Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Nonparametric Bayesian Bioassay

Master thesis

Author: Lidia Burzala

Delft, The Netherlands
July 2007

Delft University of Technology

Abstract

Nonparametric Bayesian bioassay

by Lidia Burzala

In many scientific publications the analysis of biological assays has received the attention of statisticians. This thesis presents a review of the nonparametric Bayesian techniques with the aim to explain the relation between the response probability and the dosage in quantal bioassays. The approach for nonparametric dose response curves with Dirichlet process priors as well as with product of beta priors is proposed. We show here how the problem of estimating potency curves may be handled by using two techniques: MLE-type algorithms and Markov Chain Monte Carlo methods. We extend our investigation to an analysis of prior constraints on the shape of the potency curve. The nonparametric Bayesian bioassay including ordered polytomous response will be discussed. Our interest is focused also on inference about the unknown dose level or efficacy dose level corresponding to prespecified rate. The illustrative examples with different design strategies are provided to emphasize the importance of selecting optimal parameters. Applications consist of data analysis using the Gibbs sampler method.

Key words: nonparametric Bayesian bioassay, ordered Dirichlet process prior, Gibbs sampling, Effective dose

Members of the Committee:

Chairperson of Graduate Committee:

Prof. dr. Thomas A. Mazzuchi

Graduate Committee:

Prof. dr. Thomas A. Mazzuchi
George Washington University, United States
Delft University of Technology, the Netherlands

Prof. dr. Roger M. Cooke
Delft University of Technology, the Netherlands
Resources for Future, United States

Dr. Dorota Kurowicka
Delft University of Technology, the Netherlands

Dr. Eric A.Cator
Delft University of Technology, the Netherlands

Acknowledgments

The two years at Delft University of Technology, have been a valuable experience of my life. I wish to express sincere appreciation to Prof. Roger M. Cooke for giving me the opportunity to take part in the MSc program “Risk and Environmental Modelling”. I would like to also thank Dr. Dorota Kurowicka for good care of the students in our group.

I would like to thank my supervisors: Prof. Tom Mazzuchi and Prof. Roger M. Cooke for supervising my work and providing me with very helpful comments and corrections. I am greatly indebted for their encouragement in the moment of need. I am very grateful to Prof. Tom Mazzuchi and his family for offering me their house and great time in Washington DC.

I’d like to also thank all students from my group for unforgettable two years at TUDelft, especially I would like to thank my friends: Iwona and Maryia for their friendship, support and help.

I dedicate the present work to my parents and Marcin.

Delft, The Netherlands
July 15, 2007

Lidia Burzala

Table of Contents

List of tables

List of figures

Chapter 1

Introduction

The use of animals in scientific investigation has been traced back to several centuries BC. For instance, the writings of Aristotle (384-322 BC) and Erasitratus (304-258 BC) indicate that they had studied the anatomy of various animals. Early investigations such as these were the beginning of the basic sciences that today form the foundations for new drug development. Nowadays, in the epoch of many incurable, fatal diseases, the pharmaceutics laboratories conduct extensive studies in order to find suitable drugs or vacines. The experiments on living animals are a very important part of a drug development. They enable to test new drugs in order to understand how they work. This type of scientific experiment is called biological assay. Finney[1] introduces the biological assay as "a set of techniques relevant to … the measurement of the potency of a stimulus; physical, chemical or biological, …by means of the reactions that it produces in living matters". By conducting this kind of experiment, scientists can make more rapid progress than would otherwise have been possible. However, it is important to know how to use the information these experiments provide. As it will be shown in this thesis, statistical methods play a crucial role in bioassay. The issues facing statistician include:

[1]D.J.Finney (1971) "Probit analysis"

- estimation of the tolerance distribution (potency of the stimulus) $P(x)$ for any dose level x;
- distribution of an effective dose $\hat{x}$ such that for a fixed p, $P(\hat{x}) = p$;
- the design of biological experiment to accomplish in an optimal way the potency curve and the distribution of an effective dose;

We can distinguish different types of biological assay. In this thesis we emphasize an indirect assay based on quantal response, recorded as the response "all or nothing" for example whether a death does or does not occur. More generally, each experimental subject possesses a threshold or tolerance level. If the dose is less than the tolerance level, the subject does not respond; otherwise a response is observed. The relationship between stimulus and response can be used to study the potency of a dose from produced response. Note that many dose – response studies are conducted during the early phase evaluation of a new drug, and therefore little is know about the behavior of this drug at that stage. The statistical inference can be sensitive to the choice of the parametric form. A nonparametric approach with the flexible and adaptive modeling property to estimate the dose – response relation is therefore of considerable interest.

1.1 Objectives of the thesis

The main purpose of this thesis is:

- to review the nonparametric Bayesian approach to bioassay problems with emphasis on the techniques for estimating the tolerance distribution
- to present the application of one of the Markov Chain Monte Carlo methods, namely Gibbs sampler, to bioassay data.

1.2 Outline of the thesis

The thesis is organized as follows. In chapter 2 the terminology of typical bioassay will be described and the prior of the response curve will be proposed. The main interest focuses on family of the Dirichlet process prior (Ferguson, 1973), however we also present the alternative class of priors based on the product-Beta family (Gelfand and Kuo, 1991). The study about properties of prior we will extend to a discussion about different shapes of the potency curve (Ramgopal, Laud and Smith, 1993). We provide also a useful extension of the quantal bioassay model which allows for ordered and polytomous response arising from the stochastically ordered potency curves (Gelfand, Kuo, 1991). Chapter 3 deals with the problem of estimating potency curve. We shall illustrate here how Markov Chain Monte Carlo procedures such as Gibbs sampling and the MLE method can be used to carry out Bayesian inference in order to explain the relation between the response probability and the dosage in a quantal bioassay. The reader interested in obtaining the posterior distribution of an effective dose (ED) is intended to read the next chapter. Chapter 5 provides few examples of different design strategies. We present here discussion about parameters: the “best” thresholds for testing and the optimal number of experimental subjects required to test at each threshold. The application part is provided in chapter 6, where the bioassay data are analyzed by using Gibbs sampler. Conclusions are presented in chapter 7.

Chapter 2

Prior specification

2.1 Experiment

In a typical experiment a set of doses, the numbers signifying the level of applied stimulus (for example, a vitamin, a hormone or a drug) are selected and administered to experimental subjects. Denote these dose levels by $x_1, x_2, \ldots, x_M$ assuming that they are listed in the order of increasing values. Suppose that n_i subjects receive dose x_i ($i = 1, ..., M$). The results are measured by quantal responses, the numbers of affected subjects at each dose. Let us denote by s_i the number of affected subjects at dose level x_i. Finally define $P(x)$ - the tolerance distribution which can be viewed as the expected proportion of the experimental subjects with tolerance level less than or equal to the dose level x. We consider potency curves P (often called response curves), which are nondecreasing and such that $P(0) = 0$ and $P(\infty) = 1$. Furthermore, we assume that P is right-continuous so that P is a distribution function. Biological experiments provide very useful information about the potency curve. If $p_i = P(x_i)$ denotes the true 'response rate' at dose x_i, the number of events at x_i follows the binomial distribution $Bi(n_i, p_i)$. We shall further assume that the subjects respond independently. Therefore the joint

likelihood function of $p = (p_1, \ldots, p_M)$ at the observational doses $x = (x_1, \ldots, x_M)$ can be expressed as the product:

$$L(s \mid p) = \prod_{i=1}^{M} \binom{n_i}{s_i} p_i^{s_i} (1 - p_i)^{n_i - s_i} , \qquad (2.1)$$

2.2 Families of priors

As we have already mentioned in chapter 1, this work provides a study of the nonparametric Bayesian approach. An essential ingredient of Bayesian analysis is the prior distribution. It plays a major role in determining the final results. Traditional parametric inference considers models for the unknown *P* that can be indexed by the finite dimensional parameters (for example; mean and covariance matrix for a multivariate normal distribution). However, in many cases, constraining inference to a specific parametric form may limit the range and type of inferences that can be drawn from such model. It is often felt that statistical models based on conventional parametric distributions are not sufficiently flexible to provide a realistic description of the data.,especially in case of bioassay data provided from the studies which are conducted during the early phase evaluation of a new drug. This leads to widespread use of nonparametric estimators. In such an inference, we assume that the unknown *P* belongs to a nonparametric class of right continuous, non-decreasing functions taking values in $[0,1]$ (no specific parametric form is assumed) with the requirement that our prior reflects prior information as accurately as possible. Thus a prior may be thought of as a stochastic process taking values in the given function space. Ferguson (1973) states two important desirable properties for this class of priors:

- their support should be large
- posterior inference should be “analytically manageable”

Gelfand and Kuo (1991) proposed two families of such priors. The first one is frequently referenced in the literature as 'Ferguson's Dirichlet process prior'. The second class of priors, which is less popular than the previous one, is the product-Beta family. Below we focus on aforementioned classes of priors.

2.2.1 Dirichlet process prior

The Dirichlet process is an important tool for the treatment of nonparametric statistical problems from a Bayesian point of view. This class of prior distributions has been studied by for example: Ramsey (1972), Antoniak (1974), Disch (1981), Ammann (1984). However, the most comprehensive and extensive studies about Dirichlet processes (*DP*) were presented by Ferguson (1973). He introduce *DP* as a prior distribution on the collection of all probability measures. We define it as follow: a random probability distribution *P* is generated by *DP* if for any partition $B_1,\ldots,B_M$ of the sample space, the vector of random probabilities $P(B_i)$ follows a Dirichlet distribution: $(P(B_1),\ldots,P(B_M)) \sim D(\alpha P_0(B_1),\ldots,\alpha P_0(B_M))$. Note that in order to specify DP prior; a precision parameter $\alpha > 0$ and base distribution P_0 are required. Here P_0 defines the expectation: $E[P(B_i)] = P_0(B_i)$, whereas parameter α appears in the expression of variance: $\text{var}[P(B_i)] = [P_0(B_i)(1-P_0(B_i))]/[\alpha+1]$.

In our study we assume that the *P* has an Ordered Dirichlet distribution, with density at $(p_1,\ldots,p_M)$ as stated below:

$$\pi_D(p_1,\ldots,p_M) = \frac{\Gamma(\sum_{i=1}^{M+1}\alpha\xi_i)}{\prod_{i=1}^{M+1}\Gamma(\alpha\xi_i)} p_1^{\alpha\xi_1-1}(p_2-p_1)^{\alpha\xi_2-1}\ldots(p_M-p_{M-1})^{\alpha\xi_M-1}(1-p_M)^{\alpha\xi_{M+1}-1}, \qquad (2.2)$$

where $0 \le p_1 \le \ldots \le p_M \le 1$ and $\xi_i = P_0(x_i) - P_0(x_{i-1})$ ($i = 1, \ldots M$). Note that the constraints on $P_0(x_i)$ ($P_0(x_0) \equiv 0$ and $P_0(x_{M+1}) \equiv 1$ where $x_0 = 0, x_{M+1} = \infty$) require that $\xi = (\xi_1, \ldots, \xi_{M+1}) > 0$ and $\sum_{i=1}^{M+1} \xi_i = 1$.

The reason for which this family of priors is so attractive from the Bayesian point of view, is the very useful properties of *DP*. For example, the marginal distributions are known. Namely, $P(B_i)$ for any i is a Beta distribution, $P(B_i) \sim Beta(\alpha P_0(B_i), \alpha(1 - P_0(B_i)))$. Moreover, the Dirichlet distribution is conjugate family for the multinomial distribution. Note that these properties are used in the next chapter in order to estimate the potency curve. Now let us stress the interpretation of the precision parameter and base distribution. The question arises how to select them in an efficient, smart way and how sensitive our results will be for different selections. Gelfand and Kuo (1991) suggest that the P_0 distribution should be taken to be the standard distribution whose median agrees with our prior guess for the ED50 (a dose for which probability of respond is 50%) and whose spread provides rough agreement with our prior expectation at other dosage levels. To aid in selecting a precision parameter, it is helpful to observe the density functions for marginals. We plot one of these densities with different values of precision parameter for $\xi_i = \xi_{i+1} = 0.25$. The results are presented in the figure 2.1. As can be verified, this density function approaches to the uniform as $\alpha \to \frac{1}{0.25}$ (we plot then pdf of $Beta(1,1)$), whereas as α increases, the density concentrates at the interior of the interval $(0,1)$. In general, in case where $P(x_i)$ is restricted to the interval $(P(x_{i-1}), P(x_{i+1}))$, this parameter controls the probability that $P(x_i)$ will be close to $P(x_{i-1})$ or $P(x_{i+1})$ and thus it controls the degree of smoothness to be expected for the posterior estimates of $P(x)$. This parameter can be thought of as the assessor's strength of belief in the prior guess. A large value of it reflects that *P* is tightly concentrated about P_0. In chapter 6, results from the analysis of bioassay data confirm all of the theoretical interpretations.

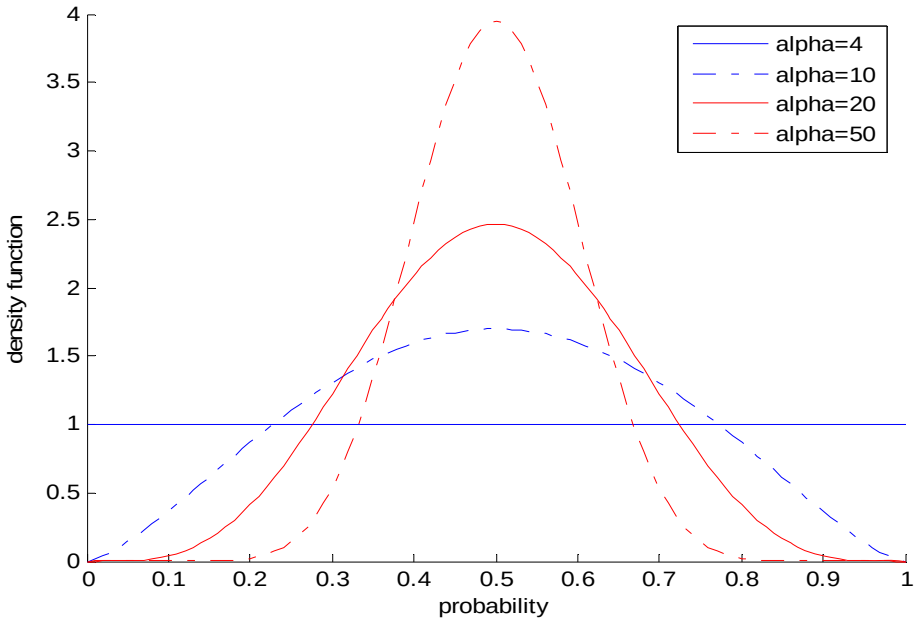

Figure 2.1. Density functions $Beta(\alpha\xi_i, \alpha\xi_{i+1})$ for varying value α and $\xi_i = \xi_{i+1} = 0.25$.

As we have already mentioned, the *DP* is the most popular prior in nonparametric Bayesian study. Consequently, our Bayesian inference about the potency curve will be mostly based on that class of priors. However we shall also introduce an alternative family of priors discussed by Gelfand and Kuo (1991).

2.2.2 Product – Beta priors

We shall now to investigate another class of priors, the product – Beta family. The density of *p* can be expressed as:

$$\pi_B(p) = c_M(\alpha, \beta)\prod_{i=1}^{M} p_i^{\alpha_i - 1}(1 - p_i)^{\beta_i - 1}, \qquad (2.3)$$

where $\alpha = (\alpha_1, \ldots, \alpha_M)$, $\beta = (\beta_1, \ldots, \beta_M)$ and c_k is the normalizing constant under restriction to $S^M = \{p : 0 \le p_1 \le \ldots \le p_M \le 1\}$.

Although the product - Beta family is a very flexible class of priors, it is not obvious how to select parameters α and β in accordance with prior information. Gelfand and Kuo (1991) propose some intuitive specification of these parameters. Let us first define $e^{(i)}$, a row vector having *1* at the *i*-th coordinate and zero elsewhere. The expectation involving the p_i's may be formally given as:

$$E_B(p_i) = \frac{c_k(\alpha, \beta)}{c_k(\alpha + e^{(i)}, \beta)}.$$

This suggests equating $E_B(p_i) = P_o(x_i)$ for $i = 1, \ldots, M$. Moreover, authors define a precision parameter $M_i = \alpha_i + \beta_i$ analogous to α from *DP* class of priors. The value of this parameter reflects our confidence in the value $P_0(x_i)$. Unfortunately, an explicit calculation of $E_B(p_i)$ may be infeasible as indicated by Gelfand and Kuo. However, the conditional distribution of $p_i \mid p_j (i \neq j)$ is evidently a Beta distribution, $Beta(\alpha_i, \beta_i)$ restricted to $[p_{i-1}, p_{i+1}]$. Though, again, the mean of this conditional distribution is not available explicitly, its mode is $\rho_i = \frac{\alpha_i - 1}{M_i - 2}$ provided $\alpha_i > 1$, $\beta_i > 1$ and $p_{i-1} \leq \rho_i \leq p_{i+1}$.

Finally, taking ρ_i as an approximation to the marginal mode for p_i we obtain

$$\rho_i = \frac{\alpha_i - 1}{M_i - 2} = P_0(x_i).$$

This implies that

$$\alpha_i = (M_i - 2)P_0(x_i) + 1,$$

and

$$\beta_i = M_i - \alpha_i.$$

Gefland and Kuo (1991) provide a discussion about advantages and disadvantages of using the aforementioned families of priors. For example, they note that π_B is convenient being the conjugate with respect to *(2.1)* while π_D is not. However, to specify a product – Beta prior we shall encounter some difficulties in the cases of selecting appropriate parameters.

2.3 Prior constraint on the shape of the potency curve

The motivation for extending our studies about *DP* prior was the discussion provided by Ramgopal, Laud and Smith (1993). They propose to consider a prior constraint on the shape of potency curve. Note that convexity and concavity constraints are commonly employed in the field of bioassay as reasonable shape constraint on a dose – response relation. Shaked and Singpurwalla (1990) present extensive discussion about these two cases. Moreover, Ramgopal, Laud and Smith (1993) propose one more shape constraint, called ogive, which is based on the previous two.

Since all of the aforementioned cases occur frequently in pharmacodynamic studies (where the response of subjects to varying levels of some stimulus is examined) we shall discuss each of them separately. Note that our investigation is based on Ramgopal, Laud and Smith (1993) and the following three cases are considered.

- $P(x)$ is convex, corresponding to non-decreasing density function f;
- $P(x)$ is concave, corresponding to non-increasing density function f;
- $P(x)$ is ogive, corresponding to f which, is non-decreasing in $[x_0, x']$ for some unknown x' and then non-increasing in $(x', x_{M+1}]$;

Let $x_0 \equiv 0$ and x_{M+1} be the right endpoint of the support of P such that $x_{M+1} = \sup\{x; P(x) < 1\}$; note that . x_{M+1} may be ∞.

Before we start to describe each of the cases above let us first define 'slope parameters' z_i given as follow:

$$z_i = \frac{P(x_i) - P(x_{i-1})}{x_i - x_{i-1}}, \qquad \text{for } i = 1, \ldots, M+1.$$

Note that if $x_{M+1} = \infty$, then $x_{M+1} - x_M = \infty$ thus we assume in this case that $z_{M+1} = 0$.

2.3.1 Convex priors

For the description of convex priors, we start with the definition of convex function.

***Definition**:* A function *g* is convex on an interval $[a,b]$ if for any two points x_1 and x_2 in $[a,b]$ and any λ $(0<\lambda<1)$,

$$g[\lambda x_1+(1-\lambda)x_2]<\lambda g(x_1)+(1-\lambda)g(x_2).$$

***Remark**:* If *g* has a second derivative in $[a,b]$, then a necessary and sufficient condition for it being convex on that interval is that the second derivative $g(x)''>0$ for all *x* in $[a,b]$.

It follows from the remark above that convex shape of *P(x)*, corresponds to a non - decreasing density function.

Define now the following parameters:

- $y_i=z_i-z_{i-1}$, for $i=1,\ldots,M+1$, with $z_0=0$;

  Notice that the convexity assumption implies that $y_i\geq 0$ for all *i*.
- $U_i=(x_{M+1}-x_{i-1})y_i$, for all *i*, with $U_1+\ldots U_{M+1}=1$;

  It is straightforward to verify that $U_i\geq 0$ for all *i*

From the definition above it follows that $U=(U_1,\ldots,U_{M+1})$ can be regarded as a vector of probabilities. Therefore, Ramgopal, Laud and Smith (1993) assign a Dirichlet prior to *U* analogously to Shaked and Singpurwalla (1990), with the density at $u=(u_1,\ldots,u_{M+1})$ as follows.

$$\pi(u) = \frac{\Gamma(\sum_{i=1}^{M+1} \alpha_i)}{\prod_{i=1}^{M+1} \Gamma(\alpha_i)} u_1^{\alpha_1 - 1}(u_2 - u_1)^{\alpha_2 - 1} \ldots (u_M - u_{M-1})^{\alpha_M - 1}(1 - u_M)^{\alpha_{M+1} - 1}, \qquad (2.4)$$

for some $\alpha_i > 0$.

An attractive property of this prior is the knowledge about the conditional distribution of each U_i. Note that it can be expressed as:

$$U_i \mid U_1, \ldots U_{i-1} \sim Beta(\alpha_i, \sum_{j=i+1}^{M+1} \alpha_j; 0, 1 - \sum_{j=1}^{i-1} u_j), \qquad (2.5)$$

where $u_0 = 0$ and $v \sim Beta(a, b; c, d)$ means that $v = c + (d - c)w$, for $c < d$, where w is a standard $Beta(a, b)$.

By writing $p = (p_1, \ldots, p_{M+1})$ as a transformation of u it follows from *(2.5)*, after some algebra, that, for $i = 1, \ldots, M$,

$$p_i \mid p_1, \ldots p_{i-1} \sim Beta(\alpha_i, \sum_{j=i+1}^{M+1} \alpha_j; \bar{c}_i, \bar{d}_i), \qquad (2.6)$$

where

$$\bar{c}_i = p_{i-1} + \frac{x_i - x_{i-1}}{x_{i-1} - x_{i-2}}(p_{i-1} - p_{i-2}), \qquad \bar{d}_i = p_{i-1} + \frac{x_i - x_{i-1}}{x_{M+1} - x_{i-1}}(1 - p_{i-1}),$$

with $p_{-1} = 0$.

As it will be pointed out later on, the conditional distribution *(2.6)* allows us to avoid to a large extent, difficult computations in estimation of the potency curve.

2.3.2 The concave prior

We start here with the definition of a concave function.

Definition: A function *g* is concave on an interval $[a,b]$ if for any two points x_1 and x_2 in $[a,b]$ the function -*g* is convex.

Remark: If *g* has a second derivative in $[a,b]$, then a necessary and sufficient condition for it to be concave on that interval is that the second derivative $g''(x) < 0$ for all *x* in $[a,b]$, which corresponds to a non – increasing density function.

Define analogous to the previous case

- $y_i = z_i - z_{i+1}$, for $i = 1,\ldots,M+1$, with $z_{M+2} = 0$;
  Note that the concavity assumption implies $y_i \geq 0$ for all *i*.
- $U_i = (x_i - x_0)y_i$, for all *i*, with $U_1 + \ldots U_{M+1} = 1$;
  It is straightforward to verify that $U_i \geq 0$ for all *i*.

Analogous to the previous case we assign a Dirichlet prior to *U*. Then the conditional distributions for $i = M+1,\ldots,2$ can be expressed as below

$$u_i \mid u_{i+1},\ldots u_{M+1} \sim Beta(\alpha_i, \sum_{j=1}^{i-1}\alpha_j; 0, 1-\sum_{j=i+1}^{M+1} u_j), \qquad (2.7)$$

where $u_{M+2} = 0$. It follows now from *(2.7)*, after some algebra and manipulations that, for $i = M,\ldots,1$,

$$p_i \mid p_{i+1},\ldots, p_M \sim Beta(\sum_{j=1}^{i}\alpha_j, \alpha_{i+1}; c_i, d_i), \qquad (2.8)$$

where

$$c_i = p_{i+1} - \frac{x_{i+1} - x_i}{x_{i+1} - x_0}(p_{i+1} - p_0), \qquad d_i = p_{i+1} - \frac{x_{i+1} - x_i}{x_{i+2} - x_{i+1}}(p_{i+2} - p_{i+1}),$$

with $p_{M+2} = 1$.

2.3.3 The ogive prior

The ogive case is the most general case. The appropriate prior forms are defined, conditional on some unknown i^* $(0 \le i^* \le M+1)$. Note that i^* represents the index of the dose level acting as the turning point, The ogive shape is simply based on the combination of the convex formulation for U_i with $i = 1,\ldots,i^*$, and the concave formulation for U_i with $i = M+1,\ldots,i^*+1$. Observe that the case with $i^* = M+1$ and $i^* = 0$ corresponds to the convex and concave assumptions, respectively. If the first derivative of an ogive function P exists, then the ogive will be unimodal, with a maximum at the x_{i^*}.

For the case $1 \le i^* \le M$ the prior for the u_i's written in terms of successive conditioning are presented below:

$$u_i \mid u_1,\ldots u_{i-1} \sim Beta(\alpha_i, \sum_{j=i+1}^{M+1} \alpha_j; 0, 1 - \sum_{j=1}^{i-1} u_j), \qquad (2.9)$$

where $u_0 = 0$, and $i = 1,\ldots,i^*$ and

$$u_i \mid u_1,\ldots u_{i^*}, u_{M+1},\ldots,u_{i+1} \sim Beta(\alpha_i, \sum_{j=1}^{i-1} \alpha_j - \sum_{j=1}^{i^*} \alpha_j; 0, 1 - (\sum_{j=1}^{i^*} u_j + \sum_{j=i+1}^{M+1} u_j)), \qquad (2.10)$$

for $i = M+1,\ldots,i^*+1$

The exact derivations of the forms *(2.9)* and *(2.10)* can be found in Appendix A. Moreover, after some algebra and manipulations, it follows that the prior induced for the p_i's, conditional on the known i^*, implies the conditional distributions (see Appendix A):

$$p_i \mid i^*, p_1, \ldots p_{i-1} \sim Beta(\alpha_i, \sum_{j=i+1}^{M+1} \alpha_j; \bar{c}_i, \bar{d}_i), \qquad (2.11)$$

for $i = 1, \ldots, i^*$, with $\bar{c}_i$ and $\bar{d}_i$ defined as below

$$\bar{c}_i = p_{i-1} + \frac{x_i - x_{i-1}}{x_{i-1} - x_{i-2}}(p_{i-1} - p_{i-2}), \qquad \bar{d}_i = p_{i-1} + \frac{x_i - x_{i-1}}{x_{M+1} - x_{i-1}}(1 - p_{i-1}),$$

and also the conditional distribution

$$p_i \mid i^*, p_1, \ldots p_{i^*}, p_{M+1}, \ldots, p_{i+1} \sim Beta(\sum_{j=1}^{i} \alpha_j - \sum_{j=1}^{i^*} \alpha_j, \alpha_{i+1}; c_i, d_i), \qquad (2.12)$$

for $i = M, \ldots, i^* + 1$, with c_i and d_i defined as below

$$c_i = p_{i+1} - \frac{x_{i+1} - x_i}{x_{i+1} - x_0}(p_{i+1} - p_0), \qquad d_i = p_{i+1} - \frac{x_{i+1} - x_i}{x_{i+2} - x_{i+1}}(p_{i+2} - p_{i+1}),$$

In case of the unknown i^*, the prior specification is completed by assigning $P(i^* = i) = \pi_i$, where $\pi_i \geq 0$ for $i = 0, \ldots, M+1$ and $\pi_0 + \ldots + \pi_{M+1} = 1$.

2.4 Ordered potency curves

Gelfand and Kuo (1991) present a useful extension of our studies about *DP* priors which treats the case of ordered polytomous response. The idea is based on the nested events. Let us present the illustrative example from Gelfand and Kuo (1991). Suppose that event *A* is identified by 'patient died' while the event *C* means that 'patient's condition has been worsened'. Note that $A \subset C$. Now suppose that at some dose level x_i we observe n_i subjects. During this experiment event *A* occurred Z_i times whereas event *C*, Y_i times. Since the events are nested we have implies that $Z_i \le Y_i$. Gelfand and Kuo proposed to model this situation with two underlying potency curves $P_A(x)$ and $P_C(x)$ which are stochastically ordered, namely $P_A(x) \le P_C(x)$. Moreover they assume that the joint distribution of Z_i and Y_i is specified via

$$(Z_i, Y_i - Z_i) \sim Mult(n_i; p_i, q_i - p_i, 1 - q_i), \qquad (2.13)$$

where $p_i = P_A(x_i)$ and $q_i = P_C(x_i)$.

Hence, with $p = (p_1, \ldots, p_M)$ and $q = (q_1, \ldots, q_M)$ the likelihood at $Z = z, Y = y$ is

$$\prod_{i=1}^{M} \frac{n_i!}{z_i!(y_i - z_i)!(n_i - y_i)!} p_i^{z_i} (q_i - p_i)^{y_i - z_i} (1 - q_i)^{n_i - y_i}, \qquad (2.14)$$

Next, they consider two families of priors: Dirichlet process priors and Beta product priors (described in more details in section 2.2.1 and 2.2.2). We assume that the prior specifications are restricted to the set:

$$T^M = \{(p,q) : 0 \le p_1 \le \ldots \le p_M \le 1, \;\; 0 \le q_1 \le \ldots \le q_M \le 1, \;\; p_i \le q_i \text{ for all } i\}.$$

π_D the Dirichlet process prior is replaced by a product of Dirichlet process prior with stochastic order:

$$\pi_D(p,q) = c(\gamma,\eta)\prod_{i=1}^{M+1}\Delta_i^{\gamma_i-1}\varepsilon_i^{\eta_i-1}, \qquad (2.15)$$

where $\Delta_i = p_i - p_{i-1}$, $\varepsilon_i = q_i - q_{i-1}$, $\gamma_i = M_A\{P_{A,0}(x_i) - P_{A,0}(x_{i-1})\}$,

$\eta_i = M_C\{P_{C,0}(x_i) - P_{C,0}(x_{i-1})\}$,

with base distributions $P_{A,0}$, $P_{C,0}$ and precision parameters M_A , M_C and the standardizing constant $c(\gamma,\eta)$.

Whereas π_B is extended to produce a form over T^M which is conjugate with *(2.14)*

$$\pi_B(p,q) = c(\alpha,\beta,\delta)\prod_{i=1}^{M} p_i^{\alpha_i-1}(q_i - p_i)^{\beta_i-1}(1-q_i)^{\delta_i-1}, \qquad (2.16)$$

where $c(\alpha,\beta,\delta)$ is the standardizing constant.

The parameters α,β *and* δ can be selected by using the conditional modes of p_i and $q_i - p_i$. Assuming $\alpha_i + \beta_i + \delta_i = M_i$, where M_i is specified and $\alpha_i, \beta_i, \delta_i > 1$ we obtain

$$\alpha_i = (M_i - 3)P_{A,0}(x_i),\ \ \beta_i = (M_i - 3)(P_{C,0}(x_i) - P_{A,0}(x_i)),\ \ \delta_i = M_i - \alpha_i - \beta_i.$$

Chapter 3

Bayesian inference

This section concerns the main problem of our study. As we have pointed out at the beginning, one of the most important problems of the biologist is to predict the tolerance curve, in order to explain the probability of a positive response (death, tumor, etc.) as a function of the concentration of the toxic substance. Bayes' theorem, one of the most powerful tools, provides the mechanism to combine the information in the data with the prior to produce an updated probability distribution. Recall that in our study, data are provided from the biological experiments whereas prior is assumed to be Dirichlet process or product beta as an alternative. Unfortunately, in this case Bayesian approach has suffered from the difficulties of analytical, intractable form of the posterior. All of these problems were considered by numerous authors. In this work we propose two methodologies; MLE-type methods and Markov Chain Monte Carlo techniques, in order to estimate the unknown potency curve P.

3.1 Form of the posterior distribution

Bayesian methods propose an optimal way to make consistent decisions in the face of uncertainty. The reason behind this is that Bayesian statistics seek to optimally combine information from two sources: the information that we have or believe at the start of the research (in our case Dirichlet Process prior) and the information in the observed data. Bayes' theorem provides the mechanism to combine these sources of information in order to improve our statistical inference. According to the Bayesian rule (see appendix B), multiplication of the likelihood *(2.1)* and the prior density (described in chapter 2) gives us the kernel of the posterior density of $p = (p_1, \ldots, p_M)$. For example; if we assign the Dirichlet process as a prior then the posterior takes the form

$$f(p_1, \ldots, p_M \mid s) = C\{\prod_{i=1}^{M} p_i^{s_i}(1-p_i)^{n_i - s_i}\}\prod_{i=1}^{M+1}(p_i - p_{i-1})^{\alpha\xi_i - 1}, \qquad (3.1)$$

where C is a normalizing constant .

An interesting study about the nature of a posterior based on the class of Dirichlet process as a prior was provided by Antoniak (1974). He noticed that the exact form of *(3.1)* has an extremely complicated support region, due to inequality constraints on the $p_i's$ induced by the prior ($0 \le p_1 \le \ldots \le p_M \le 1$). Moreover, he showed that the posterior distribution of *P* given data is a mixture of Dirichlet process distributions. Unfortunately this posterior becomes increasingly intractable when the number of stimulus levels increases. To illustrate this problem we follow two simple examples which were presented by Antoniak.

First, let us consider the simplest case where data are available only from one stimulus level. Suppose that *P* is chosen to be a Dirichlet process with parameter α and distribution P_0, and the dosage level *x* is selected in order to administer this dosage to *n* animals. Obviously, the prior distribution of *P(x)* is $Beta(\alpha P_0(x), \alpha(1-P_0(x)))$ and the posterior distribution of *P(x)* given *s* positive responses in n trials is $Beta(\alpha P_0(x) + s, n - s + \alpha(1-P_0(x)))$ (see appendix C). The case with one dosage level does not cause us any problems. However, things become more complicated as the

number of thresholds which we include to our analysis increases. The study with two levels confirms this conjecture. Following Antoniak, assume now that $x_1 < x_2$ and let s_i be the number of successes among the n_i trials at x_i . Then the joint density of $S_1, S_2, P(x_1), P(x_2)$ is easily expressed in terms of $Y_i's$, where $Y_1 = P(x_1)$, $Y_2 = P(x_2) - P(x_1)$, $Y_3 = 1 - P(x_2)$ and $\beta_i's$ where $\beta_1 = \alpha P_0(x_1)$, $\beta_2 = \alpha(P_0(x_2) - P_0(x_1))$ and $\beta_3 = \alpha(1 - P_0(x_2))$. Then the joint density of $S_1, S_2, P(x_1), P(x_2)$ takes the following form

$$\binom{n_1}{s_1}\binom{n_2}{s_2} y_1^{s_1}(1-y_1)^{n_1-s_1}(y_1+y_2)^{s_2}(1-y_1-y_2)^{n_2-s_2}\frac{\Gamma(M)}{\Gamma(\beta_1)\Gamma(\beta_2)\Gamma(\beta_3)} y_1^{\beta_1-1} y_2^{\beta_2-1} y_3^{\beta_3-1}, \qquad (3.2)$$

over the set $S = \{y_1, y_2, y_3 \mid y_1 \geq 0, y_2 \geq 0, y_3 \geq 0\}$

As it was noticed in Antoniak (1974) the form *(3.2)* can be recalculated in order to obtain more meaningful interpretation of its structure. Indeed, he transformed this expression into a mixture of Dirichlet distribution by making the substitutions $1 - y_1 = y_2 + y_3$, and expanding $(y_2 + y_3)^{n_1 - s_1}$ and $(y_1 + y_2)^{s_2}$ using the Binomial formula. This leads finally to an expression for the conditional distribution of Y_1, Y_2 given $S_1 = s_1,\ S_2 = s_2$ as

$$\sum_{i=0}^{s_2}\sum_{j=0}^{n_1-s_1} a_{ij} D(\beta_1 + s_1 + i, \beta_2 + n_1 - s_1 + s_2 - i - j, \beta_3 + n_2 - s_2 + j), \qquad (3.3)$$

where $$a_{ij} = \frac{b_{ij}}{\sum_{i=0}^{s_2}\sum_{j=0}^{n_1-s_1} b_{ij}},$$

and $$b_{ij} = \binom{n_1 - s_1}{j}\binom{s_2}{i}\frac{\Gamma(\beta_1 + s_1 + i)\Gamma(\beta_2 + n_1 - s_1 + s_2 - i - j)\Gamma(\beta_3 + n_2 - s_2 + j)}{\Gamma(\beta_1)\Gamma(\beta_2)\Gamma(\beta_3)}.$$

Summarizing the investigation of Antoniak; he found the form of posterior distribution of the tolerance curve (which is mixture of Dirichlets) but examples, which are presented above, showed how sensitive this form is to increasing number of dose levels.

3.2 Estimation methods

The previous study showed us that the joint posterior distribution which is based on Dirichlet process prior, is too complicated to allow an analytical solution, especially for obtaining the marginals. This forces us to use some approximating techniques. Therefore, in this part we propose two methodologies; the MLE – type algorithm and the Markov Chain Monte Carlo algorithm (MCMC). The former is deterministic in nature and design to obtain the modes of the posterior distributions in Bayesian contexts. MCMC algorithms aim to approximate the full distribution, which is a more ambitious task than the point estimation in case of MLE algorithm. There is an extensive literature on the subject of the aforementioned techniques. Therefore the next sections provide a detailed discussion.

3.2.1 MLE approach

The primary contributions in the nonparametric Bayesian inference have come from Ramsey. His publication (1972) pioneers the study of estimating the potency curve based on Dirichlet process prior. In this part we illustrate the method of estimating potency curve, which consists of maximizing the likelihood. Note that in order to adapt this procedure to the Bayesian situation, the prior distribution is treated as 'prior' observations. Therefore, we maximize the joint posterior density given by *(3.1)* and as a result the joint mode $\left(\widehat{P}(x_1),\ldots,\widehat{P}(x_M)\right)$ (the most probable value) is calculated. Ramsey suggests to use this joint mode to summarize the posterior distribution. Add that, the joint mode $\left(\widehat{P}(x_1),\ldots,\widehat{P}(x_M)\right)$ is understood as the *k*-dimensional point which for any set of doses

$x_1,\ldots,x_M$ maximizes the joint posterior density of $\left(P(x_1),\ldots,P(x_M)\right)$. Therefore, our problem of estimating the potency curve is reduced to the problem of finding the mode function. Ramsey stressed that different cases with respect to different values of precision parameter α should be considered. The reason for that are different ways of obtaining modes. Before we present each of these cases it is important to note that prior which is used by Ramsey differs from the prior expressed by *(2.2)*. Indeed, he assumes that the successive differences in potency have a Dirichlet distribution with density function:

$$f(p_1,\ldots,p_M) \propto \{\prod_{i=1}^{M+1}(p_i - p_{i-1})^{\xi_i}\}^{\alpha}, \qquad (3.4)$$

for $0 \le p_1 \le \ldots \le p_M \le 1$.

Below we present three different ways of obtaining modes with respect of different values of α.

- If $\alpha \to \infty$ the posterior distribution is dominated by the prior distribution and the prior distribution approaches a degenerate distribution giving probability one to the prior mode;

- If $\alpha = 0$ the mode of the joint density is the isotonic regression estimator;
  Ramsey referred here to Ayer (1955) and suggested that the solution may be written as follows

$$\hat{P}(x) = 0, \qquad \textit{for } x < x_1$$

$$\hat{P}(x) = \min_{i \le s \le M} \max_{1 \le r \le i} \frac{\sum_{i=r}^{s} s_i}{\sum_{i=r}^{s} n_i}, \qquad \textit{for } x_i \le x < x_{i+1} \quad (\text{where } i = 1,\ldots,M)$$

  Notice that this modal function is uniquely defined only at the observational doses. The interpolation procedure is arbitrary, subject to the constraint of monotonicity.

- If $0 < \alpha < \infty$ the joint posterior densities are convex and unimodal;
  $\hat{P}(x)$ is an uniquely defined non-decreasing function; such that for any collection of $\{x_1, \ldots, x_M\}$ it is the mode of the joint posterior density.

According to Govindarajulu (1988) "the estimates based on the mode of the posterior are easier to compute than for example those based on the quadratic loss and seem to give estimates very close to those obtained from the mean of the posterior".

Let us now present the method of estimating the mode function for $0 < \alpha < \infty$ proposed by Ramsey. The posterior density based on the prior *(3.4)* is expressed as:

$$f(p_1, \ldots, p_M) = C\{\prod_{i=1}^{M} p_i^{s_i}(1-p_i)^{n_i - s_i}\}\{\prod_{i=1}^{M+1}(p_i - p_{i-1})^{\xi_i}\}^{\alpha}, \qquad (3.5)$$

Observe that the constant C, does not affect the location of the mode and therefore it plays no role in the optimization schema and it can be omitted. Note also that the logarithm of the joint posterior *(3.5)* is proportional to the expression below

$$\ln\{f(p_1, \ldots, p_M)\} \propto \sum_{i=1}^{M} s_i \ln(p_i) + \sum_{i=1}^{M}(n_i - s_i)\ln(1-p_i) + \alpha \sum_{i=1}^{M+1} \xi_i \ln(p_i - p_{i-1}), \qquad (3.6)$$

Now, we can reparametrize the posterior density *(3.5)* by setting $\Theta_i = p_i - p_{i-1}$ into *(3.6)* for each $i = 1, \ldots, M+1$ with the assumption; $p_0 = 0$ and $p_{M+1} = 1$. Observe that p_i can be expressed by terms of Θ_j 's as $p_i = \sum_{j=1}^{i} \Theta_j$ and the Θ_i 's are a subject to the constraints $\sum_{i=1}^{M} \Theta_i = 1$. Now, we maximize the natural logarithm of the posterior density, subject to the constraint by using the Lagrange method. Let us consider the following set of equations:

$$\begin{cases} \ln(f(\Theta)) \propto \{\sum_{i=1}^{M} s_i \ln(\sum_{j=1}^{i} \Theta_i)\} + \{\sum_{i=1}^{M} (n_i - s_i)\ln(1-\sum_{j=1}^{i}\Theta_j)\} + \{\alpha \sum_{i=1}^{M+1} \xi_i \ln(\Theta_i)\}, \\ \Theta_1 + \ldots + \Theta_M = 1, \\ \sum_{i=1}^{M} \Theta_i - 1 = g(\Theta), \\ \Phi(\Theta) = \ln(f(\Theta)) + \lambda g(\Theta), \end{cases}$$

The partial derivatives with respect to Θ_k and Θ_{k+1} yield

$$\frac{\partial \Phi(\Theta)}{\partial \Theta_k} = \sum_{j=k}^{M} s_j \frac{1}{\sum_{m=1}^{j} \Theta_m} - \sum_{j=k}^{M} (n_j - s_j) \frac{1}{1-\sum_{m=1}^{j}\Theta_m} + \frac{\alpha \xi_k}{\Theta_k} + \lambda = 0, \qquad \textit{(3.7a)}$$

and

$$\frac{\partial \Phi(\Theta)}{\partial \Theta_{k+1}} = \sum_{j=k+1}^{M} s_j \frac{1}{\sum_{m=1}^{j} \Theta_m} - \sum_{j=k+1}^{M} (n_j - s_j) \frac{1}{1-\sum_{m=1}^{j}\Theta_m} + \frac{\alpha \xi_{k+1}}{\Theta_{k+1}} + \lambda = 0, \qquad \textit{(3.8a)}$$

Setting $p_i = \sum_{j=1}^{i} \Theta_j$ into *(3.7a) and (3.8a)* we obtain

$$\sum_{j=k}^{M} s_j \frac{1}{p_j} - \sum_{j=k}^{M} (n_j - s_j) \frac{1}{1-p_j} + \frac{\alpha \xi_k}{\Theta_k} + \lambda = 0, \qquad \textit{(3.7b)}$$

$$\sum_{j=k+1}^{M} s_j \frac{1}{p_j} - \sum_{j=k+1}^{M} (n_j - s_j) \frac{1}{1-p_j} + \frac{\alpha \xi_{k+1}}{\Theta_{k+1}} + \lambda = 0, \qquad \textit{(3.8b)}$$

As a result, the formulas *(3.7b)* and *(3.8b)* give us the relation between the corresponding potencies, which maximizes the joint posterior density *(3.5)*.

$$\frac{n_k}{\hat{p}_k(1-\hat{p}_k)}\left[\frac{s_k}{n_k} - \hat{p}_k\right] = \alpha\left[\frac{\xi_{k+1}}{\hat{p}_{k+1} - \hat{p}_k} - \frac{\xi_k}{\hat{p}_k - \hat{p}_{k-1}}\right], \qquad \textit{(3.9)}$$

where for each $i = 1, \ldots, M$, $\hat{p}_k = \hat{P}(x_k)$, with the assumptions that $\hat{p}_0 = 0$ and $\hat{p}_{M+1} = 1$. Note that in order to determine $\hat{P}(x)$, we need to solve *M* simultaneous equations *(3.9)* for the $\hat{p}_k$. A way to accomplish this is for example Newton – Raphson method.

Moreover, Ramsey proposes also estimation of the tolerance curve at the non-observationed doses. Notice that, if we take into consideration the fact that for some *x* such that $x_i < x < x_{i+1}$ the prior is still defined (by replacing the interval $(x_i; x_{i+1})$ by the two intervals $(x_i; x)$ and $(x; x_{i+1})$) and that the likelihood does not have the term associated with *x*; we can still derive a formula similar to *(3.9)*. The relation *(3.10)* can be thought of as the way of interpolation and extrapolation.

$$\frac{\xi^*}{\hat{p} - \hat{p}_i} - \frac{\xi^{**}}{\hat{p}_{i+1} - \hat{p}} = 0 \qquad (3.10)$$

where $\hat{p}$ is the mode at the non-observationed dose *x* and $\xi^* = P_0(x) - P_0(x_i)$, $\xi^{**} = P_0(x_{i+1}) - P_0(x)$. After some calculations relation *(3.10)* takes the following form

$$\hat{p} = \frac{\xi^* \hat{p}_{i+1} + \xi^{**} \hat{p}_i}{\xi^* + \xi^{**}}, \qquad (3.11)$$

which describes the way for calculating the mode at the non-observed dose *x*. Notice that formula *(3.9)* and *(3.11)* enables us to obtain only a point estimate for tolerance curve. Further studies (section 3.2.2) show us more advanced investigations.

Since we have already known one of the techniques to estimate the potency curve we can extend the discussion from section 2.2.1 about the strategy of choosing appropriate parameters which are required to specify a DP prior. Ramsey (1972) and Disch (1981) consider the following investigation. The main idea of both authors is based on the so called a 'Bayes two step' rule. Let us first describe the methodology presented by Ramsey. Suppose that previous experiment was run with n_0 observations at each dose, and that no prior information was available $(\alpha = 0)$. The following question arises: what prior should be assessed for a second experiment. Ramsey's answer for this question is formalized for our purpose as a *rule1*.

Rule 1. If a previous experiment based on n_0 observations per dose gave a posterior mode of $P_0(x)$, then adopt $P_0(x)$ as the current prior mode and use $\alpha = n_0$.

To illustrate this rule we present example 1 taken from Ramsey (1972).

Example 1. The results from the previous experiment are placed in the table below:

doses	1	2	3	4	5
affected	5	6	7	8	9
Sample size	10	10	10	10	10
Posterior mode	0.5	0.6	0.7	0.8	0.9

Table 3.1. The results from the previous experiment with $\alpha = 0$

In the current experiment the same doses are used and 50 observations at each dose produce the following results:

doses	1	2	3	4	5
affected	5	10	15	20	25
Sample size	50	50	50	50	50

Table 3.2. The results from the current experiment

Ramsey proposed two possible ways to proceed. The first one is based on pooling data set and the next is related to Bayes two steps rule. Assume first that the data sets are pooled and the posterior mode for $\alpha = 0$ is determined. The results are placed in the table below.

doses	1	2	3	4	5
affected	10	16	22	28	34
Sample size	60	60	60	60	60

Posterior mode	0.1667	0.2667	0.3667	0.4667	0.5667

Table 3.3. The results from the pooled data set

An alternative way is provided by the *rule 1*. We adopt the posterior mode from the previous experiment as the prior mode for use with the second set with assumption that $\alpha = n_0$ in order to reflect the strength belief in prior information. The next table presents results which are obtained from this methodology. Additionally we compare posteriors for other choices of α .

doses		1	2	3	4	5
Prior mode		0.5	0.6	0.7	0.8	0.9
Posterior modes	$\alpha = 0$	0.1	0.2	0.3	0.4	0.5
	$\alpha = 10$	**0.1474**	**0.2138**	**0.3051**	**0.4077**	**0.5301**
	$\alpha = 20$	0.1721	0.2331	0.3180	0.4209	0.5529
	$\alpha = 30$	0.1899	0.2501	0.3325	0.4356	0.5627
	$\alpha = 40$	0.2044	0.2652	0.3469	0.4503	0.5906
	$\alpha = 50$	0.2170	0.2788	0.3605	0.4645	0.6069

Table 3.4. The results obtained by using *Rule1*

Ramsey pointed out that the prior distribution is not a 'conjugate' prior and it is not, therefore, equivalent to a posterior distribution from a previous experiment. Thus the procedure of rule 1 leads to the conclusion that the parameters of the prior distribution may result in loss of information from the first experiment. Another problem, which was noticed by Disch is that the choice of α in *rule 1* is too conservative in its use of the prior information. Since precision parameter can be regarded as the assessor's prior strength of belief measured in number of pieces of data at a particular dose level, Dish suggests making a modification to this rule. We call it *rule 2*.

Rule 2. Approximate the posterior of the p_i by a Dirichlet distribution with parameters $\xi = (\xi_1, \ldots, \xi_{M+1})$, calculated from the posterior mode and

$\alpha = prior\alpha + \min_i(D_i)$, where $D_i = n_i +$ number of successes at the dose levels less than $x_i +$ the number of failures at dose levels greater than x_i.

The next two examples, proposed by Ramsey, illustrate the advantage of having a smoothing procedure. Ramsey provides interesting experiments of estimating the potency curve with respect to different values of precision parameter. The results and conclusions are presented below.

Example 2. Suppose that the actual effective curve is

$$P(x) = \Phi(\frac{x-0.6}{0.25}),$$

Where Φ is the standard cumulative normal. Moreover, we selected for the experiment *40* equally spaced doses as follows

$$x_i = 0.025i \text{ for } i = 1,\ldots,40.$$

Where at each dose, one subject is put on trial. The data consist of 40 independent observations $Y_i = 1$ *or* 0 according to whether a randomly chosen observation from a *Uniform*(0,1) did not or did exceed $P(x_i)$. The prior mode is assumed as below

$$P_0(x) = \begin{cases} 0, & \text{if } x < 0, \\ 2x, & \text{if } 0 \le x \le 0.025, \\ 0.05 + 1.6(x - 0.025), & \text{if } 0.025 \le x \le 0.075, \\ 0.13 + 1.2(x - 0.075), & \text{if } 0.075 \le x \le 0.125, \\ 0.19 + 0.8(x - 0.125), & \text{if } 0.125 \le x \le 0.9, \\ 0.81 + 1.2(x - 0.9), & \text{if } 0.9 \le x \le 0.950, \\ 0.87 + 1.6(x - 0.95), & \text{if } 0.950 \le x \le 1.000, \\ 0.95 + 2(x - 1), & \text{if } 1 \le x \le 1.025, \\ 1, & \text{if } 1.025 \le x, \end{cases}$$

The prior mode and the actual curve obtained by Ramsey are displayed in the figure 3.1 together with posterior with different choices of α. The effect of smoothing with the Bayesian approach is summarized in the table 3.5.

α	0	1	5	10	$+\infty$
K-S distance	0.325	0.274	0.184	0.159	0.220
$\lvert \hat{ED}50 - ED50 \rvert$	0.100	0.095	0.055	0.010	0.0875

Table 3.5. The effect of smoothing with the Bayesian approach

Note that *K-S distance* relates to Kolmogorov-Smirnov distance and it is based on the maximum distance between two curves. Whereas '*ED*50' is understand as a dose level *x* for which $P(x) = 0.5$ (more about effective dose reader can find in chapter 4). As we can observe from the table above, some degree of smoothing the isotonic regression improves both the K-S distance and the error in the ED50 estimate. Moreover, Ramsey showed also that advantages of smoothing procedure are relatively insensitive to the choice of prior parameters. This property is illustrated by the next example.

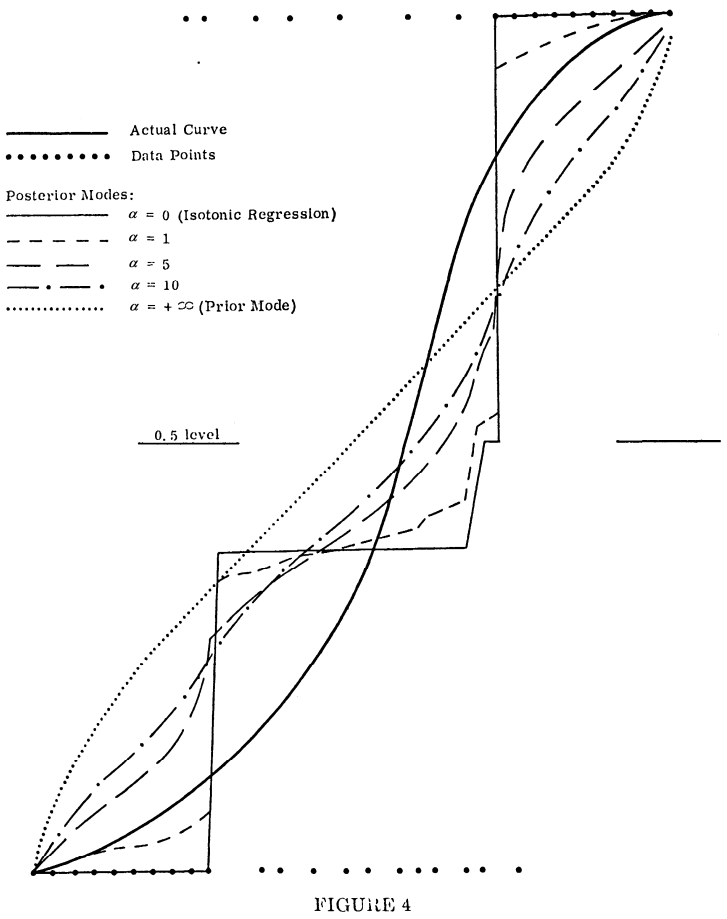

Figure 3.1. Smoothing procedure.

Example 3. Let us repeat the example 2 but with the worst prior such as it is presented below

$$P_0(x) = \begin{cases} 0, & \text{if } x < 0, \\ 0.4x, & \text{if } 0 \le x \le 0.025, \\ 1, & \text{if } x \ge 1,, \end{cases}$$

Figure 3.1 illustrates the results obtained from this experiment. We summarize them also in table 3.6.

α	0	5	$+\infty$
K-S distance	0.325	0.240	0.600

$\lvert \hat{ED}50 - ED50 \rvert$	0.100	0.090	1.250

Table 3.6. The effects of smoothing with the Bayesian approach

From table above we can draw the conclusion that an improvement in both K-S distance and ED50 estimation error is still observed.

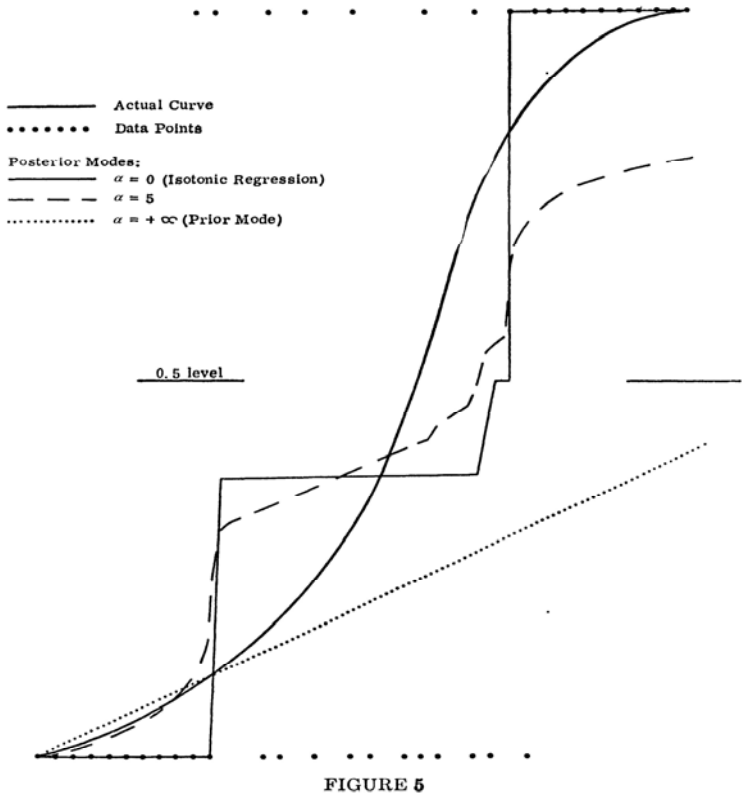

Figure 3.2. Smoothing procedure

3.2.2 Markov Chain Monte Carlo techniques

An important computational issue in many statistical problems is the calculation of the marginal distributions. The previous section showed us that the joint posterior distribution is too complicated to allow an analytical solution, especially for obtaining the marginals. Derivations of the exact marginal posteriors, based on the ordered Dirichlet prior were presented by Disch (1981). He noticed that the integrations needed to calculate

marginals are extremely difficult to perform, either analytically or numerically. The goal of this part is to introduce the Markov Chain Monte Carlo technique which allows estimation of the complex models that are difficult to estimate with classical methods. In particular, Gibbs sampling, the most commonly used MCMC method, can be a useful tool in bioassay analysis for the reconstruction of the marginal posterior densities. For an extensive discussion of the MCMC approach we refer also to Castella and George (1992). Gibbs sampling provides a method for calculating the marginals without having their exact form. Through the use of this technique, we are able to avoid difficult calculations, replacing them instead with a sequence of easier calculations. The main idea of this methodology is based on generating a large sample of independent draws from the full conditional distributions, which are in many cases well known. The procedure of a sampler proceeds as follows

- Specify the initial values $p^{(0)} = (p_1, \ldots, p_M)$,
- Generate draws $p^{(1)} = (p_1, \ldots, p_M)$ from the conditional distributions:

  - $p_1^{(1)}$ from $\pi(p_1 \mid p_2^{(0)}, \ldots, p_M^{(0)})$,
  - $\vdots$
  - $p_i^{(1)}$ from $\pi(p_i \mid p_1^{(1)}, \ldots, p_{i-1}^{(1)}, p_{i+1}^{(0)}, \ldots, p_M^{(0)})$,
  - $\vdots$
  - $p_M^{(1)}$ from $\pi(p_M \mid p_1^{(1)}, \ldots, p_{M-1}^{(1)})$,

As we can observe from the above schema, we generate a sample from the marginals by sampling instead from the conditional distributions. A result of a single iteration is a vector $p = (p_1, \ldots, p_M)$ which represents a transition from initial values $p^{(0)} = (p_1, \ldots, p_M)$ to $p^{(1)} = (p_1, \ldots, p_M)$. If this iteration is repeated r times, the Gibbs sequence $p^{(0)}, p^{(1)}, \ldots, p^{(k)}$ is generated. It turns out that under reasonably general conditions, the distribution of $p_i^{(r)}$ converges to $\pi(p_i)$ (the marginal of p_i) as $r \to \infty$. Thus, for k large enough, the final result $p^{(r)} = (p_1, \ldots, p_M)$, is effectively a sample from marginals. The

convergence of Gibbs sequence can be exploited in different ways to obtain an approximate sample from the marginals. An interesting idea was provided by Gelfand and Kuo (1991). They conducted the Gibbs sampling with *v* independent parallel replications each taken to *r* iterations. Whereas the choice of *v* determines how close our density estimate is to the exact density at the *r*th iteration. Choice of *r* determines how close the latter density is to the actual marginal posterior density.

Our further study provides detailed explanation of sampling approach for ordered Dirichlet prior. We will follow the investigation of Gelfand and Kuo (1991) who proposed Gibbs sampling as a very successful method in estimating the tolerance distribution in a quantal bioassay. After it we will extend also our study about this technique to the case of Dirichlet prior with shape constraint. Note that comprehensive discussion which refers to this subject can be found in Ramgopal, Laud, Smith (1993).

Ordered Dirichlet prior

We begin our investigation of Gibbs sampling procedure with estimation of a potency curve based on the Dirichlet prior. Since the Dirichlet distribution is conjugate family for the multinomial distribution, the posterior distribution is again a Dirichlet distribution. This attractive property enables us to use standard sampling method in order to estimate the potency curve.

As we have already noticed, in order to use Gibbs sampling, we need to first specify the complete conditional distributions. Thus, let us define:

- $p \mid S, Z_1, \ldots, Z_M, n$
- $Z_i \mid S, p, Z_j (i \neq j)$.

To specify the second aforementioned conditional distribution, let us introduce the set of unobserved multinomial variables in order to simplify a sampling procedure.

Let

$$Z_i = (Z_{i1},\ldots,Z_{ij},\ldots,Z_{i,M+1}) \sim Mult(n_i,\lambda), \qquad \text{for } i = 1,\ldots,M$$

where $\lambda = (\lambda_1,\ldots,\lambda_j,\ldots,\lambda_{M+1})$ with $\lambda_j = p_j - p_{j-1}$, $p_0 \equiv 0$ and $p_{M+1} \equiv 1$. The variable Z_{ij} denotes, amongst the n_i individuals receiving dosage level x_i, the unobserved number who would have responded to dosage level x_j but not to dosage level x_{j-1}. Notice that Z_i is a concatenation of two multinomials $Z_i = \{Z_i(1), Z_i(2)\}$ such that:

$$Z_i(1) = (Z_{i1},\ldots,Z_{ii}) \text{ and } Z_i(2) = (Z_{i,i+1},\ldots,Z_{i,M+1}),$$

where $Z_i(1) \sim Mult\{s_i, p_i^{-1}\lambda(1)\}$ and $Z_i(2) \sim Mult\{n_i - s_i, (1-p_i)^{-1}\lambda(2)\}$,

with $\lambda(1) = (\lambda_1,\ldots,\lambda_i)$ and $\lambda(2) = (\lambda_{i+1},\ldots,\lambda_{M+1})$.

Whereas the former conditional distribution is an ordered Dirichlet updating *(2.2)*

$$\frac{\Gamma(\sum_{j=1}^{k+1}\overline{\xi}_j)}{\prod_{j=1}^{k+1}\Gamma\left(\overline{\xi}_j\right)}\prod_{j=1}^{k+1}(p_j - p_{j-1})^{\overline{\xi}_j - 1}, \qquad \textit{(3.12)}$$

where $\overline{\xi}_j = \xi_j + \sum_{i=1}^{M} Z_{ij}$.

Thus, the complete conditional density for p_i, over the set $[p_{i-1}, p_{i+1}]$, denoted by $g_D(p_i \mid S, Z_1,\ldots,Z_M, p_j, j \neq i)$, is $Beta(\overline{\xi}_i, \overline{\xi}_{i+1})$. In other words, $p_i \sim p_{i-1} + (p_{i+1} - p_{i-1})Beta(\overline{\xi}_i, \overline{\xi}_{i+1})$.

Given the aforementioned conditional distributions we can execute the Gibbs sampling as follow:

- specify the initial value $p^{(0)} = (p_1, \ldots, p_M)$,
- for each $i = 1, \ldots, M$, sample from:

$$Z_i^{(0)}(1) \sim Mult\left(s_i, (p_i^{(0)})^{-1}\lambda(1)\right) \text{ with } \lambda(1) = (p_1^{(0)}, p_2^{(0)} - p_1^{(0)}, \ldots, p_i^{(0)} - p_{i-1}^{(0)}),$$

$$Z_i^{(0)}(2) \sim Mult\left(n_i - s_i, (1 - p_i^{(0)})^{-1}\lambda(2)\right) \text{ with } \lambda(2) = (p_{i+1}^{(0)} - p_i^{(0)}, \ldots, 1 - p_M^{(0)}),$$

- for each $i = 1, \ldots, M$, sample from

$$p_i^{(1)} \sim p_{i-1}^{(1)} + (p_{i+1}^{(0)} - p_{i-1}^{(1)})Beta(\overline{\xi}_j, \overline{\xi}_{j+1}),$$

where $\overline{\xi}_j = \xi_j + \sum_{i=1}^{M} Z_{ij}^{(0)}$ and with $p_0 = 0$ and $p_{M+1} = 1$.

Then we sample again from Multinomials (using $p^{(1)} = (p_1^{(1)}, \ldots, p_M^{(1)})$) in order to calculate $p^{(2)} = (p_1^{(2)}, \ldots, p_M^{(2)})$. This schema is repeated *r* times and as a result we obtain $(p^{(r)}, Z_1^{(r)}, \ldots, Z_M^{(r)})$.

Moreover, Gelfand and Kuo (1991) proposed to calculate *v* replications of this procedure (each to the *r*-th iteration, with the same initial value of $p^{(0)}$), such that the results from it: $(p_s^{(r)}, Z_{1s}^{(r)}, \ldots, Z_{Ms}^{(r)})$ for $s = 1, \ldots, v$, can be used to calculate the estimate of the marginal posterior density of each p_i as stated below:

$$\hat{f}_D(p_i \mid S) = v^{-1}\sum_{s=1}^{v} g_D(p_i \mid S, Z_{1s}^{(r)}, \ldots, Z_{Ms}^{(r)}, p_{js}^{(r)}, j \neq i), \qquad (3.13)$$

Similarly the posterior mean of p_i is estimated using the mean of g_D leading to

$$E_D(p_i \mid S) = v^{-1}\sum_{s=1}^{v}[p_{i-1,s}^{(r)} + (p_{i+1,s}^{(r)} - p_{i-1,s}^{(r)})\{\bar{\xi}_{is}/(\bar{\xi}_{is} + \bar{\xi}_{i+1,s})\}], \qquad (3.14)$$

where $\bar{\xi}_{js} = \xi_j + \sum_{i=1}^{M} Z_{ijs}^{(r)}$, for $j = 1,\ldots,M+1$.

Moreover, Gelfand and Kuo (1991) provide also the posterior density estimation for non-observed dose level x^* ($x_i \le x^* \le x_{i+1}$) by including $P(x^*)$ as an additional model parameter. More precisely we revise the prior to include $P(x^*)$ and to take the form $\pi_D(p)h_D(P(x^*) \mid p)$. Where h is naturally the density of $p_i + \Delta^*$, with $\Delta^* \sim (p_{i+1} - p_i)Be(\xi^*, \xi_{i+1} - \xi^*)$ and $\xi^* = M(P_0(x^*) - P_0(x_i))$. Since there are no data at dosage level x^*, the complete conditional distribution for $P(x^*)$ is $h_D(P(x^*) \mid p)$. And therefore the posterior density estimate for $P(x^*)$ is as follow

$$\hat{f}_D(P(x^*)) = v^{-1}\sum_{s=1}^{v} h_D(P(x^*) \mid p_s^{(r)}), \qquad (3.15)$$

Whereas the expected value of the conditional density is expressed by

$$E_D(p^* \mid S) = E_D\{p_i + (p_{i+1} - p_i)\xi^*/\xi_{i+1} \mid S\} = \frac{\xi_{i+1} - \xi^*}{\xi_{i+1}} E_D(p_i \mid S) + \frac{\xi^*}{\xi_{i+1}} E_D(p_{i+1} \mid S), \qquad (3.16)$$

Note that the analysis of data sets in chapter 6 is based on the procedure which was described above.

The ogive prior

In this part we would like to present how the Gibbs sampler technique can be used in order to estimate the posterior in case where the prior is assumed to have ogive shape.

Recall that this shape is a generalization of convex and concave case. Therefore the procedure of sampling also suits to these two shapes.

Notice that in the previous case the property about conjugate family was used, which simplify significantly the computations. Here the Gibbs sampler needs to be combined with one of the sampling – importance – resampling technique. Let us first describe a Gibbs sampling approach to simulating from the joint posterior for p and i^*, where i^* is the turning point from convex shape to concave. Given arbitrary starting values, a long iteration of successive random variate generations from

$$\begin{aligned} &p_i \mid i^*, s, n, p_j,\ \mathrm{j} \neq \mathrm{i}, \\ &i^* \mid p, s, n, \end{aligned} \qquad \textit{(3.17)}$$

results in eventual (p, i^*) realizations which are close to being drawings from the joint posterior. Below we present the conditional forms *(3.17)*.

Let us start from the conditional distribution for the turning point. Observe that $i^* \mid p, s, n$ does not, in fact, depend on s and n. Therefore the distribution of i^*, can be expressed as follow:

$$pr(i^* = i \mid p) \propto \pi_i f_i(p) I(i \mid \{i_0 - 1, i_0\}), \qquad \textit{(3.18)}$$

where

$$i_0 = \min\{i; 1 \le i \le M+1 \text{ and } \frac{p_i - p_{i-1}}{x_i - x_{i-1}} > \frac{p_{i+1} - p_i}{x_{i+1} - x_i}\},$$

$$f_i(p) = \prod_{j=1}^{i} g(p_j \mid \alpha_j, \bar{\xi}_{j+1}, \bar{c}_j, \bar{d}_j) \prod_{j=i+1}^{M} g(p_j \mid \xi_j - \xi_i, \alpha_{j+1}, c_j, d_j),$$

for $i = 1, \ldots, M$, where $g\left(p_j \mid \alpha_j, \bar{\xi}_{j+1}, \bar{c}_j, \bar{d}_j\right)$ and $g\left(p_j \mid \xi_j - \xi_i, \alpha_{j+1}, c_j, d_j\right)$ define the conditional densities *(2.11)* and *(2.12)*. In case of $i = 0$ or $i = M+1$, $f_0(p)$, $f_{M+1}(p)$ are given by the product of the beta densities corresponding to *(2.6)* and *(2.8)*, respectively. The form *(3.20)* specifies a simple discrete distribution over the two points $i_0 - 1$ and i_0,

which are easily seen to be the only possible 'switch points' from convex to concave if, given p, i_0 is the smallest stimulus value at which the 'slope' is subsequently decreasing.

Simulation from $p \mid i^*, s, n$, an M-dimensional joint density is clearly not so directly straightforward. With $\pi(\cdot)$ as generic notation for probability density functions, we calculate the following formulas for:

- $i = 1, \ldots, i^* - 1, (i^* \geq 2)$

$$\pi(p_i \mid i^*, s, n, p_j, j \neq i) \propto p_i^{s_i}(1-p_i)^{n_i - s_i} I_{(\bar{a}_i, \bar{b}_i)}(p_i)\{\prod_{j=i}^{i+2} g(p_j \mid \alpha_j, \bar{\xi}_{j+1}, \bar{c}_j, \bar{d}_j)\}, \qquad (3.19)$$

where

$$\bar{a}_i = \max\{p_{i-1} + \frac{\Delta_i}{\Delta_{i-1}}(p_{i-1} - p_{i-2}), p_{i+1} - \frac{\Delta_{i+1}}{\Delta_{i+2}}(p_{i+2} - p_{i+1})\},$$

$$\bar{b}_i = \Delta_{i+1} p_{i-1} + \Delta_i \frac{p_{i+1}}{\Delta_i + \Delta_{i+1}},$$

$$\Delta_i = x_i - x_{i-1},$$

and $\bar{\xi}_{i+1} = \sum_{j=i+1}^{M+1} \alpha_j$.

- $i = i^* + 1, \ldots, M, (i^* \leq M - 1)$

$$\pi(p_i \mid i^*, s, n, p_j, j \neq i) \propto p_i^{s_i}(1-p_i)^{n_i - s_i} I_{(a_i, b_i)}(p_i)\{\prod_{j=i-2}^{i} g(p_j \mid \xi_j - \xi_{i^*}, \alpha_{j+1}, c_j, d_j)\}, \qquad (3.20)$$

where

$$a_i = \Delta_i p_{i+1} + \Delta_{i+1} \frac{p_{i-1}}{\Delta_i + \Delta_{i+1}},$$

$$b_i = \min\{p_{i+1} - \frac{\Delta_{i+1}}{\Delta_{i+2}}(p_{i+2} - p_{i+1}), p_{i-1} + \frac{\Delta_i}{\Delta_{i-1}}(p_{i-1} - p_{i-2})\};$$

- At i^*

$$\pi(p_{i^*} \mid i^*, s, n, p_j, j \neq i^*) \propto p_{i^*}^{s_{i^*}} (1-p_{i^*})^{n_{i^*}-s_{i^*}} I_{(a^*,b^*)}(p_{i^*})$$
$$\times g(p_{i^*} \mid \alpha_{i^*}, \overline{\xi}_{i^*+1}, \overline{c}_{i^*}, \overline{d}_{i^*}) \{ \prod_{j=i^*+1}^{M} g(p_j \mid \xi_j - \xi_{i^*}, \alpha_{j+1}, c_j, d_j) \}, \qquad (3.21)$$

where

$$a^* = p_{i^*-1} + \frac{\Delta_{i^*}}{\Delta_{i^*-1}}(p_{i^*-1} - p_{i^*-2}), \quad b^* = p_{i^*+1} + \frac{\Delta_{i^*+1}}{\Delta_{i^*+2}}(p_{i^*+2} - p_{i^*+1});$$

A drawing from $p \mid i^*, s, n,$ is then obtained by successively drawing $p_1, \ldots, p_M$ from the forms defined by *(3.19) - (3.21).* It remains only to sample efficiently from each of these one-dimensional truncated forms. As Ramgopal, Laud and Smith noticed a number of methods are available and easily implemented, including classical rejection and ratio – of – uniforms techniques. For details, the authors refer to for example Devroye (1986).

Now given the forms *(3.19) - (3.21)* we can execute simulations. Numerous methods are available and easily implemented, including the classical rejection and the ratio-of-uniforms techniques.

Moreover the studies of Ramgopal, Laud and Smith (1993) involves also inference about $p_* = P(x_*)$ for a specified x_* with $x_i < x_* < x_{i+1}$. The relationship below enables us to compute the potency curve at any non-observed dose level.

$$\pi(p_* \mid s, n) = \sum_{i^*} \int \pi(p_* \mid i^*, p) \pi(i^*, p \mid s, n) dp, \qquad (3.22)$$

Whereas the left – hand side can be approximated by averaging the form $\pi(p_* \mid i^*, p)$ over the random sample of (i^*, p) pairs from the posterior. For any given i^*,

$$\pi(p_* \mid i^*, p) = \frac{\pi(p^* \mid i^*)}{\pi(p \mid i^*)},$$

where $p^* = (p_1, \ldots, p_i, p_*, p_{i+1}, \ldots, p_M)$, so that the required form can be identified from the forms given in parts 2.3.1. - 2.3.3 and analogous forms for the extended p^*.

Chapter 4

Effective Dose

In bioassays, different concentrations of stimuli are applied to experimental animals, and the all – or – none reaction of the animals are then recorded. For example, in pharmacology, the effective action of a drug or vaccines is treated by an animal experiment, where death or other all – or – none reactions of the animals are recorded after exposure to the drug at various levels. The previous chapter is devoted to the approximation of the distribution of this tolerance. However, the literature shows that many researchers in biostatistics are often interested in a particular dose level, at where $100\gamma\%$ (where $0 < \gamma < 1$) of the subjects react. Notice that this dose level (called effective dose) is known in the literature as $ED100\gamma$. $ED100\gamma$ summarizes the potency of the concentration of the stimulus and may subsequently form the basis of comparison between different levels of stimuli. Let us formalize definition of $ED100\gamma$ for our purpose, by the definition which was proposed by Disch (1981).

***Definition**:* For $0 < p < 1$, EDp is the (unique) pth effective dose if $EDp = \inf\{x : p \le P(x)\}$.

To conform to more common notation we will understand for example *ED(0.5)* as *ED50*, etc. Notice also that $p(\cdot)$ denotes the response curve which is strictly monotone increasing, so the functional $p^{-1}(\gamma)$, which is referred to as $ED100\gamma$, is well defined for $0<\gamma<1$. This section is devoted to the estimation of $p^{-1}(\gamma)$. We present here few approaches which are related to this problem.

4.1 Ramsey's approach

As one of the first who provoked the discussion about the nonparametric Bayesian bioassay was Ramsey (1972). In his work he provided also a small discussion about the problem of deriving the dose level which is assigned to some particular reaction of examined subjects. Let us now present his investigation.

Suppose that, given probability γ, we are interested in a dose level x such that the posterior mode at this level is $\hat{P}(x)=\gamma$. Notice that, if for some observational dose x_i, $\hat{P}(x_i)=\gamma$, then the problem is trivial. Otherwise (the case where $\gamma=\hat{P}(x)$ for some non-observed dose level x), we need to apply some interpolation technique. Ramsey proposed the following procedure.

Let us first determine the pair of observational doses (x_i, x_{i+1}), such that the inequality $\hat{P}(x_i)<\gamma<\hat{P}(x_{i+1})$ is satisfied. Note that the potency at the dose x between x_i and x_{i+1} may be included in the prior, and this has no effect on the posterior at the observational doses. According to the chapter 3.2.1, the relations *(3.10)*, stated below:

$$\frac{P_0(x_{i+1})-P_0(x)}{\hat{P}(x_{i+1})-\hat{P}(x)}=\frac{P_0(x)-P_0(x_i)}{\hat{P}(x)-\hat{P}(x_i)},$$

can be used to find the dose level x such that $\gamma = \hat{P}(x)$. Note that if we rewrite this equation as:

$$\frac{P_0(x_{i+1}) - P_0(x)}{P_0(x) - P_0(x_i)} = \frac{\hat{P}(x_{i+1}) - \hat{P}(x)}{\hat{P}(x) - \hat{P}(x_i)},$$

we are able to find the dose level x by following the schema below:

- calculate the ratio $\dfrac{\hat{P}(x_{i+1}) - \hat{P}(x)}{\hat{P}(x) - \hat{P}(x_i)} = \dfrac{a}{A} = \delta$,
- determine $P_0(x)$ on the prior mode such that $\delta = \dfrac{P_0(x_{i+1}) - P_0(x)}{P_0(x) - P_0(x_i)} = \dfrac{b}{B}$.
- finally we are able to find the dose level x such that $P_0(x)$ (see figure 4.1.)

Notice that this method requires the posterior mode $\hat{P}(x)$ to have the same shape as the prior mode $P^*(x)$ piecewise between observational doses.

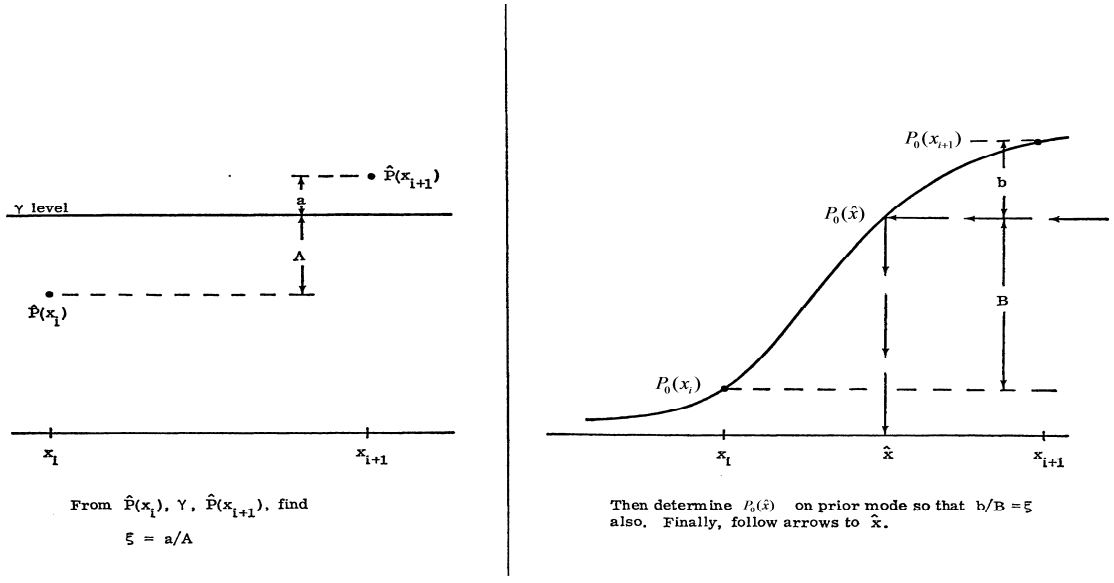

Figure 4.1.Interpolation procedure.

4.2 Disch's approach

Note that Ramsey (1972) described how to estimate any particular effective dose. However, Dish (1981) went further and he has obtained the posterior distribution of an *ED,* rather than simply a point estimate. Thus, the researcher would gain an idea of the precision of his estimate. We provide here the derivation of the prior and posterior distribution of the *ED* follows the investigation of Disch.

Let us first define for a random variable X with cumulative distribution function F_X,

$$F_X(t-) = \lim_{s \to t} F_X(s) = pr(X < t).$$

Recalling that for a fixed dose x, $P(x)$ is random, we have the following theorem

***Theorem**:* For a fixed dose x, if $q = P(x)$, then $pr\{EDp \le x\} = 1 - F_q(p-)$.

We can easily prove this theorem by using the definition of EDp and the monotonicity of P. Indeed, it follows that $EDp \le x$ if and only if $p \le P(x) = q$ and hence $pr\{EDp \le x\} = pr(p \le q) = 1 - F_q(p-)$.

The theorem above implies that the prior distribution for the EDp can be obtained from the prior distribution for $P(x)$. As it is well known, the prior distribution of $q = P(x)$, with fixed x such that $x_k < x \le x_{k+1}$, is $Beta(a,b)$, where $a = \alpha(\sum_{i=1}^{k} \xi_i + \xi^*) = \alpha P^*(x)$ and $b = \alpha(\xi^{**} + \sum_{i=k+2}^{M+1} \xi_i) = \alpha(1 - P^*(x))$, with $\xi^* = P^*(x) - P^*(x_i)$ and $\xi^{**} = P^*(x_{i+1}) - P^*(x)$. Hence the cdf of q is continuous and $F_q(p-) = F_q(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_o^p x^{a-1}(1-x)^{b-1} dx = I_p(a,b)$ Moreover, the prior cdf of the $ED(p)$ is then given by;

$$F_{EDp}(x) = pr\{EDp \le x\} = 1 - I_p(a,b) = I_{1-p}(b,a), \qquad (4.3)$$

Where, the final term in *(4.3)* was obtained by using the property of the incomplete beta function.

Moreover, we can calculate the posterior cdf of EDp by the formula

$$F_{EDp|data}(x) = 1 - \int_0^p f(q)dq, \qquad (4.4)$$

where *f(q)* is the posterior marginal density. However, as it has been pointed out in the previous section, the theoretical inference on the effective dose enjoyed limited success since, in application; the exact form of the posterior marginal becomes unmanageable. Therefore, the problem of estimating the posterior of an effective dose reduces to problem of estimation the posterior marginal of the potency curve.

4.3 Mukhopadhyay's approach

An interesting study of Bayesian nonparametric inference on the unknown dose level for a prespecified response rate was provided by Mukhopadhyay (2000). He developed a closed form for the conditional posterior distribution of the $ED\gamma$ given $p = (p_1, \ldots, p_M)$. Notice that this theorem follows from a general result that, given $s = (s_1, \ldots, s_M)$ and $p = (p_1, \ldots, p_M)$, the conditional distribution of the random segment of *P* between two adjacent dose levels follows a scaled *DP*. For our purpose we present the form of this conditional distribution in the theorem below.

***Theorem**:* The conditional posterior c.d.f. of $ED\gamma$ at x given p and s can be obtained by identifying k $(0 \le k \le M)$ such that $p_k < \gamma \le p_{k+1}$ and the exact form of it is given by

$$P(ED\gamma \le x \mid p,s) = \begin{cases} 1, & \text{if } x > x_{k+1}, \\ \dfrac{\Gamma(\alpha(\xi^* + \xi^{**}))}{\Gamma(\alpha\xi^*)\Gamma(\alpha\xi^{**})} \int_\tau^1 w^{\alpha\xi^* - 1}(1-w)^{\alpha\gamma^{**}-1} dw, & \text{if } x_k < x \le x_{k+1}, \\ 0, & \text{if } x \le x_k, \end{cases} \qquad (4.6)$$

where $\tau = \dfrac{\gamma - p_k}{p_{k+1} - p_k}$, $\xi^* = P_0(x) - P_0(x_k)$ and $\xi^* = P_0(x_{k+1}) - P_0(x)$.

All technical derivations are in the appendix D and reader is encouraged to look at this before proceeding.

Notice that the numerical inference on $ED\gamma$ is obtained in two steps. The first one is to draw a posterior sample $p = (p_1, \ldots, p_M)$ given by *(3.1)* and the second step is a conditional draw from $P(ED\gamma \le x \mid p,s)$ using *(4.6)*.

Chapter 5

Experimental design strategy

The design of experiments is an important part of scientific research. In many cases appropriate techniques are able to reduce trial duration and costs, such that the experiments still provide reliable results. The relevant issues in bioassay design are; choosing the sample size of the experimental subjects, determining the level of the stimuli and allocating the subject to different levels of these stimuli. Unfortunately like many areas of Bayesian statistics, applications to actual experiments still lag behind the theory. However there are few examples of examining an experiment in nonparametric Bayesian design framework. In this section we provide a short discussion about design strategy which was provided by Ramsey (1972).

Example 4. Assume that the prior mode is the standard cumulative normal distribution function. Further assume that the actual potency curve is a shifted standard normal c.d.f.

$$P_0(x) = \Phi(x),$$

$$P_\Theta(x) = \Phi(x - \Theta).$$

Ramsey suggests interpreting Θ as the difference between the actual and prior $ED50's$.

The idea of this example is to illustrate four different experimental designs;

- **Design A** is arranged as follows; six dose levels are spaced in such a way that $\xi_1 = \ldots = \xi_6$, where $\xi_i = P_0(x_i) - P_0(x_{i-1})$ and at each dose one experimental subject is examined;

- **Design B** is arranged as follows; six dose levels are equally spaced and at each dose one experimental subject is examined;

- **Design C** is arranged as follows; three dose levels are spaced in such a way that $\xi_1 = \ldots = \xi_3$, where $\xi_i = P_0(x_i) - P_0(x_{i-1})$ and at each dose two experimental subjects are examined;

- **Design D** is arranged as follows; two dose levels are spaced in such a way that $\xi_1 = \ldots = \xi_3$, where $\xi_i = P_0(x_i) - P_0(x_{i-1})$ and at each dose three experimental subjects are examined;

To illustrate the idea of the aforementioned design strategies we present illustrations below.

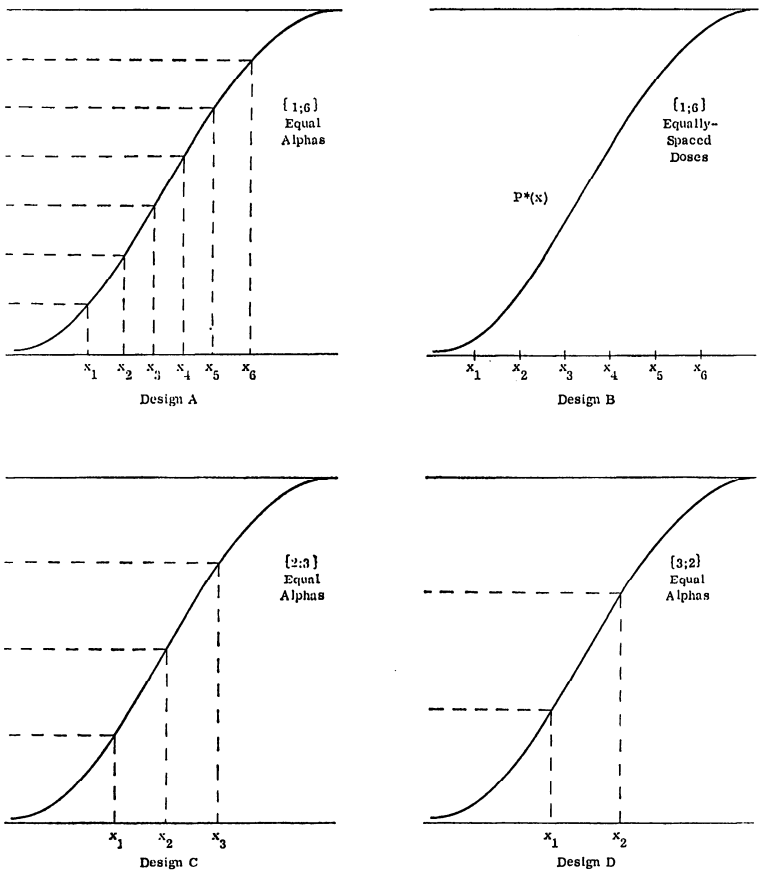

FIGURE 6

DESIGNS USING SIX EXPERIMENTAL SUBJECTS

Figure 5.1. Design strategies of experiment

Ramsey provides here a comparison of the posterior estimate of $ED50's$, for each of the strategies by using the methods which are describe in section 3.2.1. Pictures below present the bias, standard deviation and the square root of the mean square error in the estimators over a range of Θ.

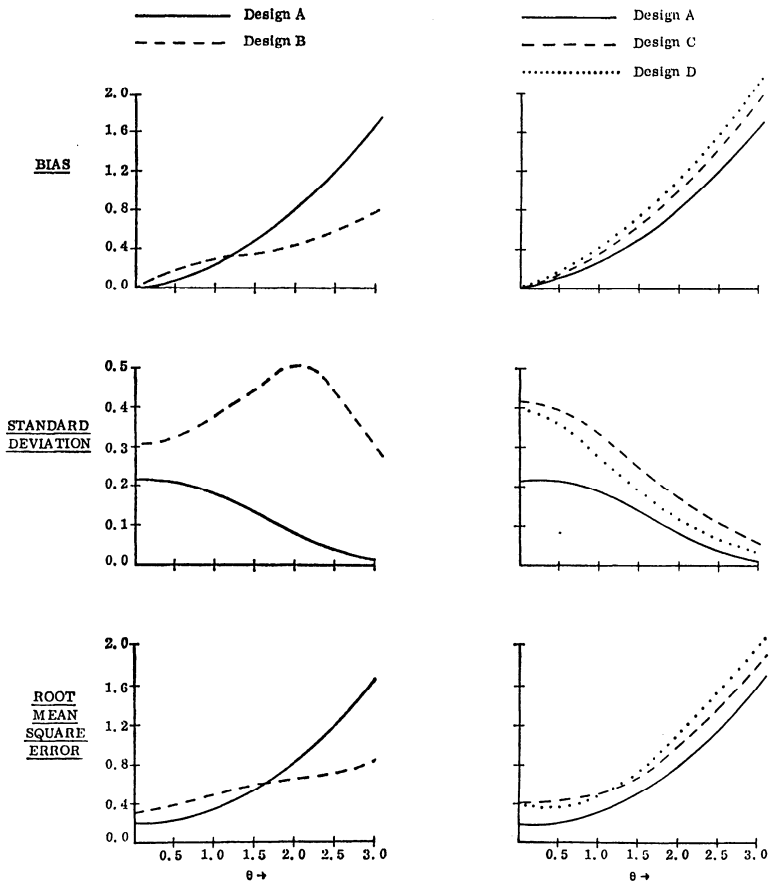

FIGURE 7
PROPERTIES OF ESTIMATORS OF THE ED50 USING DIFFERENT DESIGNS

Figure 5.2.Comparison of different strategies.

From the figure 5.2 we can draw a few conclusions. If we compare designs A and B; both with six dose levels but with different ways of spacing, it can be observed that design A is the superior design if the prior $ED50$ is a good guess of the actual $ED50$. However, if the prior $ED50$ is a poor guess, it is better to spread out the doses as in design B. The results for the comparison of design A, C and D (all using equal ξ_i) also give as very interesting conclusions. As it can be seen from the pictures it seems that the less observations per dose, the better. Recall that design A uses only one observation per dose.

Chapter 6

Applications

Motivation of using the Gibbs sampler procedure proposed by Gelfand and Kuo (1991).

This chapter is addressed to the application of the nonparametric Bayesian approach. Our analysis is based on the ordered Dirichlet process prior described in more details in subsection 2.2.1. The reasons for using this prior are its very attractive and useful properties. Firstly, Dirichlet process prior is a conjugate family for the multinomial distribution. In the Bayesian context it means that the exact form of the posterior is known. A second attractive property relates to the form of the marginal distributions. Recall that if a random probability distribution $P = \left(P(x_1),\ldots,P(x_M)\right)$ is generated by Dirichlet process with precision parameter α and base distribution P_0, then the marginal of $P(x_i)$ for $i = 1,\ldots,M$ is $Beta(\alpha P_0(x_i), \alpha(1 - P_0(x_i)))$. Moreover, the conditional distributions for $P(x_i)$ given $P(x_{i-1})$ and $P(x_{i+1})$ are Beta as well, with parameters $(\alpha(P_0(x_i) - P_0(x_{i-1})), \alpha(P_0(x_{i+1}) - P_0(x_i)))$. The aforementioned properties can be successively used in MCMC methods. In this part the Gibbs sampling technique (one of the MCMC methods) is examined, based on Gelfand and Kuo (1991) where this

approach is used in bioassay context. More details about this methodology are in subsection 3.2.2.

Description of the data

For the purpose of discussion, two bio assay data sets are provided. We present them in the tables below.

Dose	Nr subjects	Nr responses
0	50	1
21	49	15
60	45	20

Table 6.1. Data set 1

Dose	Nr subjects	Nr responses
1000	30	1
1500	30	6
1750	10	7
2000	30	20
2500	20	17
3000	10	10

Table 6.2. Data set 2.

Note that the first one consists only three dose levels at which 50, 49 and 45 experimental subjects were observed. An interesting situation we can notice at zero dose level. The data set provides the information that at this dose level, one (from among 50) experimental subject reacted. Thus, we can suspect that the experimental animals react also on some different stimuli. Recall that Dirichlet prior is based on the assumption that the expected value of response at zero level is zero. However the probability of response at this level is not so high (0.02), therefore it should not be the reason for any additional problems. Notice also that at the last dose level we can observe 20 responses from among 45 experimental subjects. This suggests us that data set 1, does not provide the information about the whole potency curve. We do not know for example what amount of

examined substance is enough to induce the response of all experimental subjects which are under the observation.

Data set 2, presented by the table 6.2, consists of 6 dose levels. Note that at the first level equal to 1000, one response from among 30 experimental subjects were observed, whereas at the last dose level equal to 3000 10 responses from among 10 experimental subjects are observed. Unlike the previous data set, in this case the information about the whole potency curve is provided. However, one worrying situation can be observed. If we look at the data presented in the table 6.2 at the dose levels 1750 and 2000, the following conclusion can be drawn: the probability of response increases as the dose level decreases. To be more precise; at dose level 1750 we observe 7 responses by testing 10 animals (the probability of response is 0.7) whereas at the dose 2000 we observe 20 from among 30 subjects (the probability of response is 0.67). This observation causes conflict with our assumption that the response curve is non-decreasing. Therefore at this dose levels we can suspect some problems. The further analyses of the data verify all of our conjectures.

Stability of the results.

Before we start to execute the Gibbs sampling method, two parameters need to be specified. As we have mentioned in subsection 3.2.2 this sampling procedure is conducted with v parallel replications each taken to r iterations. Note that the choice of v determines how close our density estimates is to the exact density at the *r-th* iteration, whereas the choice of r determines how close the density estimate at *r-th* iteration is to the actual marginal posterior density. Therefore, setting for r and v to obtain smoothly converged estimates is strictly dependent on the application and the prior information about the studied problem. In order to find suitable parameters for our case, we perform the Gibbs sampler for an extensive experimental range of the iteration-replication combinations.

Three values of parallel replications (1000, 1250, and 1500) together with four values of iterations (20, 40, 60, and 80) were used in order to find the stable estimates. The results of these experiments (posterior mean computed by the formula *(3.14)*) were stored in the tables. As an example, we present below one of the experiments for the 2nd data set with two choices of precision parameter: $\alpha = 0.1$, $\alpha = 10$ and the base distribution $P_o(x) = \frac{x^3}{3000^3}$.

Iterations and α	***replications***	***Dose levels***					
$v = 1000$ $\alpha = 0.1$	*r*=20	0.044128	0.193313	0.417062	0.659130	0.846099	1.0000
	r=40	0.042576	0.219189	0.493654	0.615683	0.794841	1.0000
	r =60	0.058682	0.19910	0.545650	0.689910	0.806278	1.0000
	r =80	0.062037	0.192590	0.567986	0.696272	0.797298	1.0000
$v = 1000$ $\alpha = 10$	*r* =20	0.04637	0.191870	0.417249	0.658330	0.848490	1.0000
	r =40	0.041967	0.216300	0.492210	0.615607	0.796974	1.0000
	r =60	0.044247	0.217422	0.506223	0.616470	0.801920	1.0000
	r =80	0.044567	0.226121	0.509834	0.614833	0.798842	1.0000

Table.6.3 Experiments with different values of replications and iterations for the 2st data set.

Iterations and α	***replications***	***Dose levels***					
$v = 1250$ $\alpha = 0.1$	r =20	0.045540	0.194101	0.419390	0.663003	0.852585	1.0000
	r =40	0.049150	0.204076	0.499200	0.685549	0.829161	1.0000
	r =60	0.058147	0.198961	0.547940	0.694183	0.810467	1.0000
	r =80	0.061973	0.193056	0.564190	0.693343	0.801165	1.0000
$v = 1250$ $\alpha = 10$	r =20	0.043136	0.199385	0.413685	0.604449	0.801443	1.0000
	r =40	0.042713	0.220199	0.488546	0.617270	0.799830	1.0000
	r =60	0.044182	0.221295	0.507928	0.617228	0.799390	1.0000
	r =80	0.042458	0.220024	0.512013	0.617808	0.800769	1.0000

Table.6.4 Experiments with different values of replications and iterations for the 2st data set.

Iterations and α	***replications***	***Dose levels***					
$v = 1500$ $\alpha = 0.1$	r =20	0.043025	0.194562	0.413692	0.657564	0.852350	1.0000
	r =40	0.050290	0.199979	0.504819	0.688928	0.828148	1.0000
	r =60	0.056635	0.195852	0.546117	0.694637	0.815175	1.0000
	r =80	0.061405	0.192079	0.557603	0.69808	0.796617	1.0000
$v = 1500$ $\alpha = 10$	r =20	0.041570	0.198264	0.414907	0.603066	0.800648	1.00000
	r =40	0.042782	0.220508	0.491382	0.617719	0.799482	1.00000
	r =60	0.043374	0.221569	0.507544	0.618003	0.799999	1.00000
	r =80	0.044146	0.219520	0.505901	0.612969	0.798149	1.00000

Table.6.5 Experiments with different values of replications and iterations for the 2st data set.

From the tables above we can draw a few conclusions which are very useful in the process of choosing values for parameters *v* and *r*. If we look more carefully on each of

the tables 6.3-6.5; experiments with the same number of parallel replications but different values of iterations of the Gibbs sampler, we can observe that as the parameter *r* increases, the results become more stable. To be precise, the differences between the results obtained from Gibbs sampler with $r = 20 * i$, $(i = 1,2,3)$ and $r = 20 * (i + 1)$, are smaller as the value of parameter *r* increases. This can be easily seen for example in the first table, where the differences between the results obtained by using $r = 60$ and $r = 80$ are 10 times smaller than the differences between the results with $r = 40$ and $r = 60$. Moreover, we observe that these differences become even smaller as the number of parallel replications increases. Indeed, in case of the results obtained by using $v = 1500$ and $r = 80$ they are smaller or equal to 10^{-3}, whereas for $v = 1250$ and $v = 1000$ ($r = 80$) the differences are of the accuracy 10^{-2}. This suggests that stable estimate for the posterior mean could be obtained by using 1500 replications and 80 iterations of the Gibbs sampler.

For comparison, Gelfand and Kuo (1991) use the Gibbs sampler with total of $r = 20$ iterations and $v = 1000$ replications, whereas Ramgopal, Laud and Smith (1993) use $r = 25$ iterations and $v = 1500$ replications. Recall that the values of these parameters are dependent on the application and the prior information, available before the experiments. Indeed, the reason for such small numbers of iterations and replications in the aforementioned examples is connected with very good prior information. The base distribution used in these examples reflects the data quite well.

Analysis of the data.

As mentioned in subsection 2.2.1, to specify the Dirichlet process prior it is required to choose the precision parameter α and the base distribution P_0. A brief interpretation of α and P_0 is provided after the introduction of DP. Now, let us complete it by the experiments based on the Gibbs sampler technique. In order to examine their impact on the results, we perform the analysis of both data sets, each with two base distributions. Note that choice of these distributions should not be influenced by the available data which in fact we wish to analyze. Recall that the prior information should base, for

example on the prior experiments or the knowledge of the scientists of the problem. In our investigation we assume the following prior shape distributions:

- 1st data set: $P_0(x) = \frac{x}{200}$, $P_0(x) = \left(P_0(x_1), P_0(x_2), P_0(x_3)\right) = \left(\frac{1}{4}, \frac{2}{4}, \frac{3}{4}\right)$;
- 2nd data set: $P_0(x) = \frac{x^3}{3000^3}$, $P_0(x) = \frac{x}{3000}$;

Note that our choice of these distributions was motivated only by the assumption that these functions are non-decreasing and by the assumption that at dose level zero the potency is equal to zero. In order to compare how far the prior information is from the data, in the table below we present the values of the base distributions at the observational dose levels together with the maximum likelihood $\left(\frac{s_i}{n_i}\right)$. Moreover we examine also the impact of the precision parameter α. We perform the experiments with $\alpha = 0.1$, $\alpha = 10$, $\alpha = 50$, $\alpha = 100$ and the results are summarize in the tables. To make our analysis simpler and clearer we illustrate graphically these results by plots of the prior mean together with the posterior mean obtained by using different values of precision parameter. Note that the Gibbs sampling procedure is provided with 1500 replications and 80 iterations. The choice of these values is motivated by the numerous tests of iteration-replication combinations, from which it follows that the results are stable.

Analysis of data set 1

As a first we examine data set 1 with the following base distributions:

- $P_o(x) = \frac{x}{200}$;

	Dose levels		
MLE	0.02000	0.30612	0.44444
P_o	0	0.10500	0.30000

Table.6.6..MLE and prior.

α	***Dose levels***			
0.1		**0.022287**	**0.364330**	**0.404044**
	1.0e-003 *	0.139757	0.166319	0.206954
10		**0.019993**	**0.278240**	**0.434851**
	1.0e-003 *	0.112056	0.587708	0.764273
50		**0.013641**	**0.216150**	**0.403991**
	1.0e-003 *	0.063066	0.486889	0.724020
100		**0.009083**	**0.183088**	**0.378432**
	1.0e-003 *	0.034203	0.370174	0.604455

Table.6.7. Posterior mean and SD.

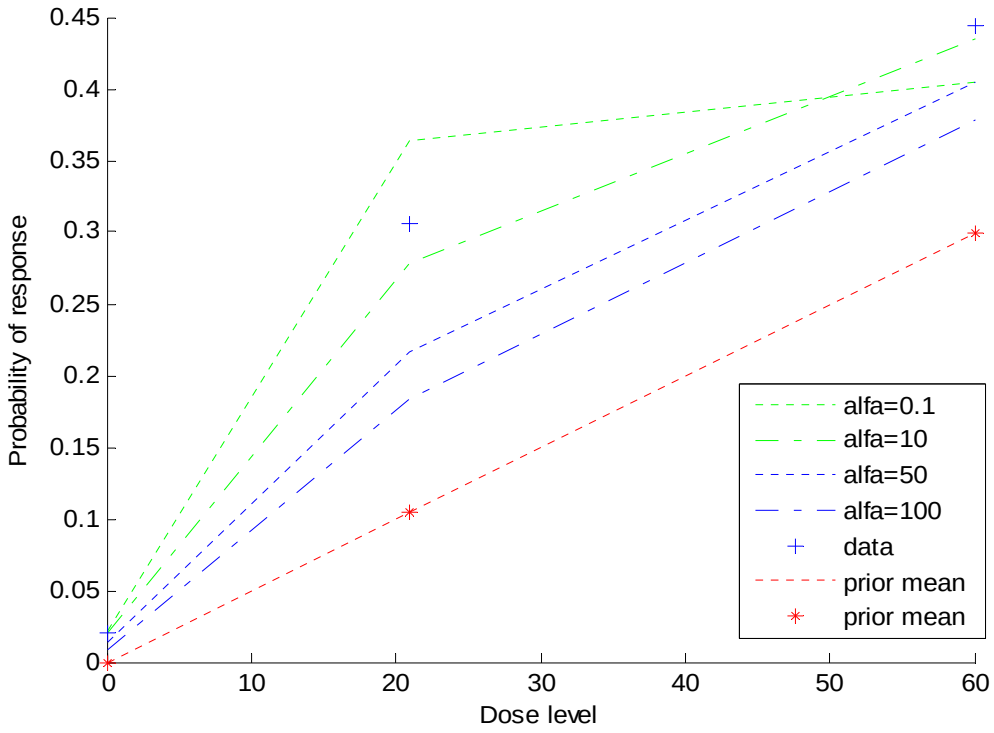

Figure 6.1 Nonparametric Bayes of potency curve for different value of precision parameter.

- $P_o = \left(P_o(x_1), P_o(x_2), P_o(x_3)\right) = \left(\frac{1}{4}, \frac{2}{4}, \frac{3}{4}\right)$;

	Dose levels		
MLE	0.02000	0.30612	0.44444
P_o	0.25000	0.50000	0.75000

Table.6.8 .MLE and prior.

α	***Dose levels***			
0.1		**0.022188**	**0.368366**	**0.406662**
	1.0e-003 *	0.137191	0.160749	0.197812
10		**0.058091**	**0.297545**	**0.482478**
	1.0e-003 *	0.285003	0.629629	0.847534
50		**0.118718**	**0.333171**	**0.571970**
	1.0e-003 *	0.385860	0.571229	0.775596
100		**0.151851**	**0.371416**	**0.620980**
	1.0e-003 *	0.363077	0.473378	0.611281

Table.6.9 .Posterior mean and SD.

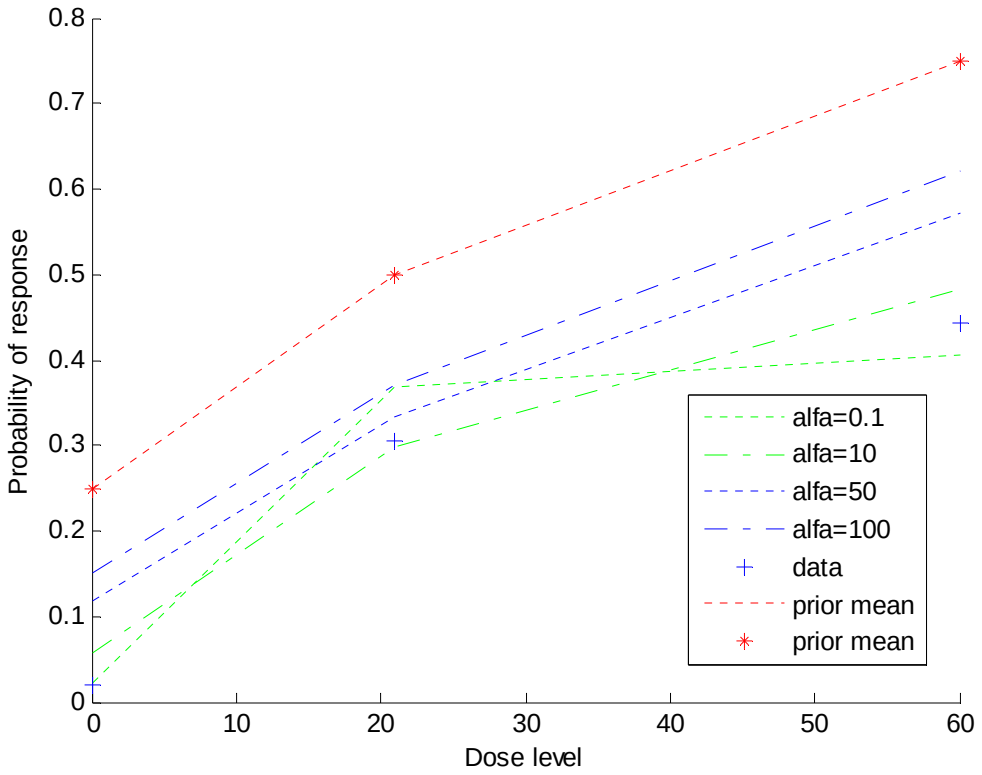

Figure 6.2 Nonparametric Bayes of potency curve for different value of precision parameter.

The simple conclusion follows from these two experiments. As the value of precision parameter decreases the posterior approaches the maximum likelihood. Recall that this parameter controls the strength of belief in the prior guess. A large value reflects that

posterior is tightly concentrated about P_0. The worrying situations which can be observe here are the results for $\alpha = 0.1$ in both choices of the base distribution. If we compare the results obtained by Gibbs sampler with $\alpha = 0.1$ and $\alpha = 10$ it follows that the interpretation of this parameter is wrong. Indeed, the posterior at the observational dose levels obtained from the nonparametric Bayesian inference with $\alpha = 10$ is closer to the maximum likelihood values than the posterior obtained by using $\alpha = 0.1$. We suspect that in case of α closer than zero, the numerical errors occurred.

Analysis of data set 2

Below we present the analysis of the second data set with the following base distributions:

- $P_o(x) = \frac{x^3}{3000^3}$;

	Dose levels				
MLE	0.033333	0.200000	0.700000	0.666667	0.850000
P_o	0.037037	0.125000	0.198495	0.296296	0.578704

Table.6.10 .*MLE* and prior.

α	***Dose levels***					
0.1		**0.061502**	**0.192479**	**0.567631**	**0.699036**	**0.800125**
	1.0e-003 *	0.135078	0.427997	0.141957	0.119417	0.278851
10		**0.045389**	**0.218772**	**0.509655**	**0.614589**	**0.797997**
	1.0e-003 *	0.212752	0.672610	0.427537	0.389943	0.630650
50		**0.050816**	**0.214445**	**0.362687**	**0.484385**	**0.715609**
	1.0e-003 *	0.198351	0.401370	0.343521	0.427121	0.694450
100		**0.050528**	**0.190148**	**0.304349**	**0.422557**	**0.672873**
	1.0e-003 *	0.154980	0.261726	0.240840	0.342140	0.611245

Table.6.11 .Posterior mean and SD

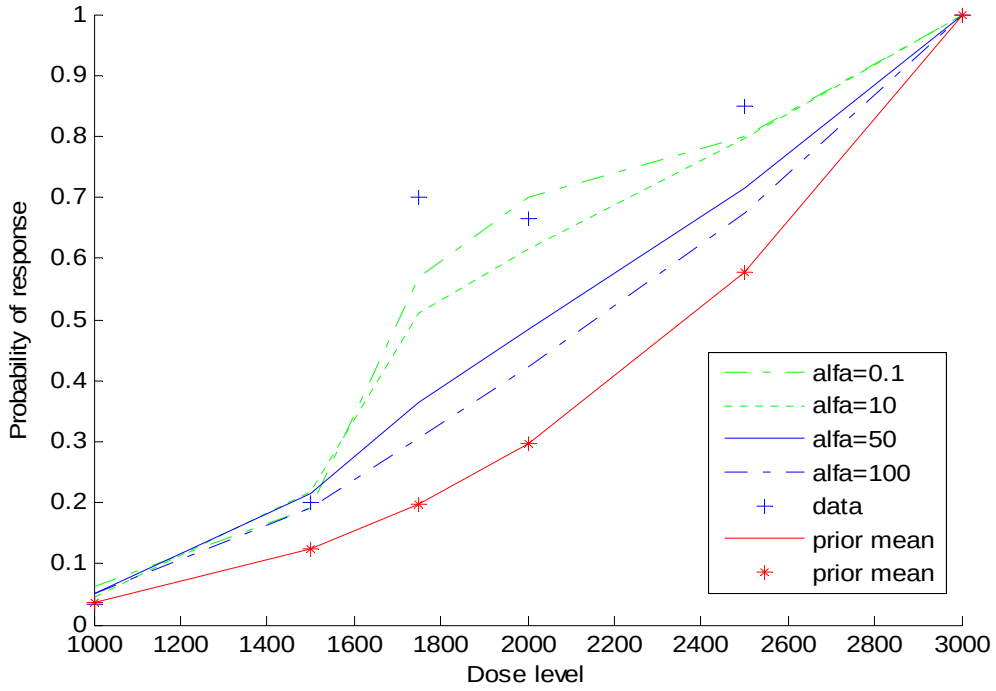

Figure 6.3 Nonparametric Bayes of potency curve for different value of precision parameter.

- $P_o(x) = \dfrac{x}{3000}$;

	Dose levels				
MLE	0.033333	0.200000	0.700000	0.666667	0.850000
P_o	0.333333	0.500000	0.583333	0.666667	0.833333

Table.6.12 *MLE* and prior.

α	***Dose levels***					
0.1		**0.087610**	**0.229636**	**0.611674**	**0.667944**	**0.790502**
	1.0e-003 *	0.136916	0.430580	0.078035	0.052970	0.337731
10		**0.105236**	**0.256820**	**0.558855**	**0.659161**	**0.834435**
	1.0e-003 *	0.390561	0.632242	0.423279	0.362307	0.524026
50		**0.204155**	**0.368270**	**0.508066**	**0.622774**	**0.820390**
	1.0e-003 *	0.491958	0.395523	0.321041	0.380531	0.500737
100		**0.247526**	**0.413180**	**0.525832**	**0.626234**	**0.817303**
	1.0e-003 *	0.429004	0.282898	0.218388	0.277797	0.397961

Table.6.13 .Posterior mean, SD

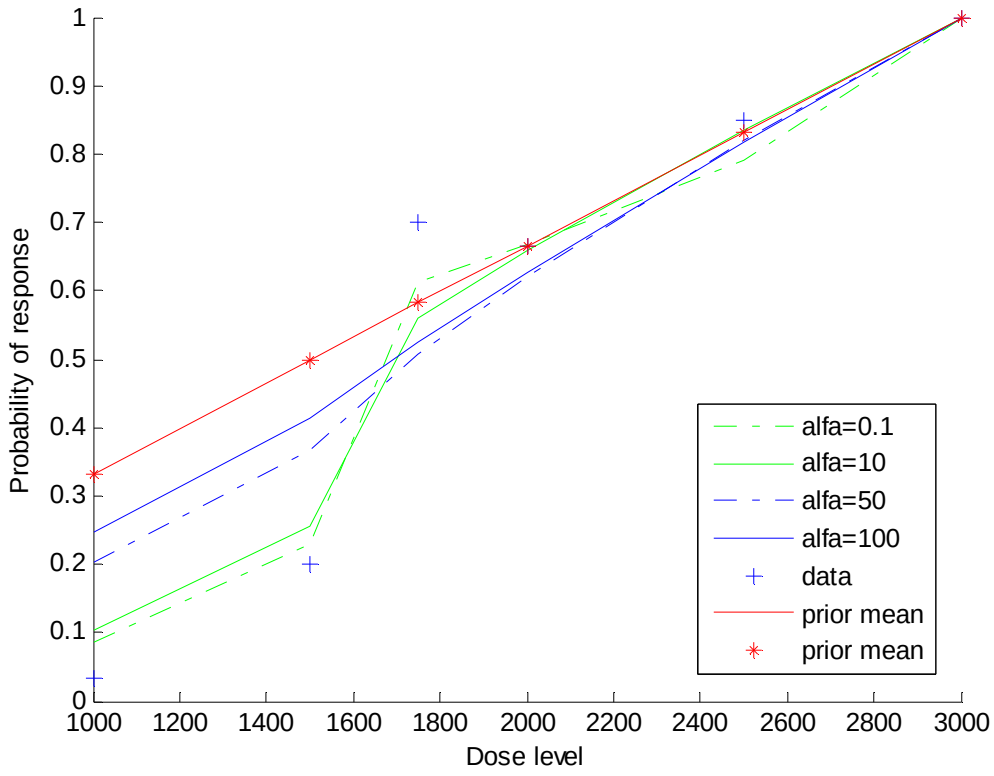

Figure 6.4 Nonparametric Bayes of potency curve for different value of precision parameter.

The role of controlling the strength of belief in the prior guess is visible also in the analysis of the 2nd data set. However, some worrying results are observed at the dose level 1750. As we have already note on the beginning of this part, the "conflict" between the data at the dose level $x = 1750$ and $x = 2000$ occurs. Recall that it follows from the

data that the probability of response at level 1750 is a little bit bigger than at the higher dose level 2000. Therefore, the posterior mean at dose 1750 is in both cases quite far from the maximum likelihood at this dose level.

During these experiments the following question has arisen: does the nonparametric Bayesian procedure can be used in case when the prior information is not available? In other words, we try to consider the situation when the analysis is based only on the data, due to the lack of prior information about the problem. The answer to this question reader can find for example in subsection 3.2.1 where the idea of Ramsey is presented. The solution for this case could be the noninformative prior such as Jefrreys prior.

How the model fits the data

The purpose of the next part of this chapter is to analyze the model suitability to recover the observational probabilities. Therefore, the results from the nonparametric Bayesian approach (*NB*) together with the results obtained from the U.S. Environmental Protection Agency (EPA) Benchmark Dose Software (*BMD*) are compared with the results from marginal likelihood method. Note that the marginal likelihood method (known also as the probabilistic inversion) seeks a distribution over parameters in a dose response model which recovers the uncertainty in bioassay data. The distribution is fit (by using probabilistic inversion with variant of the iterative proportional fitting algorithm) to the parameters so that the number of responding has a distribution which matches as closely as possible the binomial distribution with parameters from data set. Whereas the *BMD* software gives a distribution of maximum likelihood estimates of a selected model. Note that now, the *BMD* software offers models that are appropriate for the analysis of dichotomous (quantal) data (Gamma, Logistic, Log-Logistic, Multistage, Probit, Log-Probit, Quantal-Linear, Quantal-Quadratic, Weibull), continuous data (Linear, Polynomial, Power, Hill) and nested developmental toxicology data (NLogistic, NCTR, Rai & Van Ryzin). It is worth to stress that this method is not intended to recover the bioassay uncertainty, what can be easily verified by the further analysis.

In the figures below, “stairs” represents cumulative distribution functions for the binomial distribution, giving the number of individuals showing response, if the probability of response is given by the percentage observed on the experiments in table 6.1 and 6.2, and the number of subjects given in the data sets. Whereas, the solid lines are the results of sampling:

- the probabilities of responses at each observational dose level form nonparametric Bayesian model and multiplying by the number of exposed animals (red lines for both data sets);
- the parameters from the model used in the *BMD* software, substituting these into the formula describing the probability of response and multiplying by the number of exposed animals (red lines for the 1st data set and black lines for the 2nd data set).

Data set 1

The nonparametric Bayesian estimates based on the 1st data set are presented with two base distributions and two precision parameters ($\alpha = 10, \alpha = 1$). Note that in case of $\alpha = 0.1$ the previous studies indicate some worrying results (probably due to numerical errors), which in fact it has also impact on the shape of cumulative distribution function of the number of responses. Thus, in order to examine the effect of different choices of this parameter on the results we omit this value.

- $P_o(x) = \dfrac{x}{200}$;

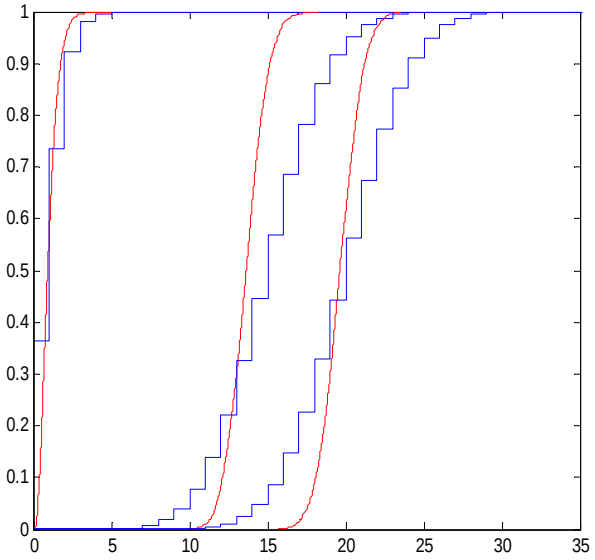

Figure 6.5 NB inference of data set 1, $\alpha = 10$.

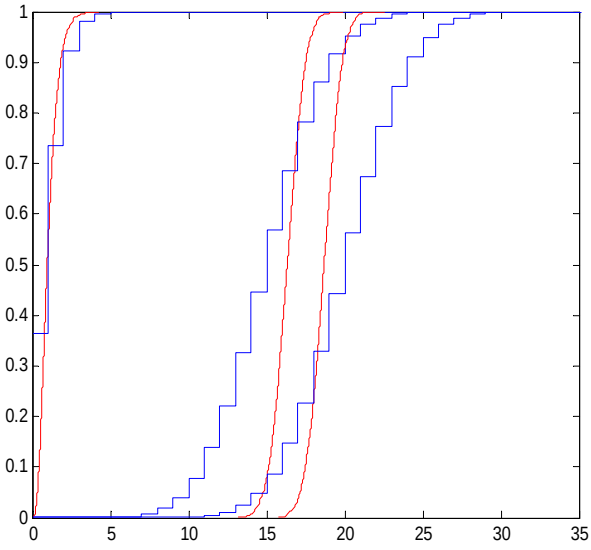

Figure 6.6 NB inference of data set 1, $\alpha = 1$.

- $P_o = \left(P_o(x_1), P_o(x_2), P_o(x_3)\right) = \left(\dfrac{1}{4}, \dfrac{2}{4}, \dfrac{3}{4}\right)$,

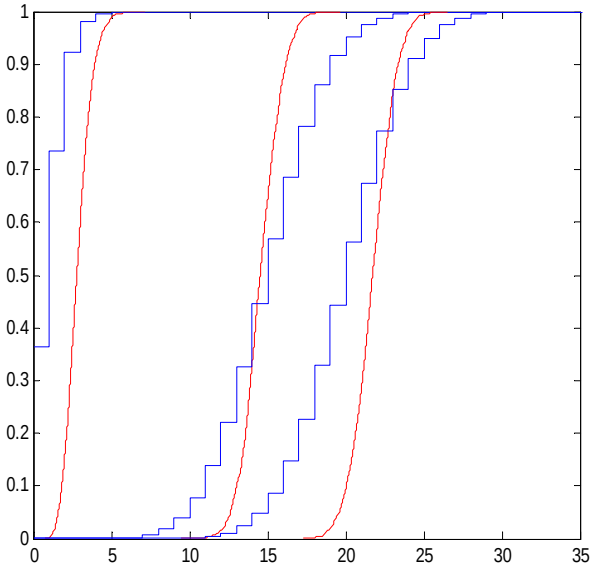

Figure 6.7 NB inference of data set 1, $\alpha = 10$.

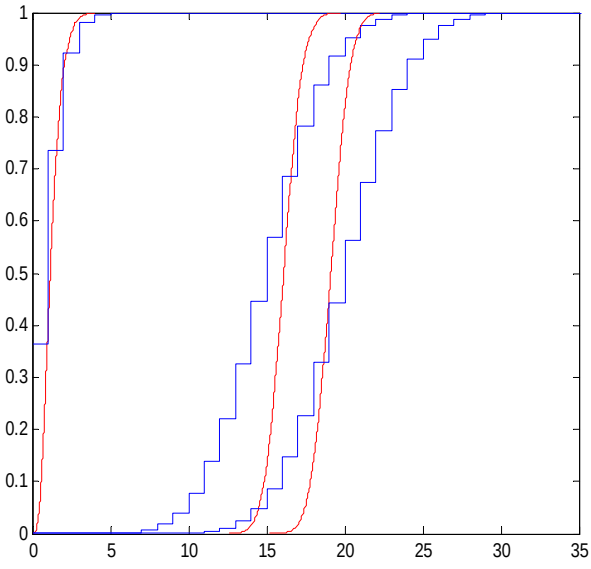

Figure 6.8 NB inference of data set 1, $\alpha = 1$.

Figure 6.9 presents the results from parametric methods: marginal likelihood (blue lines) and the results obtained by the *BMD* software (red lines). The recommended model, for the data set 1 is the log logistic model.

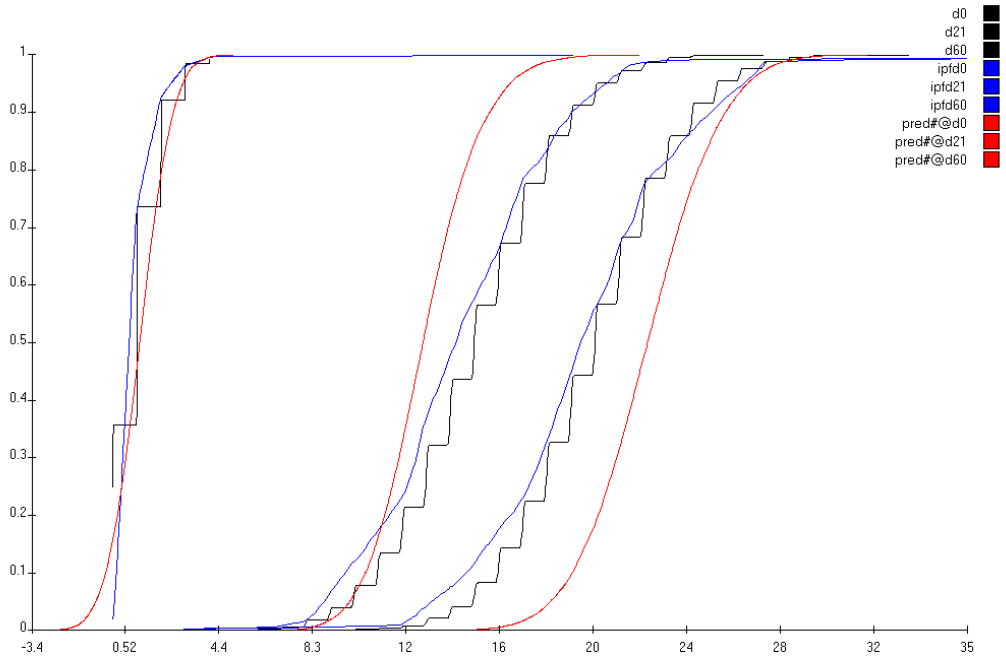

Figure 6.9 Parametric methods (PI and BMD software) used to data set 1.

The analysis of the nonparametric estimates confirms our previous conclusions about the role of the precision parameter. In each of the choices of base distribution, results obtained from the Gibbs sampler with $\alpha = 1$ fit the binomial distributions better than in case of $\alpha = 10$. As we can observe the fit of CDF of the number of responses at dose level zero to the binomial distribution with parameters in table 6.1 is decent for nonparametric Bayesian method (with $\alpha = 1$) as well as for both parametric methods. However, note that around 15% of the predicted responses are negative. Distributions obtained for dose level 21 and 60 from Gibbs sampling and the BMD software do not follow the binomial distribution so well. Note that in case of the NB method, the range of the number of responses for which the probability is bigger than zero and less than 1 is quite narrow; around 5 (the 'binomials range' around 17). Recall that the probabilities needed to calculate numbers of responses are sampled from the conditional beta distributions with very condensed density. If we compare the biggest distance between

binomial distributions and CDF obtained from sampling, it follows that better fit is obtained by the results from the *BMD* software, than from the *NB* method. We can also observe that in case of *BMD*, prediction of the number of responses at dose level 21 is more optimistic (in sense that an event is more probable) than in case of *NB*. Whereas, at dose level 60 this prediction is more pessimistic, than the results obtained from Gibbs sampler. Obvious conclusion is that the marginal likelihood fits the binomial distribution much better than the other methods.

Data set 2

Pictures 6.10-6.13 illustrate the results obtained from the analysis of data set 2 by using Gibbs sampler with two base distributions and two values of precision parameter ($\alpha = 10, \alpha = 1$). Note that in case of $\alpha = 0.1$ the cumulative distribution function of the number of responses at each dose level is almost the same as in case of $\alpha = 1$. According to our earlier observations this CDF should approach the binomial distributions with parameters in table 6.2. along with the value of precision parameter decreases. Note also that each of the pictures consists of the results obtained at the dose levels starting from the left hand side: $x = 1000$, $x = 1500$ (dashed lines), $x = 1750$, $x = 2500$ and $x = 2000$.

- $P_o(x) = \dfrac{x^3}{3000^3}$,

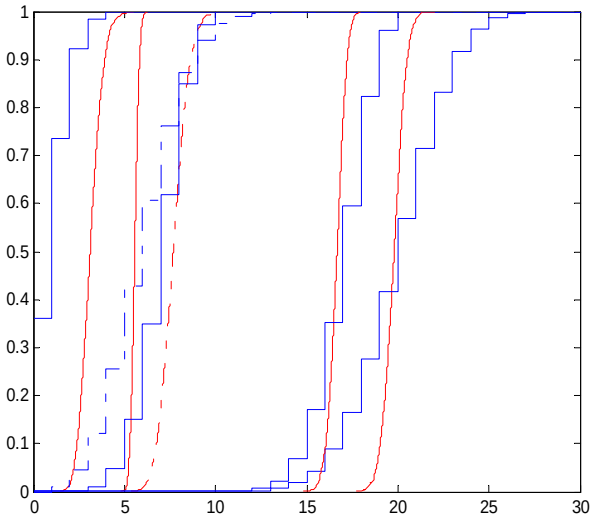

Figure 6.10. NB inference of data set 2, $\alpha = 10$.

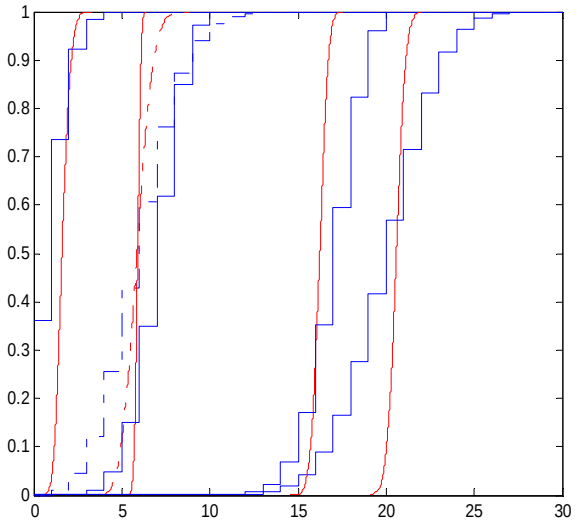

Figure 6.11 NB inference of data set 2, $\alpha = 1$

- $P_o(x) = \dfrac{x}{3000}$,

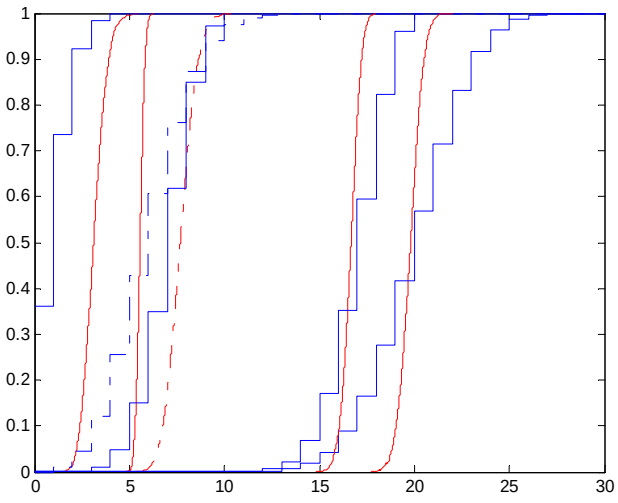

Figure 6.12 NB inference of data set 2, $\alpha = 10$.

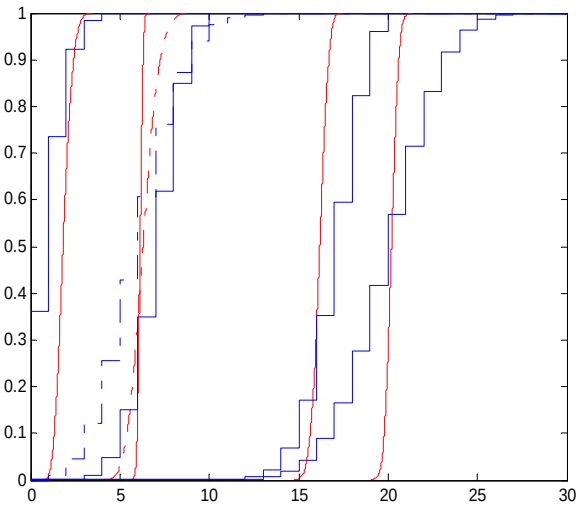

Figure 6.13 NB inference of data set 2, $\alpha = 1$

Figures 6.14-6.18 present the results from parametric methods: marginal likelihood (blue lines) and the results obtained by the BMD software (black lines). Note that the marginal

likelihood found the best fit using the loglogistic model, whereas the results from the BMD software are based on the logProbit model.

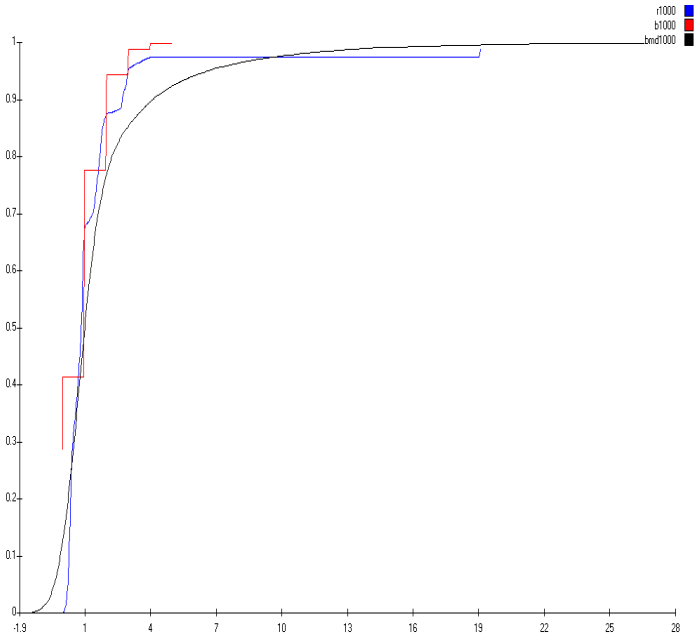

Figure 6.14 Parametric methods (PI and the BMD software) used to data set 2 at dose 1000.

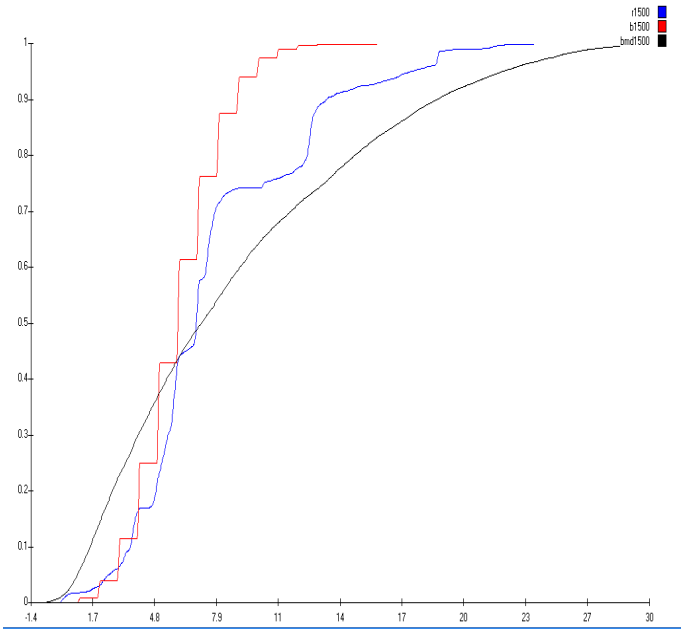

Figure 6.15 Parametric methods (PI and the BMD software) used to data set 2 at dose 1500.

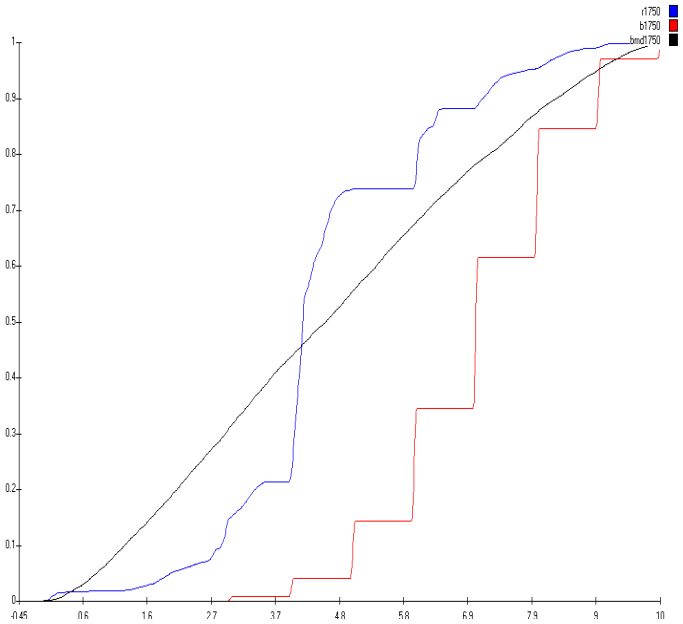

Figure 6.16 Parametric methods (PI and the BMD software)

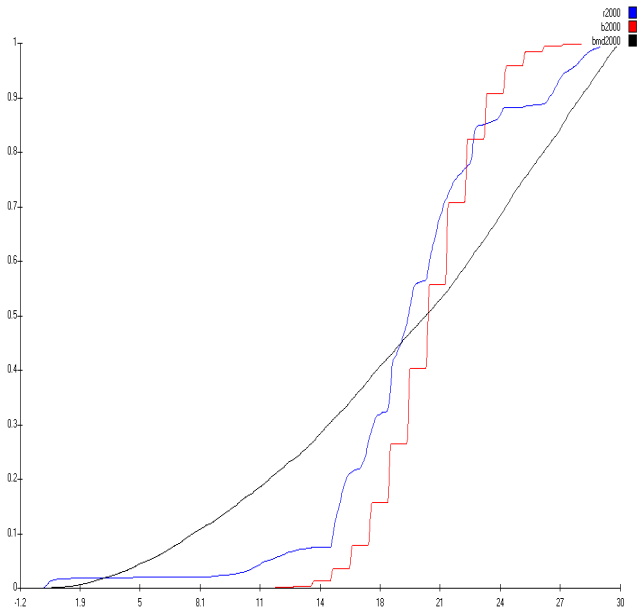

Figure 6.17 Parametric methods (PI and the BMD software)

used to data set 2 at dose 1750. used to data set 2 at dose 2000.

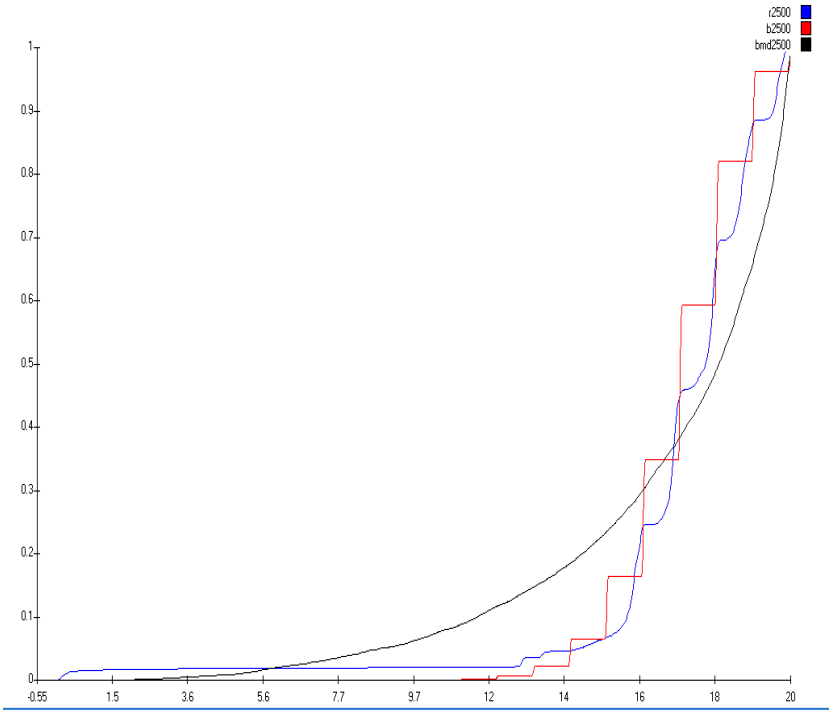

Figure 6.18 Parametric methods (PI and the BMD software) used to data set 2 at dose 2500.

Analogous to the conclusions of the results for data set 1 from the nonparametric Bayesian model, we can observe that better fits is obtained by using lower value of precision parameter. Analysis of recoverage (the ability of a parameter distribution to re-capture the uncertainty in the observed data) shows us that also in case of data set 2 the *BMD* distribution is more flexible, than the CDF obtained from the Gibbs sampling. It can be observed especially at the last two dose levels: 2000 and 2500. The ranges for the number of responses - for which the probability is bigger than zero and less than 1, are even smaller (around 1.5) than in case of the 1st data set. However, it seems that the CDF obtained from *NB* methodology at the first two dose levels (1000 and 1500) track the binomial distributions closely than the CDF obtained from the *BDM* software. The lack of fit at dose level 1750 can be easily explained by the 'conflict' in data at dose level

1750 and 2000. Note that this situation is however more visible in case of parametric method.

Chapter 7

Conclusions

The problem of estimating a distribution function without assuming a specific parametric form for the distribution is a familiar problem in various branches of statistical endeavor. The main focus of this thesis is to present a nonparametric Bayesian inference in order to explain the relation between the response probability and the dosage in quantal bioassay. Many bioassay studies are conducted during the early phase evaluation of a new drug, and therefore we have a little knowledge about the dose – response behavior of the drug which is under test. Thus, a nonparametric approach with flexible and adaptive modeling property can be considered here. This can be done by assigning the Dirichlet process or Beta product family as a prior. As we have underlined in chapter 2, the problem which can arise here could be specification of the required 'parameters'; precision parameter and based distribution. The discussion in chapter 2 gives an intuitive meaning to these parameters which could facilitate the appropriate selection of these parameters. The Bayesian approach is then used to derive the posterior distribution of the potency curve. We showed that this posterior has a form of mixed Dirichlet. However, it has an extremely complicated form, often analytically infeasible. This leads to many difficulties; for example in obtaining marginal distributions from the joint densities. Therefore some estimation techniques for exploring the tolerance distribution are called for. There is an extensive literature on this subject. In this work we illustrate the Markov Chain Monte

Carlo approach and maximum likelihood techniques. The former method seems to be more efficient, owing to the wide availability of the high – speed computations. We propose here a Gibbs sampler (one of the most commonly used MCMC method) as a technique for generating random variables from a (marginal) distribution indirectly, without having to calculate the density. It avoids difficult calculations, replacing them instead with a sequence of easier calculations. However, in order to obtain stable results by using Gibbs sampler we need to perform an extensive range of replication-iteration combinations. In some cases the stable results can be never obtained. The MLE estimation could be often difficult to perform, due to the extremely complicated form of the joint posterior distribution. Moreover, the procedure of calculating the joint mode of posterior required some numerical methods such as Newton – Raphson method. In this method, the operation of matrix inversion is performed, which in case of very bad conditioned matrix returns very inaccurate results. Useful extensions in bioassay study are constraints on shape of the potency curve. In many cases this analysis provides a real description of the bioassay data. However, the forms of the conditional posterior are very complicated. As we have mentioned in chapter 3, the Gibbs sampler needs to be combined with some sampling resampling methods. Therefore the time of computations increases considerably.

One of the contributions which statisticians can make in bioassay studies is the analysis of an effective dose. We provide here an exact form of the distribution of an effective dose conditional on the estimated potency. Moreover we present also some discussion about the design of experiment. The relevant aspects of the design strategy such as; choosing the sample of experimental subjects and system of allocating them, specification of the number and magnitudes of doses to be tested, are analyzed. We illustrate here few different strategies and show the way how they can be compared.
Summarizing, this report reviews the nonparametric Bayesian approach which can be successively used to analyzing the quantal bioassay data.

Appendix A: Proof of conditional densities in section 2.3.3

Assume that $u = (u_1, \ldots, u_{M+1})$ is Dirichlet with density function

$$\pi(u) \propto \prod_{i=1}^{M+1} u_i^{\alpha_i - 1}, \tag{A.1}$$

Then, the conditional distributions for each u_i are as follows:

- for $i = 1, \ldots i^*$,

$$u_i \mid u_1, \ldots, u_{i-1} \sim Beta(\alpha_i, \sum_{j=i+1}^{M+1} \alpha_j; 0, 1 - \sum_{j=1}^{i-1} u_j) \tag{A.2}$$

- for $i = M, \ldots i^* + 1$,

$$u_i \mid u_1, \ldots, u_{i^*}, u_{M+1}, \ldots, u_{i+1} \sim Beta(\alpha_i, \sum_{j=1}^{i-1} \alpha_j - \sum_{j=1}^{i^*} \alpha_j; 0, 1 - \left(\sum_{j=1}^{i^*} u_j + \sum_{j=i+1}^{M+1} u_j \right)), \tag{A.3}$$

Moreover, it can be shown that:

- for $i = 1, \ldots i^*$,

$$p_i \mid i^*, p_1, \ldots p_{i-1} \sim Beta(\alpha_i, \sum_{j=i+1}^{M+1} \alpha_j; \bar{c}_i, \bar{d}_i), \tag{A.4}$$

- for $i = M, \ldots i^* + 1$,

$$p_i \mid i^*, p_1, \ldots p_{i^*}, p_{M+1}, \ldots, p_{i+1} \sim Beta(\sum_{j=1}^{i} \alpha_j - \sum_{j=1}^{i^*} \alpha_j, \alpha_{i+1}; c_i, d_i), \tag{A.5}$$

with c_i, d_i and $\bar{c}_i, \bar{d}_i$ defined in section 2.3.3.

Note also that, omitting the normalizing constant, a $w \sim Beta(a, b; c, d)$ random variable has density

$$g(w \mid a, b, c, d) \propto (w - c)^{a-1} (d - w)^{b-1}, \tag{A.6}$$

Moreover,

$$\int_c^d (w - c)^{a-1} (d - w)^{b-1} \, dw \propto (d - c)^{a+b-1}, \tag{A.7}$$

Proof of conditional densities (A.2)

Notice that the conditional density can be express as follows

$$\pi(u_i \mid u_1,\ldots,u_{i-1}) = \frac{\pi(u_i,\ldots,u_1)}{\pi(u_{i-1},\ldots,u_1)},$$

where

$$\pi(u_i,\ldots,u_1) \propto \int_0^{1-\sum_{j=1}^{i} u_j} \cdots \int_0^{1-\sum_{j=1}^{M-1} u_j} u_1^{\alpha_1-1}\ldots u_M^{\alpha_M-1}(1-\sum_{j=1}^{M} u_j)^{\alpha_{M+1}-1} du_M \ldots du_{i+1}$$

$$\propto \int_0^{1-\sum_{j=1}^{i} u_j} \cdots \int_0^{1-\sum_{j=1}^{M-2} u_j} \prod_{j=1}^{M-1} u_j^{\alpha_j-1} du_{M-1}\ldots du_{i+1} \left[\int_0^{1-\sum_{j=1}^{M-1} u_j} u_M^{\alpha_M-1}(1-\sum_{j=1}^{M-1} u_j - u_M)^{\alpha_{M+1}-1} du_M \right]$$

$$\propto \int_0^{1-\sum_{j=1}^{i} u_j} \cdots \int_0^{1-\sum_{j=1}^{M-2} u_j} \prod_{j=1}^{M-1} u_j^{\alpha_j-1} \left(1-\sum_{j=1}^{M-1} u_j\right)^{\alpha_{M+1}+\alpha_M-1} du_{M-1}\ldots du_{i+1}$$

$$\propto (1-\sum_{j=1}^{i} u_j)^{\sum_{j=i+1}^{M+1}\alpha_j-1} \prod_{j=1}^{i} u_j^{\alpha_j-1},$$

and

$$\pi(u_{i-1},\ldots,u_1) \propto \int_0^{1-\sum_{j=1}^{i} u_j} \cdots \int_0^{1-\sum_{j=1}^{M-1} u_j} u_1^{\alpha_1-1}\ldots u_M^{\alpha_M-1}(1-\sum_{j=1}^{M} u_j)^{\alpha_{M+1}-1} du_M \ldots du_i$$

$$\propto (1-\sum_{j=1}^{i-1} u_j)^{\sum_{j=i}^{M+1}\alpha_j-1} \prod_{j=1}^{i-1} u_j^{\alpha_j-1},$$

Note that the property *(A.7)* is used here to each integration.
The relation above enable us to calculate the conditional density as follows

$$\pi(u_i \mid u_1,\ldots,u_{i-1}) \propto \frac{(1-\sum_{j=1}^{i} u_j)^{\sum_{j=i+1}^{M+1}\alpha_j-1} \prod_{j=1}^{i} u_j^{\alpha_j-1}}{(1-\sum_{j=1}^{i-1} u_j)^{\sum_{j=i}^{M+1}\alpha_j-1} \prod_{j=1}^{i-1} u_j^{\alpha_j-1}} \propto \frac{(1-\sum_{j=1}^{i} u_j)^{\sum_{j=i+1}^{M+1}\alpha_j-1} u_i^{\alpha_i-1}}{(1-\sum_{j=1}^{i-1} u_j)^{\sum_{j=i}^{M+1}\alpha_j-1}}$$

$$\propto (1-\sum_{j=1}^{i-1} u_j - u_i)^{\sum_{j=i+1}^{M+1}\alpha_j-1} u_i^{\alpha_i-1},$$

According to definition *(A.6)*, the expression above implies that

$$\pi(u_i \mid u_1,\ldots,u_{i-1}) \sim Beta(\alpha_i, \sum_{j=i+1}^{M+1}\alpha_j; 0, 1-\sum_{j=1}^{i-1} u_j).$$

In order to prove the expression *(A.4)* recall that:

$$u_i = (x_{M+1} - x_{i-1})\left(\frac{p_i - p_{i-1}}{x_i - x_{i-1}} - \frac{p_{i-1} - p_{i-2}}{x_{i-1} - x_{i-2}} \right),$$

Notice also that

$$u_i = (1 - \sum_{j=1}^{i-1} u_j) Beta(\alpha_i, \sum_{j=i+1}^{M+1} \alpha_i),$$

where $$(1 - \sum_{j=1}^{i-1} u_j) = 1 - p_{i-2}\left(1 - \frac{x_{M+1} - x_{i-2}}{x_{i-1} - x_{i-2}}\right) - p_{i-1}\frac{x_{M+1} - x_{i-2}}{x_{i-1} - x_{i-2}},$$

Therefore:

$$(x_{M+1} - x_{i-1})\left(\frac{p_i - p_{i-1}}{x_i - x_{i-1}} - \frac{p_{i-1} - p_{i-2}}{x_{i-1} - x_{i-2}} \right) = \left[1 - p_{i-2}\left(1 - \frac{x_{M+1} - x_{i-2}}{x_{i-1} - x_{i-2}}\right) - p_{i-1}\frac{x_{M+1} - x_{i-2}}{x_{i-1} - x_{i-2}} \right] Beta(\alpha_i, \sum_{j=i+1}^{M+1} \alpha_i),$$

And after few manipulations it follows that:

$$p_i = \frac{x_i - x_{i-1}}{x_{i-1} - x_{i-2}}(p_{i-1} - p_{i-2}) + p_{i-1} + \frac{x_i - x_{i-1}}{x_{M+1} - x_{i-1}} Beta(\alpha_i, \sum_{j=i+1}^{M+1} \alpha_i),$$

It follows from the expression above that:

$$\overline{c}_i = \frac{x_i - x_{i-1}}{x_{i-1} - x_{i-2}}(p_{i-1} - p_{i-2}) + p_{i-1},$$

and

$$\overline{d}_i - \overline{c}_i = \frac{x_i - x_{i-1}}{x_{M+1} - x_{i-1}}\left[1 - p_{i-2}\left(1 - \frac{x_{M+1} - x_{i-2}}{x_{i-1} - x_{i-2}}\right) - p_{i-1}\frac{x_{M+1} - x_{i-2}}{x_{i-1} - x_{i-2}} \right],$$

Thus,

$$\begin{aligned} \overline{d}_i &= \overline{c}_i + \frac{x_i - x_{i-1}}{x_{M+1} - x_{i-1}}\left[1 - p_{i-2}\left(1 - \frac{x_{M+1} - x_{i-2}}{x_{i-1} - x_{i-2}}\right) - p_{i-1}\frac{x_{M+1} - x_{i-2}}{x_{i-1} - x_{i-2}} \right] = \\ &= \frac{x_i - x_{i-1}}{x_{i-1} - x_{i-2}}(p_{i-1} - p_{i-2}) + p_{i-1} + \frac{x_i - x_{i-1}}{x_{M+1} - x_{i-1}}\left[1 - p_{i-2}\left(1 - \frac{x_{M+1} - x_{i-2}}{x_{i-1} - x_{i-2}}\right) - p_{i-1}\frac{x_{M+1} - x_{i-2}}{x_{i-1} - x_{i-2}} \right] = \\ &= p_{i-1} + \frac{x_i - x_{i-1}}{x_{M+1} - x_{i-1}}(1 - p_{i-1}), \end{aligned}$$

This implies the from *(A.4)*.

Proof of conditional densities (A.3)

In order to prove conditional density A3, we use the following property of Dirichlet distribution:

Property: Let C be any index set and let P_C be a subvector of $P=(P_1,\ldots,P_k)$, i.e. $P_C=(P_l, l\in C)$ and let $\overline{\alpha}_C=(\alpha_l, l\in C)$, then

$$P_C \sim Dirichlet(\overline{\alpha}_C, \sum_{i=1}^{k+1}\alpha_i - \sum_{l\in C}\alpha_i).$$

Notice that the conditional density can be express as follows

$$\pi(u_i \mid u_1,\ldots,u_{i^*},u_{M+1},\ldots,u_{i+1}) = \frac{\pi(u_1,\ldots,u_{i^*},u_{M+1},\ldots,u_i)}{\pi(u_1,\ldots,u_{i^*},u_{M+1},\ldots,u_{i+1})},$$

Where

$$\pi(u_1,\ldots,u_{i^*},u_i,\ldots,u_{M+1}) \propto u_1^{\alpha_1-1}\ldots u_{i^*}^{\alpha_{i^*}-1}u_i^{\alpha_i-1}\ldots u_{M+1}^{\alpha_{M+1}-1}\left(1-\sum_{j=1}^{i^*}u_j-\sum_{j=i+1}^{M+1}u_j-u_i\right)^{\sum_{j=1}^{M+2}\alpha_j-\sum_{j=1}^{i^*}\alpha_j-\sum_{j=i}^{M+1}\alpha_j}.$$

Note that the density above was obtained from the properties of Dirichlet distribution. Recall also that the assumption $u_{M+2}=0$, implies that $\alpha_{M+2}=0$.

Using once again aforementioned property we obtain the next density

$$\pi(u_1,\ldots,u_{i^*},u_{i+1},\ldots,u_{M+1}) \propto u_1^{\alpha_1-1}\ldots u_{i^*}^{\alpha_{i^*}-1}u_{i+1}^{\alpha_{i+1}-1}\ldots u_{M+1}^{\alpha_{M+1}-1}\left(1-\sum_{j=1}^{i^*}u_j-\sum_{j=i+1}^{M+1}u_j\right)^{\sum_{j=1}^{M+2}\alpha_j-\sum_{j=1}^{i^*}\alpha_j-\sum_{j=i+1}^{M+1}\alpha_j}.$$

Thus, the relation above enables us to calculate the conditional density as follows

$$\pi(u_i \mid u_1,\ldots,u_{i^*},u_{i+1},\ldots,u_{M+1}) \propto \frac{u_1^{\alpha_1-1}\ldots u_{i^*}^{\alpha_{i^*}-1}u_i^{\alpha_i-1}\ldots u_{M+1}^{\alpha_{M+1}-1}\left(1-\sum_{j=1}^{i^*}u_j-\sum_{j=i+1}^{M+1}u_j-u_i\right)^{\sum_{j=1}^{M+2}\alpha_j-\sum_{j=1}^{i^*}\alpha_j-\sum_{j=i}^{M+1}\alpha_j}}{u_1^{\alpha_1-1}\ldots u_{i^*}^{\alpha_{i^*}-1}u_{i+1}^{\alpha_{i+1}-1}\ldots u_{M+1}^{\alpha_{M+1}-1}\left(1-\sum_{j=1}^{i^*}u_j-\sum_{j=i+1}^{M+1}u_j\right)^{\sum_{j=1}^{M+2}\alpha_j-\sum_{j=1}^{i^*}\alpha_j-\sum_{j=i+1}^{M+1}\alpha_j}} \propto$$

$$\propto u_i^{\alpha_i-1}\left(1-\sum_{j=1}^{i^*}u_j-\sum_{j=i+1}^{M+1}u_j-u_i\right)^{\sum_{j=1}^{M+2}\alpha_j-\sum_{j=1}^{i^*}\alpha_j-\sum_{j=i}^{M+1}\alpha_j}$$

According to definition *(A.6)*, the expression above implies that

$$\pi(u_i \mid u_1,\ldots,u_{i^*},u_{M+1},\ldots,u_{i+1}) \sim Beta(\alpha_i, \sum_{j=1}^{i-1}\alpha_j - \sum_{j=1}^{i^*}\alpha_j; 0, 1-\left(\sum_{j=1}^{i^*}u_j+\sum_{j=i+1}^{M+1}u_j\right)),$$

In order to prove expression *(A.5)*, recall that:

$$u_{i+1} = (x_{i+1} - x_0)\left[\frac{p_{i+1} - p_i}{x_{i+1} - x_i} - \frac{p_{i+2} - p_{i+1}}{x_{i+2} - x_{i+1}}\right],$$

Notice also that

$$u_{i+1} = \left(1 - \sum_{j=1}^{i^*} u_j - \sum_{j=i+2}^{M+1} u_j\right) Beta(\alpha_{i+1}, \sum_{j=1}^{i} \alpha_j - \sum_{j=1}^{i^*} \alpha_j),$$

where

$$1 - \sum_{j=1}^{i^*} u_j - \sum_{j=i+2}^{M+1} u_j = \frac{x_{i^*} - x_0}{x_{i^*+1} - x_{i^*}}\left(p_{i^*+1} - p_{i^*}\right) - p_{i^*} + p_{i+2} - \frac{x_{i+2} - x_0}{x_{i+2} - x_{i+1}}\left(p_{i+2} - p_{i+1}\right) = A_{i^*} + B_i,$$

Therefore:

$$(x_{i+1} - x_0)\left[\frac{p_{i+1} - p_i}{x_{i+1} - x_i} - \frac{p_{i+2} - p_{i+1}}{x_{i+2} - x_{i+1}}\right] = \left(A_{i^*} + B_i\right)(1 - Beta(\sum_{j=1}^{i} \alpha_j - \sum_{j=1}^{i^*} \alpha_j, \alpha_{i+1})),$$

And after few manipulations it follows that:

$$p_i = -\frac{x_{i+1} - x_i}{(x_{i+1} - x_0)}\left(A_{i^*} + B_i\right) - \frac{x_{i+1} - x_i}{x_{i+2} - x_{i+1}}\left(p_{i+2} - p_{i+1}\right) + p_{i+1} + \frac{x_{i+1} - x_i}{(x_{i+1} - x_0)}\left(A_{i^*} + B_i\right) Beta(\sum_{j=1}^{i} \alpha_j - \sum_{j=1}^{i^*} \alpha_j, \alpha_{i+1}),$$

Note that:

$$c_i = -\frac{x_{i+1} - x_i}{(x_{i+1} - x_0)}\left(A_{i^*} + B_i\right) - \frac{x_{i+1} - x_i}{x_{i+2} - x_{i+1}}\left(p_{i+2} - p_{i+1}\right) + p_{i+1},$$

and

$$d_i - c_i = \frac{x_{i+1} - x_i}{(x_{i+1} - x_0)}\left(A_{i^*} + B_i\right),$$

Therefore

$$d_i = \frac{x_{i+1} - x_i}{(x_{i+1} - x_0)}\left(A_{i^*} + B_i\right) + c_i = p_{i+1} - \frac{x_{i+1} - x_i}{x_{i+2} - x_{i+1}}\left(p_{i+2} - p_{i+1}\right),$$

and

$$c_i = -\frac{x_{i+1} - x_i}{x_{i+1} - x_0}\left(A_{i^*} + B_i\right) - \frac{x_{i+1} - x_i}{x_{i+2} - x_{i+1}}\left(p_{i+2} - p_{i+1}\right) + p_{i+1} =$$

$$= -\frac{x_{i+1} - x_i}{x_{i+1} - x_0} A_{i^*} - \frac{x_{i+1} - x_i}{x_{i+1} - x_0}\left(p_{i+2} - \frac{x_{i+2} - x_0}{x_{i+2} - x_{i+1}}\left(p_{i+2} - p_{i+1}\right)\right) - \frac{x_{i+1} - x_i}{x_{i+2} - x_{i+1}}\left(p_{i+2} - p_{i+1}\right) + p_{i+1} =$$

$$= -\frac{x_{i+1} - x_i}{x_{i+1} - x_0} A_{i^*} - \left(p_{i+2} - p_{i+1}\right)\frac{x_{i+1} - x_i}{x_{i+2} - x_{i+1}}\left(-\frac{x_{i+2} - x_0}{x_{i+1} - x_0} + 1\right) + p_{i+1} - \frac{x_{i+1} - x_i}{x_{i+1} - x_0} p_{i+2} =$$

$$= -\frac{x_{i+1} - x_i}{x_{i+1} - x_0} A_{i^*} + \left(p_{i+2} - p_{i+1}\right)\frac{x_{i+1} - x_i}{x_{i+1} - x_0} + p_{i+1} - \frac{x_{i+1} - x_i}{x_{i+1} - x_0} p_{i+2} = -\frac{x_{i+1} - x_i}{x_{i+1} - x_0} A_{i^*} + p_{i+1}\frac{x_i - x_0}{x_{i+1} - x_0} =$$

$$= -\frac{x_{i+1} - x_i}{x_{i+1} - x_0}\left[\frac{x_{i^*} - x_0}{x_{i^*+1} - x_{i^*}}\left(p_{i^*+1} - p_{i^*}\right) - p_{i^*}\right] + p_{i+1}\frac{x_i - x_0}{x_{i+1} - x_0}$$

Appendix B: Bayes theorem

Let random vectors X, Y have joint density $f(x,y)$. It is well known that $f(x,y) = f_X(x) f_{Y|X}(y \mid x) = f_Y(y) f_{X|Y}(x \mid y)$ then

$$f_{X|Y}(x \mid y) = \frac{f_X(x) f_{Y|X}(y \mid x)}{f_Y(y)}$$

whenever $f_Y(y) \neq 0$.

Appendix C: Beta is a family of conjugate priors to binomial

Suppose that:

- the prior distribution of p is $Beta(a,b)$, i.e.

  $$g(p) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}$$

  where $B(a,b)$ is Beta function, $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

- Likelihood has a binomial distribution

  $$f(x \mid p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Then the posterior distribution of p given x is

$$h(p \mid x) = \frac{f(x \mid p) g(p)}{f(x)} = \frac{f(x \mid p) g(p)}{\int_0^1 f(x \mid p) g(p) dp} =$$

$$= \frac{B(a,b)\binom{n}{x} p^x (1-p)^{n-x} p^{a-1}(1-p)^{b-1}}{B(a,b)\int_0^1 \binom{n}{x} p^{x+a-1}(1-p)^{n-x+b-1} dp} =$$

$$= \frac{p^{a+x-1}(1-p)^{n+b-x-1}}{B(a+x, n+b-x)} = Beta(a+x, n+b-x)$$

Appendix D: Proof of the theorem for the conditional posterior distribution of the percentile.

For a fixed p and any fixed k $(0 \le k \le M)$, define a c.d.f $P_k^*(x)$ with the support on $(x_k, x_{k+1}]$ as $P_k^*(x) = \dfrac{P(x) - p_k}{p_{k+1} - p_k}$ if $x_k < x \le x_{k+1}$; the randomness of $P \sim DP(\alpha P_0)$ is also inherited in P_k^*. Now let us define another c.d.f $P_{0,k}^*(x)$ on the same support $(x_k, x_{k+1}]$ as $P_{0,k}^*(x) = \dfrac{P_0(x) - P_0(x_k)}{P_0(x_{k+1}) - P_0(x_k)}$ if $x_k < x \le x_{k+1}$ (assuming known base distribution P_0). With such a c.d.f. we present the following lemma

LEMMA: $P_{k_0}^* \mid s, p \sim DP(\alpha_k P_{0,k_0}^*)$.

Proof: Let $P(x)$ be denoted by p_x. Then, with the prior model $P \sim DP(\alpha P_0)$ and using the property of the Dirichlet process, the joint posterior distribution of (p, p_x) is given by

$$\pi(p, p_x \mid s) \propto L(p \mid s)\pi(p, p_x)$$
$$\propto \prod_{i=1}^{M} p_i^{s_i}(1-p_i)^{n_i - s_i} \times \prod_{i=1}^{k}(p_i - p_{i-1})^{\lambda_i - 1}(p_x - p_k)^{\lambda_x - 1} \times (p_{k+1} - p_x)^{\lambda_x' - 1} \times \prod_{i=k}^{M+1}(p_i - p_{i-1})^{\lambda_i - 1},$$

Where $0 = p_0 < p_1 < \ldots < p_k < p_x \le p_{k+1} < \ldots < p_k < p_{M+1} = 1$ and, as defined before, $\lambda_k = \alpha\{P_0(x_i) - P_0(x_{i-1})\}$. $\lambda_x = \alpha\{P_0(x) - P_0(x_k)\}$. and $\lambda_x' = \alpha\{P_0(x_{k+1}) - P_0(x)\}$. Therefore, the conditional posterior density of p_x is given by

$$\pi(p_x \mid p, s) \propto (p_x - p_k)^{\lambda_x - 1} \times (p_{k+1} - p_x)^{\lambda_x' - 1}, \qquad (A.1)$$

where $p_k < p_x \le p_{k+1}$.

Now using the transformation $p_x \to p_x^*$, where $p_x^* = \dfrac{p_x - p_k}{p_{k+1} - p_k}$, we get from *(A.1)* that

$$\pi(p_x^* \mid p, s) \propto (p_x^*)^{\lambda_x - 1}(1 - p_x^*)^{\lambda_x' - 1}, \qquad (A.2)$$

where $0 < p_x^* \le 1$.

Therefore, for any given x such that $x_k < x \le x_{k+1}$, $\pi(p_x^* \mid p, s)$ is beta density with parameters λ_x and λ_x'.

Mukhopadhyay generalized, if $A_1, \ldots, A_m$ be any measurable partition of $(x_k, x_{k+1}]$ and $p_{A_i}^* = \dfrac{P(A_i) - p_k}{p_{k+1} - p_k}$, the conditional posterior distribution of $(p_{A_1}^*, \ldots, p_{A_m}^*)$ given p and s follows a Dirichlet distribution with parameter vector $(\lambda_{A_1}, \ldots, \lambda_{A_m})$, where $\lambda_{A_i} = \alpha P_0(A_i) = \lambda_k P_{0,k}^*(A_i)$. Hence, by definition, $P_k^* \mid s, p \sim DP(\lambda_k P_{0,k}^*)$.

Proof of theorem.
Note that the absolute continuity assumed for P_0 assures $0 = p_0 < p_1 < \ldots < p_M < p_{M+1} = 1$ with probability one and therefore, for any given γ $(0 < \gamma < 1)$, there exists k $(0 \le k \le M)$ such that $p_k < \gamma \le p_{k+1}$, i.e., $x_k < ED\gamma \le x_{k+1}$ with probability one. Thus, if $x \le x_k$, then $P(ED\gamma \le x \mid s, p) = 0$, and if $x > x_{k+1}$, then $P(ED\gamma \le x \mid s, p) = 1$. Finally, when $x_k < x \le x_{k+1}$,

$$P(ED\gamma \le x \mid s, p) = P(P(x) \ge \gamma \mid s, p) = P(P_k^*(x) \ge \frac{\gamma - p_k}{p_{k+1} - p_k} \mid s, p) = P(p_x^* \ge \gamma_k \mid s, p) =$$

$$= \frac{\Gamma(\lambda_x + \lambda_x')}{\Gamma(\lambda_x)\Gamma(\lambda_x')} \int_\tau^1 p^{\lambda_k - 1}(1-p)^{\lambda_x' - 1} dp$$

(by using *(A.2)* in the above lemma). Hence, the theorem follows.

References

[1] AMMANN,L.P. (1984). Bayesian nonparametric inference for quantal response data. Ann. Statist.12, 636-45

[2] ANTONIAK,C.E.(1974). Mixtures of Dirichlet process with applications to Bayesian nonparametric problems. Ann.Statist. 2,1152-74.

[3] AYER,M.,BRUNK,H.D.,EWING,G.M.,REID,W.T.,SiILVERMAN,E.(1955). An empirical distribution function for sampling with incomplete information. Ann.Math.Statist.,26,641-7

[4] BHATTACHARYA,P.K.(1981). Posterior distribution of a Dirichlet process from quantal response data. Ann.Statist.9,803-11

[5] DISCH,D.(1981). Bayesian nonparametric inference for effective doses in quantal response experiment. Biometrics 37,713-22

[6] FERGUSON.T.S.(1973). A Bayesian analysis of some nonparametric problems. Ann.Statist.1,209-230

[7] FINNEY, D.J. (1978). Statistical Methods in Biological Assay

[8] GEFLAND, A.E., KUO, L. (1991). Nonparametric Bayesian bioassay including ordered polytomus response. Biometrica,78,3,pp.657-66

[9] GEFLAND, A.E., SMITH, A.F.M.(1990). Sampling based approach to calculating marginals densities. J.Amer.Statist. Assoc.,85,398-409

[10] GOVINDARAJULU,Z. (1988) Statistical techniques in bioassay. S. Karger Publishers (USA) (2nd edition)

[11] MUKHOPADHYAY,S.(2000). Bayesian nonparametric inference on the dose level with specified response rate. Biometrics 56,220-226

[12] MÜLLER P. QUINTANA F.A. (2004). Nonparametric Bayesian data analysis. Satatist.Sci.19, no.1, 95-110

[13] RAMGOPAL,P., LAUD,P.W., SMITH,A.F.(1993). Nonparametric Bayesian bioassay with prior constrains on the shape of the potency curve. Biometrika,80,3,pp.489-98

[14] RAMSEY,F.L. (1972). A Bayesian approach to bio-assay. Biometrics 28, 841-58

[15] SHAKED,M. SINGPURWALLA,N.D.(1990). A Bayesian approach for quantile and response probability estimation with applications to reliability. Ann.Inst.Statist.Math.42,1-19