Rivista di Statistica Applicata, Vol. 17, N. 2 (Giugno) 1984

Bayesian inference for change-point problems

Pietro Muliere, Istituto di Scienze Economiche e Statistiche, Università di Pavia
Marco Scarsini, Istituto Metodi Quantitativi, Università Bocconi, Milano

0. Introduction

A sequence of n random variables (*r.v.*) is considered, the first r of which $(X_1, \ldots, X_r)$ are generated by a model M_1, while the second $(n - r)$, $(X_{r+1}, \ldots, X_n)$ are generated by a model M_2. r is called the change-point and is assumed to be unknown $(0 < r < n)$. When $r = 0$ or $r = n$ there is no change, that is all the observations are generated by the same model (M_2 if $r = 0$, M_1 if $r = n$).

Recently, increasing interest has been shown in the problem of making inference about the change-point. Previous nonbayesian works include Page (1954, 1955, 1957), Gardner (1969), Hinkley (1970, 1972). The classical nonparametric approach has been developed by Page (1955), Bhattacharya and Jonhson (1968), Sen and Srivastava (1973). These authors consider tests for no change against change and the problem of estimating the location parameter. Smith (1975), Cobb (1978), Broemeling (1972) consider posterior probabilities for the change-point. Pettit (1981) solves the same problem using ranks. For the linear model Ferreira (1975), Smith (1977), Smith and Cook (1980), Chin Choy and Broemeling (1980, 1981) Guttman and Menzefriche (1982) have given Bayesian results.

The change-point problem can arise in many different situations. For instance, a production process may be subject to a sudden deterioration in the quality of the output. A social program (e.g. advertising campaign) has its effects after an unknown time. In some cases a chemical becomes abruptly toxic for a biological system. In all these situations we are facing change-point problems.

We now recall some important applications.

Chin Choy and Broemeling (1981) consider the problem of evaluating the effectiveness of various materials and methods of applications for sealing cracks in flexible pavements.

Smith and Cook (1980) deal with the problem of detecting the rejection after a renal transplant.

Carter and Blight (1981) suggest a method for the prediction and detection of ovulation in women.

Silvey (1958) and Smith (1980) solve a slightly different problem (the

Lindisfarne Scribes problem): detecting the number of change-points in a sequence.

In this study we shall consider two problems:

a) inference about the change-point,

b) inference about the parameters of the models.

Usually the assumption that the two models M_1 and M_2 are independent is made. The aim of this work is to solve problem *a*) whithout a so strict assumption: the dependence between the models will be expressed through a hierarchical procedure (section 1). Problem *b*) will be solved in the framework of the exponential family by using some known results about linear Bayes estimates (section 2).

An application to the normal model will then be considered (section 3).

A numerical example is finally reported. The subject we deal with is the one considered by Cobb (1978): the problem of the Nile. A sequence of data about the annual volume of discharge from the Nile River at Aswan for the years 1871 to 1970 is examined. The aim is to determine whether there actually was a change in rainfall regime near the turn of the last century.

Cobb assumed the two models M_1 and M_2 to be independent, but in this application the hypothesis of dependence of the two models seems particularly suitable.

1. Inference about the change-point

We consider a sequence of *r.v.*'s $X_1, X_2, \ldots X_n$ such that:

$$P\{X_1 \leq x_1, \ldots X_n \leq x_n \mid r, \theta_1, \theta_2\} = \Pi_{i=1}^{r} F(x_i \mid \theta_1)\, \Pi_{i=r+1}^{n} F(x_i \mid \theta_2) \qquad r = 1, 2, \ldots, n-1$$

$$= \Pi_{i=1}^{n} F(x_i \mid \theta_2) \quad r = 0$$

$$= \Pi_{i=1}^{n} F(x_i \mid \theta_1) \quad r = n.$$

If the distribution functions $F(\cdot \mid \theta)$ $(\theta \text{ e } \Theta)$ admit a density $f(\cdot \mid \theta)$ with respect to some dominating measure, the joint density of $X_1, X_2, \ldots, X_n$ conditional on r, θ_1 and θ_2 is given by:

$$p(x_1, x_2, \ldots, x_n \mid r, \theta_1, \theta_2) = \Pi_{i=1}^{r} f(x_i \mid \theta_1)\, \Pi_{i=r+1}^{n} f(x_i \mid \theta_2) \qquad r = 1, 2, \ldots, n-1 \quad (1.1)$$

We indicate as $p_0(r)$ the prior distribution of r (the unknown change-point). Obviously:

$$\Sigma_{r=0}^{n} p_0(r) = 1.$$

When θ_1 and θ_2 are known, the posterior distribution of the change-point is

$$p_n(r \mid x^{(n)}, \theta_1 \theta_2) \propto p_0(r)\,\Pi_{i=1}^{r} f(x_i \mid \theta_1)\,\Pi_{i=r+1}^{n} f(x_i \mid \theta_2)$$

$$r = 1, 2, \ldots, n-1$$

where $x^{(n)} = (x_1, x_2, \ldots, x_n)$.

If $r = 0$, $r = n$, we have respectively:

$$p_n(0 \mid x^{(n)}, \theta_1, \theta_2) \propto p_0(0)\,\Pi_{i=1}^{n} f(x_i \mid \theta_2).$$

$$p_n(n \mid x^{(n)}, \theta_1, \theta_2) \propto p_0(n)\,\Pi_{i=1}^{n} f(x_i \mid \theta_1).$$

The parameters θ_1 and θ_2 may be unknown, therefore it is necessary to assign them a prior density $h(\theta_1, \theta_2) = h(\theta_1, \theta_2 \mid r)$.

The posterior distribution of r is:

$$p_n(r \mid x^{(n)}) \propto p_0(r) \int\int p(x_1, \ldots, x_n \mid r, \theta_1, \theta_2)\, h(\theta_1, \theta_2)\, d\theta_1\, d\theta_2.$$

The posterior density of (θ_1, θ_2), conditional on r is:

$$h_n(\theta_1, \theta_2 \mid x^{(n)}, r) \propto p(x_1, \ldots, x_n \mid r, \theta_1, \theta_2)\, h(\theta_1, \theta_2).$$

By averaging over r we obtain the marginal posterior density:

$$h_n(\theta_1, \theta_2 \mid x^{(n)}) = \Sigma_{r=0}^{n} h_n(\theta_1, \theta_2 \mid r, x^{(n)})\, p_n(r \mid x^{(n)}).$$

It is sometimes possible to assume the mechanism generating the data to change completely in r, in such a way that, if r is known, no information about the mechanism before the change is useful for the inference about what happens after the change.

This fact may be represented assuming θ_1 and θ_2 independent (see e.g. Broemeling (1972), Smith (1975)).

It is often unreasonable to make such an assumption.

Indeed, whenever the data $(X_1, \ldots, X_r)$ affect the opinion about $(X_{r+1}, \ldots, X_n)$ even if the change point is known to be r, it is necessary to assume θ_1 and θ_2 dependent.

We shall consider one way to express dependence between θ_1 and θ_2: θ_1 and θ_2 will be still independent, but only conditionally on a hyperparameter λ, which will be given a probability density $g(\lambda) = g(\lambda \mid r)$ (see e.g. Lindley and Smith (1972)).

We shall henceforth indicate:

$X^{(r)} = (X_1, \ldots, X_r)$; $\quad X_{(n-r)} = (X_{r+1}, \ldots, X_n)$.

We summarize our assumptions (the symbol $\perp\!\!\!\perp$ denotes independence)

$X_i \perp\!\!\!\perp X_j \mid \theta_1, \theta_2, \lambda \qquad i \neq j \qquad i, j = 1, \ldots, n$

$X^{(r)} \perp\!\!\!\perp \theta_2, \lambda \mid \theta_1, r$

$X_{(n-r)} \perp\!\!\!\perp \theta_1, \lambda \mid \theta_2, r$

$\theta_1 \perp\!\!\!\perp \theta_2 \mid \lambda$

$\theta_1, \theta_2\, \lambda \perp\!\!\!\perp r.$

These assumptions imply:

$\theta_2 \perp\!\!\!\perp X^{(r)} \mid X_{(n-r)}, \lambda, r$

$\theta_1 \perp\!\!\!\perp X_{(n-r)} \mid X^{(r)}, \lambda, r$

$X^{(r)} \perp\!\!\!\perp X_{(n-r)} \mid \lambda, r.$

The posterior density of θ_2, conditional on r is:

$$h_n(\theta_2 \mid x^{(n)}, r) = \int h_n(\theta_2 \mid x^{(n)}, r, \lambda)\, g_n(\lambda \mid x^{(n)}, r)\, d\lambda$$

As

$\theta_2 \perp\!\!\!\perp X^{(r)} \mid X_{(n-r)}, \lambda, r$

this result becomes:

$$h_n(\theta_2 \mid x^{(n)}, r) = \int h_n(\theta_2 \mid x_{(n-r)}, r, \lambda)\, g_n(\lambda \mid x^{(n)}, r)\, d\lambda.$$

We have:

$$g_n(\lambda \mid x^{(n)}, r) = \frac{p(x^{(n)} \mid \lambda, r)\, g(\lambda)}{\int p(x^{(n)} \mid \lambda, r)\, g(\lambda)\, d\lambda}$$

where

$$p(x^{(n)} \mid \lambda, r) = \int \Pi_{i=1}^{r} f(x_i \mid \theta_1)\, \Pi_{i=r+1}^{n} f(x_i \mid \theta_2)\, h(\theta_1 \mid \lambda)\, h(\theta_2 \mid \lambda)\, d\theta_1\, d\theta_2$$
$$= p(x^{(r)} \mid \lambda, r)\, p(x_{(n-r)} \mid \lambda, r)$$

and

$$h_n(\theta_2 \mid x_{(n-r)}, r, \lambda) = \frac{p(x_{(n-r)} \mid \theta_2, r, \lambda)\, h(\theta_2 \mid \lambda)}{\int p(x_{(n-r)} \mid \theta_2, r, \lambda)\, h(\theta_2 \mid \lambda)\, d\theta_2}.$$

Therefore the posterior density of θ_2, given r, is:

$$h_n(\theta_2 \mid x^{(n)}, r) =$$
$$= \frac{\int h_n(\theta_2 \mid x_{(n-r)}, r, \lambda)\, p(x^{(r)} \mid \lambda, r)\, p(x_{(n-r)} \mid \lambda, r)\, g(\lambda)\, d\lambda}{\int p(x^{(r)} \mid \lambda, r)\, p(x_{(n-r)} \mid \lambda, r)\, g(\lambda)\, d\lambda}. \quad (1.2)$$

Another form of (1.2) is

$$h_n(\theta_2 \mid x^{(n)}, r) \propto \Pi_{i=r+1}^{n} f(x_i \mid \theta_2) \int h(\theta_2 \mid \lambda)\, g_n(\lambda \mid x^{(r)})\, d\lambda,$$

which separates the role of $X_{(n-r)}$ and $X^{(r)}$.

Analogous results have been obtained by Deely and Lindley (1981) dealing with empirical Bayes problem.

Inference on θ_2 is based on the posterior marginal density

$$h_n(\theta_2 \mid x^{(n)}) = \Sigma_{r=0}^{n} h_n(\theta_2 \mid x^{(n)}, r)\, p_n(r \mid x^{(n)})$$

We see that the posterior distribution of θ_2 conditional on r is influenced by all the observations: by $X_{(n-r)}$ through $f(x_i \mid \theta_2)$ $(i = r+1, \ldots, n)$, by $X^{(r)}$ only through the posterior density of the hyperparameter.

Obviously

$$h_n(\theta_1 \mid x^{(n)}, r) \propto \Pi_{i=1}^{r} f(x_i \mid \theta_1) \int h(\theta_1 \mid \lambda)\, g_n(\lambda \mid x_{(n-r)})\, d\lambda.$$

2. Exponential family

Here we deal with the one parameter exponential family, that is with distribution functions wich admit densities (w.r.t. some dominating measure) of the form:

$$f(x \mid \theta) = \frac{a(x)\, e^{-\theta x}}{c(\theta)} \tag{2.1}$$

where

$$c(\theta) = \int e^{-\theta x} a(x)\, d\mu(x).$$

It is well known that

$$E(X \mid \theta) \equiv m(\theta) = -\frac{\partial log\, c(\theta)}{\partial \theta}$$

$$V(X \mid \theta) = -\frac{\partial m(\theta)}{\partial \theta} = \left\{ \frac{\partial^2 (-log\, c(\theta))}{\partial \theta^2} \right\}$$

When θ_1 and θ_2 are known, the posterior probabilities of the change-point are given by:

$$p_n(r \mid \theta_1, \theta_2, x^{(n)}) \propto p_0(r)\, exp\, \{r x^{(r)} (\theta_2 - \theta_1)\} \left(\frac{c(\theta_2)}{c(\theta_1)} \right)^r$$

where

$$x^{(r)} = \frac{1}{r} \Sigma_{i=1}^{r} x_i$$

We note that the sufficient statistic for the unknown parameter r is the full set $X_1, X_2, \ldots, X_n$ and no reduction is possible.

If the parameters θ_1 and θ_2 are unknown and independent we can chose a natural conjugate prior for them, i.e.

$$h(\theta_1, \theta_2) = h(\theta_1)\, h(\theta_2)$$

where

$$h(\theta_i) = \frac{(c(\theta_i))^{-n_i} e^{-\theta_i z_i}}{b(n_i, z_i)} \qquad i = 1, 2;\; n_i > 0 \tag{2.2}$$

with n_i, z_i such that these distributions are proper (conditions on n_i and z_i are delineated by Diaconis and Ylvisaker (1979)).

As an immediate consequence of Bayes, theorem we see that:

$$p_n(r \mid x^{(n)}) \propto p_0(r) \int\int e^{-\theta_1(r\overline{x}^{(r)}+z_1)}[c(\theta_1)]^{-(r+n_1)}$$
$$\cdot e^{-\theta_2((n-r)\overline{x}_{(n-r)}+z_2)}[c(\theta_2)]^{-((n-r)+n_2)} d\theta_1 d\theta_2$$
$$= p_0(r)\, b(r+n_1, r\overline{x}^{(r)}+z_1)\, b((n-r+n_2),(n-r)\overline{x}_{(n-r)}+z_2).$$

Bayes' theorem gives:

$$h_n(\theta_1 \mid x^{(n)}, r) = \frac{[c(\theta_1)]^{-(n_1+r)} e^{-\theta_1}(z_1 + \Sigma_{i=1}^r x_i)}{b(n_1+r, z_1 + \Sigma_{i=1}^r x_i)}$$

Analogously for θ_2.

If we want to estimate θ_1 (or θ_2) and if we choose a squared error loss function, the Bayes estimate is the expectation of the posterior distribution. In our case, i.e. when the model belongs to the exponential family and the prior is conjugate, the posterior expectation for the mean $m(\theta_1)$ is linear:

$$E\{m(\theta_1) \mid x^{(n)}, r\} = \alpha \overline{x}^{(r)} + \beta \tag{2.3}$$

(see Jewell (1974, 1975), Diaconis and Ylvisaker (1979)).

Ericson (1969) has shown that, if (2.3) holds, α and β can be specified through the first two moments of $m(\theta_1)$ and X given θ_1.

$$E\{m(\theta_1) \mid x^{(n)}, r\} = \frac{\overline{x}^{(r)} V(m(\theta_1)) + E(m(\theta_1)) E(V(\overline{X}^{(r)} \mid \theta_1))}{V(m(\theta_1)) + E(V(\overline{X}^{(r)} \mid \theta_1))} \tag{2.4}$$

If we assume regularity conditions, it follows that, in our case, (2.4) may also be written:

$$E\{m(\theta_1) \mid x^{(n)}, r\} = \frac{r}{n_1+r}\overline{x}^{(r)} + \frac{n_1}{n_1+r} E(m(\theta_1)).$$

Hence the expectation of the marginal posterior distribution of $m(\theta_1)$ is:

$$E\{m(\theta_1) \mid x^{(n)}\} = E\{E[m(\theta_1) \mid x^{(n)}, r] \mid x^{(n)}\}$$
$$= \Sigma_{r=0}^n E\{m(\theta_1) \mid x^{(n)}, r\}\, p_n(r \mid x^{(n)}). \tag{2.5}$$

The expression can be interpreted as the expected value of $E\{m(\theta_1) \mid x^{(n)}, r\}$ under the posterior distribution of r.

(2.5) can also be written:

$$E\{m(\theta_1) \mid x^{(n)}\} = \Sigma_{i=1}^n x_i q_i + E(m(\theta_1))\, n_1 q_0$$

where

$$q_i = \Sigma_{r=1}^{n} \frac{1}{n_1 + r} p_n(r \mid x^{(n)}). \qquad i = 0, 1, \ldots, n$$

We note that $q_{i-1} \geq q_i$, so the Bayes estimate of θ_1 is a weighted average of the prior mean and the observations $x_1, \ldots, x_n$ with decreasing weights. Evidently, the Bayes estimate of θ_2 will have the same structure, but the weights of $x_1, \ldots, x_n$ will be increasing.

We now briefly consider a hierarchical model, as in section one.

Let $h(\theta_i \mid \lambda)$ be the simplest family, indexed by λ, conjugate to (2.1). Let θ_1 and θ_2 be independent given λ with densities:

$$h(\theta_i \mid \lambda) = \frac{e^{-\lambda_1 \theta_i} (c(\theta_i))^{-\lambda_2}}{b(\lambda_1, \lambda_2)} \qquad i = 1, 2$$

where

$$b(\lambda_1, \lambda_2) = \int e^{-\lambda_1 \theta} (c(\theta))^{-\lambda_2} d\theta$$

and

$$\lambda = (\lambda_1, \lambda_2) \sim g(\lambda_1, \lambda_2)$$

From (1.2), after some simplications, we have:

$$h_n(\theta_2 \mid x^{(n)}, r) \propto e^{-\theta_2 \Sigma_{i=r+1}^{n} x_i} (c(\theta_2))^{-(n-r)} \cdot A(\theta_2)$$

where

$$A(\theta_2) = \iint \frac{e^{-\lambda_1 \theta_2} (c(\theta_2))^{-\lambda_2} b(\lambda_1 + \Sigma_{i=1}^{r} x_i, \lambda_2 + r)}{b^2(\lambda_1, \lambda_2)} g(\lambda_1, \lambda_2) \, d\lambda_1 \, d\lambda_2.$$

It is not possible, in general, to obtain simple formulae for the Bayes estimates.

3. Application to the normal model

In this section we apply the previous results to the normal model: We consider a sequence of independent *r.v.* (given r, θ_1, θ_2) such that:

$$X_i \mid r, \theta_1, \theta_2 \sim N(\theta_1, \sigma^2) \qquad i = 1, 2, \ldots, r$$

$$X_i \mid r, \theta_1, \theta_2 \sim N(\theta_2, \sigma^2) \qquad i = r + 1, \ldots, n$$

σ^2 known.

If θ_1 and θ_2 are known, too, then

$$p_n(r \mid x^{(n)}, \theta_1, \theta_2) \propto p_0(r) \exp\left\{-\frac{1}{2\sigma^2} [\Sigma_{i=1}^{r} (x_i - \theta_1)^2 + \right.$$

$$\left. + \Sigma_{i=r+1}^{n} (x_i - \theta_2)^2]\right\}$$

If θ_1 and θ_2 are unknown, independent and normally distributed:

$\theta_1 \mid r \sim N(\mu_1, \tau_1^2)$

$\theta_2 \mid r \sim N(\mu_2, \tau_2^2)$

we obtain the distribution of $X^{(n)}$, conditional on r, by integrating over θ_1 and θ_2.

Given r, $X^{(r)}$ and $X_{(n-r)}$ are independent and

$X^{(r)'} \mid r \sim N_r(\underline{1}_r \mu_1, \Sigma_r)$

where

$\underline{1}_r$ is the vector $(1, 1, \ldots, 1)'$,

$\Sigma_r = \sigma^2 I_r + \tau_1^2 J_r$

and J_r is a $r \times r$ matrix, every element of which is unity.

Analogously

$X'_{(n-r)} \mid r \sim N_{(n-r)}(\underline{1}_{(n-r)} \mu_2, \Sigma_{(n-r)})$

where

$\Sigma_{(n-r)} = \sigma^2 I_{(n-r)} + \tau_2^2 J_{(n-r)}$

The posterior distribution of r is

$p_n(r \mid x^{(n)}) \propto p_0(r)\, N_r(\underline{1}_r \mu_1, \Sigma_r)\, N_{(n-r)}(\underline{1}_{(n-r)} \mu_2\, \Sigma_{(n-r)})$

The posterior distribution of θ_1, conditional on r is

$h_n(\theta_1 \mid r, x^{(n)}) \sim N(m_r, \phi_r^2)$

where

$$m_r = \frac{r\bar{x}^{(r)} \tau_1^2 + \mu_1 \sigma^2}{r\tau_1^2 + \sigma^2}$$

and

$$\phi_r^2 = \frac{\tau_1^2 \sigma^2}{r\tau_1^2 + \sigma^2}$$

The marginal posterior distribution of θ_1 is obtained by averaging over r.

If we consider a hierarchical model, no problem arises as long as the distribution of the hyperparameter is proper.

From

$X_i \mid \theta_1, \theta_2, \sigma^2, r \sim N(\theta_1, \sigma^2) \qquad i = 1, 2, \ldots, r$

$X_i \mid \theta_1, \theta_2, \sigma^2, r \sim N(\theta_2, \sigma^2) \qquad i = r+1, \ldots, n$

$\theta_1 \mid \mu, r \sim N(\mu, \tau^2)$

$\theta_2 \mid \mu, r \sim N(\mu, \tau^2)$

$\mu \sim N(\nu, \psi^2)$

By integrating with respect to the parameters θ_1, θ_2 and μ we obtain the distribution:

$$X^{(n)} \mid r \sim N(\underline{1}_n \nu, \Sigma)$$

where the covariance matrix Σ has this form:

$$\Sigma = \begin{pmatrix} A & \vdots & B \\ \cdots & \cdot & \cdots \\ B' & \vdots & C \end{pmatrix}$$

with

$$A = \sigma^2 I_r + (\tau^2 + \psi^2) J_r$$
$$B = \psi^2 J_{r,(n-r)}$$
$$C = \sigma^2 I_{(n-r)} + (\tau^2 + \psi^2) J_{(n-r)}$$

Obviously the posterior distribution of r is:

$$p_n(r \mid x^{(n)}) \propto p_0(r) \exp\{-1/2\,[(x - \underline{1}_n \nu)' \Sigma^{-1} (x - \underline{1}_n \nu)]\}$$

The prior distribution of $(\theta_1, \theta_2)' \mid r$ is

$$(\theta_1, \theta_2)' \mid r \sim N(\underline{1}_2 \nu, \tau^2 I_2 + \psi^2 J_2)$$

The posterior distribution is given by Bayes' theorem

$$(\theta_1, \theta_2)' \mid r, x^{(n)} \sim N(t, V_r)$$

where

$$V_r = \{D_r + (\tau^2 I + \psi^2 J)^{-1}\}^{-1}$$
$$\underline{t} = V_r \{\underline{d} + (\tau^2 I + \psi^2 J)^{-1} \underline{1}\,\nu\}$$

with

$$\underline{d} = \begin{pmatrix} \dfrac{r\bar{x}^{(r)}}{\sigma^2} \\[2ex] \dfrac{(n-r)\,\bar{x}_{(n-r)}}{\sigma^2} \end{pmatrix} \qquad D_r = \begin{pmatrix} r\,\sigma^{-2} & 0 \\ 0 & (n-r)\,\sigma^{-2} \end{pmatrix}$$

If, on the contrary, we assume a diffuse prior for μ, then the distribution of $X^{(n)} \mid r$ does not admit any density *w.r.t.* Lebesgue measure In fact the whole probability mass is concentrated on the subspace $X_1 = X_2 = \ldots X_r = X_{r+1} = \ldots = X_n$. Anyway, it is possible to compute $p_n(r \mid x^{(n)})$ and $h_n(\theta_1, \theta_2 \mid x^{(n)}, r)$, by considering the diffuse prior foı

μ as the limit of a proper conjugate prior, when the variance $\psi^2 \to +\infty$.

In this case we have

$$p_n(r \mid x^{(n)}) \propto p_0(r) \exp\left\{ -\frac{1}{2} [(x - \underline{1}\,\nu)' H (x - \underline{1}\,\nu)] \right\} \tag{3.1}$$

where (see Appendix)

$$H = \left(\begin{array}{ccc|ccc} a & \cdot & b & & c & \\ & \cdot & & & & \\ b & & \cdot a & & & \\ \hline & c & & d & \cdot & e \\ & & & & \cdot & \\ & & & e & & \cdot d \end{array} \right)$$

with

$$a = \frac{2(r-1)(n-r)\tau^2 + (n-1)\sigma^2}{\sigma^2(2r(n-r)\tau^2 + n\sigma^2)} \qquad b = \frac{-2(n-r)\tau^2 - \sigma^2}{\sigma^2(2r(n-r)\tau^2 + n\sigma^2)}$$

$$c = \frac{-1}{2r(n-r)\tau^2 + n\sigma^2}$$

$$d = \frac{-2(n-r-1)r\tau^2 + (n-1)\sigma^2}{\sigma^2(2r(n-r)\tau^2 + n\sigma^2)} \qquad e = \frac{-2r\tau^2 - \sigma^2}{\sigma^2(2r(n-r)\tau^2 + n\sigma^2)}$$

Now we compute the posterior distribution of the parameters:

$$h_n(\theta_1, \theta_2 \mid x^{(n)}, r) \sim N_2(\underline{m}_r, C_r)$$

where

$$\underline{m}_r = \begin{pmatrix} m_{1,r} \\ m_{2,r} \end{pmatrix}$$

$$m_{1,r} = \frac{r\,\bar{x}^{(r)}\tau^2 + \bar{x}\,\sigma^2}{r\tau^2 + \sigma^2} \qquad m_{2,r} = \frac{(m-r)\,\bar{x}_{(n-r)}\tau^2 + \bar{x}\,\sigma^2}{(n-r)\tau^2 + \sigma^2}$$

$$\bar{x} = \frac{\dfrac{r}{r\tau^2 + \sigma^2}\bar{x}^{(r)} + \dfrac{(n-r)}{(n-r)\tau^2 + \sigma^2}\bar{x}_{(n-r)}}{\dfrac{r}{r\tau^2 + \sigma^2} + \dfrac{(n-r)}{(n-r)\tau^2 + \sigma^2}}$$

The Bayes estimate $m_{1,r}$ is a weighted average of the partial sample mean $\bar{x}^{(r)}$ and the total sample mean $\bar{x}$. Analogously for $m_{2,r}$.

$$C_r = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

$$C_{11} = \frac{r\tau^2 + \sigma^2}{\tau^2\sigma^2} - \frac{1}{2\tau^2}$$

$$C_{22} = \frac{(n-r)\tau^2 + \sigma^2}{\tau^2\sigma^2} - \frac{1}{2\tau^2}$$

$$C_{12} = C_{21} = -\frac{1}{2\tau^2}$$

4. Numerical example

We now reconsider a problem examined by Cobb (1978): the annual volume of the Nile river from 1871 to 1970. Cobb assumes that the annual volume of discharge at Aswan is normally distributed with mean θ_1 and standard deviation σ, before the unknown change-point and with mean θ_2 and standard deviation σ after the change-point. He considers all the parameters $(\theta_1, \theta_2, \sigma)$ known, in particular $\theta_1 = 1100$, $\theta_2 = 850$, $\sigma = 125$. He assumes an uniform prior for r (so do we for all following models).

We weaken Cobb's hypotheses and assume first θ_1 and θ_2 random, independent and normally distributed, with means respectively $\mu_1 = 1100$, $\mu_2 = 850$ and standard deviations $\tau_1 = \tau_2 = 50$.

We compute the posterior distribution $p_n(r \mid x^{(n)})$ of the change-point and the Bayes estimates for θ_1 and θ_2 with quadratic loss function.

We observe that the posterior distribution of r is highly concentrated on the 28th observation (corresponding to 1898). The probability of this point is only slightly less than the one obtained in Cobb's model.

Next we enrich our model with hierarchical structure: we assume θ_1 and θ_2 independent, equally and normally distributed given μ, with mean μ and standard deviation $\tau = 50$. Then we assign a noninformative prior to μ. The posterior distribution we obtain for r is not significantly different from the previous one.

We note that:

a) the first model is a particular case of the second one, and is obtained by taking $\tau = 0$.

b) The posterior distribution of r is rather insensitive with respect to different assumption about τ, θ_1, θ_2 in any of the three models.

c) The posterior for r is sensitive to misspecification of σ in the sense that if σ is assumed large, it becomes flat. In fact, in our models, σ is considered known, so that the experiment does not affect the opinion about it. If it is large, even large deviation of the X's around their mean are to be considered in the norm and not due to a change in the model, so that the likelihood doesn't give any evidence of a change-point in the sequence of observations.

d) In the posterior distribution of r there are only few values, around the 28th, whose probability is significantly different from zero.
e) The posterior distribution of θ_1 and θ_2 may be strongly affected by the prior assumptions in the usual Bayesian way.

Tab. I - Bayes posterior.

Years	Cobb's model $\theta_1=1100$ $\theta_2=850$ $\sigma=125$	θ_1 and θ_2 independent $\mu_1=1100$ $\mu_2=850$ $\sigma=125$ $\tau=50$	Hierarchical model $\mu=975$ $\sigma=125$ $\tau=50$
⋮	⋮	⋮	⋮
1893	0.000000	0.000000	0.000001
1894	0.000009	0.000017	0.000046
1895	0.000899	0.001151	0.001832
1896	0.045333	0.047946	0.051043
1897	0.109294	0.110569	0.119760
1898	0.807567	0.796876	0.749449
1899	0.032396	0.036621	0.059787
1900	0.003736	0.005342	0.012949
1901	0.000742	0.001410	0.004660
1902	0.000008	0.000032	0.000212
1903	0.000005	0.000025	0.000190
⋮	⋮	⋮	⋮

Tab. II - Bayes estimates.

Parameters	θ_1, θ_2 independent	Hierarchical model
θ_1	$\hat{\theta}_1$ = 1097,79	$\hat{\theta}_1$ = 1072,91
θ_1	$\hat{\theta}_2$ = 850,63	$\hat{\theta}_2$ = 860,00
$\mu_1 = \int x dF_1(x)$		
$\mu_2 = \int x dF_2(x)$		

Appendix

We show the method for obtaining matrix H (see 3.1). Matrix Σ may be inverted by means of the following formula (Rao 1973)

$$(D + E F E')^{-1} = D^{-1} - D^{-1} E (E' D^{-1} E + F^{-1})^{-1} E' D^{-1}.$$

We have:

$$\Sigma^{-1} = \left(\begin{array}{ccc|ccc} \alpha & \cdot & \beta & & \gamma & \\ \beta & \cdot & \alpha & & & \\ \hline & & & \delta & \cdot & \varepsilon \\ & \gamma & & \varepsilon & \cdot & \delta \end{array} \right)$$

with

$$\alpha = \frac{2(r-1)(n-r)\psi^2\tau^2 + (n-1)\psi^2\tau^2 + (r-1)(n-r)\tau^4 + (n-1)\tau^2\sigma^2 + \sigma^4}{\sigma^2 (2 r (n-r) \psi^2 \tau^2 + r (n-r) \tau^4 + n \psi^2 \sigma^2 + n \tau^2 \sigma^2 + \sigma^4)}$$

$$\beta = \frac{-2 (n-r) \psi^2 \tau^2 - \psi^2 \sigma^2 - (n-r) \tau^4 - \tau^2 \sigma^2}{\sigma^2 (2 r (n-r) \psi^2 \tau^2 + r (n-r) \tau^4 + n \psi^2 \sigma^2 + n \tau^2 \sigma^2 + \sigma^4)}$$

$$\gamma = \frac{-\psi^2 \sigma^2}{\sigma^2 (2 r (n-r) \psi^2 \tau^2 + r (n-r) \tau^4 + n \psi^2 \sigma^2 + n \tau^2 \sigma^2 + \sigma^4)}$$

$$\delta = \frac{2(n-r-1) r \psi^2\tau^2 + (n-1)\psi^2\sigma^2 + (n-r-1) r \tau^4 + (n-1)\tau^2\sigma^2 + \sigma^4}{\sigma^2 (2 r (n-r) \psi^2 \tau^2 + r (n-r) \tau^4 + n \psi^2 \sigma^2 + n \tau^2\sigma^2 + \sigma^4)}$$

$$\varepsilon = \frac{-2 r \psi^2 \tau^2 - \psi^2 \sigma^2 - r \tau^4 - \tau^2 \sigma^2}{\sigma^2 (2 r (n-r) \psi^2 \tau^2 + r (n-r) \tau^4 + n \psi^2 \sigma^2 + n \tau^2 \sigma^2 + \sigma^4)}$$

By taking the limit we have:

$$\lim_{\psi^2 \to +\infty} \Sigma^{-1} = H.$$

Work performed while the authors were members of CNR-GNAFA.
This paper takes origin from the close collaboration of the authors. However the first author has written section 1 and 3, the second one sections 2 and 4.
The authors wish to thank dr. Corrado Giannasca for some suggestions about the computer programs, and Professor S.D. Silvey for some helpful comments on an earlier version of the paper. However, the authors remain responsible for any errors and omissions.

References

Bhattacharya, G.K. & Johnson, R.A. (1968) Nonparametric tests for shift at an unknown time point, *Ann. Math. Statist.* **39**, 1731-1743.

Broemeling, L.D. (1972) Bayesian procedures for detecting a change in a sequence of random variables, *Metron* **30**, 214-227.

Broemeling, L.D. & Chin Choy, J.H. (1981) Detecting structural change in linear models, *Commun. Statist.-Theor. Meth.* **A10**, 2551-2561.

Carter, R.L. & Blight, B.J.N. (1981) A Bayesian change point problem with an application to the prediction and detection of ovulation in women, *Biometrics* **37**, 743-751.

Chernoff, H. & Zacks, S. (1964) Estimating the current mean of a normal distribution wich is subjected to changes in time, *Ann. Math. Statist.* **35**, 999-1018.

Chin Choy, J.H. & Broemeling, L.D. (1980) Some Bayesian Inferences for a changing linear model, *Technometrics* **22**, 71-78.
Cobb, G.W. (1978) The problem of the Nile: conditional solution to a change-point problem. *Biometrika* **65**, 243-251.
Deely, J.J. & Lindley, D.V. (1981) Bayes empirical Bayes, *J. Amer. Statist. Assoc.* **76**, 833-841
Diaconis, P. & Ylvisaker, D. (1979) Conjugate priors for exponential families, *Ann. Statist.* **7**, 269-281.
Ericson, W.A. (1969) A note on the posterior mea nof a population mean. *J. Roy. Statist. Soc. Ser.* **B31**, 332-334.
Ferreira, P.E. (1975) Bayesian analysis of a switching regression model: Known number of regimes, *J. Amer. Statist. Assoc.* **70**, 370-374.
Gardner, L.A. (1969) On detecting changes in the mean of normal variables, *Ann. Math. Statist.* **40**, 116-126.
Guttman, I. & Menzefriche, U. (1982) On the use of loss-functions in the change-point problem. *Ann. Inst. Statist. Math.* **34**, 319-326.
Hinkley, D.V. (1972) Time-ordered classification, *Biometrika* **59**, 509-523.
Hinkley, D.V. (1970) Inference about the change-point in a sequence of random variables, *Biometrika* **57**, 1-17.
Jewell, W.S. (1974) Credible means are exact Bayesian for exponential families, *Astin Bull* **8**, 77-90.
Jewell, W.S. (1975) Regularity conditions for exact credibility, *Astin Bull,* **8**, 336-341.
Kander, Z. & Zacks, S. (1966) Test procedures for possible changes in parameters of statistical distributions occurring at an unknown time point, *Ann. Math. Statist.* **37**, 1196-1210.
Lindley, D.V. & Smith, A.F.M. (1972) Bayes estimates for the linear model (with discussion), *J. Roy. Statist. Soc. Ser.* **B34**, 1-41.
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.
Pettit, A.N. (1979) A nonparametric approach to the change point problem, *Appl. Statist.* **28**, 126-135.
Pettit, A.N. (1981) Posterior probabilities for a change-point using ranks, *Biometrika* **68**, 443-450.
Rao, C.R. (1973) *Linear Statistical Inference and its Applications*, Wiley, New York.
Sen, A.K. & Srivastava, M.S. (1973) On multivariate tests for detecting change in mean, *Sankhya* **A35**, 173-186.
Sen, A.K. & Srivastava, M.S. (1975) On tests for detecting change in mean, *Ann. Statist.* **3**, 98-108.
Silvey, S.D. (1958) The Lindisfarne scribes problem, *J. Roy. Statist. Soc. Ser.* **B20**, 93-101.
Smith, A.F.M. (1975) A Bayesian approach to inference about a changepoint in a sequence of random variables, *Biometrika* **62**, 407-416.
Smith, A.F.M. (1977) A Bayesian analysis of some time-varying models. In *Recent Developments in Statistics*, eds. Barra, J.R. et al. 257-267, North-Holland, Amsterdam.
Smith, A.F.M. (1980) Change-point problems: approaches and applications, *Trab. Estadist.* **31**, 83-98.
Smith, A.F.M. & Cook, D.G. (1980) Straight-lines with a change-point: a Bayesian analysis of some renal transplant data, *Appl. Statist.* **29**, 180-189.

Riassunto

Si considera una successione finita di n variabili aleatorie, le prime r delle quali sono generate da un modello M_1, mentre le seconde $(n - r)$ sono generate da un modello M_2. Il punto di cambiamento r è incognito.

Si calcola la distribuzione finale del punto di cambiamento nell'ipotesi di un modello a struttura gerarchica. Si forniscono le distribuzioni finali e le stime bayesiane dei parametri dei modelli M_1 e M_2 qualora questi appartengano alla famiglia esponenziale. Il lavoro si conclude con una applicazione al modello normale ed un esempio numerico.