An Objective Bayesian Analysis of the Change Point Problem

Elías Moreno, George Casella and Antonio Garcia-Ferrer

Universidad de Granada, University of Florida and
Universidad Autónoma de Madrid

August 3, 2004

Abstract

The Bayesian literature on the change point problem deals with the inference of a change in the distribution of a set of time-ordered data based on a sample of fixed size. This is the so-called "retrospective or off-line" analysis of the change point problem. A related but different problem is that of the "sequential" change point detection, mainly analyzed from a frequentist viewpoint.

While the former typically focuses on the estimation of the position in which the change point occurs, the latter is a testing problem which has a natural formulation as a Bayesian model selection problem. In this paper we provide such a Bayesian formulation, which generalizes previous formulations such as the well-known CUSUM stopping rule.

We show that the conventional improper priors (also called non-informative, objective or default), cannot be used either for sequential detection of the change or for retrospective estimation. Then, we propose objective intrinsic prior distributions for the unknown model parameters. The normal and Poisson cases are studied in detail and examples with simulated and real data are provided.

Keywords: change point, intrinsic priors, model selection, sequential detection, retrospective estimation.

1 Introduction

There is an extensive literature on the subject of quick detection of changes in the parameters of a model of a stochastic system on the basis of sequential observations of the system, see for instance Lai (1995), and references therein. The simplest case of this problem can be described as follows. For a given experiment a sample of independent sequential observations $\{x_n, n \geq 1\}$ is available. When the experiment is under control the observations come from a specific density $f_0(x)$, and when the experiment is out-of-control at a certain point, the following observations are then coming from a known but different density $f_1(x)$. Therefore, the sampling density of the first n observations $\mathbf{x} = (x_1, ..., x_n)$ is given by

$$f_n(\mathbf{x}|r) = \begin{cases} \prod_{i=1}^{r} f_0(x_i) \prod_{i=r+1}^{n} f_1(x_i), & \text{if } 1 \leq r \leq n-1, \\ \\ \prod_{i=1}^{n} f_0(x_i), & \text{if } \quad r = n, \end{cases} \tag{1}$$

where the discrete unknown parameter r indicates a change point in the sample.

Two related problems are here of interest. One is that of testing sequentially whether the experiment is under control, i.e., for each n we want to test the null hypothesis of no change $H_0 : f_n(\mathbf{x}|n) = \prod_{i=1}^{n} f_0(x_i)$, *versus* the alternative of a change point $H_1 : \{f_n(\mathbf{x}|r), 1 \leq r \leq n-1\}$. As a consequence, we stop the experiment at the first value of n for which H_0 is rejected. The other one is the retrospective estimation of a change point based on the whole dataset.

From a Bayesian viewpoint we need to solve a testing problem which is formulated as a model selection problem between

$$\begin{aligned} M_0 &: f_n(\mathbf{x}|n), \\ M_1 &: \{f_n(\mathbf{x}|r), \pi(r|n)\}, \end{aligned} \tag{2}$$

where $\pi(r|n)$ is a prior distribution on the set $\{1, 2, ..., n-1\}$.

The paper is organized as follows. In Section 2 we formulate the sequential detection and the retrospective estimation of the change point problem. Section 3 deals with the analysis of the change point problem for parametric families of sampling distributions and presents our proposal of using intrinsic priors for the model parameters. Section 4 and 5 provide formulae for the Poisson and normal cases, respectively. Section 6 illustrates the findings using five sets of data: the British coal-mining disaster data of Jarret (1979), the simulated data of Page (1955), the industrial data analyzed in Pettitt (1979), the stock market returns studied by Hsu (1979), and the annual volume of the Nile river for the years 1871 to 1970 taken from the work of Cobb (1978). Section 7 provides concluding remarks.

2 Bayesian stopping rules

Consider a loss function $L(d_i, M_j) = c_{ij}$, where d_i represents the decision of choosing model M_i, and c_{ij} is the cost associated to the decision d_i

when the underlying model is M_j for $i, j = 0, 1$. Assuming that the cost of a correct decision is zero, that is $c_{00} = c_{11} = 0$, the optimal decision is to reject M_0 if the posterior risk $R(d_1|\mathbf{x})$ is smaller than $R(d_0|\mathbf{x})$. If we denote $P(M_0)$ and $P(M_1)$ as the prior probabilities of the two model involved, then simple calculations show that the optimal decision is to reject M_0 when the inequality

$$\frac{P(M_1|\mathbf{x})}{P(M_0|\mathbf{x})} \geq c$$

holds, where $c = c_{10}/c_{01}$ and

$$P(M_1|\mathbf{x}) = \frac{\sum_{r=1}^{n-1} f_n(\mathbf{x}|r)\pi(r|n)\ P(M_1)}{f_n(\mathbf{x}|n)P(M_0) + \sum_{r=1}^{n-1} f_n(\mathbf{x}|r)\pi(r|n)\ P(M_1)}.$$

The above inequality can be written as

$$\frac{\sum_{r=1}^{n-1} f_n(\mathbf{x}|r)\pi(r|n)}{f_n(\mathbf{x}|n)} \geq c_1,$$

where $c_1 = c\ P(M_0)/P(M_1)$.

An anonymous referee has pointed out that it might not be reasonable to maintain the loss c_{ij} constant in value during the course of an experiment. We are in complete agreement. When this is the case, the Bayesian stopping rules below can be easily adapted to this setting.

A default (also called objective) choice for $P(M_0)$ is 1/2 for which the resulting Bayesian stopping rule is

$$N^{\pi} = \inf\left\{n : \frac{\sum_{r=1}^{n-1} f_n(\mathbf{x}|r)\pi(r|n)}{f_n(\mathbf{x}|n)} \geq c\right\}. \tag{3}$$

The prior $\pi(r|n)$ is typically chosen as the uniform distribution $\pi(r|n) = 1/(n-1)$ (Chernoff and Zacks 1965, Smith 1975, Carlin et al. 1992), for which the Bayesian stopping rule becomes

$$N^{U} = \inf\left\{n : \frac{1}{(n-1)}\frac{\sum_{r=1}^{n-1} f_n(\mathbf{x}|r)}{f_n(\mathbf{x}|n)} \geq c\right\}. \tag{4}$$

If $\pi(r|n)$ in (3) is chosen to be a degenerate distribution on the maximum likelihood estimator of r (this is a non-orthodox data dependent prior), then the likelihood ratio test is obtained and the corresponding stopping rule would then be

$$N^{MLE} = \inf\left\{n : \max_{1\leq r\leq n-1} \prod_{i=r+1}^{n} \frac{f_1(x_i)}{f_0(x_i)} \geq c\right\}. \tag{5}$$

This is the CUSUM stopping rule introduced by Page (1954).

An extension of CUSUM to allow the single density f_1 to describe the out-of-control distribution to belong to the exponential family $f_\theta(x)$ have been considered by Lorden (1971). The proposed stopping rule has the form

$$N^L = \inf\left\{ n : \max_{1\le r\le n-1} \sup_{\theta\in\Theta} \prod_{i=r+1}^{n} \frac{f_\theta(x_i)}{f_0(x_i)} \ge c \right\}.$$

More general situations where $f_0(x)$ and $f_1(x)$ belong to a family of densities have already been considered, see, for instance, Sen and Srivastava (1973), Hsu (1979), Pettitt (1979), Worsley (1986), Siegmund (1988), Pollak and Siegmund (1991) and Müller (1992), among others.

2.1 Retrospective estimation of a change point

Suppose that a sample of fixed size n is drawn from the sampling model with a change point $f_n(\mathbf{x}|r)$, where $1 \le r \le n-1$. In this setting, the so-called retrospective analysis refers to the problem of estimating r, the position of the change point. For a prior $\pi(r|n)$, the Bayesian estimation of r is based on the posterior distribution

$$\pi(r|\mathbf{x}) = \frac{f_n(\mathbf{x}|r)\pi(r|n)}{\sum_{r=1}^{n-1} f_n(\mathbf{x}|r)\pi(r|n)}, \quad 1 \le r \le n-1.$$

Sometimes, we may assume that several changes might occur. Then, model (2) should be modify accordingly. Note, however, that in the sequential detection of a first change no assumption of this nature is required.

3 The parametric change point problem

The Bayesian literature on the change point problem focuses on the retrospective analysis for situations where $f_0(x)$ and $f_1(x)$ belong to a parametric family of densities; see, for instance, Ferreira (1975), Smith (1975), Choy and Broemeling (1980), Smith and Cook (1980), Menzefrike (1981), Raftery and Akman (1986), Carlin et al.(1992), among others. When $f_0(x)$ and $f_1(x)$ are densities belonging to a parametric family to be denoted as $f(x|\theta_1)$ and $f(x|\theta_2)$ respectively, where θ_1 and θ_2 are unknown points of a parameter space Θ, the sampling density of the first n observations is

$$f_n(\mathbf{x}|r,\theta_1,\theta_2) = \prod_{i=1}^{r} f(x_i|\theta_1) \prod_{i=r+1}^{n} f(x_i|\theta_2), \quad \text{if} \quad 1 \le r \le n-1, \tag{6}$$

and

$$f_n(\mathbf{x}|n,\theta) = \prod_{i=1}^{n} f(x_i|\theta), \quad \text{if} \quad r = n. \tag{7}$$

We notice that $f_n(\mathbf{x}|n,\theta)$ is nested in $f_n(\mathbf{x}|r,\theta_1,\theta_2)$ for any $1 \le r \le n$, a useful property that we will use later.

For some priors distributions $\pi_0(\theta)$, $\pi_1(\theta_1,\theta_2)$, and the loss function stated in Section 2, the Bayesian stopping rule is now given by

$$N^{\pi_0,\pi_1} = \inf\left\{ n : \frac{\sum_{r=1}^{n-1} \pi(r|n) \int f_n(\mathbf{x}|r,\theta_1,\theta_2)\pi_1(\theta_1,\theta_2)d\theta_1 d\theta_2}{\int f_n(\mathbf{x}|n,\theta)\pi_0(\theta)d\theta} \ge c \right\},$$

where as before $c = c_{10}/c_{01}$. This stopping rule can also be written as

$$N^{\pi_0,\pi_1} = \inf\left\{ n : T_n(\mathbf{x}) \ge c \right\}, \tag{8}$$

where $T_n(\mathbf{x}) = \sum_{r=1}^{n-1} B_{rn}(\mathbf{x})\pi(r|n)$ and

$$B_{rn}(\mathbf{x}) = \frac{\int f_n(\mathbf{x}|r,\theta_1,\theta_2)\pi_1(\theta_1,\theta_2)d\theta_1 d\theta_2}{\int f_n(\mathbf{x}|n,\theta)\pi_0(\theta)d\theta}, \tag{9}$$

is the Bayes factor for comparing model (6) and (7) for a fixed r.

Given a sample of size n, and assuming that one change point has occurred before n with prior probability $\pi(r|n)$, the posterior probability that the change point is located at position r is computed as

$$\pi(r|\mathbf{x}) = \frac{\pi(r|n) \int f_n(\mathbf{x}|r,\theta_1,\theta_2)\pi_1(\theta_1,\theta_2)d\theta_1 d\theta_2}{\sum_{r=1}^{n-1} \pi(r|n) \int f_n(\mathbf{x}|r,\theta_1,\theta_2)\pi_1(\theta_1,\theta_2)d\theta_1 d\theta_2}, \tag{10}$$

for $1 \le r \le n-1$.

Sometimes we are interested in the magnitude of the change. For instance, when θ_1 and θ_2 are either scale or location parameters then the posterior density of $\varphi(\theta_1,\theta_2) = \theta_1/\theta_2$ or $\varphi(\theta_1,\theta_2) = \theta_1 - \theta_2$ might be of interest. The posterior density of φ can be obtained from the posterior density

$$\pi(\theta_1,\theta_2|\mathbf{x},n) = \frac{\sum_{r=1}^{n-1} f_n(\mathbf{x}|r,\theta_1,\theta_2)\pi_1(\theta_1,\theta_2)\pi(r|n)}{\sum_{r=1}^{n-1} \pi(r|n) \int f_n(\mathbf{x}|r,\theta_1,\theta_2)\pi_1(\theta_1,\theta_2)d\theta_1 d\theta_2}, \tag{11}$$

with an appropriate change of variables.

We note that while the Bayesian stopping rule (8) depends on the prior distributions $\pi_0(\theta)$ and $\pi_1(\theta_1,\theta_2)$, the posterior distributions (10) and (11) only depend on the prior $\pi_1(\theta_1,\theta_2)$. This makes sense since in the derivation of (10) and (11) we have assumed that before n a change has occurred.

3.1 Objective Bayesian methods

Both the retrospective analysis and the sequential detection of a change point, share the difficulty of assessing prior distributions to the unknown parameters. On parameter r a uniform prior is the common choice, and for the sampling densities parameters either conjugate priors or vague priors, which are obtained as a limit of conjugate priors with respect to some of the hyperparameters, are typically used.

However, the use of such a priors is far from simple: conjugate priors need to assess values for the hyperparameters, so that some sort of subjective input or empirical Bayes estimation is necessary. Further, the use of vague priors hides the fact that they are improper. On the other hand, the Jeffreys's or the reference (Berger and Bernardo 1992) priors are typically improper. Unfortunately, for improper priors the resulting marginals of the data are not well-defined, in fact they are defined up to an arbitrary positive constant. Thus, when using such a type of priors neither the posterior distribution of r nor the Bayesian stopping rule do exist. An alternative approach is the use of recent methods (called objective) based on partial Bayes factors or in intrinsic priors, which have become very popular (Berger 2000, Clyde 2001, Kim and Sun 2000, Wasserman 2000, Berger et al. 2001, Sweeting 2001, Casella and Moreno 2003).

In this paper, we propose the use of intrinsic priors for the unknown parameters of the models involved. They are derived from the arithmetic intrinsic Bayes factor (Berger and Pericchi 1996) plus an asymptotic argument (Moreno *et al.* 1998). Interestingly, among the existing objective procedures this seems to be the only one that can be employed for the change point problem. This assertion follows from the fact that most of the objective methods use real samples for training the improper prior, and such a real training samples might not exist for the change point model. Indeed, in a retrospective analysis, conditional on a sample of fixed size n, a change point might occur at a position smaller in value than the minimal training sample size and hence no training sample exists. On the other hand, for the sequential detection of the change point, the stopping rule assumes that a change may occur at any position, and therefore the same difficulty arises.

Intrinsic priors, however, do not use real training samples, but theoretical training ones, and hence the difficulty with the absence of a real training sample dissappear. Furthermore, reasons for using intrinsic priors include (i) they are free of hyperparameters, (ii) they provide a well-defined Bayesian solution for testing problems, and (iii) under the intrinsic prior distribution, the parameters in the alternative model are not independent and they are "centered" at the null, a condition widely required in testing scenarios (Jeffreys 1961, Berger and Sellke 1987, Casella and Berger 1987, Morris 1987). Moreover, intrinsic priors have proved to behave extremely well for a wide variety of problems (see, for instance, Moreno *et al.* 1999, 2000, 2003, Moreno and Liseo, 2003. Kim and Sun 2000, Casella and Moreno 2002, 2003a, 2003b). The main inconvenient of the intrinsic priors for complex models is that they are difficult to compute. In particular, a closed form for them is typically not available.

3.2 Intrinsic priors

The essential feature of the change point problem is that the observations come from either model $f(x|\theta)$ or $f(x|\theta_1)f(x|\theta_2)$. Hence the canonical form of the problem is that of testing the null $H_0 : f(x|\theta)$ *versus* $H_1 : f(x|\theta_1)f(x|\theta_2)$. Let us denote $\pi^D(\theta)$ the conventional improper prior for θ. For nested scenarios, as is our case, an standard solution to the difficulties mentioned in subsection 3.1 is to replace the improper prior $\pi^D(\theta_1)\pi^D(\theta_2)$ with the intrinsic prior $\pi^I(\theta_1, \theta_2)$, so that the model selection problem becomes that of choosing between M_0 and M_1

$$M_0 : \{f_n(\mathbf{x}|n, \theta), \pi^D(\theta)\},$$

$$M_1 : \{f_n(\mathbf{x}|r, \theta_1, \theta_2), \pi^I(\theta_1, \theta_2)\pi(r|n)\}.$$

The intrinsic prior (Berger and Pericchi 1996; Moreno et al. 1998) for the parameters θ_1, θ_2 is given by

$$\pi^I(\theta_1, \theta_2) = \pi^D(\theta_1)\pi^D(\theta_2)E_{X|\theta_1,\theta_2}B^D_{01}(X_1(\ell), X_2(\ell)), \tag{12}$$

where

$$B^D_{01}(X_1(\ell), X_2(\ell)) = \frac{m^D_0(X_1(\ell), X_2(\ell))}{m^D_1(X_1(\ell), X_2(\ell))}.$$

In this expression $X = (X_1(\ell), X_2(\ell))$ is a random vector with 2ℓ i.i.d. components, and ℓ is the minimum size such that the marginal

$$m^D(X_1(\ell)) = \int f(X_1(\ell)|\theta_1)\pi^D(\theta_1)d\theta_1$$

is positive and finite a.s.. The pair $\{\pi^D(\theta), \pi^I(\theta_1, \theta_2)\}$ is called intrinsic priors.

The use of intrinsic priors has many advantages that we briefly summarize.

i) The Bayes factor (9) for the intrinsic priors $\{\pi^D(\theta), \pi^I(\theta_1, \theta_2)\}$,

$$B^{IP}_{rn}(\mathbf{x}) = \frac{\int f_n(\mathbf{x}|r, \theta_1, \theta_2)\pi^I(\theta_1, \theta_2)d\theta_1 d\theta_2}{\int f_n(\mathbf{x}|n, \theta)\pi^D(\theta)d\theta}, \ 1 \le r \le n-1, \tag{13}$$

is well defined as it does not depend on any arbitrary constant.
ii) The parameters θ_1 and θ_2 are dependent under the intrinsic prior $\pi^I(\theta_1, \theta_2)$, which seems to be reasonable.
iii) The intrinsic prior $\pi^I(\theta_1, \theta_2)$ can also be written as the mixture

$$\pi^I(\theta_1, \theta_2) = \int \pi^I(\theta_1, \theta_2|\theta)\pi^D(\theta)d\theta.$$

This means that the intrinsic prior is "centered at the null" (Casella and Moreno, 2002).
iv) The intrinsic priors are fully determined, so there is no parameter to be adjusted.

3.3 Intrinsic stopping rule and posterior probability of a change point

For the intrinsic priors $\{\pi^D(\theta), \pi^I(\theta_1, \theta_2)\}$ and the uniform $\pi(r|n) = 1/(n-1)$ the Bayesian stopping rule (8) becomes

$$N^{IP} = \inf\left\{n : \frac{1}{n-1}\sum_{r=1}^{n-1} B_{rn}^{IP}(\mathbf{x}) \geq c\right\}, \tag{14}$$

where the Bayes factor inside the bracket is given in (13) .

On the other hand, the posterior probability that a change occurs at position r, conditional on a sample of size n, when using the intrinsic $\pi^I(\theta_1, \theta_2)$ and the uniform $\pi(r|n) = 1/(n-1)$ priors, is given by

$$\pi(r|\mathbf{x}, n) = \frac{\int f_n(\mathbf{x}|r, \theta_1, \theta_2)\pi^I(\theta_1, \theta_2)d\theta_1 d\theta_2}{\sum_{r=1}^{n-1}\int f_n(\mathbf{x}|r, \theta_1, \theta_2)\pi^I(\theta_1, \theta_2)d\theta_1 d\theta_2}, \tag{15}$$

for $1 \leq r \leq n-1$.

4 The Poisson case

Let us assume that the sampling density is a Poisson distribution. For a sample $\mathbf{x} = (x_1, ..., x_n)$ the likelihood function under the null of no change is

$$f_n(\mathbf{x}|n, \theta) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} \exp\{-\theta\}, \; \theta > 0,$$

and is

$$f_n(\mathbf{x}|r, \theta_1, \theta_2) = \prod_{i=1}^{r} \frac{\theta_1^{x_i}}{x_i!} \exp\{-\theta_1\} \prod_{i=r+1}^{n} \frac{\theta_2^{x_i}}{x_i!} \exp\{-\theta_2\}, \; if \;\; 1 \leq r \leq n-1,$$

under the alternative that a change occurs. The Jeffreys prior for the Poisson parameter is the improper prior $\pi^D(\theta) = k\ \theta^{-1/2}$, where k is an arbitrary positive constant.

Lemma 1. The intrinsic prior for θ_1, θ_2, conditional on θ, is

$$\pi^I(\theta_1, \theta_2|\theta) = \prod_{i=1}^{2} \theta_i^{-1/2} \frac{\exp\{-(\theta+\theta_i)\}}{\Gamma(1/2)}\ F_0^1(1/2, \theta\theta_i), \tag{16}$$

where F_0^1 denotes the hypergeometric function (Abramowitz and Stegun 1970).

Proof. See Appendix 1.

By construction this prior is a probability density for any value of θ, so that $\int \pi^I(\theta_1, \theta_2|\theta) d\theta_1 d\theta_2 = 1$. This density factorizes as $\pi^I(\theta_1, \theta_2|\theta) = \pi^I(\theta_1|\theta)\pi^I(\theta_2|\theta)$, so that the parameters are independent, conditional on θ. The marginal distribution of θ_i, conditional on θ, is a unimodal continuous density with mode at 0 for $i = 1, 2$.

The unconditional intrinsic prior for θ_1, θ_2,

$$\pi^I(\theta_1, \theta_2) = k \int \pi^I(\theta_1, \theta_2|\theta) \frac{1}{\theta^{1/2}} d\theta,$$

and they are not independent. It is an improper distribution, as it is typically the case for a non subjective prior, although the Bayes factor $B_{rn}^{IP}(x)$ for the intrinsic priors $\{\pi^D(\theta), \pi^I(\theta_1, \theta_2)\}$ is well defined.

4.1 Bayes factor for intrinsic priors

Using some algebra it can shown that the marginal of the sample $\mathbf{x}$ under model $\{f_n(\mathbf{x}|r, \theta_1, \theta_2), \pi^I(\theta_1, \theta_2|\theta)\}$, for $1 \le r \le n-1$, can be written as

$$m_r(\mathbf{x}|\theta) = k(r, \mathbf{x})\ exp\{-2\theta\}$$

$$\times F_1^1(r\bar{x}_1(r) + 1/2, 1/2, \frac{\theta}{r+1}) F_1^1((n-r)\bar{x}_2(r) + 1/2, 1/2, \frac{\theta}{n-r+1}), \tag{17}$$

where F_1^1 denotes the Kummer function (Abramowitz and Stegun 1970),

$$k(r, \mathbf{x}) = \frac{1}{\prod_{i=1}^n x_i!} \left(\frac{1}{r+1}\right)^{r\bar{x}_1(r)+1/2} \left(\frac{1}{n-r+1}\right)^{(n-r)\bar{x}_2(r)+1/2}$$
$$\times \frac{\Gamma(r\bar{x}_1(r) + 1/2)\Gamma((n-r)\bar{x}_2(r) + 1/2)}{\Gamma(1/2)\Gamma(1/2)},$$

and

$$\bar{x}_1(r) = \frac{1}{r}\sum_{i=1}^{r} x_i, \quad \bar{x}_2(r) = \frac{1}{n-r}\sum_{i=r+1}^{n} x_i.$$

On the other hand, under no change the marginal $m_n(\mathbf{x}|\theta)$ is simply

$$m_n(\mathbf{x}|\theta) = \frac{1}{\prod_{i=1}^n x_i!}\theta^{n\bar{x}} \exp\{-n\theta\}, \tag{18}$$

where $\bar{x} = \sum_{i=1}^n x_i/n$. From (17) and (18) the Bayes factor for intrinsic prior $B_{rn}^{IP}(\mathbf{x})$, is

$$\begin{aligned} B_{rn}^{IP}(\mathbf{x}) &= \frac{\int_0^\infty m_r(\mathbf{x}|\theta)\theta^{-1/2}d\theta}{\int_0^\infty m_n(\mathbf{x}|\theta)\theta^{-1/2}d\theta} \\ &= \left(\prod_{i=1}^n x_i!\right) \frac{n^{n\bar{x}+1/2}}{\Gamma(n\bar{x}+1/2)} \int_0^\infty m_r(\mathbf{x}|\theta)\theta^{-1/2}d\theta. \end{aligned} \tag{19}$$

We note that in this expression the statistic $\prod_{i=1}^n x_i!$ cancels out, so that the Bayes factor only depends on the sufficient statistics $(\bar{x}_1(r),\ \bar{x}_2(r), \bar{x})$. We also note that the computation of $B_{rn}^{IP}(\mathbf{x})$ requires a one-dimensional numerical integration, which can be done with the help of a standard package.

Once $\{B_{rn}^{IP}(\mathbf{x}),\ 1 \leq r \leq n-1\}$ is found the Bayesian stopping rule (14) for the Poisson case follows immediately. Likewise, the posterior probability (15) turns out to be

$$\pi(r|\mathbf{x}, n) = \frac{\int_0^\infty m_r(\mathbf{x}|\theta)\theta^{-1/2}d\theta}{\sum_{r=1}^{n-1}\int_0^\infty m_r(\mathbf{x}|\theta)\theta^{-1/2}d\theta}, 1 \leq r \leq n-1. \tag{20}$$

5 The normal case

For the normal family of densities and assuming that the observations $\mathbf{x} = (x_1, ..., x_n)$ are independent, the likelihood function under the null of no change is

$$f(\mathbf{x}|\theta, \tau, n) = \prod_{i=1}^n N(x_i|\theta, \tau^2),$$

where the location θ and scale τ parameters are unknown. Under the alternative of a change point at position r,

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}, r) = \prod_{i=1}^r N(x_i|\mu_1, \sigma_1^2) \prod_{i=r+1}^n N(x_i|\mu_2, \sigma_2^2),\ if\ \ 1 \leq r \leq n-1, \tag{21}$$

is the likelihood function for the parameters $\boldsymbol{\mu} = (\mu_1, \mu_2)$, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)$ and r.

Intrinsic priors for the change point problem are derived by comparing the two models

$$M_0 : \left\{N(x|\theta, \tau^2),\ \pi^D(\theta, \tau) = \frac{k_0}{\tau}\right\},$$

and

$$M_1 : \left\{ N(x|\mu_1, \sigma_1^2) N(y|\mu_2, \sigma_2^2),\ \pi^D(\boldsymbol{\mu}, \boldsymbol{\sigma}) = \frac{k_1}{\sigma_1 \sigma_2} \right\},$$

where π^D represents the reference prior, and c_0 and c_1 are arbitrary positive constants.

Lemma 2. In comparing M_0 versus M_1 the intrinsic prior for (μ, σ), conditional on (θ, τ), is

$$\pi^I(\boldsymbol{\mu}, \boldsymbol{\sigma}|\theta, \tau) = \prod_{i=1}^{2} N(\mu_i|\theta, \frac{\sigma_i^2 + \tau^2}{2}) HC^+(\sigma_i|0, \tau),$$

where HC^+ denotes a half Cauchy distribution on the positive part of the real line. The unconditional intrinsic prior is

$$\pi^I(\boldsymbol{\mu}, \boldsymbol{\sigma}) = \int \pi^I(\boldsymbol{\mu}, \boldsymbol{\sigma}|\theta, \tau)\ \pi^D(\theta, \tau) d\theta d\tau.$$

Proof. See Appendix 2.

By construction the intrinsic prior $\pi^I(\boldsymbol{\mu}, \boldsymbol{\sigma}|\theta, \tau)$, conditional on (θ, τ), is a probability density. The marginal intrinsic prior for μ_i conditional on θ and τ, is a unimodal distribution with mode at θ, mean θ, and it has very heavy tails. Indeed, it does not have moments of order greater than one; this follows from the fact that all moments of the mixing distribution are infinite. This behaviour is close to that advocated by Jeffreys (1961) when testing that the mean of a normal distribution has a zero value. He proposed a Cauchy prior for the mean and the conventional default improper prior, $1/\sigma$ say, for the nuisance parameter σ. Note, however, that the marginal intrinsic prior for σ_i, conditional on τ, is a proper distribution.

However, although the unconditional intrinsic prior $\pi^I(\boldsymbol{\mu}, \boldsymbol{\sigma})$ is an improper density, the pair $(\pi^D(\theta, \tau), \pi^I(\boldsymbol{\mu}, \boldsymbol{\sigma}))$ depends on the same arbitrary positive constant k_0 so that the corresponding Bayes factor is well defined (some authors refer this as to well-calibrated distributions).

5.1 Bayes factor for intrinsic priors

Given the sample $\mathbf{x} = (x_1, ..., x_n)$, we set

$$\bar{x}_1(r) = \frac{1}{r} \sum_{i=1}^{r} x_i, \quad \bar{x}_2(r) = \frac{1}{n-r} \sum_{i=r+1}^{n} x_i$$

$$s_1^2(r) = \frac{1}{r} \sum_{i=1}^{r} (x_i - \bar{x}_1(r))^2, \quad s_2^2(r) = \frac{1}{n-r} \sum_{i=r+1}^{n} (x_i - \bar{x}_2(r))^2.$$

The Bayes factor $B_{rn}^{IP}(\mathbf{x})$ for the intrinsic priors $\{\pi^I(\boldsymbol{\mu},\boldsymbol{\sigma}),\ \pi^D(\theta,\tau)\}$ is computed as follows. The marginal of $\mathbf{x}$ under model M_r with respect to $\pi^I(\boldsymbol{\mu},\boldsymbol{\sigma})$ is given by

$$m(\mathbf{x}|r) = k(r)\ I(r,\mathbf{x}),$$

where

$$k(r) = k_0\ \frac{2\Gamma(n/2)}{\pi^2(2\pi)^{(n-1)/2}(r(n-r))^{1/2}},$$

$$I(r,\mathbf{x}) = \int_0^{\pi/2}\int_0^{\pi/2} \frac{(\cos\psi)^{-(n-2)}\sin(\psi)(\cos\varphi)^{-(r-2)}(\sin\varphi)^{-(n-r-2)}}{\left(\frac{\cos^2\psi\cos^2\varphi}{r}+\frac{\cos^2\psi\sin^2\varphi}{n-r}+\frac{\cos^2\psi}{2}+\sin^2\psi\right)^{1/2}}$$

$$\times\frac{A(r,\mathbf{x},\psi,\varphi)^{-n/2}}{(\cos^2\psi\cos^2\varphi+\sin^2\psi)(\cos^2\psi\sin^2\varphi+\sin^2\psi)}d\varphi d\psi,$$

and

$$A(r,\mathbf{x},\psi,\varphi) = \frac{rs_1^2(r)}{2\cos^2\psi\cos^2\varphi}+\frac{(n-r)s_2^2(r)}{2\cos^2\psi\sin^2\varphi}$$

$$+\frac{(\bar{x}_1(r)-\bar{x}_2(r))^2}{2(\cos^2\psi\cos^2\varphi/r+\cos^2\psi\sin^2\varphi/(n-r)+\cos^2\psi/2+\sin^2\psi)}.$$

The marginal $m(\mathbf{x}|n)$ with respect to $\pi^D(\theta,\tau)$ is given by

$$m(\mathbf{x}|n) = k_0\frac{\Gamma((n-1)/2)}{2\pi^{(n-1)/2}n^{n/2}s_1(n)^{(n-1)}}.$$

The Bayes factor for intrinsic prior $B_{rn}^{IP}(\mathbf{x})$ is the ratio

$$B_{rn}^{IP}(\mathbf{x}) = \frac{m(\mathbf{x}|r)}{m(\mathbf{x}|n)},\quad 1\le r\le n-1,$$

We remark that the computation of $B_{rn}^{IP}(\mathbf{x})$ requires a two-dimensional numerical integration on the square $[0,\pi/2]^2$. With the values $B_{rn}^{IP}(\mathbf{x})$ in (14) the Bayesian stopping rule for the normal family is obtained. The retrospective analysis, conditional to a sample size n, is obtained immediately from (20).

5.2 The homoscedastic case

For some applications it is reasonable to assume that $\sigma_1=\sigma_2$. Then, the sampling model $f(\mathbf{x}|\ \boldsymbol{\mu},\boldsymbol{\sigma},r)$ in (21) becomes

$$f_1(\mathbf{x}|\boldsymbol{\mu},\sigma,r) = \prod_{i=1}^{r} N(x_i|\mu_1,\sigma^2)\prod_{i=r+1}^{n} N(x_i|\mu_2,\sigma^2). \tag{22}$$

This likelihood involves a parameter space with lower dimension than that of (21), so that it is of interest to test whether or not the homoscedastic condition can be accepted.

For testing homoscedasticity the Bayesian nested models to be compared are

$$\{f_1(\mathbf{x}|\boldsymbol{\eta},\sigma,r),\pi_1^N(\boldsymbol{\eta},\sigma,r)=\frac{c_0}{\sigma}\frac{1}{n-1}\},\tag{23}$$

$$\{f(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma},r),\pi^N(\boldsymbol{\mu},\boldsymbol{\sigma},r)=\frac{c_1}{\sigma_1\sigma_2}\frac{1}{n-1}\},\tag{24}$$

where c_0 and c_1 are arbitrary positive constants.

Lemma 3. When comparing model (23) and (24) the intrinsic prior for $(\boldsymbol{\mu},\boldsymbol{\sigma})$, conditional on (η,σ), is

$$\pi^I(\boldsymbol{\mu},\boldsymbol{\sigma}|\boldsymbol{\eta},\sigma)=\prod_{i=1}^{2}N(\mu_i|\eta_i,\frac{\sigma_i^2+\sigma^2}{2})HC^+(\sigma_i|0,\sigma).\tag{25}$$

The unconditional intrinsic prior for $(\boldsymbol{\mu},\boldsymbol{\sigma})$ is

$$\pi^I(\boldsymbol{\mu},\boldsymbol{\sigma})=c_0\int\left\{\prod_{i=1}^{2}HC^+(\sigma_i|0,\sigma)\right\}\frac{1}{\sigma}d\sigma.$$

Proof.- The proof is similar to that of the Lemma 3 and hence omitted.

Note that under the unconditional intrinsic prior, σ_1 and σ_2 are independent and $\boldsymbol{\mu}$ is uniformly distributed on the plane. The Bayes factor for comparing model (23) and (24) when using the intrinsic prior is

$$B_{12}^{IP}(\mathbf{x})=\frac{m_1(\mathbf{x})}{m_2(\mathbf{x})}=\frac{\sum_{r=1}^{n-1}m_1(\mathbf{x}|r)}{\sum_{r=1}^{n-1}m_2(\mathbf{x}|r)},$$

where

$$m_1(\mathbf{x}|r)=\frac{c_0}{2\pi^{(n-2)/2}\sqrt{r(n-r)}}\frac{\Gamma(\frac{n-2}{2})}{(rs_1^2(r)+(n-r)s_2^2(r))^{(n-2)/2}},$$

$$m_2(\mathbf{x}|r)=k(r)I_3(r,\mathbf{x}),$$

and

$$I_3(r,\mathbf{x})=\int_0^{\pi/2}\int_0^{\pi/2}\frac{(\cos\psi)^{-(n-3)}\sin(\psi)(\cos\varphi)^{-(r-1)}(\sin\varphi)^{-(n-r-1)}}{(\cos^2\psi\cos^2\varphi+\sin^2\psi)}$$
$$\times\frac{A_1(r,\mathbf{x},\psi,\varphi)^{-(n-1)/2}}{(\cos^2\psi\sin^2\varphi+\sin^2\psi)}d\varphi d\psi,$$

with

$$A_1(r,\mathbf{x},\psi,\varphi)=\frac{rs_1^2(r)}{2\cos^2\psi\cos^2\varphi}+\frac{(n-r)s_2^2(r)}{2\cos^2\psi\sin^2\varphi}.$$

5.3 Bayesian stopping rule and posterior probability of a change

Conditional on the homoscedastic condition, the likelihood function of the sample $\mathbf{x}$ is either $f_1(\mathbf{x}|\boldsymbol{\mu},\sigma,r)$ as given in (22) for some $r=1,...,n-1$, or $f(\mathbf{x}|\theta,\tau,n)$. The Bayesian stopping rule needs the Bayes factor for comparing models

$$M_r : \{f(\mathbf{x}|\theta,\tau,n), \pi^D(\theta,\tau)=c_0/\tau\}, \tag{26}$$

and

$$M_n : \{f_1(\mathbf{x}|\boldsymbol{\mu},\sigma,r), \pi^D(\boldsymbol{\mu},\sigma)=c_1/\sigma\}, \tag{27}$$

where π^D represents the reference prior distribution (Berger and Bernardo 1992). As before, we compute the intrinsic priors for the models involved.

Lemma 4. When comparing model (26) versus (27) the intrinsic prior for $(\boldsymbol{\mu},\sigma)$, conditional on (θ,τ), is given by

$$\pi^I(\boldsymbol{\mu},\sigma|\theta,\tau) = N(\mu_1|\theta,\frac{\sigma^2+\tau^2}{2})N(\mu_2|\theta,\sigma^2+\tau^2)HC^+(\sigma|0,\tau).$$

Proof. The proof is similar to that of Lemma 3 and hence omitted.

We note that under the conditional intrinsic prior μ_1 and μ_2 are independent but they are not identically distributed. The reason is that the theoretical minimal training sample we have considered involves three independent random variables, two of them with normal distribution $N(x|\mu_1,\sigma^2)$, and the third one distributed as $N(x|\mu_2,\sigma^2)$. If, instead, we consider instead two of them as $N(x|\mu_2,\sigma^2)$ distributed and the third one as $N(x|\mu_1,\sigma^2)$, then we obtain the conditional intrinsic prior as

$$\pi(\boldsymbol{\mu},\sigma|\theta,\tau) = N(\mu_1|\theta,\sigma^2+\tau^2)N(\mu_2|\theta,\frac{\sigma^2+\tau^2}{2})HC^+(\sigma|0,\tau).$$

However, the resulting Bayes factors by using either $\pi^I(\boldsymbol{\mu},\sigma|\theta,\tau)$ or $\pi(\boldsymbol{\mu},\sigma|\theta,\tau)$ are exactly the same.

For comparing model $\{f_1(\mathbf{x}|\boldsymbol{\mu},\sigma,r), \pi^I(\boldsymbol{\mu},\sigma|\theta,\tau)\pi^N(\theta,\tau)\}$ and $\{f(\mathbf{x}|\theta,\tau,n), \pi^N(\theta,\tau)\}$, the Bayes factor is $B_{rn}^{IP}(\mathbf{x}) = m^{IP}(\mathbf{x}|r)/m^{IP}(\mathbf{x}|n)$ where

$$m^{IP}(\mathbf{x}|n) = k_0\frac{\Gamma((n-1)/2}{2\pi^{(n-1)/2}n^{n/2}s_1(n)^{(n-1)}},$$

and

$$m^{IP}(\mathbf{x}|r) = k_1(r)I_4(r,\mathbf{x}),$$

with

$$k_1(r) = c_0\frac{\Gamma(\frac{n-1}{2})}{2^{(n-1)/2}\pi^{(n+1)/2}\sqrt{r(n-r)}},$$

$$I_4(r,\mathbf{x}) = \int_0^{\pi/2} \frac{C(r,\mathbf{x},\varphi)^{-(n-1)/2}}{(\sin\varphi)^{n-2}\sqrt{\frac{n}{r(n-r)}\sin^2\varphi + \frac{3}{2}}} d\varphi,$$

and

$$C(r,\mathbf{x},\varphi) = \frac{rs_1^2(r) + (n-r)s_2^2(r)}{2\sin^2\varphi} + \frac{(\bar{x}_1(r) - \bar{x}_2(r))^2}{2\left(\frac{n}{r(n-r)}\sin^2\varphi + \frac{3}{2}\right)}.$$

The Bayesian stopping rule follows from expression (14) with the above Bayes factor. The posterior distribution of the change point follows from expression (20) with the marginals $\{m^{IP}(\mathbf{x}|r),\ 1 \leq r \leq n\}$.

6 Some examples

We apply the results given in Sections 3 and 4 to five datasets: the intervals of time between coal-mine disasters (Jarret 1979), the simulated data of Page (1955), the industrial data considered by Pettitt (1979), the stock rates of return data given in Hsu (1979), and the Nile River flows of Cobb (1978).

Example 1.

Consider the data of the number of intervals between British coal-mining disasters during the 112-year period, between 1851-1962 (Jarret 1979). The observed annual counts given in Table 1, have been taken from Carlin *et al.* (1992).

Table 1: British coal-mining disaster data by year, 1851-1962. C=Count

Year	**C**	**Year**	**C**	**Year**	**C**	**Year**	**C**	**Year**	**C**	**Year**	**C**
1851	4	1871	5	1891	2	1911	0	1931	3	1951	1
1852	5	1872	3	1892	1	1912	1	1932	3	1952	0
1853	4	1873	1	1893	1	1913	1	1933	1	1953	0
1854	1	1874	4	1894	1	1914	1	1934	1	1954	0
1855	0	1875	4	1895	1	1915	0	1935	2	1955	0
1856	4	1876	1	1896	3	1916	1	1936	1	1956	0
1857	3	1877	5	1897	0	1917	0	1937	1	1957	1
1858	4	1878	5	1898	0	1918	1	1938	1	1958	0
1859	0	189	3	1899	1	1919	0	1939	1	1959	0
1860	6	1880	4	1900	0	1920	0	1940	2	1960	1
1861	3	1881	2	1901	1	1921	0	1941	4	1961	0
1862	3	1882	5	1902	1	1922	2	1942	2	1962	1
1863	4	1883	2	1903	0	1923	1	1943	0		
1864	0	1884	2	1904	0	1924	0	1944	0		
1865	2	1885	3	1905	3	1925	0	1945	0		
1866	6	1886	4	1906	1	1926	0	1946	1		
1867	3	1887	2	1907	0	1927	1	1947	4		
1868	3	1888	1	1908	3	1928	1	1948	0		
1869	5	1889	3	1909	2	1929	0	1949	0		
1870	4	1890	2	1910	2	1930	2	1950	0		

Frequentist change point analyses of these data are found in Jarret (1979), Rudemo (1982), Worsley (1986) and Siegmund (1988). Bayesian analyses have been given by Raftery and Akman (1986) and Carlin *et al.* (1992).

We assume that the data come from the Poisson family of distributions as suggested by Jarret (1979) and Rudemo (1982). Figure 1 shows the values of the stopping rule T_n given by expression (8), when using the intrinsic priors, for $n \leq 46$. These values are quite small for n below 46, but afterwards T_n increases considerably, giving strong evidence for a change point in the year 1896.

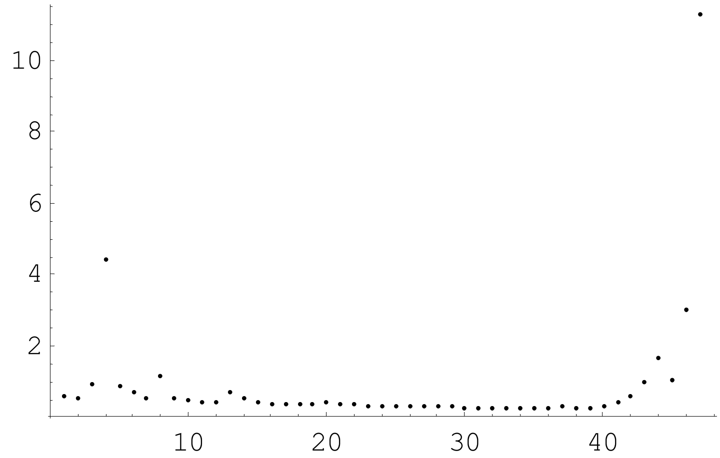

Figure 1. Values of the intrinsic stopping rule

On the other hand, a retrospective analysis on the position of the change point based on the total number of observations, shows that the posterior probability $\pi(r|\mathbf{x}, n = 112)$ is close to zero for all r except for those r between $r = 36$ and $r = 46$. A plot of the posterior probability of r is given in Figure 2. The mode of the posterior distribution of r is attained at point $r = 41$, which corresponds to the year 1891.

While the retrospective analysis, which uses the whole set of data, finds 1891 as the most plausible candidate for a change, the sequential detection procedure finds evidence for a change in the year 1896. It suggests that the sequential procedure detects a change point with some delay. This is a sensible result, however, since the model needs some data to recognize a change.

The posterior probability of the mode is 0.24 and the posterior mean of r is 39.9. The expectation of the magnitude of the change is $E(\lambda_1/\lambda_1|\mathbf{x}) =$ 3.38. Raftery and Akman (1986) and Carlin *et al.* (1992) obtained similar results. The estimate of the change point found by Worsley (1986) was 1890.

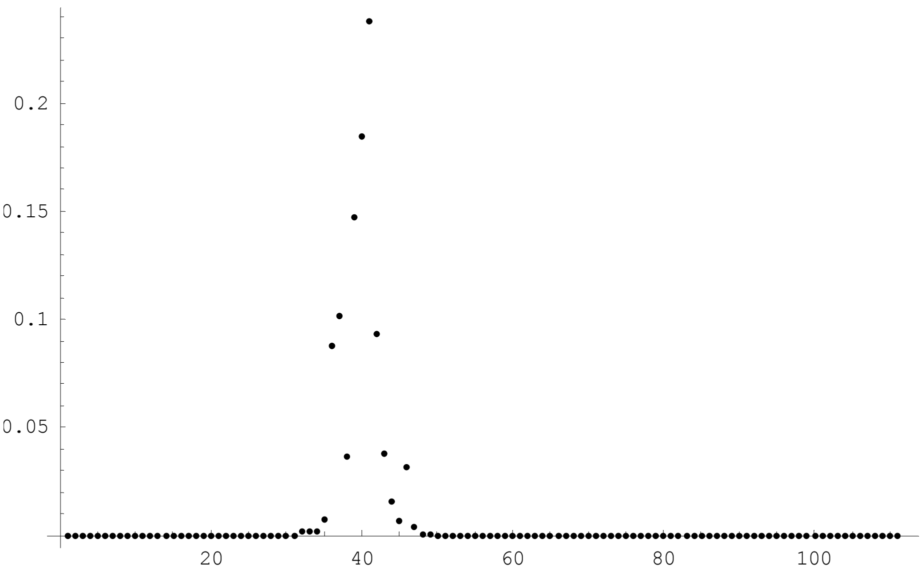

Figure 2. Intrinsic posterior probabilities of the changepoint

Example 2 (Page's data).

The data is based on a simulation with forty observations with a change point at $r = 20$. The first twenty observations are taken from a normal distribution with mean 0 and standard deviation 1, and the remaining ones from a normal with mean 1 and standard deviation 1. The data are plotted in Figure 3.

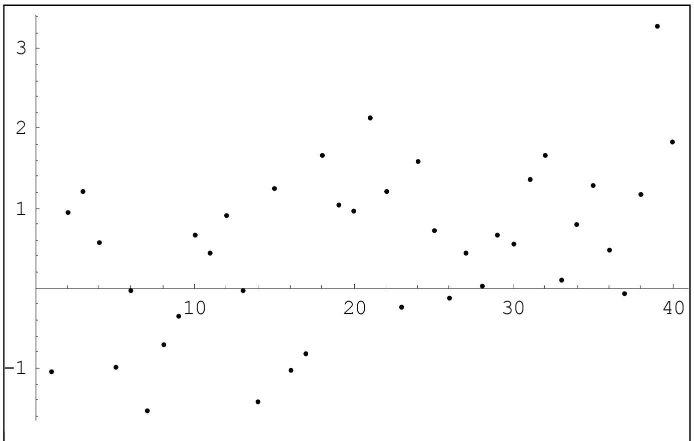

Figure 3: Simulated normal data (Page,1955)

Assuming a normal model, we first test the homoscedastic constraint. The Bayes factor for testing homoscedasticity (HO) against heteroscedasticity (HE) has a value of 5.2 so conveying enough evidence for rejecting heteroscedasticity. In fact, for the prior Pr(HO) = Pr(HE) = 0.5, the posterior probabilities of the homoscedastic and heteroscedastic models turn out to be

$$\Pr(\text{HO}|\mathbf{x}) = 0.84, \ \Pr(\text{HE}|\mathbf{x}) = 0.16.$$

Under the homoscedastic normal model, the sequence of values of the stopping rule T_n is plotted in Figure 4. These values provide evidence of a change point at $r = 20$, a very accurate inference.

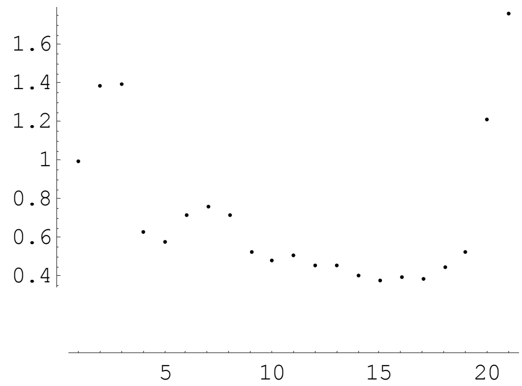

Figure 4. Values of the intrinsic stopping rule

Based on the whole sample and the evidence of a change point, the posterior distribution of the change point is plotted in Figure 5.

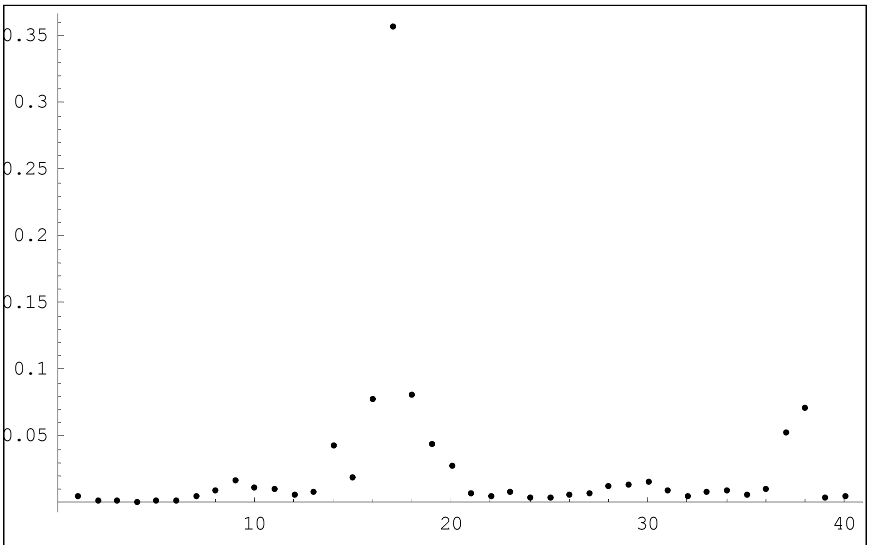

Figure 5: Intrinsic posterior probabilities of the changepoint

The mode of this distribution is $r = 17$ and the probability of the mode is 0.36. The mean is $E(r|\mathbf{x},\text{HO}) = 20.6$, and the posterior expected value of the change is $E(\mu_1 - \mu_2|\mathbf{x},\text{HO}) = -1.02$. So that we expect a shift of the location of 1.02. The posterior standard deviation turns out to be $SD((\mu_1 - \mu_2|\mathbf{x},\text{HO}) = 0.43$.

Under the heteroscedastic normal model the distribution of r is very close to that presented in Figure 1. The mode is at point $r = 17$ and the posterior mean is $E(r|\mathbf{x},\text{HE}) = 23.4$. The expectation of the magnitude of the change is found to be $E(\mu_1 - \mu_2|\mathbf{x},\text{HE}) = -1.39$, the standard deviation is $SD((\mu_1 - \mu_2|\mathbf{x},\text{HE}) = 0.59$, and $E(\sigma_1/\sigma_2|\mathbf{x},\text{HE}) = 1.07$. Therefore, a summary of the inference conditional on the data is

$$E(r|\mathbf{x}) = 20.6 \times 0.84 + 23.4 \times 0.16 = 21,$$

$$E(\mu_1 - \mu_2|\mathbf{x}) = -\,1.02 \times 0.84 - 1.39 \times 0.16 = -1.08,$$

$$SD(\mu_1 - \mu_2|\mathbf{x}) = 0.43 \times 0.84 + 0.59 \times 0.16 = 0.45$$

which seems to be an accurate inference.

A nonparametric analysis of the above data has been done by Petitt (1979) who obtained the estimation of the change point $\hat{r} = 17$. Smith (1975) also found for the posterior mode the value 17 and 17.72 for the posterior mean; he used conjugate priors with subjective values for the hyperparameters.

Example 3 (Pettitt's data).

We consider the data given in Table 3 in Pettitt (1979) presenting the observed percentage of a given material in 27 batches. A plot of these data is given in Figure 6.

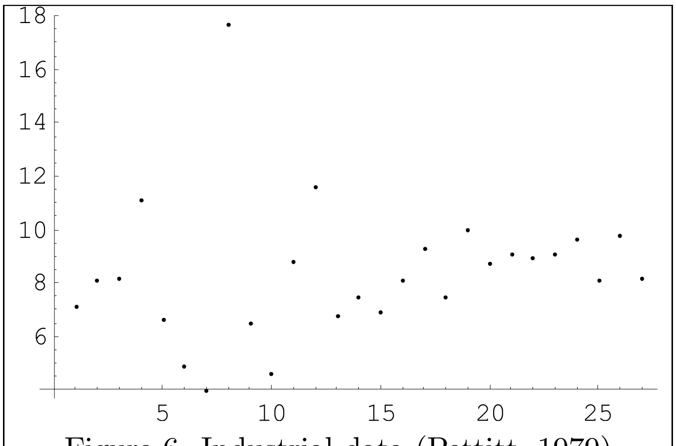

Figure 6. Industrial data (Pettitt, 1979)

Following Menzefricke (1981) we will assume that the data follow a normal distribution with a possible change point. We do not impose any restrictions on the normal mean and variance. The values for the Bayesian stopping rule T_n obtained under this model are plotted in Figure 7. For

$n \leq 20$ the values of T_n are quite small, and they considerably increase as n increases giving evidence for a change point at position $r = 21$.

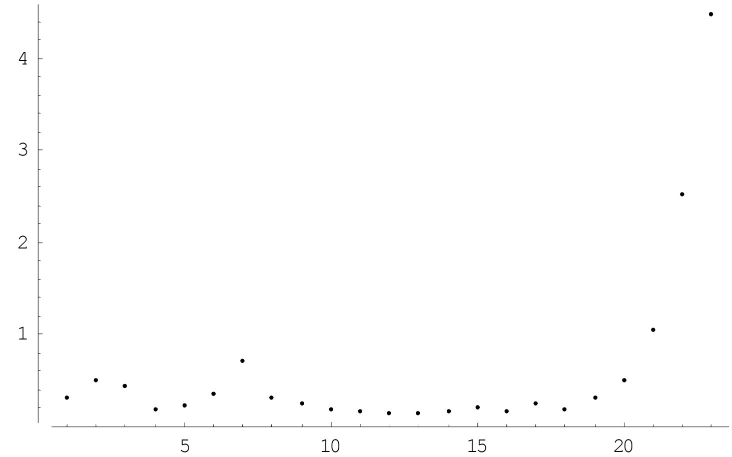

Figure 7. Values of the intrinsic stopping rule

Based on the whole dataset and the prior $\Pr(\mathrm{HO}) = \Pr(\mathrm{HE}) = 0.5$ the test for homoscedasticity renders

$$\Pr(\mathrm{HO}|\mathbf{x}) = 0.004, \ \Pr(\mathrm{HE}|\mathbf{x}) = 0.995.$$

This implies a strong evidence against homoscedasticity. Under the heteroscedastic normal model, the resulting posterior distribution of a change point conditional on the whole dataset is plotted in Figure 8. We note that there is no a single abrupt change point but many points between $r = 10$ and $r = 20$ supporting positive probability masses. The mode of $\pi(r|\mathbf{x})$ is located at $r = 15$ and the mean value is $E(r|\mathbf{x}) = 14.2$.

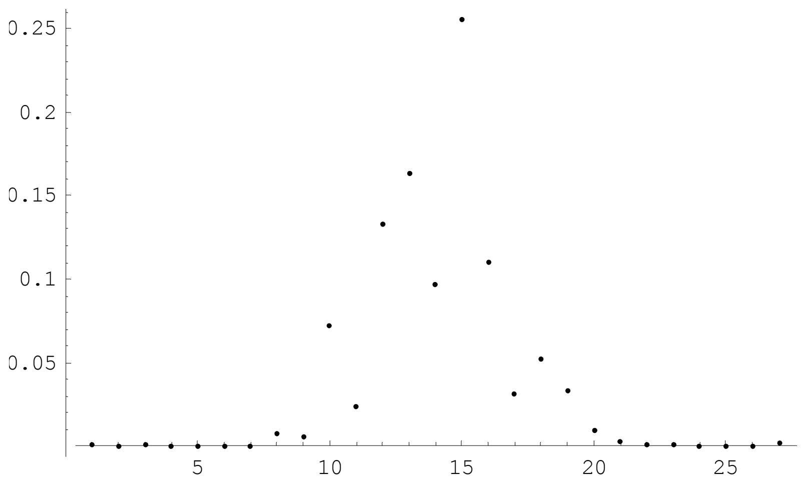

Figure 8: intrinsic posterior probabilities of the changepoint

The posterior expectation of the change in location and scale, and the posterior standard deviation of $\mu_1 - \mu_2$ are

$$E(\mu_1 - \mu_2|\mathbf{x}) = -0.67, \quad E(\frac{\sigma_1}{\sigma_2}|\mathbf{x}) = 3.78, \; SD(\mu_1 - \mu_2|\mathbf{x}) = 1.09.$$

These numbers suggest a small shift of the location, and a reduction of the scale parameter by the factor 1/3.78 .

Using a conjugate analysis on the normal model Menzefricke (1981) obtained the same mode for $\pi(r|\mathbf{x})$. He also detects a change in variance, while the nonparametric analysis by Pettitt (1979) only gave the estimation of $r = 16$.

Example 4 (Stock rates of return data).

We consider the weekly closing price values P_i of the Dow-Jones Industrial Average from July 1, 1971 through August 2, 1974, given in Hsu (1979). This is a rather peculiar period of time in which we assume, as in Hsu (1979), Menzefricke (1981) and Worsley (1986), that the data follow a normal distribution. As a matter of fact, these dataset do not represent current stock rates of return research . Nevertheless, it is considered here as an illustrative example of how the intrinsic Bayesian procedure is able to detect a change point for the variance of the underlying normal distribution without imposing any restriction on the mean. This data has been analyzed by Hsu (1979), Menzefricke (1981) and Worsley (1986) under the assumption that the mean does not change.

The respective rates of return are defined as

$$x_i = \frac{P_{i+1} - P_i}{P_i}, \; i = 1, ..., 161.$$

A plot of these returns is given in Figure 9.

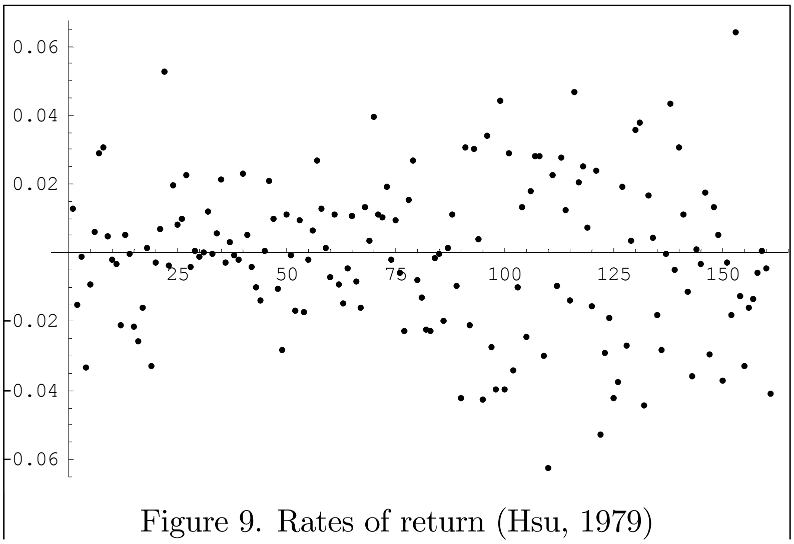

Figure 9. Rates of return (Hsu, 1979)

Figure 10 shows the values of the stopping rule T_n for $n = 2$ to $n = 101$. For $n \leq 99$ the values of T_n are smaller than 1, and for $n = 100, 101, 102$ the values of T_n are 2.85, 2.84, 5.25, respectively, suggesting a change point located at position 100.

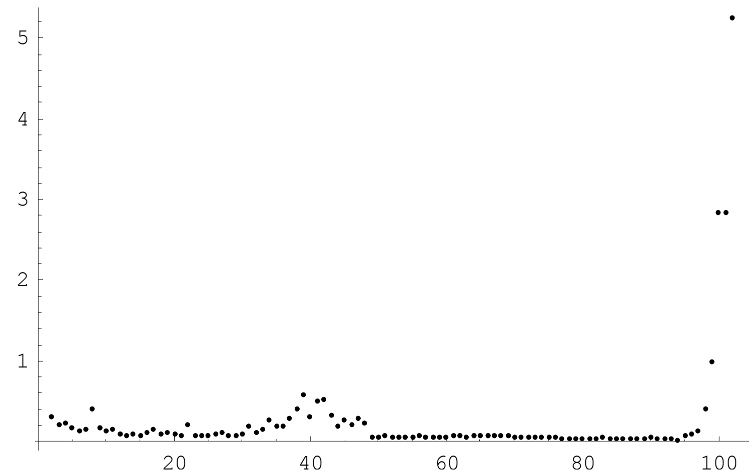

Figure 10. Values of the intrinsic stopping rule

On the other hand, the homoscedastic constraint is strongly refuted by the data. Under the heteroscedastic model, the distribution of the change point is plotted in Figure 11.

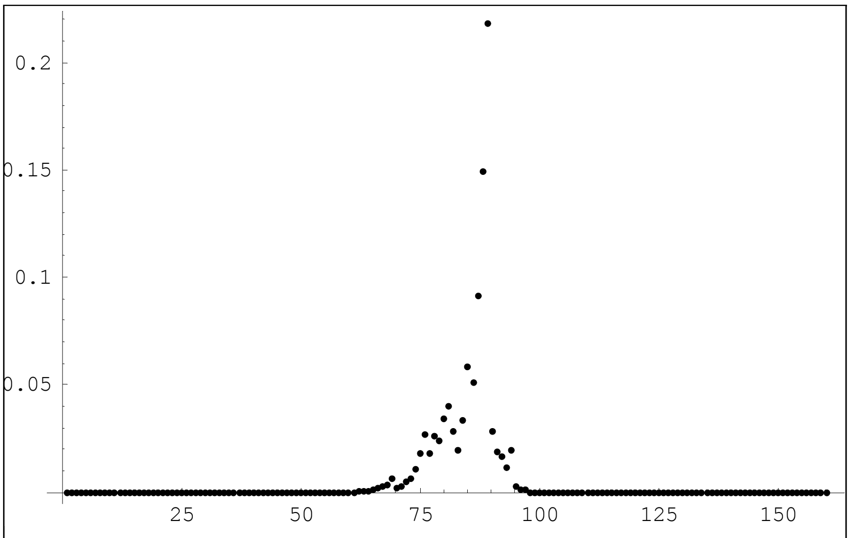

Figure 11. Intrinsic posterior probabilities of the changepoint

The mode of this distribution is $r = 89$ and the mean $E(r|\mathbf{x}) = 85.1$. The posterior expectation of the change in location and scale are

$$E(\mu_1 - \mu_2|\mathbf{x}) = 0.008, \;\; E(\frac{\sigma_1}{\sigma_2}|\mathbf{x}) = 1.6.$$

The posterior standard deviation of $\mu_1 - \mu_2$ has the value $SD(\mu_1 - \mu_2|\mathbf{x}) = 0.003$.

This implies that the location does not change and the scale decreases approximately to $1/1.6$ times its value before the change. Notice that we have not imposed the restriction that the means do not change although the empirical evidence favours that assumption.

Example 5. (The Nile River Data)

The data appearing in Figure 12 are measurements of the annual volume of discharge, 10^{10} m^3, from the Nile River at Aswan for the years 1871 to 1970.

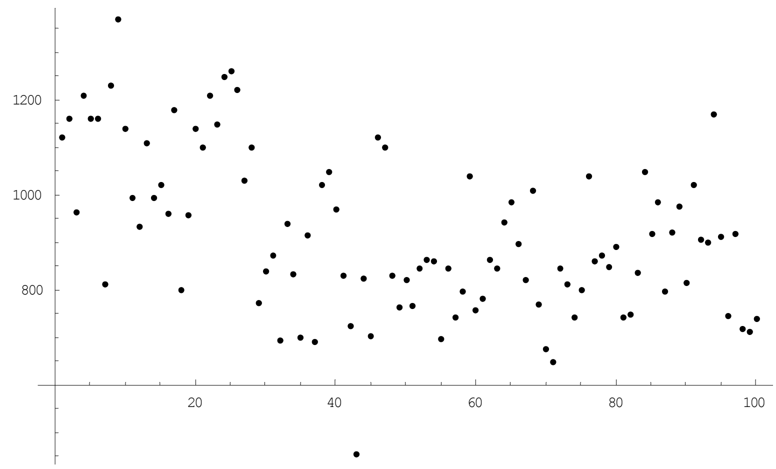

Figure 12. Annual volume (1871 to 1970).

This series was examined by Cobb (1978), Carlstein (1988), Dümbgen (1991) and Balke (1993), and the plot of the data reveals a marked and long-recognized decrease in annual volume after 1898. Some authors have associated the drop to the presence of a dam that began operation in 1902, but Cobb (1978) cited independent evidence on tropical rainfall records to support the decline in volume. For the purpose of illustration, we assume that the observations are independent and coming from a normal distribution with a possible change point, as in Cobb (1978), although we do not impose the homocedastic condition. Figure 13 shows the values of the stopping rule T_n for several values of n.

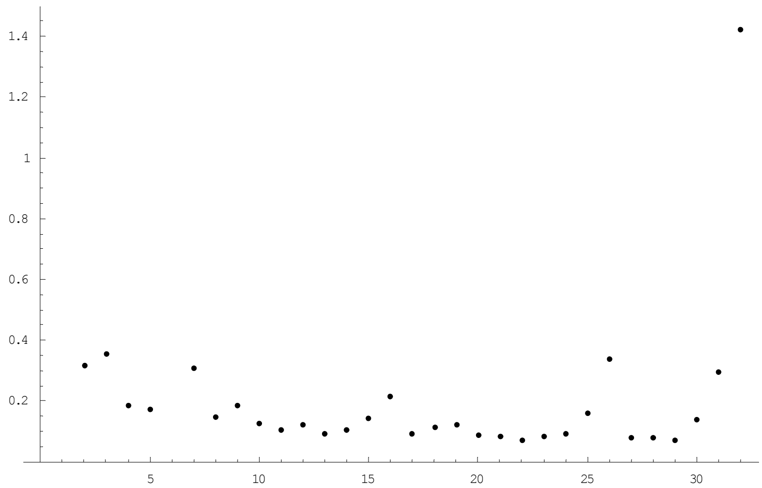

Figure 13. Values of the sequential stopping rule

For $n \leq 32$ the values of T_n are smaller than one, and they increase as n increases providing evidence of a change point in 1901. On the other hand, a retrospective analysis shown in Figure 14 indicates that the mode of the posterior distribution of r is attained at point $r = 28$, which corresponds to the year 1898, and the posterior probability of the mode is 0.736. Again, we find that the change point is detected by the sequential procedure with some delay.

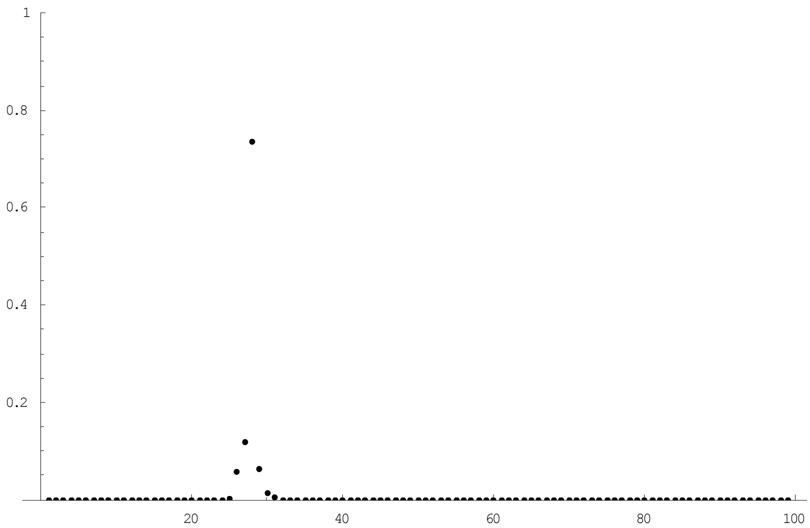

Figure 14. Intrinsic posterior distribution of the changepoint

The posterior mean of the position of the change is $E(r|\mathbf{x}) = 28$. The posterior expectation of the change in location and scale are

$$E(\mu_1 - \mu_2|\mathbf{x}) = 245.36, \;\; E(\frac{\sigma_1}{\sigma_2}|\mathbf{x}) = 2.36,$$

and the posterior standard deviation of $\mu_1 - \mu_2$ is $SD((\mu_1 - \mu_2|\mathbf{x}) = 31.2$. Hence, after the year 1898 the normal mean decreases 245.36 units and the normal standard deviation decreases 2.36 times the value it had before the change point.

7 Conclusions

For independent sampling and under the assumption that subjective prior information on the sampling models parameters is not available, Bayesian analyses for a sequential detection of a change point and for estimating the position where a change occurs have been given. The reference prior, which is the prior commonly used in the absence of prior information, has to be ruled out either for the sequential detection of a change point or for the estimation of the position of the change since they are improper. Methods based on real training samples have been proposed to overcome this difficulty. However, we have argued that such a methods do not apply for the model selection problem involved in the change point analysis. The reason is that real training samples might not exist.

We have proposed the use of intrinsic prior distributions in the Bayesian analyses that not only overcome the above difficulties but they are also free of hyperparameters. This can be seen as a fully default (or objective) Bayesian procedure. The intrinsic priors are derived with the help of the arithmetic intrinsic Bayes factor and an asymptotic argument, as usual.

Closed form for the conditional intrinsic priors for the Poisson and normal models have been obtained and the analysis of the change point problem has been carried out with no numerical difficulties. In particular, numerical Monte Carlo approximations or Gibb sampling are not required. For most of the illustrations considered the sequential stopping rule seems to detect the change point with a certain delay. This is suggested from the comparison of the stopping time value with that of the retrospective estimation of the position of the change point, in which the information provided for whole sample is processed.

We remark that the aim of the paper has been methodological, and the theory has been developed for the simple case of independent sampling. No attempt to considering the change point problem for more complex situations of linear or dynamic models where covariates or some form or conditioning are present has been undertaken. However, we feel that the ideas presented here can be extended to cover these situations but it is nowadays an open problem.

Acknowledgment. We are grateful to two anonymous referees for their comments which have improved an earlier version on the paper. This work has been partially supported by Ministerio de Educación y Ciencia under grant SEJ2004-02447.

References

Abramowitz, M. and Stegun I.A. (1970). *Handbook of Mathematical Functions*. New York: Dover Publications Inc.

Balke, N. S. (1993). Detecting level shifts in time series. *Journal of Business and Economic Statistics*, **11**, 81-92

Berger, J.O. (2000). Bayesian analysis: A look at today and thoughts of tomorrow. *Journal of the American Statistical Association*, **95**, 1269–1276.

Berger, J. O., De Oliveira, V. and Sansó, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, **96**, 1361–1374.

Berger, J.O. and Bernardo, J.M. (1992). On the development of the reference prior method. In *Bayesian Statistics 4*, (J.M. Bernardo et al. eds.), London: Oxford University Press, 35-60.

Berger, J.O. and Pericchi, L.R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of American Statistical Association*, **91**, 109-122.

Berger, J.O. and Pericchi, L.R. (1997). On the justification of objective and intrinsic Bayes factor. In: *Modelling and Prediction*, eds. J.C. Lee at al., New York: Springer, pp. 276-293.

Berger, J.O. and Sellke, T. (1987). Testing a point null hypothesis: the irreconcilibility of p-values and evidence (with discusion). *Journal of American Statistical Association*, **82,** 112-122.

Carlstein, E. (1988). Nonparametric change-point estimation. *The Annals of Statistics*, **16**, 188-197.

Casella, G. and Berger, R.L. (1987). Reconciling Bayesian and frequentist evidence in the one-sided testing problem (with discusion). *Journal of American Statistical Association*, **82,** 106-111.

Casella, G. and Moreno, E. (2002). Objective Bayesian variable selection, Technical Report, University of Granada.

Casella, G. and Moreno, E. (2003a). Objective Bayesian analysis of contingency tables, Technical Report, University of Granada.

Casella, G. and Moreno, E. (2003b). Intrinsic meta analysis of contingency tables. *Statistics in Medicine* (to appear).

Carlin, B.P., Gelfand, A.E. and Smith, A.F.M. (1992). Hierarchical Bayesian Analysis of change point Problems. *Applied Statistics*, **41**, 389-405.

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Annals of Mathematical Statistics*, **35**, 999-1018.

Choy, J.H. and Broemeling, L.D. (1980). Some Bayesian inferences for a changing linear model, *Technometrics*, **22**, 71-78.

Clyde, M. (2001). Discussion to Chipman, H., George, E. and McCulloch, R.E. (2001). *IMS Lecture Notes-Monograph Series*, Vol 38, 67-134.

Cobb, G.W. (1978). The problem of the Nile: conditional solution to a change-point problem. *Biometrika*, **65**, 243-251.

Dümbgen, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *The Annals of Statistics*, **19**, 1471-1495.

James, B., James, K.L. and Siegmund, D. (1987). Test for a change-point. *Biometrika*, **74**, 71-83.

Ferreira, P.E. (1975). A Bayesian analysis of a switching regression model: known number of regimes. *Journal of American Statistical Association*, **70**, 370-374.

Hinkley, D.V. (1970). Inference about a change-point in a sequence of random variables. *Biometrika*, **57**, 1-17.

Hsu, D.A. (1979). Detecting shifts in parameters in gamma sequences with application to stock price and air traffic flow analysis. *Journal of American Statistical Association*, **74**, 31-40.

Jeffreys, H. (1961). *Theory of Probability*. London: Oxford University Press.

Jarret, R.G. (1979). A note on the intervals between coal-mining disasters. *Biometrika*, **66**, 191-193.

James, B., James, K.L. and Siegmund, D. (1987). Test for a change-point. *Biometrika*, **74**, 71-83.

Kander, Z. and Zacks, S. (1966). Tets procedures for possible changes in parameters of statistical distributions occurring at unknown time point. *Annals of Mathematical Statistics*, **37**, 1196-1210.

Kim, S. and Sun, D. (2000). Intrinsic priors for Model Selection Using an Encompassing Model. *Life Time Data Analysis*, **6**, 251–269.

Lai, T. L. (1995). Sequential change point detection in quality control and dynamical systems. *Journal of the Royal Statistical Society, Series B*, **57**, 613-658.

Lorden, G. (1971). Procedures for reacting to a change in distribution. *Annal of Mathematical Statistics*, **41**, 520-527.

Moreno, E., Bertolino, F. and Racugno, W. (1998). An intrinsic limiting procedure for model selection and hypothesis testing. *Journal of American Statistical Association*, **93**, 1451-1460.

Moreno, E., Bertolino, F. and Racugno, W. (1999). Default Bayesian analysis of the Behrens-Fisher problem. *Journal of Statistical Planning and Inference*, **81**, 323-333.

Moreno, E., Bertolino, F. and Racugno, W. (2000). Bayesian Model Selection Approach to Analysis of Variance under Heterocedasticity, *Journal of the Royal Statistical Society, Series* D, 2000, 49, 1-15.

Moreno, E., Bertolino, F. and Racugno, W. (2003). Bayesian inference under partial prior information. *Scandinavian Journal of Statistics.* 2003*,* 30, 565-580.

Moreno, E. and Liseo, B. (2003). Default prior for testing the number of components of a mixture, *Journal of Statistical Planning and Inference*, **111**, 129-142.

Morris, C.N. (1987). Discussion of Berger/Sellke and Casella/Berger. *Journal of American Statistical Association*, **82,** 112-122.

Menzefrike, U. (1981).A bayesian analysis of a change in the precision for a sequence of independent normal random variables at an unknown time point. *Applied Statistics*, **30**, 141-146.

Müller, H.G. (1992). Change-points in nonparametric regression analysis. *Annals of Statistics*, **20**, 737-761.

Page, E.S. (1954). Continuous inspection schemes. *Biometrika*, **41**, 100-114.

Page, E.S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, **42**, 523-527.

Page, E.S. (1957). On problems in which a change in parameter occurs at an unknown point. *Biometrika*, **44**, 248–252.

Pettitt, A.N. (1979). A non-parametric approach to the change-point problem. *Applied Statistics*, **28**, 126-135.

Pollak, M. and Siegmund, D. (1991). Sequential detection of a change in a normal mean when the initial value is unknown. *The Annals of Statistics*, **19**, 394-416.

Raftery, A. E. and Akman, V.E. (1986). Bayesian analysis of a Poisson process with a change-point. *Biometrika*, **73**, 85-89.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, **9**, 65-78.

Sen, A.K. and Srivastava, M.S. (1973). On multivariate test for detecting change in mean. *Sankhyā* A, **35**, 173-186.

Siegmund, D. (1988). Confidence sets in change point problems. *International Statistical Review*, **56**, 31-48.

Smith, A.F.M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, **62**, 407-416.

Smith, A.F.M. and Cook, D.G. (1980). Straight lines with a change point: a Bayesian analysis of some renal ransplant data. *Applied Statis-*

tics, **29**, 180-189.

Sweeting, T. J. (2001). Coverage probability bias, objective Bayes and the likelihood principle. *Biometrika*, **88**, 657–675.

Wasserman, L (2000). Asymptotic inference for mixture models using data-dependent priors. *Journal of the Royal Statistical Society, Series B,* **62**, 159–180.

Worsley, K.J. (1986). Confidence regions and tests for a change-point in a sequence of exponential random variables. *Biometrika*, **73**, 91-104.

Appendix

Appendix 1. Proof of Lemma 1.

For the models $M_0 : P(x_1|\theta) = \frac{\theta^{x_1}}{x_1!}\exp\{-\theta\}$, and $M_1 : \{P(x_1|\theta_1)\, P(x_2|\theta_2), \pi^D(\theta_1,\theta_2) = k\ \theta_1^{-1/2}\theta_2^{-1/2}\}$, where θ is an arbitrary but fixed value, and k is an arbitrary positive constant. The minimal training sample is a pair of independent random variables X_1, X_2 such that under model M_1, $X_i \backsim P(x_i|\theta_i)$, and under M_0, $X_i \backsim P(x_i|\theta)$, $i = 0, 1$. Then, simple calculations give

$$B_{01}^N(x_1,x_2) = \frac{\theta^{x_1+x_2}\exp\{-2\theta\}}{k\ \Gamma(x_1+1/2)\Gamma(x_2+1/2)}.$$

Furthermore,

$$E^{M_1}_{x_1,x_2|\theta_1,\theta_2} B_{01}^N(x_1,x_2) = \frac{\exp\{-(\theta_1+\theta_2+2\theta)\}}{k}\sum_{x_1=0}^{\infty}\frac{(\theta\ \theta_1)^{x_1}}{\Gamma(x_1+1/2)\ x_1!}\times$$

$$\sum_{x_2=0}^{\infty}\frac{(\theta\ \theta_2)^{x_2}}{\Gamma(x_2+1/2)\ x_2!}.$$

Using the equality $\sum_{x=0}^{\infty}\frac{(\theta\ \theta_1)^x}{\Gamma(x+1/2)\ x!} = F_0^1(1/2, \theta\ \theta_1)/\Gamma(1/2)$ and then substitution in (12), Lemma 1 follows.

Appendix 2. Proof of Lemma 2.

Consider the model

$$M_0^* : N(x|\theta,\tau^2),$$

and

$$M_1 : \left\{N(x|\mu_1,\sigma_1^2)N(y|\mu_2,\sigma_2^2),\ \pi_1^N(\boldsymbol{\mu},\boldsymbol{\sigma}) = \frac{c_1}{\sigma_1\sigma_2}\right\}.$$

The minimal training sample is a random vector (X_1, X_2, Y_1, Y_2) with independent components such that under model M_1, $X_i \sim N(x_i|\mu_1,\sigma_1^2)$,

$Y_i \sim N(y_i|\mu_2, \sigma_2^2)$, and under M_0^*, X_i, $Y_i \sim N(x|\theta, \tau^2)$, $i = 1, 2$. We recall that a minimal training sample is a random vector of minimal size for which the marginal density is greater than zero and finite (except for a null set with respect to the Lebergue measure). Then,

$$B_{01}^N(x, y) = \frac{1}{m_1(\mathbf{x}, \mathbf{y})} \prod_{i=1}^{2} N(x_i|\theta, \tau^2) N(y_i|\theta, \tau^2),$$

where

$$m_1(x, y) = c_1 \frac{1}{2^2 |x_1 - x_2| \, |y_1 - y_2|}.$$

Therefore,

$$\pi^I(\boldsymbol{\mu}, \boldsymbol{\sigma}|\theta, \tau) = \frac{1}{4\sigma_1^3 \sigma_2^3 \tau^4}$$

$$\times \int |x_1 - x_2| \exp\left\{ -d_x^2(\tau^{-2} + \sigma_1^{-2}) - \frac{(m_x - \theta)^2}{\tau^2} - \frac{(m_x - \mu_1)^2}{\sigma_1^2} \right\} dx_1 dx_2$$

$$\times \int |y_1 - y_2| \exp\left\{ -d_y^2(\tau^{-2} + \sigma_2^{-2}) - \frac{(m_y - \theta)^2}{\tau^2} - \frac{(m_y - \mu_2)^2}{\sigma_2^2} \right\} dy_1 dy_2,$$

where

$$d_x^2 = \frac{(x_1 - x_2)^2}{4}, \qquad m_x = \frac{x_1 + x_2}{2},$$

$$d_y^2 = \frac{(y_1 - y_2)^2}{4}, \qquad m_x = \frac{y_1 + y_2}{2}.$$

Changing to the new variables

$$u_1 = x_1 - x_2, \qquad v_1 = x_1 + x_2,$$

$$u_2 = y_1 - y_2, \qquad v_2 = y_1 + y_2,$$

the result in Lemma 1 follows.