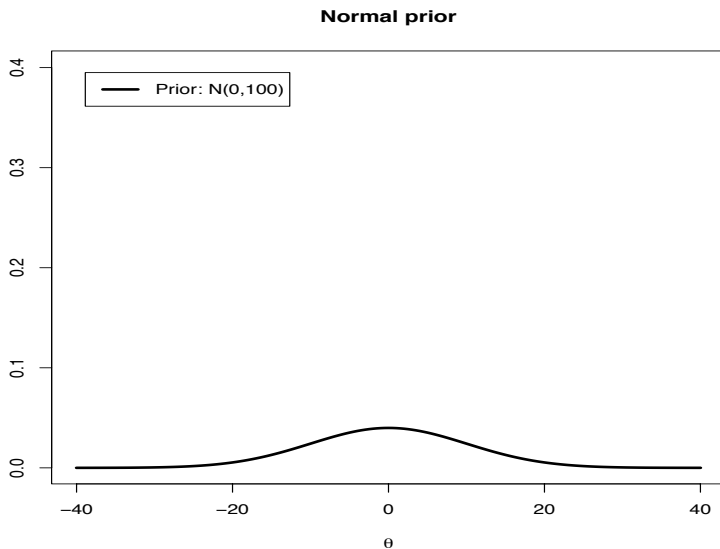
Plot of $p(\theta|x)$, when $n = 1, 10, 100$ with $\pi(\theta) \sim N(0, 1)$:

Normal prior and likelihood with various sample sizes n and $\bar{x}=4$



Sensitivity Analysis

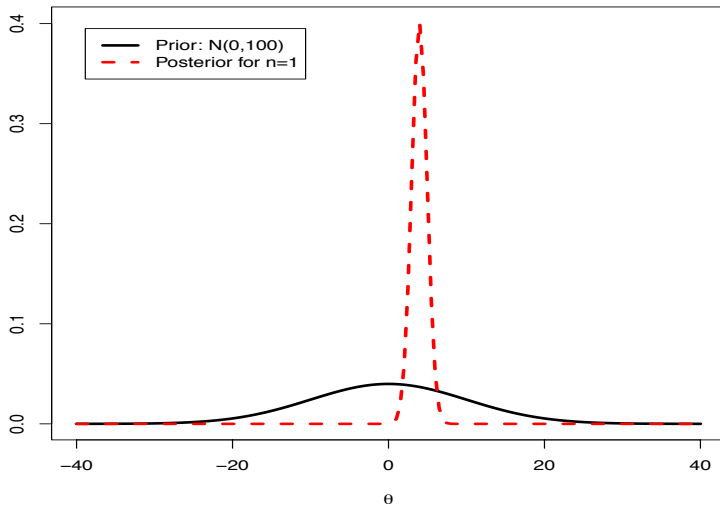
Plot of the $N(0, 10^2)$ prior distribution:



Sensitivity Analysis

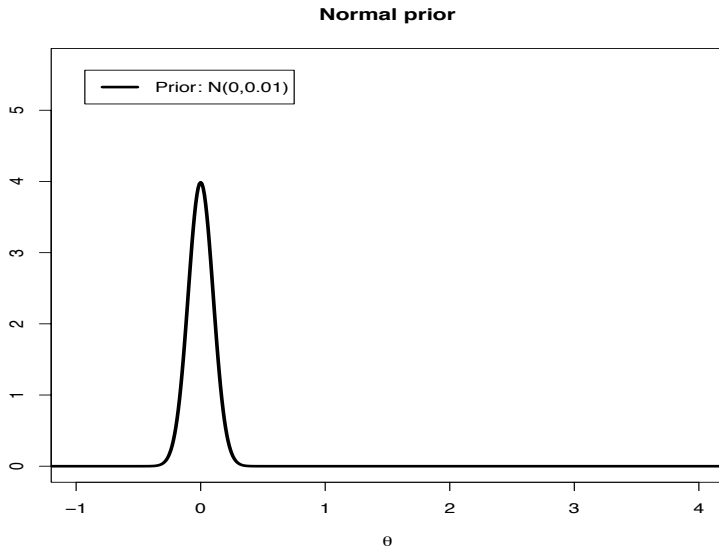
Plot of $p(\theta|x)$, when $n = 1$ with $\pi(\theta) \sim N(0, 10^2)$

Normal prior and likelihood with sample size $n=1$ and $\bar{x}=4$



Sensitivity Analysis

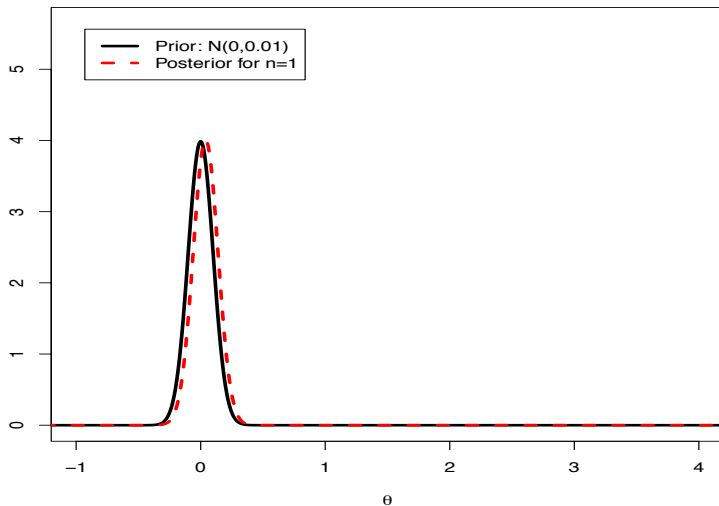
Plot of the $N(0, 0.1^2)$ prior distribution:



Sensitivity Analysis

Plot of $p(\theta|x)$, when $n = 1$ with $\pi(\theta) \sim N(0, 0.1^2)$

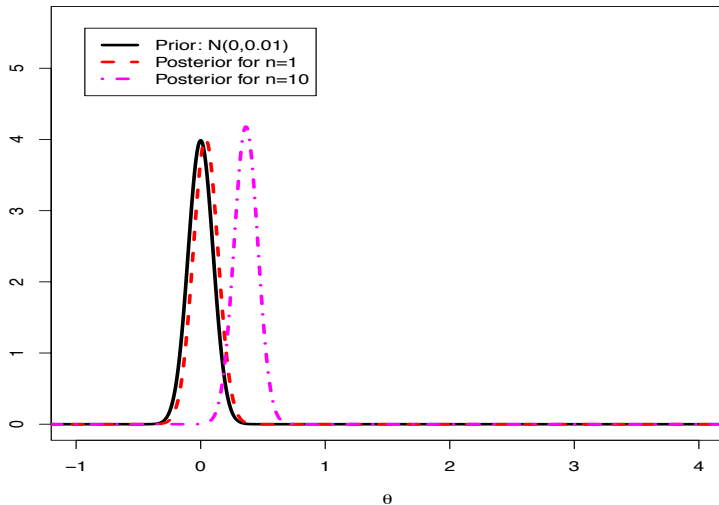
Normal prior and likelihood with various sample sizes n and $\bar{x}=4$



Sensitivity Analysis

Plot of $p(\theta|x)$, when $n = 1, 10$ with $\pi(\theta) \sim N(0, 0.1^2)$

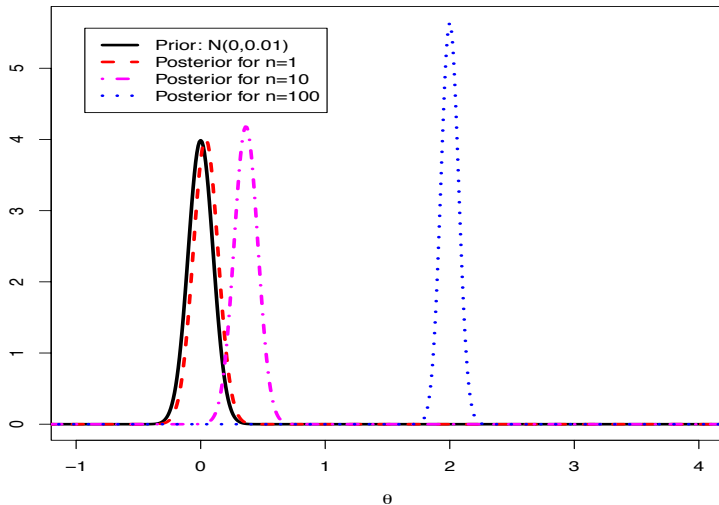
Normal prior and likelihood with various sample sizes n and $\bar{x}=4$



Sensitivity Analysis

Plot of $p(\theta|x)$, when $n = 1, 10, 100$ with $\pi(\theta) \sim N(0, 0.1^2)$

Normal prior and likelihood with various sample sizes n and $\bar{x}=4$



The Bayesian approach rocks!

Thomas Bayes



Bayesian Statistical Process Control & Monitoring (Quality Control)

Why Bayesian SPC/M?

- Within the Bayesian approach the unknown parameter(s) θ can be integrated out, deriving the **predictive** distribution.
- Use of Bayes theorem, updates the (power) prior $\pi(\theta)$ to posterior $p(\theta|\mathbf{X})$ and then for future observable(s) $\mathbf{Y} = X_{n+1}$ we get:

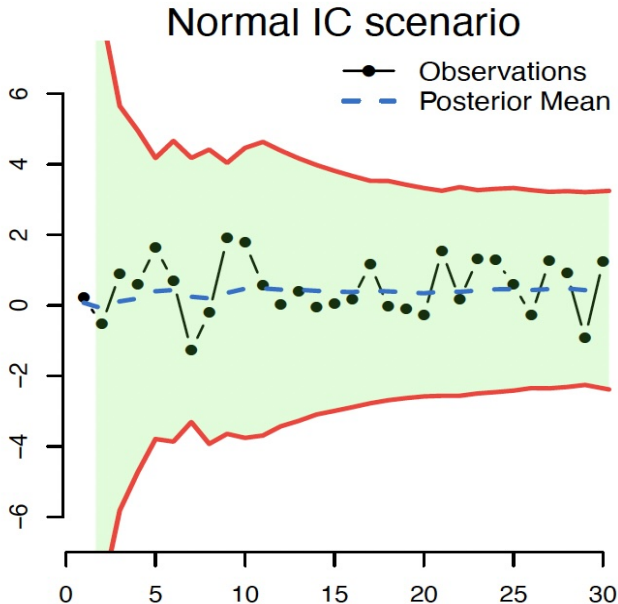
$$f(\mathbf{Y}|\mathbf{X}) = \int f(\mathbf{Y}|\theta) p(\theta|\mathbf{X}) d\theta$$

- Based on the predictive distribution we will derive two monitoring schemes:
 - PCC:** Predictive Control Charts, for detecting **transient** shifts of large magnitude (outliers).
 - PRC:** Predictive Residual Cusum: for detection of **persistent** shifts of medium/small size (extended to Predictive Ratio Cusum for any distribution in the exponential family).

PCC formulation

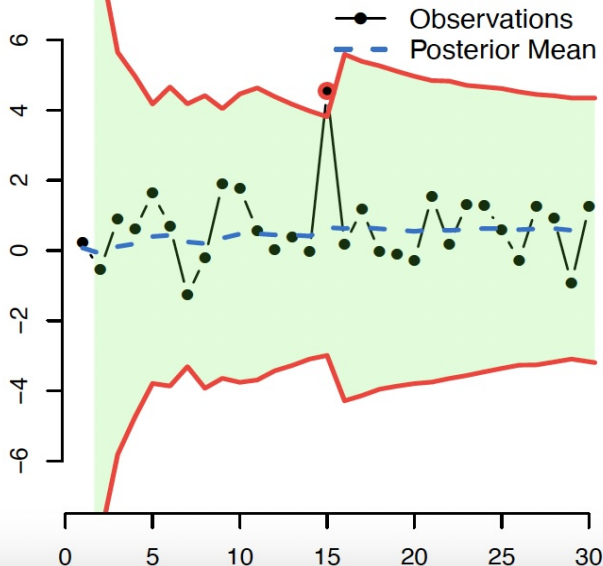
- The Predictive Control Chart (**PCC**) construction will be based on the predictive distribution and it can start as soon as $n = 2$ and is based on the sequentially updated form of the predictive distribution.
- Precisely, we will determine an IC region, R_{n+1} , where the future observable (X_{n+1}) will most likely be, as long as the process is stable (i.e. no changes occurred).
- The limits of the predictive IC region, R_{n+1} , will be established based on either the overall False Alarm Probability (FAP) or the Average Run Length (ARL).
- Remember that thanks to the (sequentially updated) Bayesian approach, at each time point we test if the process is IC or OOC (i.e. we **control** the process), but we can also draw inference (i.e. **monitor** the process), having a Bayes optimal point estimate of the unknown parameter(s).

PCC Illustration and Decision Making



PCC Illustration and Decision Making

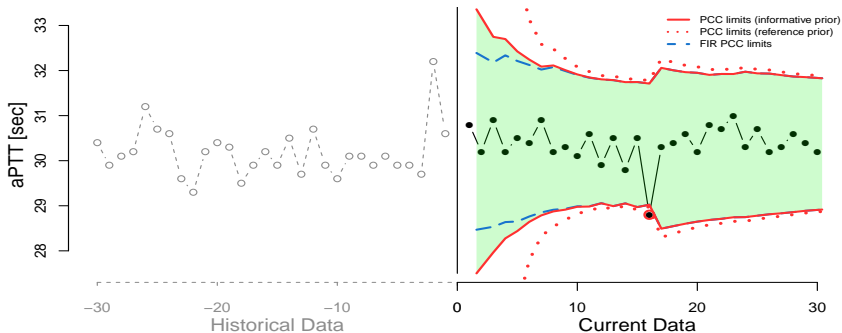
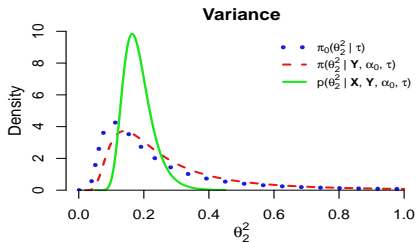
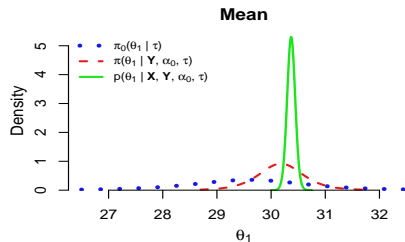
Normal OOC scenario



PCC Real Data Application (Medical Lab)

- We will use data that come from the daily Internal Quality Control (IQC) routine of a medical laboratory, monitoring “activated Partial Thromboplastin Time” (aPTT), measured in seconds.
- We gathered 30 daily normal IQC observations (X_i) from a medical lab. Notice that these data are based on control samples and in regular practice will become available sequentially.
- The goal is to accurately detect any transient parameter shift of large size, as this will have an impact on the reported patient results. Thus, it is of major importance to perform on-line monitoring of the process without a phase I exercise.
- We elicit the prior $\pi_0(\theta_1, \theta_2^2 | \tau) \sim \text{NIG}(29.6, 1/7, 2, 0.56^2)$ and we had $n_0 = 30$ historical data (from a different reagent).

PCC Real Data Application (Medical Lab)



- In PRC we will guard against informative OOC scenarios, focusing on detecting **persistent** parameter shifts of either parameter of the Normal distribution. Precisely, we developed:
 - **loc-PRC**: Location PRC that aims to identify either upward or downward persistent mean shifts of moderate/small size.
 - **sc-PRC**: Scale PRC that aims to identify persistent variance increase of moderate/small size.
- Again, thanks to the (sequentially updated) Bayesian approach, at each time point we test if the process is IC or OOC (i.e. we **control** the process), but we can also draw inference (i.e. **monitor** the process), having a Bayes optimal point estimate of the unknown mean and variance parameters.

PRC Real Data Application (Medical Lab)

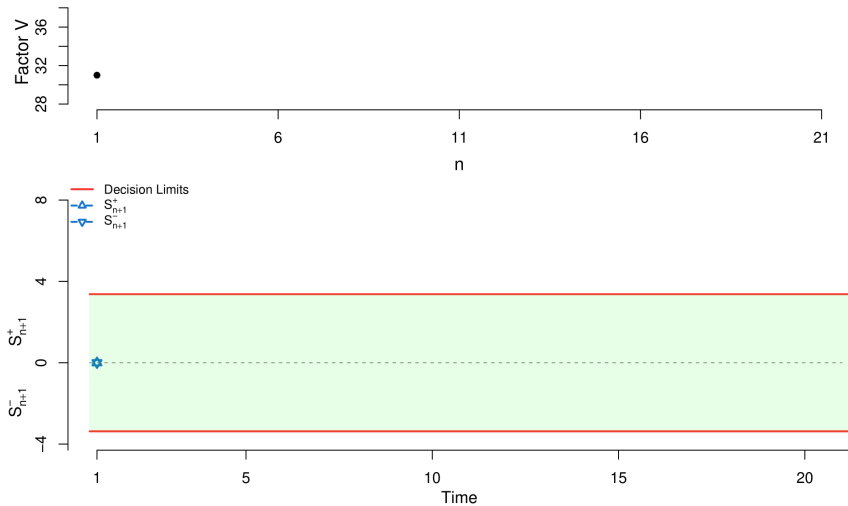
- We will use hemostasis data that come from the daily Internal Quality Control (IQC) routine of a medical laboratory.
- We are interested in the variable “Factor V”, measured in % regarding the international standards in clinical hemostasis, whose deficiencies can induce bleeding disorders of varying severity.
- A change of reagent batch in a lab can introduce a step change to the measurement of Factor V. It is crucial to identify such a change point, especially at the start of the process to avoid impacting clinically the patients care.
- We sequentially gathered 21 normally distributed IQC observations (X_i) from the control samples of a medical lab, where $X_i | (\theta_1, \theta_2^2) \sim N(\theta_1, \theta_2^2)$. We choose the parameter $k = \pm 1$ and we tune the PRC in detecting mean step changes, in either upward or downward direction, of one standard deviation size (i.e., $\pm \hat{\theta}_2$).

loc-PRC Real Data Application (Medical Lab)

- We elicit the initial prior $\pi_0(\theta_1, \theta_2^2 | \tau) \sim \text{NIG}(31.8, 1/2, 2, 4.41)$.
- Furthermore, we have $n_0 = 37$ IC historical data with $\bar{y} = 31.73$ and $\text{var}(\mathbf{y}) = 3.31$. We set $\alpha_0 = 1/37$ to convey the weight of a single data point to these.
- Combining the two sources of information we obtain:
 $\pi(\theta_1, \theta_2^2 | \mathbf{Y}, \alpha_0, \tau) \sim \text{NIG}(31.75, 3/2, 5/2, 6.02)$.
- We derive the PRC's decision limits h^+ and h^- to achieve $\text{FWER} = 5\%$ for 21 observations in a two-sided loc-PRC.

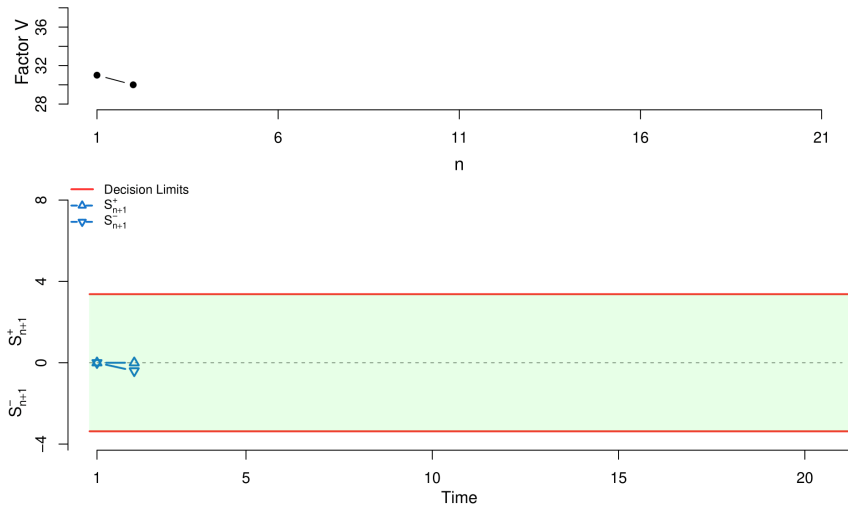
loc-PRC Real Data Application (Medical Lab)

Data



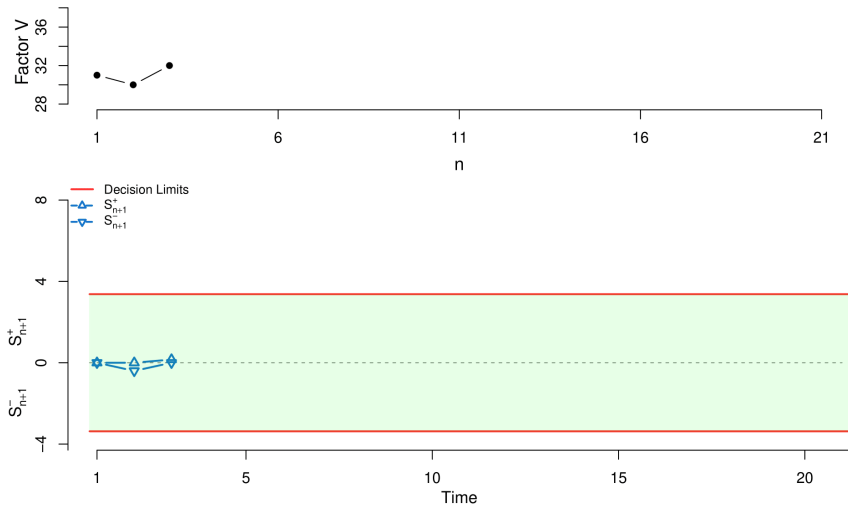
loc-PRC Real Data Application (Medical Lab)

Data



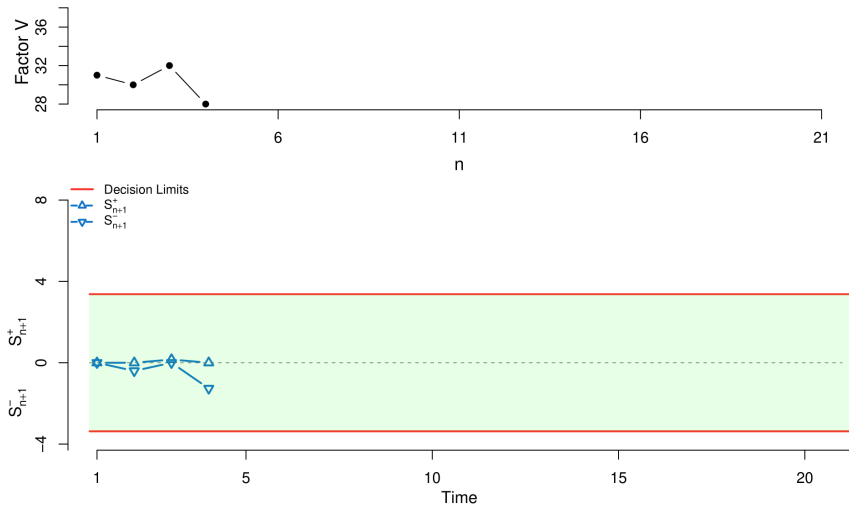
loc-PRC Real Data Application (Medical Lab)

Data



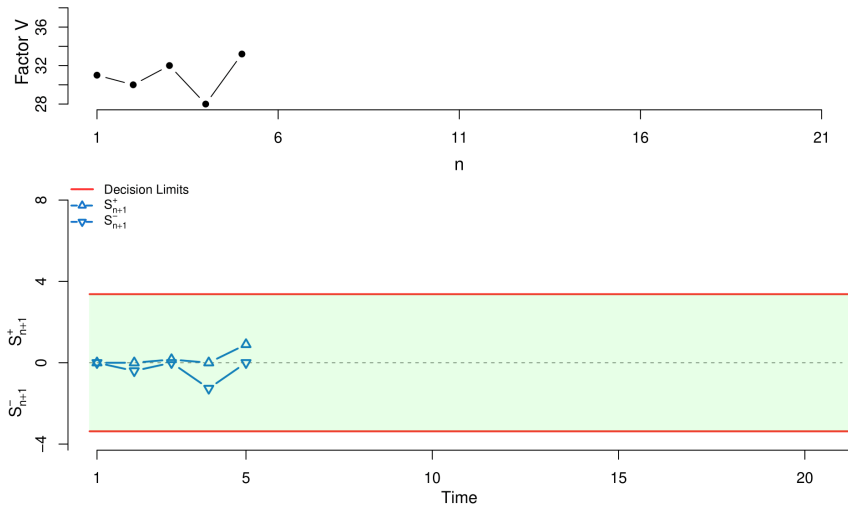
loc-PRC Real Data Application (Medical Lab)

Data



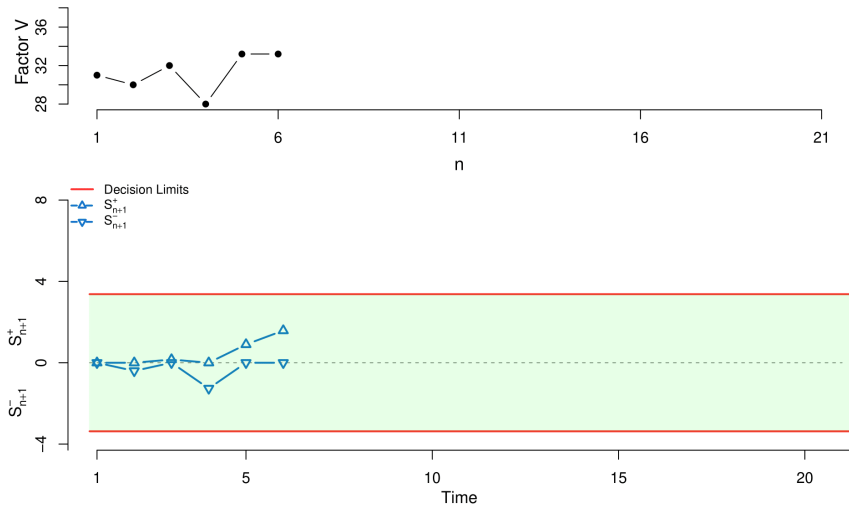
loc-PRC Real Data Application (Medical Lab)

Data



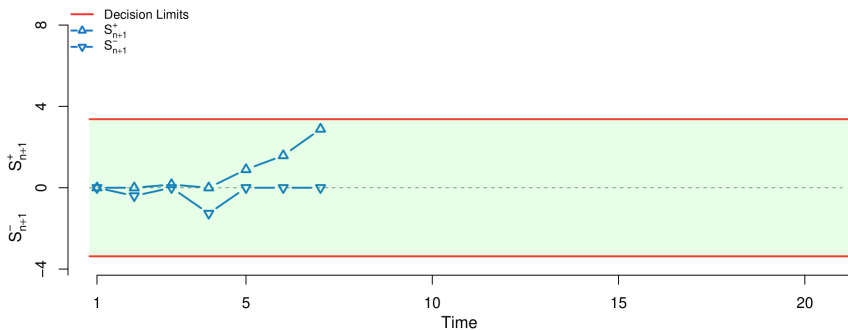
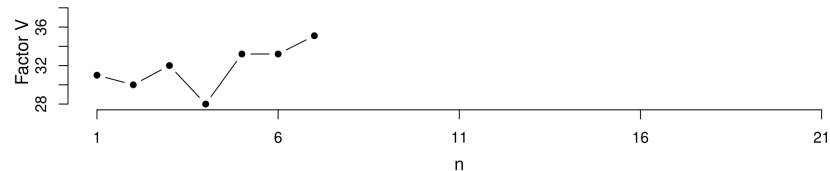
loc-PRC Real Data Application (Medical Lab)

Data



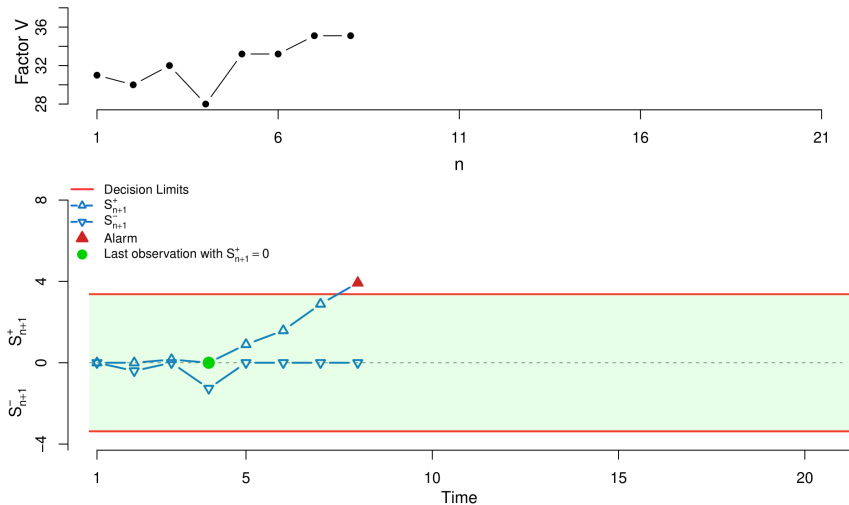
loc-PRC Real Data Application (Medical Lab)

Data



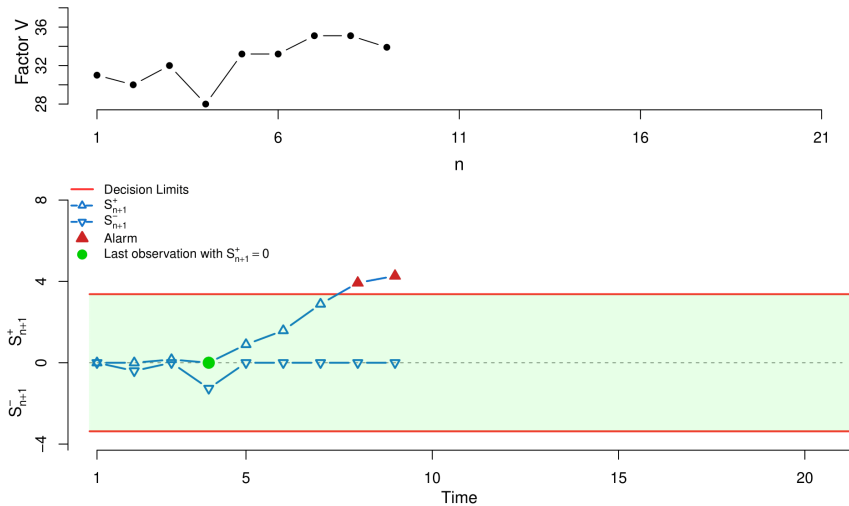
loc-PRC Real Data Application (Medical Lab)

Data

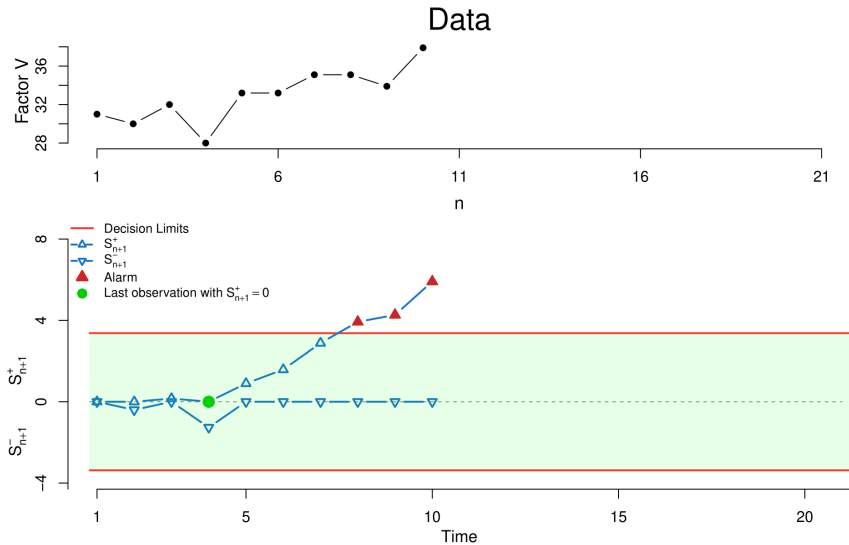


loc-PRC Real Data Application (Medical Lab)

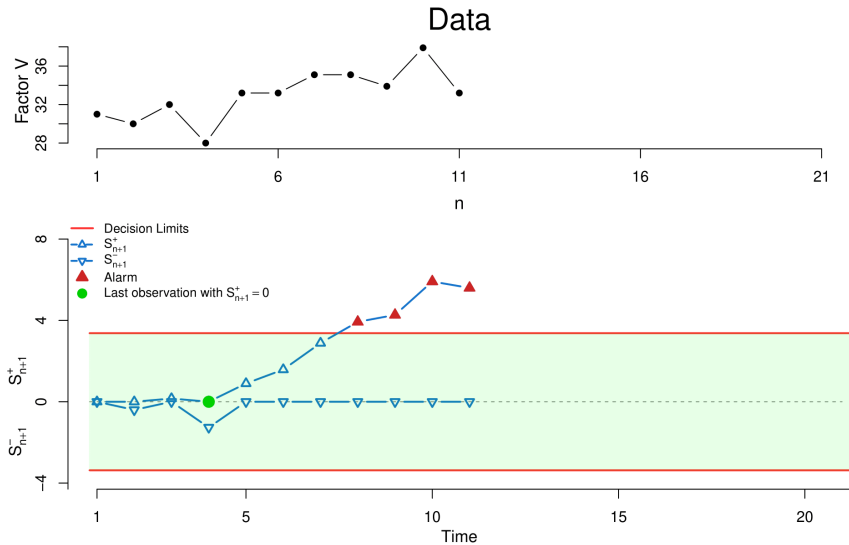
Data



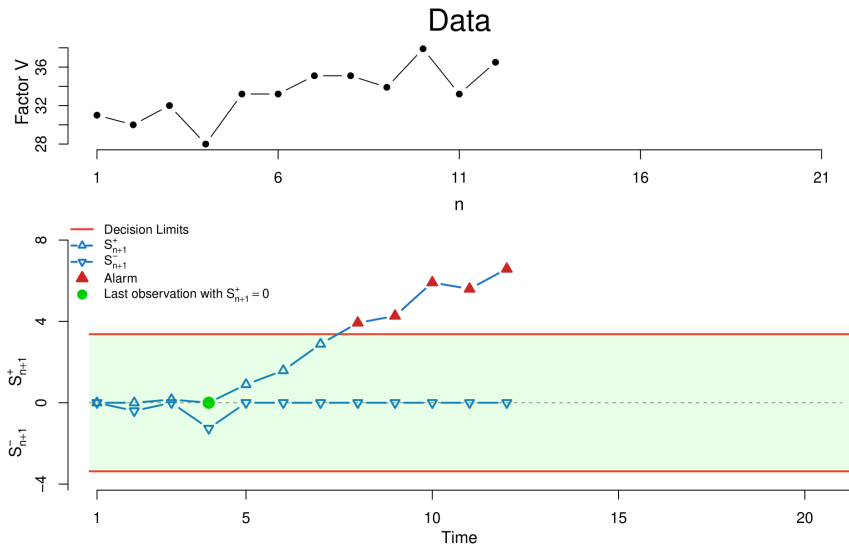
loc-PRC Real Data Application (Medical Lab)



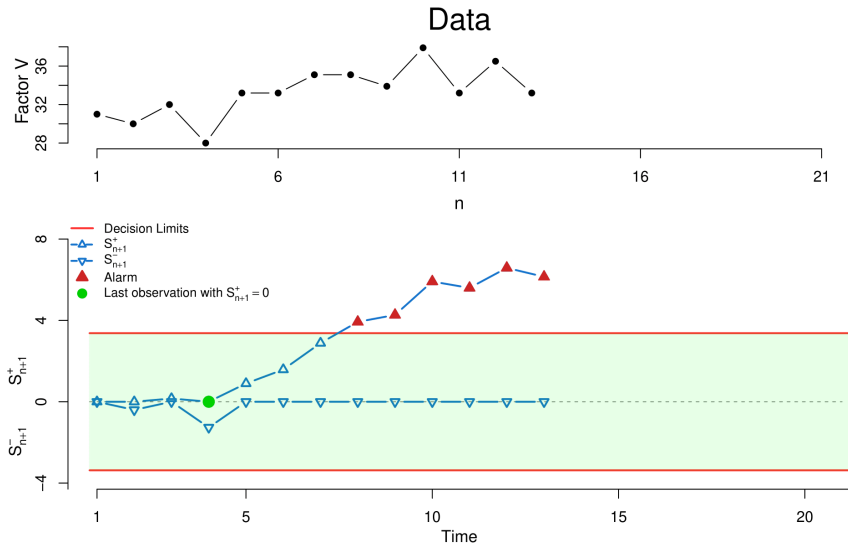
loc-PRC Real Data Application (Medical Lab)



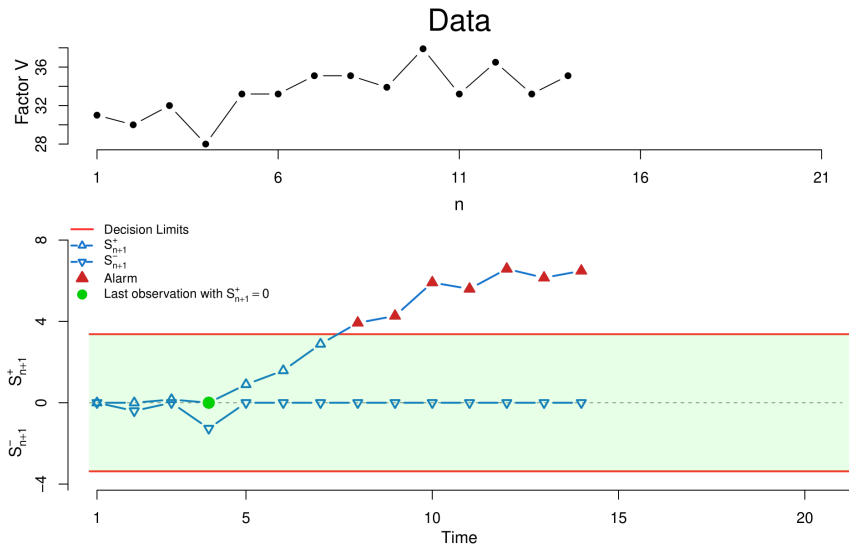
loc-PRC Real Data Application (Medical Lab)



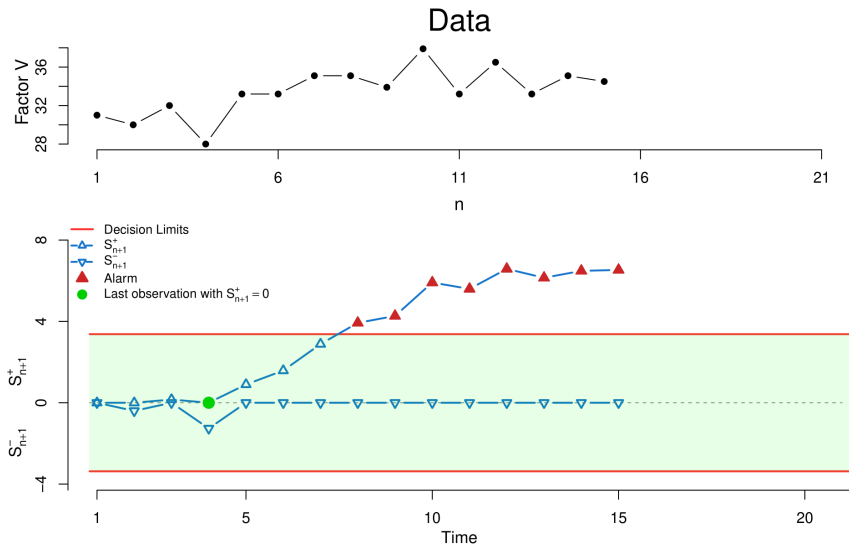
loc-PRC Real Data Application (Medical Lab)



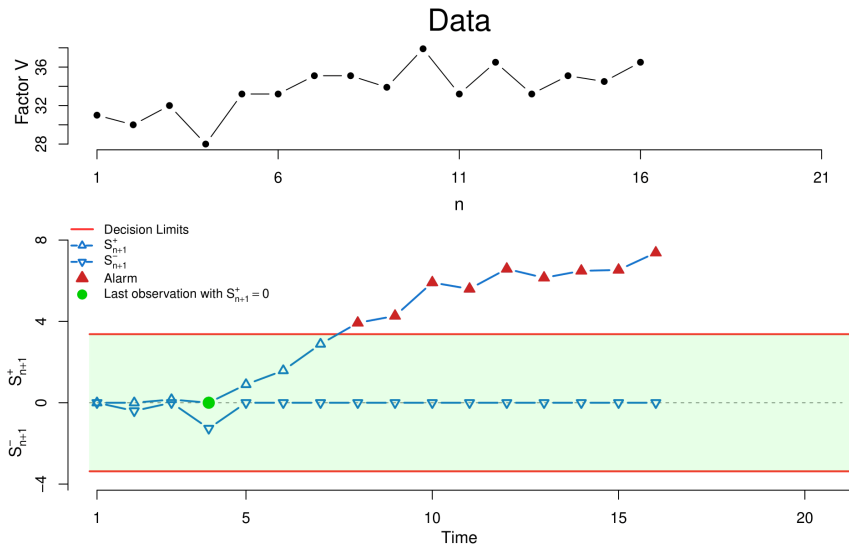
loc-PRC Real Data Application (Medical Lab)



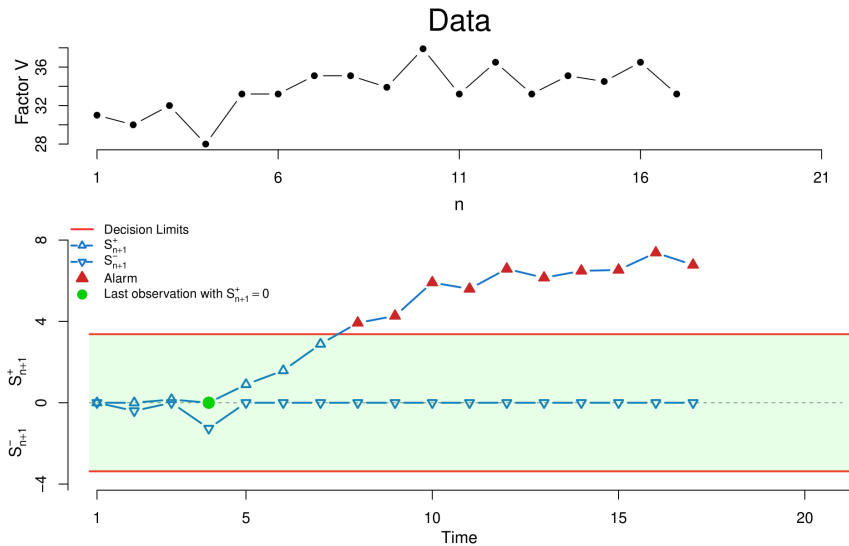
loc-PRC Real Data Application (Medical Lab)



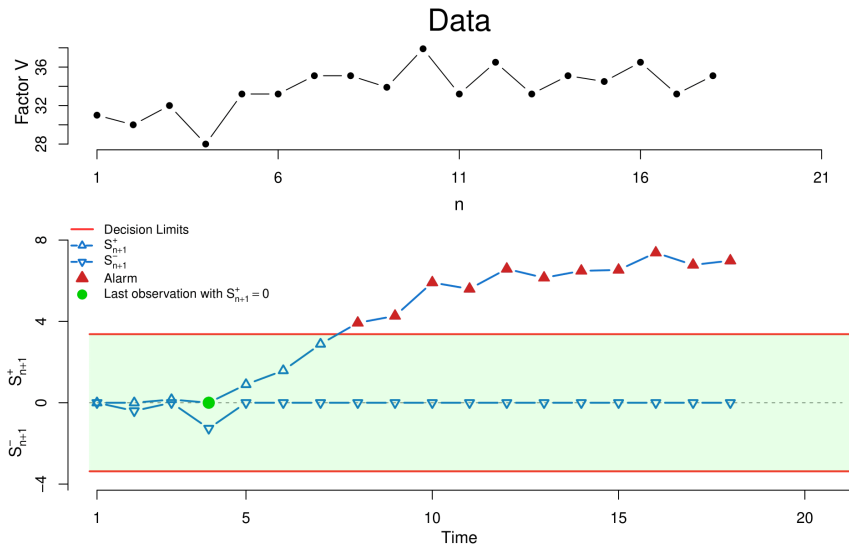
loc-PRC Real Data Application (Medical Lab)



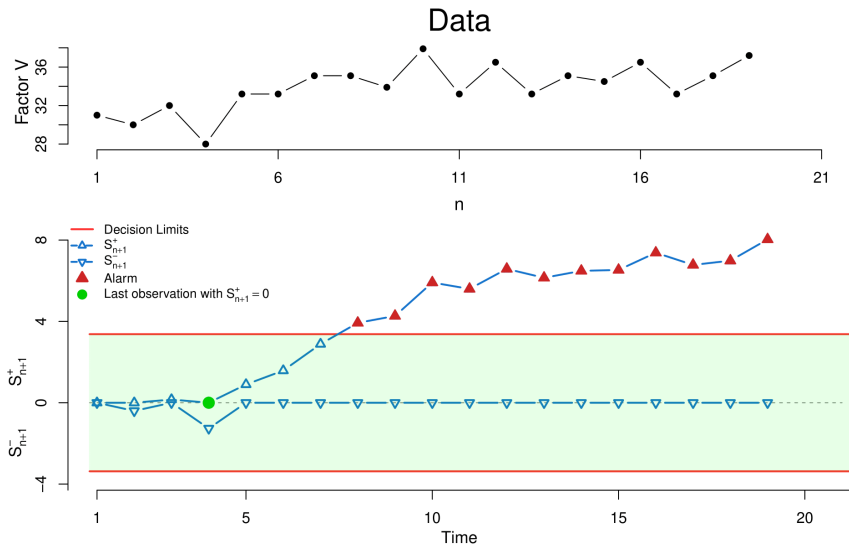
loc-PRC Real Data Application (Medical Lab)



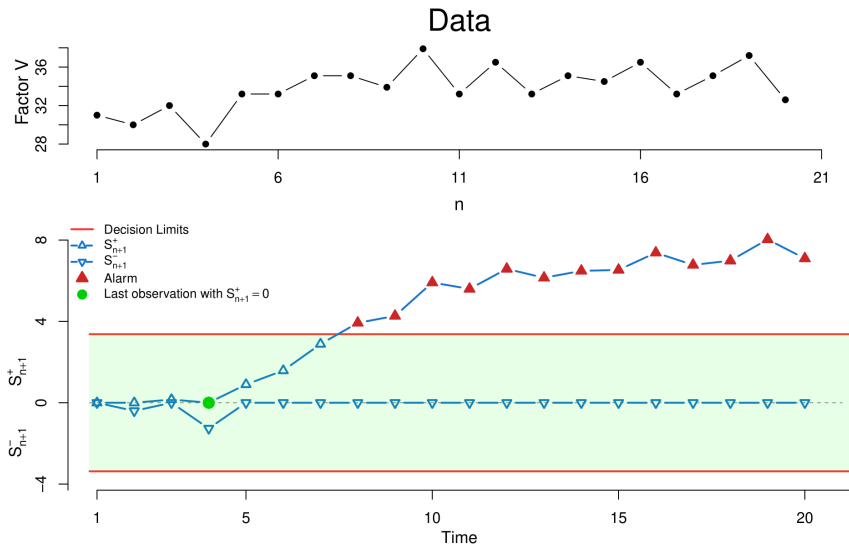
loc-PRC Real Data Application (Medical Lab)



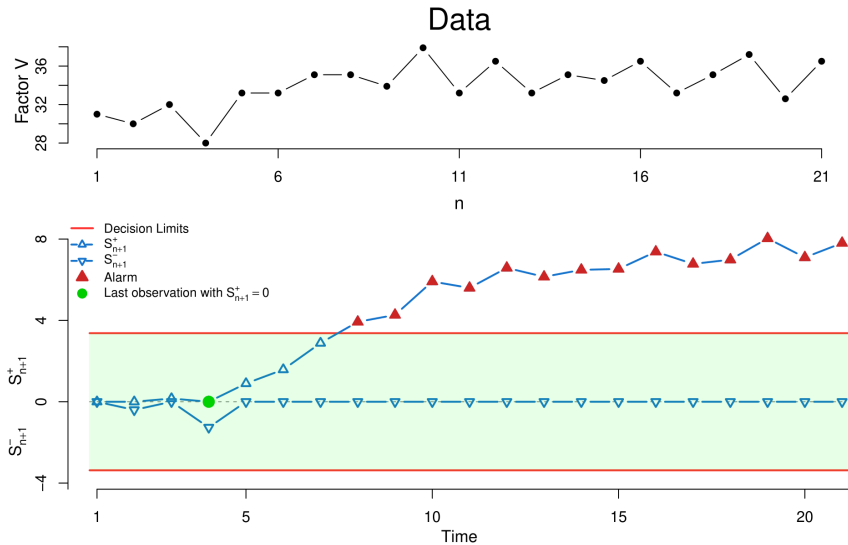
loc-PRC Real Data Application (Medical Lab)



loc-PRC Real Data Application (Medical Lab)



loc-PRC Real Data Application (Medical Lab)



Conclusions

In Bayesian SPC/M we introduced the **Predictive Control Chart (PCC)** and the **Predictive Residual (Ratio) Cusum (PRC)** mechanisms which:

- they are provided in a general form allowing their use for any **discrete** or **continuous** data from the regular exponential family.
- Any **prior information** and/or **historical data** can be incorporated, boosting performance, but **objective** priors are also available.
- At any stage of the process, apart from **outlier & change point** detection, **posterior inference** for the unknown parameter(s) is also available.
- Both PCC and PRC **outperforms** frequentist and Bayesian alternatives and their are found to be **robust** to various misspecifications.

Acknowledgements

- I would like to thank **Laurence Pieroni** and **Magali Annette Reisch** for the kind invitation to present the Bayesian perspective at CNBH.

Acknowledgements

- I would like to thank **Laurence Pieroni** and **Magali Annette Reisch** for the kind invitation to present the Bayesian perspective at CNBH.
- Special thanks to **Werfen** for their support in developing Bayesian QC tools. Parts of the PRC research were partially funded by Instrumentation Laboratory (IL), Bedford, USA.

Acknowledgements

- I would like to thank **Laurence Pieroni** and **Magali Annette Reisch** for the kind invitation to present the Bayesian perspective at CNBH.
- Special thanks to **Werfen** for their support in developing Bayesian QC tools. Parts of the PRC research were partially funded by Instrumentation Laboratory (IL), Bedford, USA.
- Warm thanks to **Piet Meijer**, ECAT director, for our collaboration and his vision of introducing state of the art statistical tools in EQA.

Acknowledgements

- I would like to thank **Laurence Pieroni** and **Magali Annette Reisch** for the kind invitation to present the Bayesian perspective at CNBH.
- Special thanks to **Werfen** for their support in developing Bayesian QC tools. Parts of the PRC research were partially funded by Instrumentation Laboratory (IL), Bedford, USA.
- Warm thanks to **Piet Meijer**, ECAT director, for our collaboration and his vision of introducing state of the art statistical tools in EQA.
- I am grateful to my former students **Konstantinos Bourazas** and **Dimitris Kiagias** for the collaboration on the Bayesian SPC/M.

Acknowledgements

- I would like to thank **Laurence Pieroni** and **Magali Annette Reisch** for the kind invitation to present the Bayesian perspective at CNBH.
- Special thanks to **Werfen** for their support in developing Bayesian QC tools. Parts of the PRC research were partially funded by Instrumentation Laboratory (IL), Bedford, USA.
- Warm thanks to **Piet Meijer**, ECAT director, for our collaboration and his vision of introducing state of the art statistical tools in EQA.
- I am grateful to my former students **Konstantinos Bourazas** and **Dimitris Kiagias** for the collaboration on the Bayesian SPC/M.
- This work would not have been existed without **Frederic Sobas** from Hospices Civils de Lyon, France, who provided an actual problem setting that inspired this work, he supported us with different ways, but most importantly he provided invaluable feedback from using **PCC & PRC** at the daily Internal Quality Control routine in the medical labs of Hospices Civils de Lyon.

- Bourazas, K., Kiagias, D., and Tsiamyrtzis, P. (2022). **“Predictive Control Charts (PCC): A Bayesian approach in online monitoring of short runs”**. Journal of Quality Technology, Vol. 54 (4):367–391. <https://doi.org/10.1080/00224065.2021.1916413>
- Bourazas K., Sobas F. and Tsiamyrtzis, P. (2023). **“Predictive ratio CUSUM (PRC): A Bayesian approach in online change point detection of short runs”**. Journal of Quality Technology, Vol. 55(4):391-403. <https://doi.org/10.1080/00224065.2022.2161434>
PRC won the **2024 Brumbaugh Award** from **ASQ** which is presented for *“the paper making the largest single contribution to the development of industrial application of quality control”*.
- Bourazas K., Sobas F. and Tsiamyrtzis, P. (2023) **“Design and properties of the predictive ratio cusum (PRC) control charts”**. Journal of Quality Technology, Vol. 55(4):404-421. <https://doi.org/10.1080/00224065.2022.2161435>

Merci beaucoup!

Merci beaucoup!

Questions? Comments? Concerns?