**République Algérienne Démocratique et Populaire**
**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique**

**UNIVERSITÉ MOULOUD MAMMERI, TIZI-OUZOU FACULTÉ DES**

**SCIENCES DÉPARTEMENT DE MATHÉMATIQUES**

Mémoire Présenté En Vue De L'obtention Du

**DIPLÔME DE MAGISTER (ÉCOLE DOCTORALE)**

**Spécialité**:MATHÉMATIQUES

**Option**: Statistique

Par

**Fatiha TALBI**

Titre

# Bayesian Robustness For Nonhomogeneous Poisson Processes

| | | | |
|---|---|---|---|
| M.FALLAG Hocine | Professeur | UMMTO | Président |
| M.RUGGERI Fabrizio | Professeur | CNR IMATI MILAN | Rapporteur |
| M.HAMADOUCHE Djamel | Professeur | UMMTO | Examinateur |
| M.BOUDIBA Mohand Arezki | M.C.A | UMMTO | Examinateur |
| Melle.ATIL Lynda | M.C.B | UMMTO | Examinatrice |

29 Octobre 2012

# Contents

# Abstract

Robustness with respect to modeling assumptions is a relevant aspect in each statistical analysis. In particular, Bayesians are interested in robustness with respect to changes in prior distributions/sampling models/loss functions.

A Poisson process is one of the most significant random processes in probability theory. It is widely used to model random points in time and space such as the arrival times of customers at a service center. A Non-Homogeneous process is a process with rate parameter $\lambda(t)$ such that the rate parameter of the process is a function of time. The NHPP is probably the best known generalization of the Poisson process .

In This work, we focused on replacing a single prior distribution by a class of priors of the parameters of a given Poisson processes, and developing methods of computing the range of the ensuing answers as the prior varied over the class. This approach, called "global robustness".

**Key words**: Bayesian Robustness, Nonhomogeneous Poisson processes, Prior Robustness, Model Robustness, Loss Robustness, Global Sensitivity, Local Sensitivity.

# Résumé

La robustesse est un aspect approprié dans chaque analyse statistique. En particulier, les Bayesiens sont intéressés par la robustesse en ce qui concerne les changements des distributions a priori/des modèles/des fonctions de perte.
Le processus de Poisson est un processus aléatoire le plus significatif dans la théorie de probabilité. il est largement utilisé pour modeler les points aléatoires dans le temps et dans l'espace tel que les temps d'arrivée des clients à un service. Un processus de Poisson non homogène est un processus avec le paramètre $\lambda(t)$ et ce paramètre est une fonction de temps. Un processus de Poisson non homogène est probablement la meilleure généralisation connue d'un processus de Poisson .
Dans ce travail on s'intéresse à remplacer une loi a priori par une classe de lois a priori des paramètres d'un processus de Poisson donné, et de développer le calcul du rang quand la loi a priori change dans cette classe, cette approche, s'appelle " la robustesse globale ".

**Mots clés** : Robustesse Bayesienne, Processus de Poisson Nonhomogène, Robustesse des lois a priori, Robustesse des Modèles, Robustesse des Fonctions de perte, sensitivité globale, sensitivité locale.

## LIST OF TABLES

# Introduction

Statistics is the science that relates data to specific questions of interest. This includes devising methods to gather data relevant to the question, methods to summarize and display the data to shed light on the question, and methods that enable us to draw answers to the question that are supported by the data. Data almost always contain uncertainty. This uncertainty may arise from selection of the items to be measured, or it may arise from variability of the measurement process. Drawing general conclusions from data is the basis for increasing knowledge about the world, and is the basis for all rational scientific inquiry. Statistical inference gives us methods and tools for doing this despite the uncertainty in the data. The methods used for analysis depend on the way the data were gathered. It is vitally important that there is a probability model explaining how the uncertainty gets into the data.

There are two main philosophical approaches to statistics. The first is often referred to as the frequentist approach. Sometimes it is called the classical approach. The frequentist approach to statistics considers the parameter to be a fixed but unknown constant. The only kind of probability allowed is long-run relative frequency. These probabilities are only for observations and sample statistics, given the unknown parameters. Statistical procedures are judged by how they perform in an infinite number of hypothetical repetitions of the experiment.

The alternative approach is the Bayesian approach. The Reverend Thomas Bayes first discovered the theorem that now bears his name. It was written up in a paper An Essay Towards Solving a Problem in the Doctrine of Chances. This paper was found after his death by his friend Richard Price, who had it published posthumously in the Philosophical Transactions of the Royal Society in 1763. Bayes showed how inverse probability could be used to calculate probability of antecedent events from the occurrence of the consequent event. His methods were adopted by Laplace and other scientists in the 19th century, but had largely fallen from favor by the early 20th century. By the middle of the 20th century, interest in Bayesian methods had been renewed by De Finetti, Jeffreys, Savage, and Lindley, among others. They developed a complete method of statistical inference based on Bayes' theorem.

The Bayesian approach to statistics allows the parameter to be considered a random variable. Probabilities can be calculated for parameters as well as observations and sample statistics. Probabilities calculated for parameters are interpreted as "degree of belief' and must be subjective.

Informally, to make inference about $\theta$ is to learn about the unknown $\theta$ from data $X$, i.e., based on the data, explore which values of $\theta$ are probable, what might be plausible numbers as estimates of different components of $\theta$ and the extent of uncertainty associated with such estimates. In addition to having a model $f(x/\theta)$ and a likelihood function, the Bayesian needs a distribution for $\theta$. The distribution is called a prior distribution or simply a prior because it quantifies her uncertainty about $\theta$ prior to seeing data. The prior may represent a blending of her subjective belief and knowledge, in which case it would be a subjective prior. Alternatively, it could be a conventional prior supposed to represent small or no information. Such a prior is called an objective prior.

Given all the above ingredients, the Bayesian calculates the conditional probability

density of $\theta$ given $X = x$ by Bayes formula

$$\pi(\theta/x) = \frac{f(x/\theta)\pi(\theta)}{\int_\Theta f(x/\theta)\pi(\theta)d(\theta)}$$

The conditional density $\pi(\theta/x)$ of $\theta$ given $X = x$ is called the posterior density, a quantification of our uncertainty about $\theta$ in the light of data. The transition from $\pi(\theta)$ to $\pi(\theta/x)$ is what we have learnt from the data.

A Bayesian can simply report her posterior distribution, or she could report summary descriptive measures associated with her posterior distribution. For example, for a real valued parameter $\theta$, she could report the posterior mean and the posterior variance or the posterior standard deviation. Finally, she could use the posterior distribution to answer more structured problems like estimation and testing.

Robust Bayesian Analysis is concerned with the sensitivity of the results of a Bayesian analysis to the inputs for the analysis. Intuitively, robustness means lack of sensitivity of the decision or inference to assumptions in the analysis that may involve a certain degree of uncertainty. In an inference problem, the assumptions usually involve choice of the model and prior, whereas in a decision problem there is the additional assumption involving the choice of the loss or utility function. An analysis to measure the sensitivity is called sensitivity analysis. Clearly, robustness with respect to all three of these components is desirable. That is to say that reasonable variations from the choice used in the analysis for the model, prior, and loss function do not lead to unreasonable variations in the conclusions arrived at.

In this work, the model proposed is the nonhomogeneous Poisson processes(NHPPs), which have relevant applications in many fields, e.g. in reliability, capturing different behaviors of the problem they are modeling and providing different results (estimations and forecasts). A thorough study of robustness associated to NHPPs is still missing and the proposed work aims to pursue such goal.

The objective of this work is then to study the robustness of the classes of priors of the parameters of a given NHPP where the priors changes over this proposed class and computation of some measures of robustness. Computations with NHPP are difficult. We start with homogeneous Poisson processes and then we will do it later with NHPP.

In particular, this work contains three chapters. Chapter 1 provides an introduction to Bayesian inference and the study of literature in Bayesian robustness, a quick review of counting process, homogeneous Poisson process and non homogeneous poisson process (definitions and proprieties) are given in Chapter 2. Identification of possible classes of priors for the parameters of a given NHPP and assessment of robustness through computation of some measures are also introduced here. Prior distributions used include the positive uniform, the Jeffreys' prior, and the gamma prior, a detailed discussion on Bayesian robustness of the classes proposed of homogeneous and nonhomogeneous poisson processes is provided in Chapter 3. Finally, a general conclusion and some recommendations suggested finish this work.

# Chapter 1

# Bayesian Robustness

Robustness with respect to modeling assumptions is a relevant aspect in each statistical analysis. In particular, Bayesians are interested in robustness with respect to changes in prior distributions/sampling models/loss functions. Most of the research, as described in the reference book by Rios Insua and Ruggeri (Springer, 2000), has concentrated on robustness with respect to the choice of the prior distribution, since it is both a characterizing and a weak aspect of the Bayesian approach. Furthermore, computation of robustness measures is easier under classes of priors, unlike for classes of sampling distributions.

In the early '90s, there was an explosion of publications focusing on studying sensitivity to the prior distribution, in part because non-Bayesians often view this sensitivity to be the major drawback of Bayesian analysis. This work focused on replacing a single prior distribution by a class of priors, and developing methods of computing the range of the ensuing answers as the prior varied over the class. This approach, called "global robustness", was soon supplemented by "local robustness" techniques, which focused on studying local sensitivity (in the sense of derivatives) to prior distributions. Interest naturally expanded into study of robustness with respect to the likelihood and loss function as well, with the aim of having a general approach to sensitivity towards all the ingredients of the Bayesian paradigm (model/prior/loss).

An overview of the robust Bayesian approach is presented. Common types of robustness analyses are illustrated, including global and local sensitivity analysis and loss and likelihood robustness.

In this Chapter, we shall concentrate on some issues related to robustness of Bayesian inference. The notations used will be introduced here as needed. Suppose X has (model) density f($x/\theta$) and $\theta$ has (prior) probability density $\pi(\theta)$. Then the joint density of (X,$\theta$), for x in the sample space and $\theta \in \Theta$ ($\Theta$ is the parameter space ), is

$$h(x, \theta) = f(x/\theta)\pi(\theta)$$

The marginal density of X corresponding with this joint density is

$$m_\pi(x) = \int_\Theta f(x/\theta)\pi(\theta)d(\theta).$$

If X is continuous, and

$$m_\pi(x) = \Sigma_\Theta f(x/\theta)\pi(\theta).$$

If X is discrete.

The posterior density of $\theta$ given x is given by

$$\pi(\theta/x) = \frac{h(\theta, x)}{m_\pi(x)} = \frac{f(x/\theta)\pi(\theta)}{m_\pi(x)}$$

The posterior mean and posterior variance with respect to prior $\pi$ will be denoted by $E^\pi(\theta/x)$ and $V^\pi(\theta/x)$, respectively. Similarly, the posterior probability of a set $A \subset \Theta$ given x will be denoted by $P^\pi(A/x)$.

## 1.1 Different approaches to Bayesian robustness

There are three main approaches to Bayesian robustness. We illustrate them considering robustness with respect to changes in the prior, but similar issues are raised when considering likelihoods and losses:

### 1.1.1 Informal approach:

In this approach, we consider a few priors and compare the quantity of interest (e.g., the posterior mean) under them. The approach has been (and is) very popular because of its simplicity and can help, but it can easily "miss" priors that are compatible with the actually elicited prior knowledge and yet which would yield very different quantity of interest. The rationale with this informal approach is that since we shall be dealing with messy computational problems, why not analyze sensitivity by trying only some alternative pairs of losses and priors? This is a healthy practice and a good way to start a sensitivity analysis.

**Example 1.1.**

Suppose a single observation, x, is observed from a normal distribution with unknown mean $\theta$ and variance 1, to be denoted $N(\theta,\ 1)$. The classic (old) textbook Bayesian analysis would suggest selection of a conjugate, normal prior for $\theta$, chosen to match some features determined by the elicitation process. Suppose, for instance, that -.954, 0, and .954 are the elicited quartiles of the distribution of $\theta$. The unique, normal prior with these quartiles is $\Pi \sim N(0,\ 2)$. It is evident that the choice of a normal prior is mathematically convenient but otherwise arbitrary. In the informal approach to robustness, one would try a few different priors, compatible with the elicited quartiles, and compare the answers. Three such priors are the normal $N(0,\ 2)$, double exponential $DE(0,\ \log \frac{2}{0.945})$ and Cauchy distribution $C(0,\ 0.954)$. Resulting posterior means, for different data x, are presented in Table 1.1.

| $x$ | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| $N$ | 0.333 | 0.667 | 1.000 | 1.333 | 1.667 | 2.000 | 2.333 | 2.667 | 6.667 |
| $DE$ | 0.292 | 0.606 | 0.960 | 1.362 | 1.808 | 2.285 | 2.776 | 3.274 | 9.274 |
| $C$ | 0.259 | 0.540 | 0.806 | 1.259 | 1.729 | 2.267 | 2.844 | 3.427 | 9.796 |

**TABLE 1.1**. Posterior means for different priors.

For moderate or large values of x in Table 1, the posterior mean is reasonably consistent for the Cauchy and Double Exponential distributions, but changes dramatically for the Normal distribution. For small values of x, the posterior means are all reasonably consistent. The conclusion of this informal analysis would be that robustness is likely obtained for smaller values of x, but not for larger values.

### 1.1.2 Global robustness

Global sensitivity is the most popular approach in Bayesian robustness. We consider the class of all priors compatible with the elicited prior information, and compute the range of the quantities of interest as the prior varies over the class. This range is typically found by determining the extremal priors in the class that yield the maximum and minimum quantities of interest. Such computations can become cumbersome in multidimensional problems.

### 1.1.3 Local robustness

Local sensitivity is interested in the rate of change in inferences, with respect to changes in the prior, and uses differential techniques to evaluate the rate. Local sensitivity measures are typically easier to compute in complicated situations than are global measures, but their interpretation and calibration is not always clear.

## 1.2 Priors robustness

We first show various ways of modeling imprecision in the prior through classes of priors. This is only the first part in a sensitivity analysis; it must be followed by the identification of a tool capable of measuring the effects of the uncertainty on the quantities of interest. Different approaches can be taken, depending, e.g., on the existence or not of a baseline prior and a neighborhood class around it. It is possible to consider a neighborhood class around a prior $\Pi_0$ and check what happens when allowing for either infinitesimal or more relevant departures from $\Pi_0$.

The following example illustrates why sensitivity to the choice of prior can be an important consideration.

**Example.1.2.**

Suppose we observe X, which follows Poisson($\theta$) distribution. Further, it is felt a prior that $\theta$ has a continuous distribution with median 2 and upper quartile 4. i.e.

$$P(\theta \leq 2) = 0.5 = P(\theta \geq 2) \; and \; P(\theta \geq 4) = 0.25$$

If these are the only prior inputs available, the following three are candidates for such a prior

(i) $\pi_1$ $\theta \sim$ exponential (a) with a=log(2)/2;
(ii) $\pi_2$ $log(\theta) \sim N(log(2), (log(2)/z_{.25})^2)$; and
(iii) $\pi_3$ $log(\theta) \sim Cauchy(log(2), log(2))$.

Then (i) under $\pi_1$, $\pi_1(\theta/x) \sim Gamma(a+1, x+1)$, so that the posterior mean is $E_{\pi_1}(\theta/x) = (a+1)/(x+1)$;
(ii) under $\pi_2$, if we let $\gamma = log(\theta)$, and $\tau = log(2)/z_{.25} = log(2)/0.675$, we obtain

$$E_{\pi_2}(\theta/x) = E_{\pi_2}(exp(\gamma)/x)$$

$$= \frac{\int_{-\infty}^{+\infty} exp(-e^{\gamma})exp(\gamma(x+1))exp(-(\gamma - log(2))^2/(2\tau^2))d\gamma}{\int_{-\infty}^{+\infty} exp(-e^{\gamma})exp(\gamma x)exp(-(\gamma - log(2))^2/(2\tau^2))d\gamma}$$

And (iii) under $\pi_3$, again if let $\gamma = log(\theta)$, we get

$$E_{\pi_3}(\theta/x) = E_{\pi_3}(exp(\gamma)/x)$$

$$= \frac{\int_{-\infty}^{+\infty} exp(-e^{\gamma})exp(\gamma(x+1))[1 + (\frac{\gamma - log(2)}{log(2)})^2]^{-1}d\gamma}{\int_{-\infty}^{+\infty} exp(-e^{\gamma})exp(\gamma x)[1 + (\frac{\gamma - log(2)}{log(2)})^2]^{-1}d\gamma}.$$

To see if the choice of prior matters, simply examine the posterior means under the three different priors in Table.1.2.

| $\pi/x$ | 0 | 1 | 2 | 3 | 4 | 5 | 10 | 15 | 20 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\pi_1$ | .749 | 1.485 | 2.228 | 2.971 | 3.713 | 4.456 | 8.169 | 11.882 | 15.595 | 37.874 |
| $\pi_2$ | .950 | 1.480 | 2.106 | 2.806 | 3.559 | 4.353 | 8.660 | 13.241 | 17.945 | 47.017 |
| $\pi_3$ | .761 | 1.562 | 2.094 | 2.633 | 3.250 | 3.980 | 8.867 | 14.067 | 19.178 | 49.402 |

**TABLE.1.2.** Posterior means under $\pi_1$,$\pi_2$ and $\pi_3$.

For small x ($x \leq 10$), there is robustness: the choice of prior does not seem to matter too much. For large values of x, the choice does matter. It is now clear that there are situations where it does matter what prior one chooses from a class of priors, each of which is considered reasonable given the available prior information.

## 1.3 Classes of priors

There is a vast literature on how to choose a class $\Gamma$ of priors to model prior uncertainty appropriately. The choice of the class of priors should be driven by the following goals:

. To ensure that as many "reasonable" priors as possible are included, is needed to ensure robustness;

. To try to eliminate "unreasonable" priors, to ensure that one does not erroneously conclude lack of robustness;

. To ensure that $\Gamma$ does not require prior information which is difficult to elicit, and

. To be able to compute measures of robustness without much difficulty.

### 1.3.1 Conjugate priors

The class consisting of conjugate priors is one of the easiest classes of priors to work with. If $X \sim N(\theta, \sigma^2)$ with known $\sigma^2$, the conjugate priors for $\theta$ are the normal priors $N(\mu, \tau^2)$. So one could consider

$$\Gamma_C = \{N(\mu, \tau^2), \mu_1 \leq \mu \leq \mu_2, \tau_1^2 \leq \tau^2 \leq \tau_2^2\}$$

For some specified values of $\mu_1$, $\mu_2$, $\tau_1^2$ and $\tau_2^2$. The advantage with the conjugate class is that posterior quantities can be calculated in closed form. In the above case, if $\theta \sim N(\mu, \tau^2)$, then $\theta/X \sim N(\mu^*(x), \delta^2)$ where:

$$\mu^*(x) = (\frac{\tau^2}{(\tau^2 + \sigma^2)})x + (\frac{\sigma^2}{(\tau^2 + \sigma^2)})\mu$$

And

$$\delta^2 = \frac{\tau^2\sigma^2}{(\tau^2 + \sigma^2)}$$

Minimizing and maximizing posterior quantities then becomes an easy task. The crucial drawback of the conjugate class is that it is usually "too small" to provide robustness and omit many reasonable priors.

### 1.3.2 Neighborhood Class

When a baseline prior $\Pi_0$ (often conjugate) is elicited by usual methods, a natural class of priors for studying sensitivity to the prior is the $\varepsilon$-contamination class. The popularity of the $\varepsilon$-contamination class arises, in part, from the ease of its specification and, in part,

from the fact that it is typically easily handled mathematically. The $\varepsilon$-contamination class is defined as:

$$\Gamma_\varepsilon = \{\Pi : \Pi = (1-\varepsilon)\Pi_0 + \varepsilon q, q \in Q\}$$

Where $Q$ is called the class of contaminations and $\varepsilon$ reflecting the uncertainty in $\Pi_0$. The class has been considered by Huber (1973) in classical robustness. In Bayesian robustness, the $\varepsilon$-contamination class is made of priors witch resemble the baseline prior $\pi_0$.
The major problem is the determination of $\varepsilon$ and $Q$. In particular the uncertainty in $\pi_0$. In fact, it is worth mentioning that, for any measurable subset A, it follows that

$$(1-\varepsilon)\Pi_0(A) + \underset{q\in Q}{inf} q(A) \leq \Pi(A) \leq (1-\varepsilon)\Pi_0(A) + \underset{q\in Q}{sup} q(A)$$

for any $\Pi \in \Gamma_\varepsilon$.

Researchers focused their interest in finding suitable classes of contaminations $Q$. The choice of $Q$ has a great effect. A variety of choices of Q have been considered in the literature. The most obvious choice is to let $Q$ be the class of all distributions, Unfortunately, this class is often too large to yield useful robustness bounds.

One important class of refinements of the contaminating class involves the addition of shape constraints, Such constraints can often be readily elicited, and can very significantly reduce the range of posterior quantities of interest.

Some choices for $Q$ are all unimodal distributions with mode $\theta_0$ ($\Gamma_U$) and all unimodal symmetric distributions with mode $\theta_0$ ($\Gamma_{US}$). The $\varepsilon$-contamination class with appropriate choice of $Q$ can provide good robustness.

Moreno and Cano (1992) addressed the problem of robustness in a multidimensional parameter space, considering the contamination class with partially know marginal. The problem has been addressed by Lavineet al. (1991) as well.

As a related class, Gelfand and Dey (1991) proposed the geometric $\varepsilon$-contamination class defined as

$$\Gamma_g = \{\pi :\ \pi = C_g(\varepsilon)\pi_0^{1-\varepsilon} q^\varepsilon,\ q \in Q\}$$

Where q is the density in the class of contamination $Q$.

One drawback of $\varepsilon$-contamination classes is that they are not true neighbourhood classes in a topological sense. A variety of other classes have been considered that do have a formal interpretation in this sense, such as the class based on a concentration function and the class based on distribution bands.

The distribution band class, described in Basu and DasGupta (1990,1995) and Basu (1995), is defined as

$$\Gamma_{BDG} = \{F : \ F \ is \ a \ cdf \ and \ F_L(\theta) \leq F(\theta) \leq F_U(\theta) \ \forall\theta\}$$

Where $F_L$ and $F_U$ are given cumulative density functions (cdf), with $F_L(\theta) \leq F_U(\theta)$. This class is important from a mathematical point of view since it includes, as special cases, the well known Kolmogorov and Lèvy neighbourhoods. The elicitation of this classis relatively simple and others features such as symmetry and unimodality of the distributions, can be added as in Basu (1992), making computations even harder.

### 1.3.3 Density Ratio Class :

Assuming the existence of densities for all the priors in the class, the density ratio class is defined as

$$\Gamma_{DR} = \{\pi : L(\theta) \leq \alpha\pi(\theta) \leq U(\theta) \ for \ some \ \alpha > 0\}$$

Where L and U are given non-negative functions, and $\pi$ is not required to have mass one, or even to be proper. An alternative definition of this class is

$$\Gamma_{DR} = \{\pi : \frac{L(\theta)}{U(\grave{\theta})} \leq \frac{\pi(\theta)}{\pi(\grave{\theta})} \leq \frac{U(\theta)}{L(\grave{\theta})} \ for \ all \ \theta, \grave{\theta}\}$$

From which it is clear that $\Gamma_{DR}$ specifies ranges for the ratios of the prior density between any two points.

**Example.1.3.**

Suppose $L(\theta) = 1$ and $U(\theta) = K$. Then $\Gamma_{DR}$ consists of all prior densities for which the density ratio between any two points lies between $K^{-1}$ and $K$. This class is a reasonable representation of prior vagueness, from a robust Bayesian perspective.

### 1.3.4 Priors with specified generalized moments

The generalized moments class is defined through

$$\Gamma = \{\pi : \int_{\Theta} H_i(\theta)\pi(\theta)d\theta \leq \alpha_i, i = 1, ..., n\}$$

Where $H_i$ are given $\pi$-integrable functions and $\alpha_i$, $i = 1, ..., n$ are fixed real numbers.

This class, considered first in Betrò et al. (1994) and Goutis (1994), is a very rich and contains well-known classes like the one with given moments and the quantile class. The former class has been used e.g. by Hartigan (1969) and Goldstein (1980), who considered the priors with given first two moments. Although the elicitation of the first two moments, actually mean and variance, is often very reasonable, it is difficult to specify further moments and reasonable distributions.
The generalised moments class alows for other choices of the functions $H_i$,i.e. neither powers or set functions. Betrò at al. (1994) considered the class defined by bounds on the marginal probabilities of given $K_i$ taking

$$H_i(\theta) = \int_{K_i} f(x/\theta)dx$$

So that, by Fubini's theorem, it follows that

$$\int_{\Theta} H_i(\theta)\pi(\theta)d\theta = \int_{K_i} m_{\pi}(x)dx$$

Where $m_{\pi}(x) = \int_{\Theta} f(x/\theta)\pi(\theta)d\theta$ is the marginal density of under $\pi$

### 1.3.5 Quantile classes

Often, there will be no baseline prior and the class of priors will simply be those that satisfy the features that are elicited.

An example of a quantile class is that considered in Berger and O'Hagan (1988), O'Hagan and Berger (1988) and Moreno and Cano (1989). The elicitation proceeds by determining the prior probabilities of each of the intervals $I_i,\ i = 1, 6$ in Table1.3. The probabilities indicated therein are the elicited values, and the quantile class, $\Gamma_Q$, consists of all distributions that are compatible with these assessments. Note that the N(0, 2) distribution is one such prior.

| $I_i$ | $]-\infty,-2]$ | $]-2,-1]$ | $]-1,0]$ | $]0,1]$ | $]1,2]$ | $]2,\infty[$ |
|---|---|---|---|---|---|---|
| $p_i$ | .08 | .16 | .26 | .26 | .16 | .08 |

**TABLE.1.3.** Intervals and prior probabilities .

Computation of ranges of a posterior quantity of interest is rather simple for a quantile class, since upper and lower bounds are achieved for discrete distributions giving mass to one point in each interval$I_i$.

## 1.4 Measures of sensitivity

Measures of sensitivity are needed to examine the robustness of inference procedures (or decisions) when a class $\Gamma$ of priors are under consideration. In recent years two types of these measures have been studied: global measures of sensitivity such as the range of posterior quantities and local measures such as the derivatives.

### 1.4.1 Global measures of Robustness

The most commonly used measure in global robustness is the range, i.e. the difference between upper and lower bounds on the quantity of interest. Its value measures the

variation caused by the uncertainty in the prior/likelihood/loss function. It should always be recognized that the range depends strongly on the size of the classes used to represent the uncertain inputs. Thus a small range with an excessively small class is not really comforting, and a large range with an excessively large class is not necessarily damning. This range provides, in general, a number that, in principle, should be interpreted in the following way:

- If the measure is "small", then robustness is achieved and any prior in the class can be chosen without relevant effects on the quantity of interest;
- If the measure is "large", one ideally narrows the class further or obtains additional data, recomputing the robustness measure and stopping as in the previous item;
- If the measure is "large" and the class cannot be modified, then a prior can be chosen in the class but we should be wary of the relevant influence of our choice on the quantity of interest. A natural adhoc approach is to replace the class by a single input, obtained by some type of averaging over the class. For instance, if the class is a class of priors, one might choose a hyperprior on the class, and perform an ordinary Bayesian analysis.

Given a class $\Gamma$ of prior measures, global sensitivity analysis will usually pay attention to the range of variation of a posterior (or predictive) quantity of interest as the prior ranges over the class,i e,

$$Sup_{\pi\in\Gamma}T(h,\pi) - Inf_{\pi\in\Gamma}T(h,\pi)$$

Further, as explained in Berger (1990), typically there are three categories of Bayesian quantities of interest.
(i) Linear functionals of the prior: $T(h,\pi) = \int_{\Theta} h(\theta)\pi(d\theta)$, where $h$ is a given function.
If h is taken to be the likelihood function $f(x/\theta)$, we get an important linear functional, the marginal density of data, i.e., $m_\pi(x) = \int_{\Theta} f(x/\theta).\pi(d\theta)$.

(ii) Ratio of linear functionals of the prior for some given function h:

$$T(h,\pi) = E^{\pi}(h(\theta)/x) = \frac{1}{m_{\pi}(x)}\int_{\Theta} h(\theta)f(x/\theta)\pi(d\theta)$$

If we take $h(\theta) = \theta$, $T(h,\pi)$ is the posterior mean. For $h(\theta) = I_C(\theta)$, the indicator function of the set C, we get the posterior probability of C.

(iii) Ratio of nonlinear functionals for some given $h$:

$$T(h,\pi) = \frac{1}{m_{\pi}(x)}\int_{\Theta} h(\theta,\phi(\pi))f(x/\theta)\pi(d\theta)$$

For $h(\theta,\phi(\pi)) = (\theta - \mu(\pi))^2$, where $\mu(\pi)$ is the posterior mean, we get $T(h,\pi) =$ the posterior variance.

Note that extreme values of linear functionals of the prior as it varies in a class $\Gamma$ are easy to compute if the extreme points of $\Gamma$ can be identified.

**Example1.4.**

Suppose $X \sim N(\theta,\sigma^2)$, with $\sigma^2$ Known and the class of interest is

$$\Gamma = \{\textit{all symmetric unimodal distributions with mode } \theta_0\}.$$

Then $\phi$ denoting the standard normal density,

$$m_{\pi}(x) = \int_{-\infty}^{+\infty} \frac{1}{\sigma}\phi(\frac{x-\theta}{\sigma})\pi(\theta)d\theta.$$

Note that any unimodal symmetric (about $\theta_0$) density $\pi$ is a mixture of uniform densities symmetric about $\theta_0$. Thus the extreme points of $\Gamma_{SU}$ are $U(\theta_0 - r, \theta_0 + r)$ distributions. Therefore,

$$\inf_{\pi\in\Gamma_{SU}} m_{\pi}(x) = \inf_{r>0}\frac{1}{2r}\int_{\theta_0-r}^{\theta_0+r}\frac{1}{\sigma}\phi(\frac{x-\theta}{\sigma})d\theta$$

$$= \inf_{r>0}\frac{1}{2r}\{\phi(\frac{\theta_0+r-x}{\sigma}) - \phi(\frac{\theta_0-r-x}{\sigma})\}$$

$$\sup_{\pi\in\Gamma_{SU}} m_\pi(x) = \sup_{r>0}\frac{1}{2r}\int_{\theta_0-r}^{\theta_0+r}\frac{1}{\sigma}\phi(\frac{x-\theta}{\sigma})d\theta$$

$$= \sup_{r>0}\frac{1}{2r}\{\phi(\frac{\theta_0+r-x}{\sigma}) - \phi(\frac{\theta_0-r-x}{\sigma})\}$$

To study ratio-linear functionals the following result from Sivaganesan and Berger (1989) is useful.

**Theorem1.1(see [42])**.

Consider the class $\Gamma_{SU}$ of all symmetric unimodal prior distributions with mode $\theta_0$. Then it follows that

$$\sup_{\pi\in\Gamma_{SU}} E^\pi(h(\theta)/x) = \sup_{r>0}\frac{\frac{1}{2r}\int_{\theta_0-r}^{\theta_0+r} h(\theta)f(x/\theta)d\theta}{\frac{1}{2r}\int_{\theta_0-r}^{\theta_0+r} f(x/\theta)d\theta},$$

$$\inf_{\pi\in\Gamma_{SU}} E^\pi(h(\theta)/x) = \inf_{r>0}\frac{\frac{1}{2r}\int_{\theta_0-r}^{\theta_0+r} h(\theta)f(x/\theta)d\theta}{\frac{1}{2r}\int_{\theta_0-r}^{\theta_0+r} f(x/\theta)d\theta}.$$

Not that $E^\pi(h(\theta)/x) = \frac{\int h(\theta)f(x/\theta)\pi(\theta)d\theta}{\int f(x/\theta)\pi(\theta)d\theta}$, where $f(x/\theta)$ is the density of the data x.

**Example.1.5.**

Suppose $X/\theta \sim N(\theta, \sigma^2$ and robustness of the posterior mean with respect to $\Gamma_{SU}$ is of interest. Then, range of posterior mean over this class can be easily computed using the Theorem. We thus obtain,

$$\sup_{\pi\in\Gamma_{SU}} E^\pi(\theta/x) = \sup_{r>0}\frac{\frac{1}{2r}\int_{\theta_0-r}^{\theta_0+r}\frac{\theta}{\sigma}\phi(\frac{x-\theta}{\sigma})d\theta}{\frac{1}{2r}\int_{\theta_0-r}^{\theta_0+r}\frac{1}{\sigma}\phi(\frac{x-\theta}{\sigma})d\theta}$$

$$= x + \sup_{r>0}\frac{\phi(\frac{\theta_0-r-x}{\sigma}) - \phi(\frac{\theta_0+r-x}{\sigma})}{\phi(\frac{\theta_0+r-x}{\sigma}) - \phi(\frac{\theta_0-r-x}{\sigma})},$$

$$\inf_{\pi\in\Gamma_{SU}} E^{\pi}(\theta/x) = \inf_{r>0} \frac{\frac{1}{2r}\int_{\theta_0-r}^{\theta_0+r} \frac{\theta}{\sigma}\phi(\frac{x-\theta}{\sigma})d\theta}{\frac{1}{2r}\int_{\theta_0-r}^{\theta_0+r} \frac{1}{\sigma}\phi(\frac{x-\theta}{\sigma})d\theta}$$

$$= x + \inf_{r>0} \frac{\phi(\frac{\theta_0-r-x}{\sigma}) - \phi(\frac{\theta_0+r-x}{\sigma})}{\phi(\frac{\theta_0+r-x}{\sigma}) - \phi(\frac{\theta_0-r-x}{\sigma})}.$$

### 1.4.2 Relative sensitivity.

The range is capable of measuring changes in the scale of the quantity of interest. Reporting only the range might be misleading, since a range of 1 has a different meaning, in terms of robustness, when the quantity of interest is of order $10^3$ or 10. The calibration of these measures is still an unresolved problem since it is not clear when we can say that in a problem there is more robustness than in another just by comparing two numbers. It is not even possible to find an objective criterion which tells when robustness is achieved. While interpretation of the size of the range is usually done within the specific applied context, some efforts have been made to derive certain generic measures also. Ruggeri and Sivaganesan (2000) suggest a scaled version of the range for this purpose. For a target quantity of interest $h(\theta)$, called relative sensitivity and defined as:

$$R_{\Pi} = \frac{(\rho^{\Pi} - \rho^{0})^2}{V^{\Pi}}$$

Where $\rho^{\Pi}$ and $\rho^{0}$ equal $E(h(\theta)/x)$ under $\Pi$ and the baseline prior $\Pi_0$, respectively, and $V^{\Pi}$ is the posterior variance of h($\theta$) with respect to $\Pi$.

The motivation for considering $R_{\Pi}$ is that the posterior variance $V^{\Pi}$ is a measure of accuracy in estimation of h($\theta$) and hence if the squared distance of $\rho^{\Pi}$ from $\rho^{0}$ relative to this is not too large, robustness can be expected.

**Example.1.6.**

Let X have the N($\theta$,1) distribution, and under $\pi_0$, let $\theta$ be N(0, 2). Consider the class $\Gamma$ of all N(0, $\tau^2$) priors with $1 \leq \tau^2 \leq 10$. Consider sensitivity of posterior inferences about $h(\theta) = \theta$ when x $> 0$ is observed. Because the posterior distribution (under the prior N(0, $\tau^2$)) of $\theta$ given x is normal with mean $\frac{\tau^2 x}{(\tau^2+1)}$ and variance $\frac{\tau^2}{(\tau^2+1)}$ note that

$$\rho^{\Pi}(x) - \rho^{0}(x) = (\frac{\tau^2}{\tau^2+1} + \frac{2}{3})x$$

and

$$R_{\Pi}(x) = \frac{(\tau^2-2)^2 x^2}{9\tau^2(\tau^2+1)}.$$

It can then be easily checked that the range of $\rho^{\Pi}(x) - \rho^{0}(x)$ is $8x/33$ and sup $R_{\Pi}(x) = 6.4x^2/99$. Thus, robustness can be expected when the observation X lies in the range $0 \leq x \leq 4$, but certainly not when x = 10.

### 1.4.3 Local measures of robustness

The global approach can become quite unfeasible for very complicated models. If, for example, X $\sim P_\theta$, and $\theta$ is p-dimensional, $p > 1$, then the range of posterior mean of $\theta_i$ may well depend on prior inputs on $\theta_j$ for $j \neq i$ also. If such is the case, global measures of robustness will involve computing ranges of posterior quantities of general functions g($\theta$) over classes of joint prior distributions of $\theta$.

The alternative, which has attracted a lot of attention, is that of trying to study the effects of small perturbations to the prior. This is called local sensitivity. In this approach also, one may either study the sensitivity of the entire posterior distribution or that of some specified posterior quantity. In the local sensitivity approach to Bayesian robustness, techniques from differential calculus are used and robustness is measured by the supremum of (functional) derivatives over a class. Stemming from work by Diaconis and Freedman (1986), Ruggeri and Wasserman (1993) measured the local sensitivity of

a posterior expectation with respect to the prior by computing the norm of the Fréchet derivative of the posterior with respect to the prior over several different classes of measures.

As an illustration of the approach, let consider as in Gustafson and Wasserman (1995). A different set of notations as given below are needed in this section. Let $\pi$ be a prior probability measure and let $\pi^x$ denote its corresponding posterior probability measure given the data x, let P be the set of all probability measures on the probability space. A distance function d : $P \to P$ is needed to quantify changes in prior and posterior measures. Let $v_\varepsilon$ be a perturbation of $\pi$ in the direction of a measure $v$. Then the local sensitivity of P in the direction of $v$ can be defined ( Gustafson and Wasserman (1995)) by

$$s(\pi, v; x) = \lim_{\epsilon \downarrow 0} \frac{d(\pi^x, v_\epsilon^x)}{d(\pi, v_\epsilon)}$$

Two different types of perturbations $v_\epsilon$ have been considered. The linear perturbation is defined as $v_\epsilon = (1-\epsilon)\pi + \epsilon v$, and the geometric perturbation as $dv_\epsilon \propto (\frac{dv}{d\pi})^\epsilon d\pi$. The local sensitivity $s(\pi, v; x)$ is simply the rate at which the perturbed posterior $v_\epsilon^x$ tends to the 'initial' posterior $\pi^x$ relative to the change in the prior. As a measure of overall sensitivity of a class $\Gamma$ of priors one may take

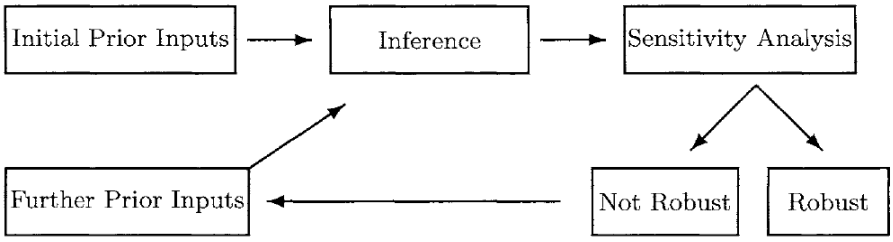
$$s(\pi, \Gamma; x) = \sup_{v \in \Gamma} s(\pi, v; x).$$

There are many possible choices for d, the distance measure.

## 1.5 Interactive Robust Bayesian Analysis

Following Berger (1994), an interactive scheme for robust Bayesian analysis can be suggested according to t he diagram Figure . The point to note is that, if lack of robustness is evident, then the class$\Gamma$ of priors obtained from initial prior inputs has to be shrunk using further prior elicitation.

**Fig.1.1.**. Interactive robust Bayesian scheme.




## 1.6 Likelihood robustness

Perturbations with respect to parametric models is more complex than changes in the prior and in the loss function, not only because of computational complexity but also because changes in the model should preserve the meaning of the parameter. The case of symmetric models is one of the most entertained since the median keeps the same meaning for all the models.

An informal approach, similar to that discussed in priors robustness, can be followed in analyzing likelihood robustness; one can simply try several models and see if the answer changes significantly.

It is difficult to develop more formal approaches to likelihood robustness, for several reasons, one of which is that parameters often change meanings in going from one model to another; hence changing models also often requires changing the prior distributions.

## 1.7 Classes of models

We review some key classes used in the literature:

### 1.7.1 Finite classes

A finite number of models can be considered. Sometimes a parameter has a clear meaning, regardless of the model under consideration, and more formal methods can be employed. An example is that of the median of a symmetric distribution. For instance, Shyamalkumar (2000) discusses the situation in which there are two competing models (normal $\mathrm{N}(\theta, 1)$ and Cauchy $\mathrm{C}(\theta, 0.675)$),

$$M = \{N(\theta, 1),\ C(\theta, 0.675)\}$$

With classes of priors on the common parameter $\theta$. The scale parameters of the two models are chosen to match the inter-quartile range. The priors on $\theta$ are specified to be in $\sigma$–contamination neighborhoods, where the contaminating classes are either $\Gamma_A$ or $\Gamma_{SU}$, containing, respectively, arbitrary and symmetric, unimodal (around zero) distributions. Shyamalkumar considered $\varepsilon = 0.1$ and computed upper and lower bounds on the posterior mean $E(\theta/x)$ of $\theta$, as shown in Table.1.4.

$$\Gamma^{A}_{0.1} = \{\pi : \pi = 0.9\pi_0 + 0.1q,\ q\ is\ arbtrary\}$$

$$\Gamma^{SU}_{0.1} = \{\pi : \pi = 0.9\pi_0 + 0.1q,\ q\ is\ a\ distribution\ symmetric\ and\ unimodal\ around\ zero\}$$

| Data | Likelihood | $\Gamma_A$ | | $\Gamma_{SU}$ | |
|---|---|---|---|---|---|
| | | $\inf \mathbb{E}(\theta\|x)$ | $\sup \mathbb{E}(\theta\|x)$ | $\inf \mathbb{E}(\theta\|x)$ | $\sup \mathbb{E}(\theta\|x)$ |
| $x = 2$ | Normal | 0.93 | 1.45 | 0.97 | 1.12 |
| | Cauchy | 0.86 | 1.38 | 0.86 | 1.02 |
| $x = 4$ | Normal | 1.85 | 4.48 | 1.96 | 3.34 |
| | Cauchy | 0.52 | 3.30 | 0.57 | 1.62 |
| $x = 6$ | Normal | 2.61 | 8.48 | 2.87 | 5.87 |
| | Cauchy | 0.20 | 5.54 | 0.33 | 2.88 |

**TABLE.1.4.**. Bounds on posterior mean for different models and classes of priors.

As noticed by Shyamalkumar, even though the width of the ranges of the posterior means are similar for the two likelihoods, the centers of the ranges are quite different when x is larger.

### 1.7.2 Parametric classes

There are many examples of parametric models. We just mention the very rich class of models due to Box and Tiao (1962), given by

$$\Lambda_{BT} = \{f(y/\theta,\sigma,\beta) = \frac{exp\{-\frac{1}{2}|\frac{y-\theta}{\sigma}|^{\frac{2}{1+\beta}}\}}{\sigma 2^{(1.5+0.5\beta)}\Gamma(1.5+0.5\beta)}; \qquad any \qquad \theta,\sigma>0,\beta\in(-0,1]\}$$

An application of this class is given in Shyamalkumar (2000).

### 1.7.3 Neighborhood classes

There are other current approaches to likelihood robustness, including use of Neighborhood classes where

**$\epsilon$-contaminations class**

This class of models have been considered in, e.g., Sivaganesan (1993), although these classes work well only when it is possible to consider the same parameter for all the entertained models,

$$\Gamma_\varepsilon = \{f(x;\theta) : f(x;\theta) = (1-\varepsilon)f_0(x;\theta) + \varepsilon p(x,\theta),\ p\in F\}$$

Where $f_0$ is the baseline density and is the class of contamination.

**Density Ratio classes**

Basu (1995) considered a density ratio class as well, i.e. all the densities $f$ for witch there exists $\alpha>0$ such that ,

$$\Gamma_{DR} = \{f : l(x-\theta_0) \le \alpha f(x/\theta_0) \le U(x-\theta_0), for\ all\ x\}$$

With $L$ and $U$ are given nonnegative functions, with $L(.) \le U(.)$.

**Class of bounded likelihoods.**

Stemming from results in Cuvevas and Sanz (1988) and similarly to the class of priors used by Ruggeri and Wasserman (1993), Ruggeri(1991, manuscript) considered the class of bounded likelihoods given by

$$\Gamma_{BL}(L,U) = \{l \in \ L_p; L(\theta) \leq l(\theta) \leq (\theta)\}$$

Where $L_p$ is the space of all integrable functions. $L$ and $U$ are are given measurable nonnegative functions, not necessarily in $L_p$. In particular he considered the special case where $l = (\frac{1}{k})\lambda$,$U = k\lambda$,$k \geq 1$ and $\lambda \in$ Ł$_p$ was given.
DeRobestis (1978) discussed one way of building nonparametric neighborhoods of the likelihood, but the first thorough illustration of a nonparametric class is due to Lavine (1991).

## 1.8 Loss robustness

When addressing Decision Theoretic issues, we immediately think of considering classes of loss functions and extend sensitivity to the prior procedures to this new setting in a straightforward manner.

Given the same decision problem, it is possible that different decision makers have different assessments for the consequences of their actions and hence may have different loss functions. In such a situation, it may be necessary to evaluate the sensitivity of Bayesian procedures to the choice of loss.

**Example.1.7.**

Suppose X is Poisson($\theta$) and $\theta$ has the prior distribution of exponential with mean 1. Suppose x = 0 is observed. Then the posterior distribution of $\theta$ is exponential with mean $\frac{1}{2}$. Therefore, the Bayes estimator of $\theta$ under squared error loss is $\frac{1}{2}$ which is the posterior mean, whereas the Bayes estimator under absolute error loss is 0.3465, the posterior median. These are clearly different, and this difference may have some significant impact

depending on the use to which the estimator is being put.

It is possible to provide a Bayesian approach to the study of loss robustness exactly as we have done for the prior distribution. In particular, if a class of loss functions is available, range of posterior expected losses can be computed and examined.

## 1.9 Classes of losses

Although some of the classes of losses are similar to those referring to priors, others arise from typical properties of the losses.

### 1.9.1 $\varepsilon$-contamination class

As shown earlier,$\varepsilon$-contamination classes are very popular in defining classes of priors. They can be used, as well, to define a neighbourhood around a loss function $L_0$ as follows

$$L^{\varepsilon}(L_0, W) = \{L : (c) = (1-\varepsilon)L_0(c) + \varepsilon M(c) \ : \ M \in W\},$$

Where $\varepsilon$ represents imprecision over $L_0$, $c$ is a consequence belonging to the state of consequences $C$ and $W$ is a class of loss functions, typically including $L_0$.

### 1.9.2 Partially Known Class

Stemming from the quantile class of priors, Martìn et al. (1998) considered a finite partition $C_1$,$C_2$,...,$C_n$ of the state of consequences $C$ and gave upper and lower bounds on the losses over each element of the partition, i.e. they considered

$$L_k = \{L : v_{i-1} \leq L(c) \leq v_i, \qquad \forall c \in C_i, \ i = 1, 2, ..., n\}$$

Where some of the sets $C_i$ could be empty. Although the specification of the class is relatively simple as for the classes of priors, the computations become here more cumbersome.

### 1.9.3 Parametric class

Various loss robustness measures have been proposed, most of which are discussed in Dey et al. (1998),The most famous parametric class of losses is LINEX, defined by

$$\Lambda = \{L_b : L_b(\theta, a) = exp\{b(a-\theta)\} - b(a-\theta) - 1, b \neq 0, b_0 < b < b_1\},$$

Where $b_0$ and $b_1$ are fixed. They (Dey et al. (1998)) consider a normal model for the observation and a conjugate prior for its mean, with specified variance, and compute the range of the posterior expected loss and the posterior regret.

### 1.9.4 Bands of convex loss functions

Martìn et al. (1998) considered losses which can be modelled as $L(\theta; a) = L(\theta - a)$. Therefore, it is possible to consider the loss as a function L(t) and study its derivative, when it exists, to detect local increase/decrease in the behavior of the loss. Let $L(t) = l(t)$ and consider two functions $\lambda(t)$ and $\mu(t)$, both positive for positive t and negative for negative t. Then a band of convex loss functions is obtained by considering

$$\mathcal{L} = \{L : \lambda(t) \leq l(t) \leq \mu(t) \ \forall t\}$$

Loss robustness studies consider a class $L$ of loss functions L($\theta$; a), defined for the action a in the action space A and $\theta$ in $\Theta$. As L varies in $L$, a measure of loss robustness is given by the range of the posterior expected loss of an action a, $sup_{L\in L}T(L; a) - inf_{L\in L}T(L; a)$.

Many studies in loss robustness concentrated on the search of the "best" loss, according to some criteria. Dey et al. (1998) proposed the following robustness measure:

$$D = sup_x|\frac{\partial}{\partial x}T(L, a^*)| - inf_x|\frac{\partial}{\partial x}T(L, a^*)|$$

Where $a^*$ is the Bayes action.

Some authors considered the influence of a loss function, defined as

$$\frac{\partial}{\partial x} inf_a T(L, a) = \frac{\partial}{\partial x} T(L, a^*)$$

Makov (1994) used the influence to order loss functions in a class. In the first way, he said that the loss $L_1$ was preferred to $L_2$ if

$$sup_x |inf_a T(L_1, a)| < sup_x |inf_a T(L_2, a)|.$$

In the second method, $L_1$ was preferred to $L_2$ if

$$E^x[T(L_1, a)] < E^x[T(L_2, a)]$$

Where the expectation was taken with respect to $m(x)$, the marginal distribution of x.

Joint robustness studies are essential, specially in decision theoretic contexts, as robustness studies with respect to only one of the elements varying, and the others fixed, may give a false sense of stability: a problem may be insensitive to changes only in the loss and changes only in the probability, but sensitive to simultaneous changes in both the loss and probability.

# Chapter 2

# A possible classes of priors for Non Homogeneous Poisson Process

The Poisson process is one of the most significant random processes in probability theory. It is widely used to model random points in time and space such as the times of radioactive emissions, the arrival times of customers at a service center and the positions of flaws in a piece of material. Several important probability distributions arise naturally from the Poisson process. The Poisson process is a collection of random variables where N(t) is the number of events that have occurred up to time t (starting from time 0). The number of events between time a and time b is given as N(b)-N(a) and has a Poisson distribution. A Non-Homogeneous process is a process with rate parameter $\lambda(t)$ such that the rate parameter of the process is a function of time e.g. the arrival rate of vehicles in a traffic light signal.

The Poisson process has found numerous applications in science, engineering, economics and other areas. The NHPP is probably the best known generalization of the Poisson process . It is characterized by a deterministic intensity function that describes how the rate of the process changes in time. For an ordinary Poisson processes, this function is a constant.

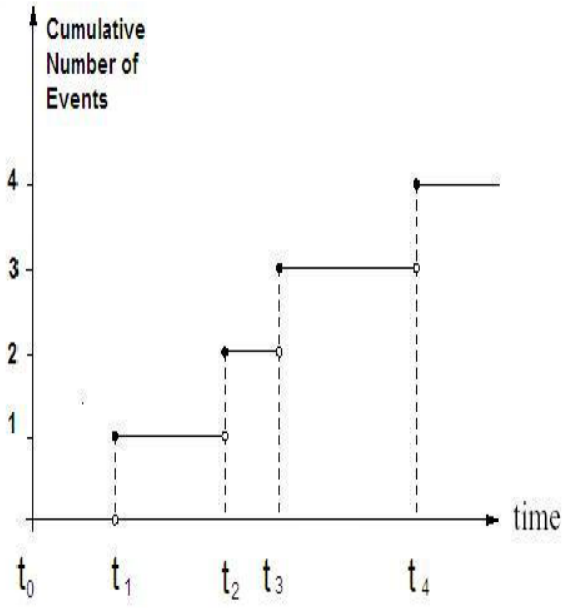
## 2.1 Counting Process

Models utilizing a counting process have played a key role in the analysis of systems composed of random occurring events. By way of motivation, suppose that we are interested in observing the occurrences of a repeatable event over a period of time. One of the simplest examples is the arrival of customers at a service station, such as a bank. Another example is the occurrences of earthquakes of a specified magnitude at a particular location over time. The example that is of interest to us here is the points in time when a system's software fails. In all such cases, the event of interest does not occur with any regularity and is therefore not exactly predictable. We are not sure about the exact times at which the events will occur and consequently about the exact number of events that will occur in any time interval.

A counting process is simply the count of the number of events that have occurred in any specified interval of time. Since N(t) is unknown for any value of t, we are facing with the problem of describing our uncertainty about an infinite collection of random variables, one for each t. Any indexed collection of random variables is called stochastic process, and when the interest is focused on counts, the process is called a counting process and is denoted by N(t), $t \geq 0$.

The sample path of a counting process is given by Figure 2.1. The horizontal line is designated to represent time; the vertical line is used to represent the total number of counts over time. It is a step function starting at zero, and taking jumps of size one at each ti, that is, the cumulative time of the ith failure.

The most commonly probabilistic models used for the counting process are homogeneous and nonhomogeneous Poisson processes.




### 2.1.1 Poisson Process

A counting process N(t) is said to be a Poisson process if

1. N(0) =0;

2. For any $a \leq b \leq c \leq d$ the random variables N(a, b] and N(c, d] are independent. This is called the independent increment property.

3. There is a function $\lambda$ such that

$$\lambda(t) \equiv \lim_{\Delta t \to 0} \frac{P\{N(t+\Delta t) - N(t) \geq 1\}}{\Delta t}$$

   The function $\lambda$ is called the intensity function of the Poisson process.

4. $$\lim_{\Delta t \to 0} \frac{P\{N(t+\Delta t) - N(t) \geq 2\}}{\Delta t} = 0$$

   This precludes the possibility of simultaneous failures.

## 2.2 Homogeneous Poisson Process (HPP)

The counting process N(t), t $\geq$ 0 is said to be a homogeneous Poisson process (HPP) if the intensity function $\lambda$(t) is a constant, that is, $\lambda(t) = \lambda$, $\lambda > 0$ and

1. N(0) =0;
2. The process has independent increments and stationary increments. A point process has stationary increments if for all k, $P[N(t+s) - N(t) = k]$ is independent of t.
3. . $P[N(t+dt) - N(t) = k] = 0 + od(t) \qquad k \geq 2$
   . $P[N(t+dt) - N(t) = k] = \lambda d(t) + od(t) \qquad k = 1$
   . $P[N(t+dt) - N(t) = k] = 1 - \lambda d(t) + od(t) \qquad k = 0$

It can be shown that the number of events in any interval of length $s = t_1 - t_2$ has a Poisson distribution with mean $\lambda_s$, that is

$$P[N(t_2) - N(t_1) = n] = \frac{e^{-\lambda_s}\lambda_s^n}{n!}, \;\; 0 \leq t_1 \leq t_2, \;\; n = 0, 1, 2, ...$$

### 2.2.1 Proprieties of Homogeneous Poisson Process

Homogeneous Poisson Process has the following properties

**Property 1.**

A process is an HPP with constant intensity function $\lambda$, if and only if the times between events are iid exponential random variables with mean $\frac{1}{\lambda}$.

**Property 2.**

The superposition of two Poisson processes with intensities $\lambda_1$ and $\lambda_2$ is a Poisson process with intensity

$$\lambda = \lambda_1 + \lambda_2.$$

**Property 3.**

If $0 < T_1 < T_2 < ... < T_n$ are the event occurred from an HPP, then the joint poisson distribution function of $T_1, T_2, ...T_n$ is

$$f(t_1, t_2, ..., t_n) = \lambda^n e^{-\lambda t_n}, \;\; 0 < t_1 < t_2 < ... < t_n.$$

**Property 4.**

The time to the nth event from a system modeled by an HPP has a gamma distribution with parameter $\alpha = n$ and $\beta = \frac{1}{\lambda}$.

**Property 5.**

For an HPP, conditional on N(t) = n, the number of events occurred at times $0 < T_1 < T_2 < ... < T_n$ are distributed as order statistics from U(0,t) distribution.

## 2.3 Nonhomogeneous Poisson Process

Nonhomogeneous Poisson process (NHPP) is a Poisson process whose intensity function is not a constant. A counting process N(t), t≥0 is a nonhomogeneous Poisson process if

1. N(0) =0;
2. The process has independent increments and stationary increments;
3. . $P[N(t+dt) - N(t) = k] = 0 + o(dt) \qquad k \geq 2;$
   . $P[N(t+dt) - N(t) = k] = \lambda(t)dt + o(dt) \qquad k = 1;$
   . $P[N(t+dt) - N(t) = k] = 1 - \lambda(t)dt + o(dt) \qquad k = 0.$

A homogeneous Poisson process may be viewed as a special case when $\lambda(t) = \lambda$, a constant rate.

The mean value function (m.v.f.) of the NHPP is defined as the nondecreasing, nonnegative, function

$$M(s,t) = E\{N(s,t)\},\ 0 \leq s < t\ with\ M(t) = E\{N(t)\},\ t \geq 0.$$

Where

$$M(y,s) = \int_y^s \lambda(t)dt.$$

Nonhomogeneous Poisson processes (NHPP) are widely used as models for failures of a repairable system. The occurrences in time will be the failure times of a repairable system. Though, sometimes we shall use the term failures instead of events.

The number of events occurred in any interval (y,s) has a Poisson distribution with mean $\int_y^s \lambda(t)dt$, that is

$$P[N(s) - N(y) = k] = \frac{M(y,s)^k}{k!}e^{-M(y,s)} = \frac{1}{k!}exp\{-\int_y^s \lambda(t)dt\}\{\int_y^s \lambda(t)dt\}^k.$$

For any integer k.

There are two different sampling protocols which provide data : (i) failure truncated case and (ii) time truncated case.

Data are said to be **failure truncated** when testing stops after a predetermined number of failures. Suppose that is observed till n events occur (fixed n), so we observe the ordered failure times $t_1 < t_2 < ... < t_n$ where $t_i$ is the time of ith failure. In this case, the number of failures is fixed and the time when the testing stops is random.

Data are said to be time truncated when testing stops at a predetermined time t. We observe a set of failure time $t_1 < t_2 < ... < t_n < t$. In this case, the time when the testing stops is fixed and the number of failures n is random.

### 2.3.1 Proprieties of Nonhomogeneous Poisson process

Analogously with homogeneous Poisson process, the non homogeneous Poisson process can be shown to have the following properties:

**Propriety 1.**

The joint Poisson distribution function of the events occurred at time $t_1, t_2, ..., t_n$ from an NHPP with intensity function $\lambda(t)$ is given by

$$f(t_1, t_2, ..., t_n) = \prod_{i=1}^{n} \lambda(t_i) exp\{-\int_0^y \lambda(t)dt\}$$

Where $y$ is the so-called stopping time: $y = t_n$ for the event truncated case, $y = t$ for the time truncated case. The two experiments lead to the same estimates in a Bayesian framework, whereas some differences are possible following a frequentist approach.

**Propriety 2.**

The superposition of two non homogeneous Poisson processes with intensities $\lambda_1(t)$ and $\lambda_2(t)$ is an non homogeneous Poisson process with intensity $\lambda(t) = \lambda_1(t) + \lambda_2(t)$

**Propriety 3.**

Given the total number of arrivals N(t) = n in the interval (0, t) from an non homogeneous Poisson process, the arrival instants of these n arrivals are distributed independently in the interval (0, t) with the density function

$$\frac{\lambda(t)}{\int_0^t \lambda(u)du}$$

**Propriety 4.**

Suppose that events are occurring according to a Poisson process with rate $\lambda$, and suppose that, independently of anything that occurred before, an event that happens at

time t is counted with probability $p(t)$. Then the process of the counted events constitutes a nonhomogeneous Poisson process with intensity $\lambda(t) = \lambda p(t)$.

A random selection from an nonhomogeneous Poisson process with intensity $\lambda(t)$ such that each arrival is selected, independent of the others, with the probability p(t) (note, may depend on time) results in an nonhomogeneous Poisson process with intensity $p(t)\lambda(t)$.

### 2.3.2 Power Law Process

A common function form for the intensity function in NHPP is

$$\lambda(t) = (\frac{\beta}{\alpha})(\frac{t}{\alpha})^{\beta-1}, \quad for\ \alpha > 0,\ \beta > 0,$$

Where $\alpha$ and $\beta$ are the scale parameter and shape parameter respectively. This nonhomogeneous Poisson process is usually called the Power Law Process. The mean value function M(t) of the process is

$$M(t) = E\{N(t)\} = \int_0^t \lambda(s)ds = \int_0^t (\frac{\beta}{\alpha})(\frac{s}{\alpha})^{\beta-1}ds = (\frac{t}{\alpha})^{\beta}$$

The power law process has been widely used in reliability growth(Crow(1982)), and software reliability models(Kyparisis and Singpurwalla(1985)), and in repairable systems (Ascher and Feingold (1984), Engelhardt and Bain(1986), Rigdon and Basu(1989)).

There has been much interest in using PLP where the value of $\beta$ varies over time. Previous research has used PLP where the value of $\beta$ is allowed to change at two fixed time-points. This model allows for three different stages of reliability: When $\beta < 1$, the intensity function $\lambda(t)$ is decreasing and the system is improving. Under this situation, the power law process can be applied as a reliability growth model.

When $\beta = 1$, the intensity function $\lambda(t)$ is constant. Under this situation, the reliability remains the same. In the third situation where the intensity function is increasing for $\beta > 1$, correspond to decay on reliability.

Let $M = \alpha^{-\beta}$. Parametric empirical Bayes procedures on the Power Law Process are easier to work with if the intensity function is parameterized as

$$\lambda(t) = M\beta t^{\beta-1},\ M,\ \beta\ >\ \ with\ M(t) = Mt^{\beta}$$

## 2.4 Some prior distribution for poisson process

We have a random sample $x_1, x_2...x_n$ from a Poisson($\lambda$) distribution. The proportional form of Bayes' theorem is given by :

$$posterior \propto prior \times likelihood$$

$$\pi(\lambda)/x_1, x_2...x_n) \propto \pi(\lambda) \times f(x_1, x_2...x_n/\lambda)$$

The parameter $\lambda$ can have any positive value, so we should use a continuous prior defined on all positive values. The proportional form of Bayes' theorem gives the shape of the posterior. The actual posterior is given by

$$\pi(\lambda) = \frac{\pi(\lambda) \times f(x_1, x_2...x_n/\lambda)}{\int_0^{\infty} \pi(\lambda) \times f(x_1, x_2...x_n/\lambda)}.$$

The likelihood of a single observation from a Poisson($\lambda$) distribution is given by

$$f(x/\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}.$$

For x = 0,1, . . . and $\lambda > 0$. The part that determines the shape of the likelihood is

$$f(x/\lambda) \propto \lambda^x e^{-\lambda}.$$

When $x_1, x_2...x_n$ is a random sample from a Poisson($\lambda$) distribution, the likelihood of the random sample is the product of the original likelihoods. This simplifies to

$$f(x_1, x_2...x_n/\lambda) = \prod_{i=1}^{n} f(x_i/\lambda)$$

$$\propto \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}.$$

In order to use Bayes' theorem, we will need the prior distribution of the Poisson parameter $\lambda$. In this section we will look at several possible prior distributions of $\lambda$ for which we can work out the posterior density without having to do the numerical integration.

### 2.4.1 Positive uniform prior density

Suppose we have no idea what the value of$\lambda$ is prior to looking at the data. In that case, we would consider that we should give all positive values of $\lambda$ equal weight. So we let the positive uniform prior density be

$$\pi(\lambda) = 1, \quad for \ \ \lambda > 0.$$

This prior density is improper since its integral over all possible values is infinite. The posterior will be proportional to prior times likelihood, so in this case the proportional posterior will be

$$\pi(\lambda)/x_1, x_2...x_n) \propto \pi(\lambda) \times f(x_1, x_2...x_n/\lambda)$$

$$\propto 1 \times \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}.$$

The posterior is the same shape as the likelihood function, that it is a Gamma$(\alpha, \beta)$, where $\alpha = \sum_{i=1}^{n} x_i + 1$ and $\beta = n$.

### 2.4.2 Jeffrey prior for Poisson

Jeffreys' method gives us priors which are objective in the sense that they are invariant under any continuous transformation of the parameter. The Jeffreys' prior for the Poisson is

$$\pi(\lambda) \propto [I(\lambda)]^{\frac{1}{2}}$$

Where

$$I(\lambda) = -E[\frac{\partial^2}{\partial\lambda^2} f(x/\lambda)]$$

is the Fisher information in the one-dimensional case

$$f(x/\lambda) = \prod_{i=1}^{n} \lambda^{x_i} \frac{e^{-\lambda}}{x_i!}$$

$$f(x/\lambda) = \lambda^{\sum_{i=1}^{n} x_i} \frac{e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

$$log\ f(x/\lambda) = \sum_{i=1}^{n} x_i log\lambda - n\lambda + \sum_{i=1}^{n} logx_i!$$

$$\frac{\partial log\ f(x/\lambda)}{\partial\lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n$$

$$\frac{\partial^2 log\ f(x/\lambda)}{\partial\lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2}$$

$$I(\lambda) = -E(-\frac{\sum_{i=1}^{n} x_i}{\lambda^2}) \;\; I(\lambda) = \frac{n}{\lambda}$$

$$\pi(\lambda) \propto [I(\lambda)]^{\frac{1}{2}}$$

$$\pi(\lambda) \propto \frac{1}{\sqrt{\lambda}} \;\; for \;\; \lambda > 0.$$

Using the Jeffreys' prior the proportional posterior will be

$$\pi(\lambda)/x_1, x_2...x_n) \propto \pi(\lambda) \times f(x_1, x_2...x_n/\lambda)$$

$$\propto \frac{1}{\sqrt{\lambda}} \times \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}$$

$$\propto \lambda^{\sum_{i=1}^{n} x_i - \frac{1}{2}} e^{-n\lambda}.$$

Which we recognize as the shape of a Gamma$(\alpha, \beta)$density, where $\alpha = \sum_{i=1}^{n} x_i + \frac{1}{2}$ and $\beta = n$.

### 2.4.3 Conjugate family for Poisson observations is the gamma family

The Gamma family of distributions is the conjugate family for Poisson observations. It is advantageous to use a prior from this family, as the posterior will also be from this family and can be found by the simple updating rules. This avoids having to do any numerical integration.

The conjugate prior for the observations from the Poisson distribution with parameter $\lambda$

$$\pi(\lambda) \sim Gamma(\alpha, \beta)$$

$$\pi(\lambda) \propto \lambda^{\alpha-1} e^{-\beta\lambda}$$

Given $T > 0$, the number N of events in the interval [0, T] is a Poisson random variable with mean equal to $\lambda T$ so that the likelihood function is

$$f(t_1, t_2, ..., t_n) = \prod_{i=1}^{n} \lambda(t_i) exp\{-\int_0^T \lambda(t)dt\}$$

$$f(t_1, t_2, ..., t_n) = \lambda^n exp^{-\lambda T}$$

By applying the Bayes theorem we obtain the posterior

$$\pi(\lambda/t) \propto \lambda^n exp^{-\lambda T}.\lambda^{\alpha-1} e^{-\beta\lambda}$$

$$\pi(\lambda/t) \propto Gamma(\alpha + n, \beta + T)$$

The posterior mean and variance are

$$E(\lambda/t) = \frac{\alpha + n}{\beta + T} \ \ and \ \ V(\lambda/t) = \frac{\alpha + n}{(\beta + T)^2}$$

## 2.5 Classes of priors for the parameters of a given NHPP

### 2.5.1 A general class

A general class of NHPP's can be described by the intensity function $\lambda(t, \alpha, \beta) = \alpha g(t, \beta)$, with $\alpha, \beta > 0$. Such that their m.v.f. is $M(t, \alpha, \beta) = \alpha G(t, \beta)$, with $G(t, \beta) = \int_0^t g(u, \beta) du$. This class contains well known processes, such as the Musa-Okumoto, the Cox-Lewis and the Power Law processes.

1. **The Musa-Okumoto process**

   This process has intensity function $\lambda(t, \alpha, \beta) = \frac{\alpha}{(t+\beta)}$ and mean value function is given by

$$M(t, \alpha, \beta) = \int_0^t (\frac{\alpha}{x + \beta}) dx$$

$$= \alpha \int_0^t (\frac{1}{x + \beta}) dx$$

$$M(t, \alpha, \beta) = \alpha log(t + \beta), \ \alpha, \beta > 0.$$

2. **The Cox-Lewis process**

   The Cox-Lewis process has intensity function $\lambda(t, \alpha, \beta) = \alpha exp\{\beta t\}$ and mean value function $M(t, \alpha, \beta)$ is given by

$$M(t,\alpha,\beta) = \int_0^t (\alpha e^{\beta x} dx)$$

$$= \alpha \int_0^t e^{\beta x} dx$$

$$= [\frac{\alpha}{\beta} e^{\beta * t} - \frac{\alpha}{\beta} e^{\beta * 0}]$$

$$M(t,\alpha,\beta) = \frac{\alpha}{\beta} e^{\beta t}(t-1),\ \alpha,\beta > 0.$$

3. **Power Law process**

   Power Law process has intensity function $\lambda(t,\alpha,\beta) = \alpha\beta t^{\beta-1}$ and mean value function is given by

$$M(t,\alpha,\beta) = \int_0^t \alpha\beta x^{\beta-1} dx$$

$$M(t,\alpha,\beta) = \alpha t^{\beta},\ \alpha,\beta > 0.$$

A unified treatment of this class from a Bayesian viewpoint is possible. Consider a NHPP with intensity function $\lambda(t)$.Suppose we observe the system up to time $y$ and let n be the number of failures, occurred at times $t_1 < t_2 ... < t_n$; then the likelihood function is given by

$$L(t_1, t_2, ..., t_n) = \prod_{i=1}^{n} \lambda(t_i) exp\{-\int_0^y \lambda(t) dt\}$$

Where $I = (t_1, t_2, ..., t_n)$.

We focus on the Power law process with intensity $\lambda(t) = \beta\alpha t^{\beta-1}$, then the likelihood function is given by

$$L(\alpha,\beta,I)=\prod_{i=1}^{n}\beta\alpha t_i^{\beta-1}exp\{-\int_0^y\beta\alpha t^{\beta-1}dt\}$$

$$L(\alpha,\beta,I)=\beta^n\alpha^n\prod_{i=1}^{n}t_i^{\beta-1}exp\{-\int_0^y\beta\alpha t^{\beta-1}dt\}$$

$$L(\alpha,\beta,I)=\beta^n\alpha^n\prod_{i=1}^{n}t_i^{\beta-1}exp\{-\alpha y^{\beta}\}$$

A possible prior distribution for $\beta$ and $\alpha$ is given by

- $\beta \sim Uniform(0,2)$
- $\alpha \sim Gamma(a,b)$

We suppose $\alpha$ and $\beta$ are independent.

By applying the Bayes theorem we obtain

$$\pi(\beta,\alpha/I)\propto L(\beta,\alpha,t)\pi(\beta)\pi(\alpha)$$

$$\pi(\beta,\alpha/I)\propto\beta^n\alpha^n\prod_{i=1}^{n}t_i^{\beta-1}e^{-\alpha y^{\beta}}\alpha^{a-1}e^{-b\alpha}I_{(0,2)}(\beta)$$

$$\pi(\beta,\alpha/I)\propto\beta^n\alpha^{n+a-1}e^{-\alpha(y^{\beta}+b)}e^{\beta\sum_{i=1}^{n}logt_i}I_{(0,2)}(\beta)$$

One can consider the following class

$$\Gamma:\{\beta\sim U(0,2)\ \sim G(a,a),\ a>0\}$$

Then

$$E(\beta/I)=\int\beta\pi(\beta,\alpha/I)d\beta$$

$$E(\beta/I)=\frac{\int\beta L(\beta,\alpha,I)\pi(\beta)\pi(\alpha)d\beta}{\int L(\beta,\alpha,I)\pi(\beta)\pi(\alpha)d\beta}$$

$$E(\beta/I) = \frac{\int \beta\alpha^{n+a-1}e^{-\alpha(y^\beta+b)}\beta^n e^{\beta\sum_{i=1}^{n} logt_i} I_{(0,2)}(\beta)d\beta}{\int \alpha^{n+a-1}e^{-\alpha(y^\beta+b)}\beta^n e^{\beta\sum_{i=1}^{n} logt_i} I_{(0,2)}(\beta)d\beta}$$

$$E(\beta/I) = \frac{\int_0^2 \beta\alpha^{n+a-1}e^{-\alpha(y^\beta+b)}\beta^n e^{\beta\sum_{i=1}^{n} logt_i} d\beta}{\int_0^2 \alpha^{n+a-1}e^{-\alpha(y^\beta+b)}\beta^n e^{\beta\sum_{i=1}^{n} logt_i} d\beta}$$

Robustness measures are computed as the prior varies in this class. Computations are the evaluation of suprema and infima of the quantity of interest (posterior mean).

$$Range = \sup_{\alpha\in\Gamma} E(\beta/I) \; - \; \inf_{\alpha\in\Gamma} E(\beta/I)$$

### 2.5.2 A class based on differential equations

Another class can be described by relation between intensity function $\lambda(t)$ and mean value function $M(t)$ which is given by

$$\lambda(t) = \lambda e^{-\theta M(t)}$$

The relation can be expressed as the first order differential equation

$$[M(t)]' = \lambda e^{-\theta M(t)};\ so$$

$$\theta[M(t)]' e^{\theta M(t)} = \theta\lambda,$$

$$[e^{\theta M(t)}]' = \theta\lambda,$$

$$e^{\theta M(t)} = \theta\lambda t + c,$$

$$M(t) = \frac{log(\theta\lambda t + c)}{\theta}$$

When $M(0) = 0$, the solution is,

$$M(t) = \frac{log(\lambda\theta t + 1)}{\theta}$$

Similar relations can be found for other NHPP's. The PLP has intensity function $\lambda(t) = \alpha\beta t^{\beta-1}$ and mean value function $M(t) = \alpha t^{\beta}$, their relation is given by $[M(t)]' = \frac{\beta M(t)}{t}$

$$[M(t)]' = \lambda(t),\ so\ \frac{\beta M(t)}{t} = \alpha\beta t^{\beta-1}$$

When $M(0) = 0$, the solution is,

$$M(t) = \alpha t^{\beta}$$

Similarly, other relations were proposed, belong to the class of first order differential equations.

# Chapter 3

# Robustness of classes of homogeneous and non homogeneous Poisson processes

The Poisson distribution is used to count the number of occurrences of rare events which are occurring randomly through time (or space) at a constant rate. The events must occur one at a time. In this Chapter, we study the robustness of the classes of priors of the parameters of a given NHPP where the prior changes over a proposed class and computation of some measures of robustness. Computations with NHPP are difficult. we start with homogeneous poisson processes and then we will do it later with NHPP.

A detailed discussion on the results of Bayesian robustness analysis of the classes proposed of homogeneous and nonhomogeneous poisson processes is provided.

## 3.1 Study of Robustness - class of prior for HPP

### 3.1.1 Class of Gamma$(\alpha, \beta)$

Let $\pi(\lambda)$ a Gamma prior corresponding $t_i$ the likelihood $f(t_1, t_2, ..., t_n)$, and define

$$\Gamma = \{\pi : \ Gamma(\alpha, \beta) : \ \ \frac{\alpha}{\beta} = k\}.$$

Class of all Gamma priors with mean k, where k is fixed, which can be written also as

$$\Gamma = \{\pi : \ Gamma(k\beta, \beta) : \ , \beta > 0\}$$

By this class $\Gamma$ of prior, the posterior distribution is a Gamma$(k\beta + n, \beta + T)$

The quantity of interest is the posterior mean;

$$E(\lambda/t) = \frac{k\beta + n}{\beta + T}$$

The uncertainty might be quantified by specifying the range spanned by the posterior mean, as the prior varies over the class.

Let

$$h(\beta) = \frac{k\beta + n}{\beta + T}$$

Then

$$h'(\beta) = \frac{k(\beta + T) - (k\beta + n)}{(\beta + T)^2} = \frac{kT - n}{(\beta + T)^2}$$

We are

$$. \ h'(\beta) > 0 \ \ if \ \ kT - n > 0 \ \Rightarrow \ \underset{\beta>0}{sup} E(\lambda/t) \underset{\beta\to\infty}{=} k \ \ and \ \ \underset{\beta>0}{inf} E(\lambda/t) \underset{\beta\to 0}{=} \frac{n}{T}$$

$$. \ h'(\beta) < 0 \ \ if \ \ kT - n < 0 \ \Rightarrow \ \underset{\beta>0}{sup} E(\lambda/t) \underset{\beta\to 0}{=} \frac{n}{T} \ \ and \ \ \underset{\beta>0}{inf} E(\lambda/t) \underset{\beta\to\infty}{=} k$$

$$. \ h'(\beta) = 0 \ \ if \ \ kT - n = 0 \ \Rightarrow \ E(\lambda/t) = k \ is \ a \ constant.$$

Then we use the global measures of sensitivity for determining the range of the posterior mean as the prior varies over the class

$$range \ = \ \underset{\pi\in\Gamma}{sup} \ E(\lambda/t) \ - \ \underset{\pi\in\Gamma}{inf} \ E(\lambda/t) = |k - \frac{n}{T}|$$

If the measure of range is small then robustness is achieved and any prior in the class can be chosen; and if the measure of range is large there is not robustness.
We consider a numerical example; data were generated by a Poisson Process assuming $\lambda = 2$ . The obtained data are shown in Table.3.1.

| n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $t_i$ | 0.17540 | 0.7840 | 1.76613 | 2.46538 | 3.09604 | 4.3574 | 4.36602 | 5.16747 | 5.19417 | 5.6102 |
| n | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| $t_i$ | 6.57309 | 7.04254 | 7.23656 | 7.84013 | 8.58444 | 9.0439 | 9.21996 | 9.9472 | 10.2435 | 10.7079 |

**Table.3.1.** Data generated by a Poisson Process assuming $\lambda = 2$.

Resulting range for different values of K are presented in Table.3.2.

| k | max | min | range | $range_{abs}$ | k | max | min | range | $range_{abs}$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.10 | 1.8182 | 0.10 | -1.7182 | 1.7182 | 1.81 | 1.8182 | 1.81 | -0.0082 | 0.0082 |
| 0.15 | 1.8182 | 0.15 | -1.6682 | 1.6682 | 1.82 | 1.82 | 1.8182 | 0.0018 | 0.0018 |
| 0.24 | 1.8182 | 0.24 | -1.5782 | 1.5782 | 1.84 | 1.84 | 1.8182 | 0.0218 | 0.0218 |
| 0.35 | 1.8182 | 0.35 | -1.4682 | 1.4682 | 1.88 | 1.88 | 1.8182 | 0.0618 | 0.0618 |
| 0.45 | 1.8182 | 0.45 | -1.3682 | .3682 | 1.90 | 1.90 | 1.8182 | 0.0818 | 0.0818 |
| 0.76 | 1.8182 | 0.76 | -1.0582 | 1.0582 | 2.05 | 2.05 | 1.8182 | 0.2318 | 0.2318 |
| 0.85 | 1.8182 | 0.85 | -0.9682 | 0.9682 | 2.36 | 2.36 | 1.8182 | 0.5418 | 0.5418 |
| 1.15 | 1.8182 | 1.15 | -0.6682 | 0.6682 | 2.55 | 2.55 | 1.8182 | 0.7318 | 0.7318 |
| 1.43 | 1.8182 | 1.43 | -0.3882 | 0.38382 | 2.85 | 2.85 | 1.8182 | 1.0318 | 1..318 |
| 1.70 | 1.8182 | 1.70 | -0.1182 | 0.1182 | 3.20 | 3.20 | 1.8182 | 1.3818 | 1.3818 |
| 1.80 | 1.8182 | 1.80 | -0.0182 | 1.0182 | 8.10 | 8.10 | 1.8182 | 6.2818 | 6.2818 |

**Table.3.2.** Range of posterior mean for Gamma prior.

For values of K inferior than 1.7 and superior than 1.9 the range is large and the

robustness is not achieved. The robust situation is when k = 1.82. Then, any prior $G(\alpha, \beta)$ in the class$\Gamma$ can be chosen and there is robustness.

**Theorem.3.1.(see[30])**

Given $h(\theta)$ the posterior quantity of interest, $L(\theta)$ the model density and $\pi(\theta)$ the prior distribution, its results

$$\sup_{\pi\in\Gamma}\frac{\int_{\Omega} h(\theta)L(\theta)\pi(\theta)d\theta}{\int_{\Omega} L(\theta)\pi(\theta)d\theta} = \sup_{\theta_i\in I_i, i=1,\ldots,m}\frac{\sum_{i=1}^{m} h(\theta_i)L(\theta_i)p_i}{\sum_{i=1}^{m} L(\theta_i)p_i}$$

### 3.1.2 Class based on median at $\lambda = 1, 2\ and\ 3$

We observe N(t), a homogeneous Poisson process with parameter $(\lambda)$ . Further, it is felt a priori that $\lambda$ has a distribution with median 1, i.e.$P^{\pi}(\lambda \leq 1) = 0.5 = P^{\pi}(\lambda \geq 1)$, we consider as well the prior with median 2. i.e.$P^{\pi}(\lambda \leq 2) = 0.5 = P^{\pi}(\lambda \geq 2)$, and with median 3 i.e.$P^{\pi}(\lambda \leq 3) = 0.5 = P^{\pi}(\lambda \geq 3)$. The classes are defined as

$$\Gamma_1 = \{\pi : \ \ median = 1\}$$

$$\Gamma_2 = \{\pi : \ \ median = 2\}$$

$$\Gamma_3 = \{\pi : \ \ median = 3\}$$

Then, under $\Gamma_1$ the posterior mean is

$$E(\lambda/t) = \frac{\int_0^{\infty}\lambda.f(t,\lambda).\pi(\lambda)d\lambda}{\int_0^{\infty} f(t,\lambda).\pi(\lambda)d\lambda}$$

$$= \frac{\int_0^{\infty}\lambda^{n+1}.e^{-\lambda T}.\pi(\lambda)d\lambda}{\int_0^{\infty}\lambda^{n}.e^{-\lambda T}.\pi(\lambda)d\lambda}$$

Then, range of posterior mean over this classes can be computed, we start with the class $\Gamma_1$

$$\sup_{\pi\in\Gamma_1} E(\lambda/t) \; - \; \inf_{\pi\in\Gamma_1} E(\lambda/t)$$

We have $P^\pi(\lambda \leq 1) = 0.5 = P^\pi(\lambda \geq 1)$, then we consider a discrete distribution in two points $\lambda_1$ and $\lambda_2$ ( a DIRAC distribution)as given by Theorem 3.1 , and we write

$$\frac{1}{2}\delta_{\lambda_1} + \frac{1}{2}\delta_{\lambda_2}$$

We obtain

$$\sup_{\lambda_1\in[0,1],\lambda_2\in[1,\infty]} E(\lambda/t) = \sup_{\lambda_1\in[0,1],\lambda_2\in[1,\infty]} \frac{\int_0^\infty \lambda^{n+1}.e^{-\lambda T}.(\frac{1}{2}\delta_{\lambda_1} + \frac{1}{2}\delta_{\lambda_2})d\lambda}{\int_0^\infty \lambda^{n}.e^{-\lambda T}.(\frac{1}{2}\delta_{\lambda_1} + \frac{1}{2}\delta_{\lambda_2})d\lambda}$$

$$= \sup_{\lambda_1\in[0,1],\lambda_2\in[1,\infty]} \frac{\frac{1}{2}\lambda_1^{n+1}e^{-\lambda_1 T} + \frac{1}{2}\lambda_2^{n+1}e^{-\lambda_2 T}}{\frac{1}{2}\lambda_1^{n}e^{-\lambda_1 T} + \frac{1}{2}\lambda_2^{n}e^{-\lambda_2 T}}$$

$$= \sup_{\lambda_1\in[0,1],\lambda_2\in[1,\infty]} \frac{\lambda_1^{n+1}e^{-\lambda_1 T} + \lambda_2^{n+1}e^{-\lambda_2 T}}{\lambda_1^{n}e^{-\lambda_1 T} + \lambda_2^{n}e^{-\lambda_2 T}}$$

By the same method of calculation we obtain :

$$\inf_{\lambda_1\in[0,1],\lambda_2\in[1,\infty]} E(\lambda/t) = \inf_{\lambda_1\in[0,1],\lambda_2\in[1,\infty]} \frac{\int_0^\infty \lambda^{n+1}.e^{-\lambda T}.(\frac{1}{2}\delta_{\lambda_1} + \frac{1}{2}\delta_{\lambda_2})d\lambda}{\int_0^\infty \lambda^{n}.e^{-\lambda T}.(\frac{1}{2}\delta_{\lambda_1} + \frac{1}{2}\delta_{\lambda_2})d\lambda}$$

$$= \inf_{\lambda_1\in[0,1],\lambda_2\in[1,\infty]} \frac{\lambda_1^{n+1}e^{-\lambda_1 T} + \lambda_2^{n+1}e^{-\lambda_2 T}}{\lambda_1^{n}e^{-\lambda_1 T} + \lambda_2^{n}e^{-\lambda_2 T}}$$

Considering the class $\Gamma_2 = \{\pi : \ \ median = 2\}$, we can obtain the range

$$\sup_{\lambda_1\in[0,2],\lambda_2\in[2,\infty]} \frac{\lambda_1^{n+1}e^{-\lambda_1 T} + \lambda_2^{n+1}e^{-\lambda_2 T}}{\lambda_1^{n}e^{-\lambda_1 T} + \lambda_2^{n}e^{-\lambda_2 T}} - \inf_{\lambda_1\in[0,2],\lambda_2\in[2,\infty]} \frac{\lambda_1^{n+1}e^{-\lambda_1 T} + \lambda_2^{n+1}e^{-\lambda_2 T}}{\lambda_1^{n}e^{-\lambda_1 T} + \lambda_2^{n}e^{-\lambda_2 T}}$$

And considering the class $\Gamma_3 = \{\pi : \ \ median = 3\}$, we can obtain the range over this class

$$\sup_{\lambda_1\in[0,3],\lambda_2\in[3,\infty]} \frac{\lambda_1^{n+1}e^{-\lambda_1 T}+\lambda_2^{n+1}e^{-\lambda_2 T}}{\lambda_1^{n}e^{-\lambda_1 T}+\lambda_2^{n}e^{-\lambda_2 T}} - \inf_{\lambda_1\in[0,3],\lambda_2\in[3,\infty]} \frac{\lambda_1^{n+1}e^{-\lambda_1 T}+\lambda_2^{n+1}e^{-\lambda_2 T}}{\lambda_1^{n}e^{-\lambda_1 T}+\lambda_2^{n}e^{-\lambda_2 T}}$$

We must have to find the max and the min of these quantities, that wants to say, we must obtain $\lambda_1$ and $\lambda_2$ which maximize this ratio and $\lambda_1$ and $\lambda_2$ which minimize this same ratio when the prior changes in each class $\Gamma_1$, $\Gamma_2$ and $\Gamma_3$ .

Considering the same data generated by Poisson process, and the classes of priors, we get the uppers and the lowers bounds of the posteriors means in each class. The results of the optimizations are presented in Table.3.3 for the class $\Gamma_1$, in Table.3.4 for the class $\Gamma_2$ and in the Table.3.5 for the class $\Gamma_3$.

| $\lambda_1$ | $\lambda_2$ | $max$ | $\lambda_1$ | $\lambda_2$ | $min$ | $range_{\Gamma 1}$ |
|---|---|---|---|---|---|---|
| 1 | 67.7393 | 67.7393 | 0.1 | 2 | 2 | 65.7393 |
| 1 | 7.9999 | 1 | 0.000 | 36.7954 | 0.0000001 | 0.9999999 |
| 1 | 25 | 1 | 0.000 | 36.6665 | 0.0000004 | 0.9999996 |
| 1 | 12 | 1 | 0.000 | 34.9552 | 0.0000006 | 0.9999994 |
| 1 | 18 | 1 | 0.000 | 34.72 | 0.0000007 | 0.9999993 |
| 1 | 8 | 1 | 0.000 | 36.5742 | 0.00000033 | 0.9999997 |
| 1 | 25 | 1 | 0.0001 | 25 | 0.000062 | 0.999938 |
| 1 | 30 | 1 | 0.000 | 30.0001 | 0.0000054 | 0.9999946 |
| 1 | 12 | 1 | 0.000 | 34.7654 | 0.00000058 | 0.99999942 |
| 1 | 17 | 1 | 0.000 | 39.6114 | 0.00000005 | 0.99999995 |
| 1 | 13 | 1 | 0.000 | 35.8118 | 0.00000031 | 0.99999969 |
| 1 | 48 | 1 | 0.000 | 48 | 0.00000007 | 0.9999773 |
| 1 | 28 | 1 | 0.000 | 36.5447 | 0.000014 | 0.999986 |
| 1 | 36 | 1 | 0.000 | 36 | 0.0000014 | 0.9999986 |
| 1 | 26 | 1 | 0.000 | 26 | 0.000038 | 0.999962 |
| 0 | 67.7393 | 67.7393 | 0.000 | 35.6434 | 0.00000035 | 67.7393 |
| 1 | 90 | 1 | 0.000 | 90 | 0.00000047 | 0.99999953 |
| 1 | 122 | 1 | 0.000 | 122 | 0.0000014 | 0.9999986 |
| 1 | 6 | 1 | 0.000 | 35.1093 | 0.00000041 | 0.00009959 |
| 1 | 280 | 1 | 0.000 | 280 | 0.00000059 | 0.99999941 |

**Table.3.3.** Range of posterior mean for the class of priors $\Gamma_1$ .

| $\lambda_1$ | $\lambda_2$ | $max$ | $\lambda_1$ | $\lambda_2$ | $min$ | $range_{\Gamma 2}$ |
|---|---|---|---|---|---|---|
| 0.000 | 67.7394 | 67.7394 | 0.000 | 3 | | 64.7394 |
| 2 | 15 | 2 | 0.000 | 35.9166 | 0.00000031 | 1.99999969 |
| 2 | 30 | 2 | 0.000 | 30 | 0.0000055 | 1.9999945 |
| 2 | 25 | 2 | 0.0001 | 25 | 0.000062 | 1.999938 |
| 2 | 24 | 2 | 0.001 | 24.0002 | 0.00010 | 1.9999 |
| 0.000 | 67.7394 | 67.7394 | 1.4355 | 2 | 1.7766 | 65.3628 |
| 2 | 20 | 2 | 0.000 | 35.0305 | 0.00000046 | 1.99999954 |
| 2 | 26 | 2 | 0.000 | 26.0001 | 0.000038 | 1.999962 |
| 2 | 25 | 2 | 0.0001 | 25 | 0.000062 | 1.999938 |
| 2 | 30 | 2 | 0.000 | 30 | 0.0000054 | 1.9999946 |
| 2 | 35 | 2 | 0.000 | 28 | 0.0000018 | 1.9999982 |
| 2 | 22 | 2 | 0.000 | 37.1315 | 0.00000016 | 1.99999984 |
| 2 | 28 | 2 | 0.000 | 28 | 0.000014 | 1.999986 |
| 2 | 2.4741 | 2.1312 | 0.000 | 53.3245 | 0.00000041 | 2.131196 |
| 2 | 50 | 0 | 0.000 | 50 | 0.00000053 | 1.99999947 |
| 0 | 67.7393 | 67.7393 | 0.000 | 34.8509 | 0.00000056 | 67.7393994 |
| 2 | 250 | 2 | 0.000 | 250 | 0.0000012 | 1.9999988 |
| 2 | 20 | 2 | 0.000 | 40.8013 | 0.000000027 | 1.999999973 |

**Table.3.4.** Range of posterior mean for the class of priors $\Gamma_2$ .

| $\lambda_1$ | $\lambda_2$ | $max$ | $\lambda_1$ | $\lambda_2$ | $min$ | $range_{\Gamma 3}$ |
|---|---|---|---|---|---|---|
| 0 | 67.7394 | 67.7394 | 0.1 | 5 | 5 | 62.7394 |
| 3 | 7 | 3 | 0.000 | 34.5216 | 0.00000082 | 2.99999918 |
| 3 | 9 | 3 | 0.000 | 35.1317 | 0.00000047 | 2.99999953 |
| 3 | 12 | 3 | 0.000 | 35.8712 | 0.00000031 | 2.99999969 |
| 2.9 | 13 | 3 | 0.000 | 35.2854 | 0.00000049 | 2.99999951 |
| 3 | 15 | 3 | 0.000 | 36.5721 | 0.00000022 | 2.99999978 |
| 3 | 16 | 3 | 0.000 | 36.7914 | 0.00000019 | 2.99999981 |
| 3 | 20 | 3 | 0.000 | 34.2235 | 0.00000074 | 2.99999926 |
| 3 | 22 | 3 | 0.000 | 35.1218 | 0.00000043 | 2.99999957 |
| 3 | 24 | 3 | 0.0001 | 24.0001 | 0.00010 | 2.9990 |
| 3 | 25 | 3 | 0.0001 | 25.0002 | 0.000062 | 2.999938 |
| 3 | 26 | 3 | 0.000 | 26 | 0.000038 | 2.999962 |
| 3 | 29 | 3 | 0.000 | 29 | 0.0000090 | 2.9999910 |
| 3 | 32 | 3 | 0.000 | 32 | 0.0000026 | 2.9999974 |
| 3 | 35 | 3 | 0.000 | 35 | 0.0000011 | 2.9999989 |
| 3 | 38 | 3 | 0.000 | 38 | 0.00000058 | 2.99999942 |
| 3 | 44 | 3 | 0.000 | 44 | 0.0000012 | 2.9999988 |
| 3 | 48 | 3 | 0.000 | 48 | 0.000000903 | 2.999999087 |
| 3 | 53 | 3 | 0.000 | 53 | 0.000001 | 2.999999 |
| 3 | 60 | 3 | 0.000 | 60 | 0.0000007 | 2.9999993 |

**Table.3.5.** Range of posterior mean for the class of priors $\Gamma_3$ .

After a choice of several values initials of $\lambda_1$ and $\lambda_2$, posterior ranges have been computed for the class of priors $\Gamma_1$, $\Gamma_2$ and $\Gamma_3$ and we have some results in the Table.3.2, Table.3. and Table 3.4.

A nice case in Table.3.3 (the best situation of robustness) is when the maximum value of the posterior mean is obtained for $\lambda_1 = 1$ and $\lambda_2 = 25$, and the minimum value of the

posterior mean is obtained for $\lambda_1 = 0.0001$ and $\lambda_2 = 25$. So that the optimal values are

$$\overline{\rho} = 1, \quad \underline{\rho} = 0.000062$$

and the range is 0.999938.

A nice case in Table.3.4 (the best situation of robustness) is when the maximum value of the posterior mean is obtained for $\lambda_1 = 2$ and $\lambda_2 = 24$, and the minimum value of the posterior mean is obtained for $\lambda_1 = 0.001$ and $\lambda_2 = 24.0002$. So that the optimal values are

$$\overline{\rho} = 2, \quad \underline{\rho} = 0.0001$$

and the range is 1.9999.

A nice case in Table.3.5 (the best situation of robustness) is when the maximum value of the posterior mean is obtained for $\lambda_1 = 3$ and $\lambda_2 = 24$, and the minimum value of the posterior mean is obtained for $\lambda_1 = 0.001$ and $\lambda_2 = 24.0001$. So that the optimal values are

$$\overline{\rho} = 3, \quad \underline{\rho} = 0.0001$$

and the range is 2.9999.

### 3.1.3 Class based on median at $\lambda = 2$ and quartiles at 1 and 3

We observe change as before. Further, it is felt a priori that $\lambda$ has a distribution with median 2, upper quartile 3 and under quartiles 1.i.e.

$$P^{\pi}(\lambda \leq 1) = P^{\pi}(1 \leq \lambda \leq 2) = P^{\pi}(2 \leq \lambda \leq 3) = P^{\pi}(\lambda \geq 3) = 0.25$$

The class is defined as

$$\Gamma_Q = \{\pi : \textit{ Quartiles at } 1 \textit{ and } 3, \textit{ median at } 2 \}$$

The quantity of interest is always the posterior mean. Then, under $\Gamma_Q$, the posterior mean is defined

$$E(\lambda/t) = \frac{\int_0^\infty \lambda.f(t,\lambda).\pi(\lambda)d\lambda}{\int_0^\infty f(t,\lambda).\pi(\lambda)d\lambda} = \frac{\int_0^\infty \lambda^{n+1}.e^{-\lambda T}.\pi(\lambda)d\lambda}{\int_0^\infty \lambda^{n}.e^{-\lambda T}.\pi(\lambda)d\lambda}$$

As it can be seen from the above class, a natural global measure of sensitivity of the Bayesian quantity to the choice of prior is the range of this quantity of interest as the prior varies in the class of priors.

$$\sup_{\pi\in\Gamma_1} E(\lambda/t) \; - \; \inf_{\pi\in\Gamma_1} E(\lambda/t)$$

we have $P^\pi(\lambda \leq 1) = P^\pi(1 \leq \lambda \leq 2) = P^\pi(2 \leq \lambda \leq 3) = P^\pi(\lambda \geq 3) = 0.25$, then we consider a discrete distribution in four points $\lambda_1$, $\lambda_2$, $\lambda_3$ and $\lambda_4$ ( a DIRAC distribution)as given by Theorem 3.1 , and we write

$$\frac{1}{4}\delta_{\lambda_1} + \frac{1}{4}\delta_{\lambda_2} + \frac{1}{4}\delta_{\lambda_3} + \frac{1}{4}\delta_{\lambda_4}$$

We obtain

$$\sup_{\lambda_1\in[0,1],\lambda_2\in[1,2],\lambda_3\in[2,3],\lambda_4\in[3,\infty]} E(\lambda/t)$$

$$= \sup_{\lambda_1\in[0,1],\lambda_2\in[1,2],\lambda_3\in[2,3],\lambda_4\in[3,\infty]} \frac{\int_0^\infty \lambda^{n+1}.e^{-\lambda T}.(\frac{1}{4}\delta_{\lambda_1} + \frac{1}{4}\delta_{\lambda_2} + \frac{1}{4}\delta_{\lambda_3} + \frac{1}{4}\delta_{\lambda_4})d\lambda}{\int_0^\infty \lambda^{n}.e^{-\lambda T}.(\frac{1}{4}\delta_{\lambda_1} + \frac{1}{4}\delta_{\lambda_2} + \frac{1}{4}\delta_{\lambda_3} + \frac{1}{4}\delta_{\lambda_4})d\lambda}$$

$$= \sup_{\lambda_1\in[0,1],\lambda_2\in[1,2],\lambda_3\in[2,3],\lambda_4\in[3,\infty]} \frac{\frac{1}{4}\lambda_1^{n+1}e^{-\lambda_1 T} + \frac{1}{4}\lambda_2^{n+1}e^{-\lambda_2 T} + \frac{1}{4}\lambda_3^{n+1}e^{-\lambda_3 T} + \frac{1}{4}\lambda_4^{n+1}e^{-\lambda_4 T}}{\frac{1}{4}\lambda_1^{n}e^{-\lambda_1 T} + \frac{1}{4}\lambda_2^{n}e^{-\lambda_2 T} + \frac{1}{4}\lambda_3^{n}e^{-\lambda_3 T} + \frac{1}{4}\lambda_4^{n}e^{-\lambda_4 T}}$$

$$= \sup_{\lambda_1\in[0,1],\lambda_2\in[1,2],\lambda_3\in[2,3],\lambda_4\in[3,\infty]} \frac{\frac{1}{4}\sum_{i=1}^{4}(\lambda_i)^{n+1}.e^{-\lambda_i T}}{\frac{1}{4}\sum_{i=1}^{4}(\lambda_i)^{n}.e^{-\lambda_i T}}$$

$$= \sup_{\lambda_1\in[0,1],\lambda_2\in[1,2],\lambda_3\in[2,3],\lambda_4\in[3,\infty]} \frac{\sum_{i=1}^{4}(\lambda_i)^{n+1}.e^{-\lambda_i T}}{\sum_{i=1}^{4}(\lambda_i)^{n}.e^{-\lambda_i T}}$$

Therefore, by the same way, we obtain

$$= \inf_{\lambda_1\in[0,1],\lambda_2\in[1,2],\lambda_3\in[2,3],\lambda_4\in[3,\infty]} \frac{\sum_{i=1}^{4}(\lambda_i)^{n+1}.e^{-\lambda_i T}}{\sum_{i=1}^{4}(\lambda_i)^{n}.e^{-\lambda_i T}}$$

We must find the values of $\lambda_1$, $\lambda_2$, $\lambda_3$ and $\lambda_4$ who minimize and maximize the quantity of the posterior mean under this class.

Considering the same data generated by Poisson process, and the class of prior $\Gamma_Q$. We get the upper and the lower bounds of the posteriors means in this class, the results of the optimizations are presented in Table.3.6.

| $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $max$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $min$ | $range$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | 1.4355 | 2 | 5.6364 | 38.2112 | 0.2 | 1.9743 | 3 | 5.4702 | 1.7766 | 36.4346 |
| 0.1999 | 1.9220 | 3 | 5.5759 | 39.5874 | 0.2 | 1.2866 | 3 | 6.4040 | 1.5066 | 38.0808 |
| 0.2517 | 1.9679 | 3 | 8 | 38.4052 | 1 | 1.4231 | 2 | 9.2154 | 1.7505 | 36.6540 |
| 0.2455 | 1.9163 | 3 | 28 | 39.7090 | 1 | 1.2454 | 3 | 29.3548 | 1.4431 | 38.2659 |

**Table.3.6.** Range of posterior mean for the class of priors $\Gamma_Q$.

After a choice of several values initials of $\lambda_1$, $\lambda_2$, $\lambda_3$ and $\lambda_4$, the difference between upper and lower bounds on the quantity of interest is too large. (Its value measures the variation caused by the uncertainty in the prior)

## 3.2 Study of Robustness - class of prior for NHPP

### 3.2.1 Class based on the Cox-Lewis process

The Cox-Lewis process is one of the general classes of a NHPP, this class is defined by the intensity function $\lambda(t, \alpha, \beta) = \alpha exp\{\beta t\}$.

The class proposed the following

$$\Gamma_C = \{All\ NHPP\ with\ \lambda(t) = \lambda e^{+\theta t},\ \theta \in [-1, 1],\ \lambda \sim G(\alpha, \beta)\}$$

Where $\lambda(t) = \lambda e^{+\theta t}$ is the intensity function of the Cox-Lewis process .

Consider a NHPP with intensity function $\lambda(t)$. Suppose we observe the system up to time T and let n be the number of occurrences, occurred at times $t_1, t_2, ..., t_n$; then the likelihood function is given by

$$L(t_1, t_2, ..., t_n) = \prod_{i=1}^{n} \lambda(t_i) exp\{-\int_0^T \lambda(t)dt\}$$

$$L(t_1, t_2, ..., t_n) = \prod_{i=1}^{n} \lambda e^{+\theta t_i} exp\{-\int_0^T \lambda e^{+\theta t}dt\}$$

$$L(t_1, t_2, ..., t_n) = \lambda^n e^{\theta \sum_{i=1}^{n} t_i} exp\{-\int_0^T \lambda e^{+\theta t}dt\}$$

$$L(t_1, t_2, ..., t_n) = \lambda^n e^{\theta \sum_{i=1}^{n} t_i} exp\{-\frac{\lambda}{\theta}(e^{\theta T} - 1)\}$$

A conjugate prior distribution for $\lambda$ is given by $\lambda \sim G(\alpha, \beta)$, then

$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta}$$

By applying the Bayes theorem we obtain

$$\pi(\lambda/t_1, t_2, ..., t_n) = \frac{\lambda^n e^{\theta \sum_{i=1}^{n} t_i} exp\{-\frac{\lambda}{\theta}(e^{\theta T} - 1)\}.\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta}}{\int_0^T \lambda^n e^{\theta \sum_{i=1}^{n} t_i} exp\{-\frac{\lambda}{\theta}(e^{\theta T} - 1)\}.\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta} d\lambda}$$

$$\pi(\lambda/\underline{t}) \propto \lambda^{\alpha+n-1}.e^{-\lambda(\beta+\frac{e^{\theta T}-1}{\theta}}$$

$$\lambda/\underline{t} \sim G(\alpha+n, \beta+\frac{e^{\theta T}-1}{\theta 1})$$

The posterior mean is

$$E(\lambda/\underline{t}) = \frac{\alpha+n}{\beta+\frac{e^{\theta T}-1}{\theta}}$$

The global robustness is based on the calculation of the range of the quantity of interest (posterior mean) when the prior change over this class

$$\underset{\pi\in\Gamma}{sup} E(\lambda/\underline{t}) - \underset{\pi\in\Gamma}{inf} E(\lambda/\underline{t})$$

We start by the computation of $\underset{\pi\in\Gamma}{sup} E(\lambda/\underline{t})$

$$\underset{\pi\in\Gamma}{sup} E(\lambda/\underline{t}) = \underset{\theta\in[-1,1]}{sup} E(\lambda/\underline{t}) = \underset{\theta\in[-1,1]}{sup} \frac{\alpha+n}{\beta+\frac{e^{\theta T}-1}{\theta}}$$

$$= \frac{\alpha+n}{\beta+\underset{\theta\in[-1,1]}{inf}\frac{e^{\theta T}-1}{\theta}}$$

We put $h_1(\theta) = \frac{e^{\theta T}-1}{\theta}$ then

$$h_1'(\theta) = \frac{T.e^{\theta T}.\theta - e^{\theta T}+1}{\theta^2}$$

We must find $\widehat{\theta}$ which minimize $h_1(\theta)$ and maximize $E(\lambda/\underline{t})$ .

After the computation of $\underset{\pi\in\Gamma}{sup} E(\lambda/\underline{t})$, we can do the same to the $\underset{\pi\in\Gamma}{inf} E(\lambda/\underline{t})$

$$\underset{\pi\in\Gamma}{inf} E(\lambda/\underline{t}) = \underset{\theta\in[-1,1]}{inf} E(\lambda/\underline{t}) = \underset{\theta\in[-1,1]}{inf} \frac{\alpha+n}{\beta+\frac{e^{\theta T}-1}{\theta}}$$

$$= \frac{\alpha+n}{\beta+\underset{\theta\in[-1,1]}{sup}\frac{e^{\theta T}-1}{\theta}}$$

We put $h_2(\theta) = \frac{e^{\theta T}-1}{\theta}$ then

$$h_2'(\theta) = \frac{T.e^{\theta T}.\theta - e^{\theta T}+1}{\theta^2}$$

We must find $\widehat{\theta}$ which maximize $h_2(\theta)$ and minimize $E(\lambda/\underline{t})$.
Considering the same data generated by Poisson process, we get the upper and the lower bounds of the posteriors means in this class, the results of the optimizations for different values of $\alpha$ and $\beta$ are presented in Table.3.7.

| $\theta_{max}$ | $supE(\lambda/t)$ | $\theta_{min}$ | $infE(\lambda/t)$ | $range$ |
|---|---|---|---|---|
| -0.9999 | 11.2219 | 1 | 0.000337 | 11.2216 |
| -0.9999 | 10.0998 | 1 | 0.000337 | 10.0994 |
| -0.9999 | 6.7332 | 1 | 0.000337 | 6.7329 |
| -0.9999 | 6.6699 | 1 | 0.000334 | 6.6696 |
| -0.9999 | 4.0002 | 1 | 0.000334 | 3.9908 |
| -0.9999 | 2.8571 | 1 | 0.000334 | 2.8568 |
| -0.9999 | 1.8182 | 1 | 0.000334 | 1.8178 |
| -0.9999 | 0.9524 | 1 | 0.000334 | 0.9520 |
| -0.9999 | 0.3922 | 1 | 0.000334 | 0.3918 |
| -0.9999 | 0.3279 | 1 | 0.000334 | 0.3275 |

**Table.3.7.** Range of posterior mean for the class of priors $\Gamma_C$.

For small values of $\alpha$ and a large values of $\beta$ of Gamma distribution , the range of posterior mean is reasonably small. The conclusion of this analysis would be that robustness likely obtains for smaller values of $\alpha$ and larger values of $\beta$.

# Conclusion

In this work, we first explore how robust or sensitive is Bayesian analysis to the choice of prior, utility, and model. In the process, we introduce and examine various quantitative evaluations of robustness.

In most cases in practice, quantification of subjective belief or judgment is not easily available. It is then common to choose from among conventional priors on the basis of some relatively simple subjective judgments about the problem and the conventional probability model for the data. Such priors have been criticized for various reasons.

Than, we proposed a Poisson process model and studied the global robustness of some class of priors proposed for the parameters of NHPP and HPP , In perspective, it will be interest to study the local sensitivity of some classes of priors for the parameters of a given NHPP.

# Bibliography


[1] .Abraham, C., et Daurés, J. (2000). Global robustness with respect to the loss function and the prior. Theory and Decision, 48(4), 359-381.

[2] .Albert, J., Delampady, M. and Polasek, W. (1991). A class of distributions for robustness studies. J. Statist. Inference, 28, 291-304.

[3] .Arnold Zellner, An Introduction to Bayesian Inference in Econometrics, Wiley 1996.

[4] .Athanasios Kottas and Bruno Sansó, Bayesian Mixture Modeling for Spatial Poisson Process Intensities, with Applications to Extreme Value Analysis, REVISED VERSION: March 16, 2006.

[5] .Athanasios Kottas, A nonparametric Bayesian approach to inference for non-homogeneous Poisson processes, University of California, 1156 High Street, Santa Cruz, CA 95064, USA, NSF grant DMS-0505085, August 23, 2006.

[6] .Basu, S., et DasGupta, A. (1995). Robust Bayesian analysis with distribution bands. Statist. Decisions, 13, 333-349.

[7] .Basu, S., et DasGupta, A. (1990). Bayesian analysis under distribution bands. Technical Report. 90-48, Departement of statistics, Purdue University, USA.

[8] .Basu, S. (1995). Ranges of posterior probability over a distribution band. Journal of Statistical Planning and Inference, 44, 149-166.

[9] .Berger, J.O, et O'Hagan, A. (1988). Ranges of posterior probabilities for unimodal priors with specified quantiles. In J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds., Bayesian statistics 3, Oxford University Press, Oxford, U.K.

[10] .Berger, J.O., Rios Insua, D. and Ruggeri, F. (2000). Bayesian Robustness. In Robust Bayesian analysis (D. Rios Insua and F. Ruggeri, eds.), Springer-Verlag, New York.

[11] .Berger.J.O. Statistical Decision Theory and Bayesian Analysis, 2nd ed. Springer Verlag, New York, 1985.

[12] .Berger.J.O. (1990). Robust Bayesian Analysis: Sensitivity to the prior. J. Statist. Plan. Inference, 25, 303-328.

[13] .Berger. J.O, et Berliner, L.M. (1986). Robust Bayes and empirical Bayes analysis with $\epsilon$-contaminated priors. Ann. Statist., 14, 461-486.

[14] .Berger.J. O. , Statistical Decision Theory and Bayesian Analysis, Second Edition, Springer-Verlag 1985.

[15] .Berger .J.O, Purdue University (1987). Robust Bayesian Analysis, Technical Report.

[16] .Bernardo, J. (1979). Reference posterior distributions for Bayesian inference (with discussion). J. Royal Statist. Soc. Series B, 41, 113-147.

[17] .Betrò, B., Meczaeski, M., et Ruggeri, F. (1994). Robust Bayesian analysis under generalized moments conditions. J. Statists. Plan. Inference, 41, 257-266.

[18] .Box, G.E.P. et Tiao, G.C. (1962). A further look at robustness via Bayes theorem. Biometrika, 49, 419-432 et 546.

[19] .Chin-Yu Huang, Michael R. Lyu, and Sy-Yen Kuo, (2003). A Unified Scheme of Some Nonhomogenous Poisson Process Models for Software Reliability Estimation.

[20] .Christine Cocozza , Processus stochastique et fiabilité des systeme, Edition, Spriger 1997.

[21] .Christian, P. Robert. (2006). Le choix Bayésien: Principes et pratique. Springer-Verlag France, Paris.

[22] .C. Mazzali and F. Ruggeri. Bayesian analysis of failure count data from repairable systems. Technical Report 98.19, CNR - IAMI, Milano, 1998.

[23] .Davide Cavallo and Fabrizio Ruggeri, Bayesian Models For Failures in a Gas Networrk, In E. Zio, M. Demichela, and N. Piccinini, editors, Safety and Reliability - ESREL 2001, pages 1963–1970, 2001.

[24] .Dominique Foata, Aimé Fuchs, Processus stochastiques: processus de Poisson, Chaines de Marcov et Martingales, Edition DUNOD Paris, 2002 2004.

[25] .Enrico Cagno, Franco Caron, Mauro Mancini and Fabrizio Ruggeri (2000), Sensitivity of Replacement Priorities for Gas Pipline Maintenace, in Robust Bayesian Analysis, (D. Rios Insua and F.Reggeri, eds.), Springer, New york.

[26] .Fabrizio Ruggeri, Rios Insua, D. et Martin, J. (2000). Robust Bayesian Analysis. In Robust Bayesian analysis (D. Rios Insua and F. Ruggeri, eds). Springer-Verlag, New York.

[27] .Fabrizio Ruggeri, (). On the Reliability of Repairable Systems: Methods and Applications

[28] .Fabrizio Ruggeri, et Sivaganesan, S. (2000). On a global sensitivity measure for Bayesian inference. Sanklya Ser. A, 62, 110-127.

[29] .Fabrizio Ruggeri, Jacinto Mart´ýn and David Riýos Insua, (February 12, 2004). Preference robustness in Bayesian Decision Theory

[30] .Fabrizio Ruggeri (1990). Posterior ranges of functions of parameters under priors with specified quantiles. Comm. Statist. Theory methods, 19, 127-144.

[31] .Fabrizio Ruggeri and S. Sivaganesan. On modeling change points in non-homogeneous Poisson processes. Tentatively accepted by Stat. Inference Stoch. Process., 2005.

[32] .Gelfand, A., et Dey, D.K. (1991). On Bayesian robustness of contaminated classes of priors. Statist. Decisions, 9, 63-80.

[33] .Gelfand, A.E., et Dey, D.K. (1994). Bayesian model choice: Asymptotics and exact calculations. Journal if the Royal Statistical Society, 56, 501-514.

[34] .Gelfand, A. et Smith, A. (1990). Sampling based approaches to calculating marginal densities. J. American statist. Assoc., 85, 398-409.

[35] .Gelman, A., Carlin, J.B., Stern, H.S., et Rubin, D.B. (1995). Bayesian data analysis. Chapman and Hall, London.

[36] .George, E.P.Box. 1980. Sampling and Bayes inference in scientific modelling and robustness. Journal of the Royal Statistical Society, 143, 383-430.

[37] .Guida,M., Calabria, R. and Pulcini,G., Bayes Inference for a Non-homogeneous Poisson Process with Power Law Intensity. IEEE Transactions on Reliability. R-38(1989)603-609.

[38] .Goutis, C. (1994). Ranges of posterior measures for some classes of priors with specified moments. In statist. Review, 62, 245-257.

[39] .Gustafson, P. (1994). Local sensitivity of posterior expectations. Ph. D. Dissertation, Departement of Statistics, Carnegie Mellon University, USA.

[40] .Gustafson, P. (2000). Local robustness in Bayesian analysis. In D. Rios Insua and F. Ruggeri, eds., Robust Bayesian analysis, Springer-Verlag, New York, USA.

[41] .Henk C. Tijms, A First Course in Stochastic Models, Wiley , England, (2003).

[42] .Jayanta K. Ghosh, Mohan Delampady and Tapas Samanta, An Introduction to Bayesian Analysis Theory and Methods. Springer (2006).

[43] .Jean-Jacques Boreux, Éric Parent et Jacques Bernier. Pratique du calcul Bayésien. Springer-Verlag France, Paris, 2010.

[44] .Lavine, M. (1991). Sensitivity in Bayesian statistics: the prior and the likelihood. Journal of the American Statistical Association, 86, 396- 399.

[45] .Madansky, A. (1990). Bayesian analysis with incompletely specified prior distributions. In S. Geisser et al, eds, Bayesian and Liklihood Methods.

[46] .Moreno, E. (2000). Global Bayesian robustness for some classes of prior distributions. In Robust Bayesian Analysis, (D.Rios Insua and F.Ruggeri,eds). New York: Springer-Verlag.

[47] .Parent, E. et Bernier, J. (2007).Le raisonnement Bayésien: Modélisation et Inférence. Springer-Verlag France, Paris.

[48] .Peter Congdon. (2003). Applied Bayesian Modelling. John Wiley et Sons Ltd, England.

[49] .Sivaganesan, S. (2000). Global and Local robustness approaches : Uses and limitations. In D. Rios Insua and F. Ruggeri, eds., Robust Bayesian Analysis. Springer-Verlag, New York, USA.

[50] .Sivaganesan, S. (1988). Range of posterior measures for priors with arbitrary contaminations. Communications in Statistics A: Theory and Methods, 17, 1591-1612.

[51] .Sivaganesan, S. (1993). Robust Bayesian diagnostics. J. Statist. Plan. Inference, 35, 171-188.

[52] .Sheldon M. Ross, Stochastic Processes, Second Edition, United staes of America, 1980.

[53] .Shyamalkumar, N.D. (2000). Likelihood robustness. In Robust Bayesian Analysis, (D. Rios Insua and F. Ruggeri, eds.). New York: Springer- Verlag.

[54] .S.E. Rigdon and Basu A.P. The Power Law process: a model for the reliability of repairable systems. Journal of Quality Technology, 10:251–260, 1989.

[55] .Tae Young Yang and Lynn Kuo, Bayesian Computation for the Superposition of Nonhomogeneous poisson processes. 1991.

[56] .William A. Massey, Geraldine A. Parker and Ward Whitt, Estimating the Parameters of a Nonhomogeneous Poisson Process with Linear Rate, AT and T Bell Laboratories Murray Hill, NJ 07974-0636 November 16, 1993.

[57] .William M. Bolstad, Introduction to Bayesian Statistics Second Edition, John Wiley and Sons, Inc. 2007.

[58] .W lodzimierz Bryc. Applied Probability and Stochastic Processes, Lecture Notes for 577/578 Class, March 21, 1996.

[59] .Yuri Goegebeur, The Poisson and the nonhomogeneous Poisson process, ST505 - Statistical Simulation, December 3, 2007.

[60] .Zhao Chen, Bayesian and Empirical Bayes Approaches to Power Law Process and Microarray Analysis, A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy Department of Mathematics College of Arts and Sciences University of South Florida, July 12th, 2004.

[61] .Zhao Chen, Empirical Bayes Analysis on the Power Law Process with Natural Conjugate Priors, Journal of Data Science 8(2010), 139-149.