Bayesian statistics: Inference and decision theory

Patric Müller und Francesco Antognini
Seminar über Statistik FS 2008

03.03.2008

Contents

1 Introduction and basic definitions

Notation: Let $X \in \chi$ be an observable random variable (the data) and χ the space of all observations. We denote with $\mathcal{P} := \{P_\theta : \theta \in \Theta\}$ the model class of distributions with unknown parameter θ.
We will assume $X \sim P \in \mathcal{P}$

Let $\mathcal{A}$ be the *action space.*

- $\mathcal{A} = \mathbb{R}$ estimating a parameter $\gamma := g(\theta) \in \mathbb{R}$
- $\mathcal{A} = \{0, 1\}$ testing
- $\mathcal{A} = [0, 1]$ randomized tests

Definition: A *decision* is a map $d : \chi \to \mathcal{A}$. $d(x)$ is the action when x is observed.

We need to define an evaluation criterion in order to compare two different decisions.

Definition: A *loss function* is a map

$$L : \Theta \times \mathcal{A} \to \mathbb{R}^+$$

with $L(\theta, a)$ being the loss when the parameter value is θ and one takes action a.

Remark: The actual determination of the loss function is often awkward in practice, in particular because the determination of the consequences of each action for each value of θ is usually impossible when $\mathcal{A}$ or Θ are large sets, for instance when they have an infinite number of elements.
The complexity of determining the subjctive loss function of the decision-maker often prompts the statistician to use *classical* or (*canonical*) losses, selected because of their simplicity and mathematical tractability.
It is still better to take a decision in a finite time using an approximate criterion rather that spending an infinite time to determine the proper loss function.

Suppose our model class $\mathcal{P} := \{P_\theta : \theta \in \Theta\}$ is dominated by ν (σ-finite). Let $X \sim P \in \mathcal{P}$. Then let $p_\theta = \frac{dP_\theta}{d\nu}$ denote the densities. We now think of p_θ as the density of X given the value of θ. We write it as

$$p_\theta(x) = p(x \mid \theta) \qquad x \in \chi.$$

A fundamental basis of Bayesian Decision Theory is that statistical inference should start with the rigorous determination of three factors:

- the distribution family for the observations, $p(x \mid \theta) = p_\theta(x)$;
- the prior distribution for the parameter, $\Pi(\theta)$;
- the loss associated with the decisions, $L(\theta, d)$.

The prior, the loss and even sometimes the sampling distribution being derived from partly sujective considerations.

Definition: The *risk* of the decision d is

$$R(\theta, d) := E_\theta \left[L\left(\theta, d(X)\right) \right] = \int_{\mathcal{X}} L\left(\theta, d(X)\right) p_\theta(x) d\nu(x)$$

Examples: some classical loss functions and their risks.

a) Estimation of $g(\theta) \in \mathbb{R}$; $\mathcal{A} = \mathbb{R}$

$$\begin{array}{rcl} L(\theta, a) & := & w(\theta) \left| g(\theta) - a \right|^r \quad , \quad r \geq 0 \\ R(\theta, d) & = & w(\theta) E_\theta \left[\left| g(\theta) - d(X) \right|^r \right] \end{array}$$

Special case: $r = 2$; $w(\theta) = 1 \forall \theta$ then $R(\theta, d) = E_\theta \left[\left| g(\theta) - d(X) \right|^2 \right]$ is the *Mean square error*. L is called *quadratic loss.*

b) $\Theta = \Theta_0 \cup \Theta_1 \quad , \quad \Theta_0 \cap \Theta_1 = \emptyset$
Testing Hypothesis: $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$
$\mathcal{A} = \{0, 1\}$

$$L(\theta, a) = \begin{cases} 1 & \text{if } \theta \in \Theta_0 \text{ and } a = 1 \\ c(> 0) & \text{if } \theta \in \Theta_1 \text{ and } a = 0 \\ 0 & \text{otherwise} \end{cases}$$

$d := \varphi$

$$R(\theta, \varphi) = \begin{cases} P_\theta(\varphi(X) = 1) & \text{if } \theta \in \Theta_0 \\ c P_\theta(\varphi(X) = 0) & \text{if } \theta \in \Theta_1 \end{cases}$$

Remark: The risk $R(\theta, d)$ is a function of the parameter θ. Therefore, the frequentist approach does not induce a total ordering on the set of procedures.

Definition: A decision d' is called *stirictly better* than d if

$$R(\theta, d') \leq R(\theta, d) \qquad \forall \theta$$

and $\exists\theta$ such that

$$R(\theta, d') < R(\theta, d).$$

When there exists a d' that is strictly better than d, then d is called *inadmissible.*

Exemple: $X = (X_1, \ldots, X_n)$ iid
$E_\theta[X_1] = \mu, \quad Var_\theta(X_1) = 1 \qquad \forall\theta$
$d(X) = \overline{X}_{n-1} = \frac{1}{n-1}\sum_{i=1}^{n-1} X_i$
$d'(X) = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
$L(\theta, a) = |\mu - a|^2$
$\Rightarrow R(\theta, d) = \frac{1}{n-1}, \quad R(\theta, d') = \frac{1}{n}$
d is inadmissible.

2 Bayes Method

Suppose the parameter space Θ is a measure space. We can then equip it with a probability measure Π. We call Π the *a priori* distribution.

Definition: The *Bayes risk* (with respect to the probability measure Π) is

$$r(\Pi, d) := \int_\Theta R(\theta, d) d\Pi(\theta).$$

A decision is called *Bayes* (with respect to Π) if

$$r(\Pi, d) = \inf_{d'} r(\Pi, d').$$

Remark: If Π has density $w := \frac{d\Pi}{d\mu}$ with respect to some dominating measure μ, we may write

$$r(\Pi, d) = \int_\Theta R(\theta, d) w(\theta) d\mu(\theta) =: r_w(d).$$

The Bayes risk may be thought of as taking a weighted average of the risks. We call $w(\theta)$ the *prior density.*

Moreover

$$p(x) := \int_\Theta p(x \mid \theta) w(\theta) d\mu(\theta).$$

Definition: The *a posteriori density* of θ is

$$w(\theta \mid x) = p(x \mid \theta)\frac{w(\theta)}{p(x)} \qquad \theta \in \Theta, x \in \chi.$$

Theorem: Given the data $X = x$, consider θ as a random variable with density $w(\theta \mid x)$. Let

$$l(x,a) := E\left[L(\theta,a) \mid X = x\right] := \int_\Theta L(\theta,a)w(\theta \mid a)d\mu(\theta)$$

and $\qquad d(X) := \arg\min_a l(x,a)$.
$\Rightarrow d$ is Bayes decision d_{Bayes}.

Proof:

$$\begin{aligned}
r_w(d') &= \int_\Theta R(\theta,d')w(\theta)d\mu(\theta) \\
&= \int_\Theta \left[\int_\chi L(\theta,d'(x))p(x \mid \theta)d\nu(x)\right] w(\theta)d\mu(\theta) \\
&\overset{\text{Fubini}}{=} \int_\chi \int_\Theta L(\theta,d'(x))w(\theta \mid x)d\mu(\theta)p(x)d\nu(x) \\
&= \int_\chi l(x,d')p(x)d\nu(x) \\
&\geq \int_\chi l(x,d)p(x)d\nu(x) \\
&= r_w(d)
\end{aligned}$$

$\square$

Notice 1: Form a strictly Bayesian point of view, only the posterior expected loss $l(x,a)$ is important.
The previous theorem give us a constructive tool for the determination of the Bayes estimators.
Notice 2: For strictly convex losses, the Bayes estimator is unique.

Remark: Let $X \in \mathbb{R}$, $E[X] = \mu$ and $Var(X) < \infty$.
If $a \in \mathbb{R}$ then

$$E\left[(X-a)^2\right] = Var(X) + (a-\mu)^2$$

since:

$$\begin{aligned}
E\left[(X-a)^2\right] &= E\left[(X-\mu-(a-\mu))^2\right] \\
&= E\left[(X-\mu)^2\right] - 2\underbrace{E\left[(X-\mu)(X-a)\right]}_{=0} + E\left[(a-\mu)^2\right] \\
&= Var(X) + (a-\mu)^2
\end{aligned}$$

$\Rightarrow \arg\min_a E\left[(X-a)^2\right] = E\left[X\right]$

Example 1: $\Theta \subseteq \mathbb{R}$, $\mathcal{A} = \mathbb{R}$

$$L(\theta, a) := (\theta - a)^2$$

Then:

$$d_{\text{Bayes}}(x) = E\left[\theta \mid X = x\right] = \int_\Theta \theta w(\theta \mid x) d\mu(\theta)$$

Example 2: An alternative solution to the quadratic loss in dimesion one is to use the absolute error loss,

$$L(\theta, a) = |\theta - a| \qquad \Theta = \mathbb{R}, \mathcal{A} = \mathbb{R}$$

$\Rightarrow d_{\text{Bayes}}(x) =$ median of $w(\theta \mid x)$

Example 3: The $0-1$ loss.
$\Theta = \{\theta_0, \theta_1\}$. $\quad H_0 : \theta = \theta_0 \quad$ vs $\quad H_1 : \theta = \theta_1 \qquad \theta_0 \neq \theta_1$
$\mathcal{A} = [0,1] \qquad d := \varphi(X)$ =probability of rejecting H_0

$$R(\theta, \varphi) = \begin{cases} E_\theta\left[\varphi(X)\right] & \text{if } \theta = \theta_0 \\ 1 - E_\theta\left[\varphi(X)\right] & \text{if } \theta = \theta_1 \end{cases}$$

Let $\nu >> P_{\theta_0}, P_{\theta_1}$ and $p_0 := \frac{dP_{\theta_0}}{d\nu}, p_1 := \frac{dP_{\theta_1}}{d\nu}$.
Let $c \geq 0$ and $q \in [0,1]$. Then

$$\varphi_{\text{NP}} = \begin{cases} 1 & \text{if } p_1 > cp_0 \\ c & \text{if } p_1 = cp_0 \\ 0 & \text{if } p_1 < cp_0 \end{cases}$$

is called *Neyman-Pearson Test*.
With the same assumption as in the Bayes method we have now

$$\begin{aligned} w_0 &= \Pi(\theta = \theta_0) \qquad 0 < w_0 < 1 \\ w_1 &= \Pi(\theta = \theta_1) = 1 - w_0 \end{aligned}$$

$\Rightarrow r_w(\varphi) = w_0 R(\theta_0, \varphi) + w_1 R(\theta_1, \varphi)$.
<u>Lemma:</u>

$$\varphi_{\text{Bayes}} = \begin{cases} 1 & \text{if } p_1 > \frac{w_0}{w_1} p_0 \\ q & \text{if } p_1 = \frac{w_0}{w_1} p_0 \\ 0 & \text{if } p_1 < \frac{w_0}{w_1} p_0 \end{cases}$$

where $q \in [0,1]$ arbitrary.
Proof:

$$
\begin{aligned}
r_w(\varphi) &= w_0 E_{\theta_0}[\varphi(X)] + w_1 (1 - E_{\theta_1}[\varphi(X)]) \\
&= w_0 \int \varphi p_0 + w_1 \left(1 - \int \varphi p_1\right) \\
&= \int (w_0 p_0(x) - w_1 p_1(x)) \varphi(x) d\nu(x) + w_1
\end{aligned}
$$

This implies

$$
\begin{aligned}
w_0 p_0(x) - w_1 p_1(x) < 0 &\Rightarrow \varphi_{\text{Bayes}} = 1 \\
w_0 p_0(x) - w_1 p_1(x) > 0 &\Rightarrow \varphi_{\text{Bayes}} = 0 \\
w_0 p_0(x) - w_1 p_1(x) = 0 &\Rightarrow \varphi_{\text{Bayes}} = q
\end{aligned}
$$

with q arbitrary value in $[0,1]$.

$\square$

Example 4: $\Theta \subset \mathbb{R}, \qquad \mathcal{A} = \Theta$

$$
L(\theta, a) = 1_{\{|\theta - a| > c\}} \qquad \text{for } c > 0 \text{ given constant}
$$

$$
\begin{aligned}
\Rightarrow l(x,a) &= \Pi(|\theta - a| > c \,|\, X = x) \\
&= \int_{|\theta - a| > c} w(\theta \,|\, x) d\mu(\theta) \\
d_{\text{Bayes}}(x) &= \arg\min_a l(x,a) \\
&= \arg\max_a (1 - l(x,a))
\end{aligned}
$$

$$
\frac{1 - l(x,a)}{2c} = \frac{1}{2c} \int_{|\theta - a| \le c} w(\theta \,|\, x) d\mu(\theta)
$$

If μ is Lebesgue measure

$$
\lim_{c \searrow 0} \frac{1 - l(x,a)}{2c} = w(a \,|\, x) \qquad \text{is called } \textit{maximum a posteriori} \text{ estimator.}
$$

<u>Conclusion:</u>

$$
d_{\text{Bayes}}(x) \approx d_{\text{MAP}}(x) := \arg\max_a w(a \,|\, x)
$$

Note: $w(a \,|\, x) = p(x \,|\, a) \frac{w(a)}{p(x)}$
$\Rightarrow \arg\max_a w(a \,|\, x) = \arg\max_\theta [p(x \,|\, \theta) w(\theta)]$
MAP = MLE if $w(\theta)$ are constant!

3 Two optimalities: minimaxity and admissibility

After having seen the concept of admissibility, to look at 'how good' or 'how bad' are our decisions, we want to introduce the concept of minimaxity:
Definition: Let d be a decision; d is called *minimax* if the minimax risk

$$\overline{R} := \inf_{\delta \in D^*} \sup_{\theta} R(\theta, \delta)$$

is equal to $\sup_\theta R(\theta, d)$.
In other words minimax means: *the best decision in the worst possible case.*
Remark: D^* is the set of randomized estimators.
Example: The first oil-drilling platforms were designed according to a minimax principle: they were so constructed to resist to the worst gale together with the worst storm ever observed, and this at the minimal temperature ever measured.
Naturally the platform is quite sure, but very expensive to build. Nowadays companies tend to use other strategies to reduce costs.
Up to now we have seen only methods to *compare* decisions, now we want to *improve* decisions.
Definition: let S be a Statistic, we say S is *Sufficient* if: $P_\theta[x \mid S]$ does not depend on θ
Thm: Rao-Blackwell

- Let $\mathcal{A} \subseteq \mathbb{R}^p$ convex
- $a \mapsto L(\theta, a)$ a convex function $\forall \theta$
- S be sufficient
- and $d : \chi \to \mathcal{A}$ be a decision
- suppose $d'(s) := E\left[d(X)\,|S = s\right]$ (assumed to exist)

then:

$$R(\theta, d') \leq R(\theta, d) \forall \theta \in \Theta \tag{1}$$

Proof: we have to use the Jensen inequality: for a convex g $E\left[g(X)\right] \geq g(E\left[X\right])$ and the iterated expectation lemma.

$$\begin{aligned} R(\theta, d) &= E\left[L\left(\theta, d(X)\right)\right] \\ &= E\left[E\left[L\left(\theta, d(X)\right)|S = s\right]\right] \\ &\geq E\left[L\left(\theta, d'(S)\right)\right] \end{aligned}$$

Remarks:

I As loss function we often use $L(\theta, a) := |\theta - a|^2$ which is convex, and $[0; 1]$ or $\mathbb{R}$ (a convex set) as action space. So it is not so difficult to satisfy the conditions of the theorem.

II The new decision is better than the oldest one, and only depend on s.

III $\underline{R} \overset{Def}{=} \sup_\Pi r(\Pi) \overset{Def}{=} \sup_\Pi \inf_{\delta \in D} r(\Pi, \delta) \leq \overline{R} = \inf_{\delta \in D^*} \sup_\theta R(\theta, \delta)$

We can now look at some relations between Bayes, minimax and admissible:

Proposition 1: If δ_0 is Bayes (wrt Π_0 a priori distribution of Θ) and if $R(\theta, \delta_0) \leq r(\Pi_0) \forall \theta \in supp(\Pi_0)$, then δ_0 is minimax.

Proposition 2: If $\exists (\Pi_n)_{n \in \mathbb{N}}$ sequence of proper prior distribution s.t.:

$$R(\theta, \delta_0) \leq \lim_{n \to \infty} r(\Pi_n) < \infty \quad \forall \theta$$

then δ_0 is minimax.

Proposition 3: If there exists a unique minimax estimator δ_0, then it is admissible.

Proof: let δ be another decision; because of minimaxity

$$\begin{aligned} & \sup_\theta R(\theta, \delta) > \sup_\theta R(\theta, \delta_0) \\ \Rightarrow \quad & \exists \theta : R(\theta, \delta) > R(\theta, \delta_0) \\ \Rightarrow \quad & \delta_0 \textit{ is admissible} \end{aligned}$$

□

Note that uniqueness is essential.

Example: $\Theta = \{0; 1\}$

$R(\theta, \delta_1) := 1$

$R(\theta, \delta_2) := \theta$

$\sup_\theta R(\theta, \delta_1) = \sup_\theta R(\theta, \delta_2)$

even if δ_i are minimax, δ_2 is strictly better than δ_1 (also δ_1 is inadmissible).

Proposition 4: If δ_0 is an admissible decision and $R(\theta, \delta_0) = R(\delta_0)$, then δ_0 is the unique minimax decision.

Proof:

$$
\begin{aligned}
\forall \theta_0 \in \Theta \quad &: \quad \sup_\theta R(\theta, \delta_0) \overset{(*)}{=} R(\theta_0, \delta_0) \\
\delta_0 \quad \text{adm.} \Rightarrow \forall \tilde{\delta} \exists \tilde{\theta} \quad &: \quad R(\tilde{\theta}, \tilde{\delta}) > R(\theta, \tilde{\delta}) \\
\Rightarrow \sup_\theta R(\theta, \tilde{\delta}) \geq R(\tilde{\theta}, \tilde{\delta}) &> R(\theta, \tilde{\delta}) \overset{(*)}{=} \sup_\theta R(\theta, \delta_0)
\end{aligned}
$$

also δ_0 is minimax! Uniqueness follows dierctly from $' >' \delta$.

□

Proposition 5: Let Π be the a priori strictly positive distribution on Θ, $R(\theta, \delta)$ continuous in θ $\forall \delta$ and $r_\Pi(\delta) < \infty$. Then Bayes implies admissible.
Proof: Assume δ_Π Bayes but inadmissible, then there exist a strictly better decision δ_1.

$$
\begin{aligned}
\Rightarrow \quad & \forall \theta R(\theta, \delta_1) \leq R(\theta, \delta_\Pi) \\
\text{and} \quad & \exists \tilde{\theta} \ : \ R(\tilde{\theta}, \delta_1) < R(\tilde{\theta}, \delta_\Pi)
\end{aligned} \tag{2}
$$

because of continuity of the risk function we can find an open set σ in which (2) holds. So we have:

$$
\begin{aligned}
r_\Pi(\delta_1) &= \int_\Theta R(\theta, \delta_1) w(\theta) d\theta \\
&= \underbrace{\int_{\Theta \setminus \sigma} R(\theta, \delta_1) w(\theta) d\theta}_{\leq \int_{\Theta \setminus \sigma} R(\theta, \delta_\Pi) w(\theta) d\theta} + \underbrace{\int_\sigma R(\theta, \delta_1) w(\theta) d\theta}_{< \int_\sigma R(\theta, \delta_\Pi) w(\theta) d\theta} \\
&< \int_\Theta R(\theta, \delta_\Pi) w(\theta) d\theta \\
&= r_\Pi(\delta_\Pi)
\end{aligned}
$$

This result is in contradiction with the fact that δ_Π is Bayes.

□

Proposition 6: If the Bayes estimator associated with a prior Π is unique, it is admissible.

Proposition 7: A Bayes estimator δ_Π (with Π prior), convex loss function and finite Bayes risk is admissible.

4 Alternatives

Up to now we have only analysed cases, where loss and distributions were given. But the reality is not always so.
Often we don't have a completly determined loss function; then we have some possibilities to solve this problem, like:

- choosing arbitrary the subjective best unknown parameter
- using a random variable ($F-$distributed) where we have unknown parameters. So the Bayes risk is: $\int_\Theta \int_\Omega L(\theta, \delta, \omega) dF(\omega) d\pi(\theta \mid x)$
- we can reduce the number of possible loss functions by taking only few of them in account, and then looking for estimators performing well for all these losses

5 Appendix

A numerical exemple to show the utility of the Bayes method.
Definition: $Z > 0$ has a *Gamma-*(k, λ) *distribution* if it has density

$$\begin{aligned} f_Z(z) &= \frac{z^{k-1} e^{-\lambda z} \lambda^k}{\Gamma(k)} \\ \Gamma(k) &= \int_0^\infty z^{k-1} e^{-z} dz \end{aligned}$$

Remark: $E[Z] = \frac{k}{\lambda}$
Definition: $Z \in \{0, 1, \ldots\}$ has a *Negative Binomial-*(k, p) *distribution* if

$$P(Z = z) = \frac{\Gamma(z + k)}{\Gamma(k) z!} p^k (1 - p)^z \qquad z = 0, 1, \ldots$$

Remarks:

- k=1: geometric distribution with parameter p.
- $E[Z] = \frac{k(1-p)}{p} \quad Var(Z) = \frac{k(1-p)}{p^2}$.

Exemple: Suppose $p(x \mid \theta) = e^{-\theta} \frac{\theta^x}{x!} \qquad x \in \{0, 1, \ldots\}$ (Poisson(θ)-distribution) and that $\Theta \sim$ Gamma(k, λ).

<u>Result 0:</u> $p(x) :=$ Negative Binomial(k, p) where $p = \frac{\lambda}{1+\lambda}$.
Proof:

$$\begin{aligned}
p(x) &= \int_\Theta p(x \mid \theta) w(\theta) d\mu(\theta) \\
&= \int_0^\infty e^{-\theta} \frac{\theta^x}{x!} \frac{\theta^{k-1} e^{-\lambda\theta} \lambda^k}{\Gamma(k)} d\theta \\
&= \frac{\lambda^k}{\Gamma(k) x!} \int_0^\infty e^{-\theta(\lambda+1)} \theta^{x+k-1} d\theta
\end{aligned}$$

We compute the last integral by partial integration

$$\begin{aligned}
\int_0^\infty \underset{\uparrow}{e^{-\theta(\lambda+1)}} \underset{\downarrow}{\theta^{x+k-1}} \, d\theta &= \underbrace{\frac{e^{-\theta(\lambda+1)}}{-(\lambda+1)} \theta^{x+k-1}}_{=0} + \int_0^\infty \frac{e^{-\theta(\lambda+1)}}{\lambda+1} (x+k-1) \theta^{x+k-2} d\theta \\
&= \frac{x+k-1}{\lambda+1} \int_0^\infty e^{-\theta(\lambda+1)} \theta^{x+k-2} d\theta \\
&= \ldots = \frac{(x+k-1) \cdot (x+k-2) \cdot \ldots 3 \cdot 2}{(\lambda+1)^{x+k-1}} \int_0^\infty e^{-\theta(\lambda+1)} d\theta \\
&= \frac{\Gamma(x+k)}{(\lambda+1)^{x+k}}
\end{aligned}$$

Finally we can write that

$$p(x) = \frac{\Gamma(x+k)}{\Gamma(k) x!} \left(\frac{\lambda}{(1+\lambda)} \right)^k \left(\frac{1}{1+\lambda} \right)^x$$

$\square$

<u>Result 1:</u> $w(\theta \mid x)$ is Gamma$(x+k, 1+\lambda)$.
Proof:

$$\begin{aligned}
w(\theta \mid x) &= p(x \mid \theta) \frac{w(\theta)}{p(x)} \\
&= e^{-\theta} \frac{\theta^x}{x!} \theta^{k-1} e^{-\lambda\theta} \frac{\lambda^k}{\Gamma(k)} \frac{1}{p(x)} \\
&\overset{\text{Result 0}}{=} e^{-\theta} \frac{\theta^x}{x!} \theta^{k-1} e^{-\lambda\theta} \frac{\lambda^k}{\Gamma(k)} \frac{1}{\frac{\Gamma(x+k)}{\Gamma(k)x!} \left(\frac{\lambda}{(1+\lambda)} \right)^k \left(\frac{1}{1+\lambda} \right)^x} \\
&= \frac{\theta^{x+k-1} e^{-(\lambda+1)\theta} (\lambda+1)^{x+k}}{\Gamma(x+k)}
\end{aligned}$$

$\square$

Remarks:

- $\widehat{\theta}_1 := E[\theta \mid X] = \frac{x+k}{1+\lambda}$ is the Bayes estimator in the case of quadratic loss.
- MLE: $\widehat{\theta}_2 := x$.
- MAP: $\arg\max_\theta e^{-\theta(\lambda+1)\theta^{x+k-1}} = \widehat{\theta}_3 = \frac{x+k-1}{1+\lambda}$.

6 Bibliographie

- Robert, Ch. *"The Bayesian Choise"*,2nd ed.
- *"Vorlesungmitschriften von Grundlagen der mathematischen Statistik"*, von Prof. Sara van de Geer