Pietro Muliere; Marco Scarsini

Change-point problems: A Bayesian nonparametric approach

Aplikace matematiky, Vol. 30 (1985), No. 6, 397–402

Persistent URL: http://dml.cz/dmlcz/104169

CHANGE-POINT PROBLEMS: A BAYESIAN NONPARAMETRIC APPROACH*

PIETRO MULIERE, MARCO SCARSINI

(Received March 12, 1984)

A change-point problem is examined from a Bayesian viewpoint, under nonparametric hypotheses. A Ferguson-Dirichlet prior is chosen and the posterior distribution is computed for the change-point and for the unknown distribution functions.

Keywords. Change-point, Dirichlet process, Bayes estimate.

1. INTRODUCTION

The change-point (c.p.) problem may be outlined as follows: consider a finite sequence $X_1, \ldots, X_n$ of random variables (r.v.'s) such that the first r of them are identically distributed according to a distribution function (d.f.) F_1, while the second $(n - r)$ ones are identically distributed according to F_2, where r is unknown.

The problem has been dealt with by many authors in a sample-theoretical framework.

A Bayesian treatment has been developed by Broemeling (1972), Smith (1975, 1977, 1980), Cobb (1978) under parametric hypotheses. Pettit (1981) used ranks to determine the (approximate) posterior distribution of the c.p.

The aim of our work is to provide a fully Bayesian procedure for deriving the posterior distribution of the c.p. when F_1 and F_2 do not belong to a parametric family. The prior distribution of F_1 and F_2 will be chosen to be a Ferguson-Dirichlet process. The Bayesian approach to c.p. problem will be briefly outlined. The posterior distributions of the c.p. and of F_1 and F_2 and the Bayes estimates of some functionals of F_1 and F_2 will be given.

2. INFERENCE ABOUT THE CHANGE-POINT

Let $X \equiv (X_1, \ldots, X_n)$ be a vector of r.v.'s such that, given r, F_1, and $F_2 : X_1, \ldots$ $\ldots, X_n$ are independent, $X_1, \ldots, X_r$ are i.i.d.r.v.'s distributed according to F_1,

*) Work performed while the authors were members of GNAFA-CNR.

$X_{r+1}, \ldots, X_n$ are i.i.d.r.v.'s distributed according to F_2. Here r, F_1, F_2 are unknown. r may assume values $0, 1, \ldots, n$. If $r = 0$ all the r.v.'s are distributed according to F_2; if $r = n$ they are all distributed according to F_1. In these two cases there is no c.p., actually.

If $F_1, F_2 \in \mathscr{F}$ where $\mathscr{F}$ is a dominated family of d.f.'s, and $\mu(\cdot, \cdot)$ is a prior probability measure on some suitable σ-field $\mathscr{S}^*$ of subsets of $\mathscr{F} \times \mathscr{F}$, if $p(r)$ is the prior distribution of r, and if (F_1, F_2) and r are *a priori* independent, then Bayes' theorem gives

$$p(r \mid x) \propto p(r) \int_{\mathscr{F} \times \mathscr{F}} l(x \mid F_1, F_2, r) \, \mathrm{d}\mu(F_1, F_2) \tag{1}$$

where $l(x \mid F_1, F_2, r)$ is the likelihood of x, given F_1, F_2, r, which exists by virtue of the dominance of $\mathscr{F}$.

If the object of the inference is (F_1, F_2), we have:

$$\mathrm{d}\mu(F_1, F_2 \mid x) \propto \mathrm{d}\mu(F_1, F_2) \sum_{r=0}^{n} l(x \mid F_1, F_2, r) \, p(r) \, . \tag{2}$$

Difficulties arise in a nonparametric model, because – generally – the family $\mathscr{F}$ is not dominated so that the posterior distributions (1) and (2) must be obtained in a different way.

Bayesian analysis of nonparametric problems started with Ferguson (1973) who provided a suitable prior measure on the space of d.f's. Ferguson's proposal was Dirichlet process (DP).

Definition. *Let $\alpha(\cdot)$ be a non-null finite measure on $(\mathbb{R}, \mathscr{B})$ (the real line endowed with the Borel σ-field), and let $P(\cdot)$ be a stochastic process indexed by the elements of $\mathscr{B}$. We say that P is a Dirichlet process with parameter α ($P \in \mathscr{D}(\alpha)$) if for every finite measurable partition $(B_1, \ldots, B_n)$ of $\mathbb{R}$, the random vector $(P(B_1), \ldots. P(B_n))$ has a Dirichlet distribution with parameter $(\alpha(B_1), \ldots, \alpha(B_n))$.*

Let $F(t) = P((-\infty, t])$; we shall indicate $F \in \mathscr{D}(\alpha)$ for $P \in \mathscr{D}(\alpha)$.

Ferguson's results have been generalized by Antoniak who proposed a class of processes called mixtures of Dirichlet processes. For the properties of DP's and mixtures of DP's, we refer to Ferguson (1973) and Antoniak (1974).

Let in the above problem F_1 be a DP with parameter $\alpha_1(\cdot)$, and F_2 be a DP with parameter $\alpha_2(\cdot)$.

The main result about the posterior distribution of r is the following:

Proposition. *Let $X_1, \ldots, X_n$ be n r.v.'s such that, given r, F_1, F_2 : $X_1, \ldots, X_n$ are independent,*

X_i are i.i.d.r.v.'s $\sim F_1$, $i = 1, 2, \ldots, r$,

X_i are i.i.d.r.v.'s $\sim F_2$, $i = r + 1, \ldots, n$.

Let $F_1 \in \mathscr{D}(\alpha_1)$, $F_2 \in \mathscr{D}(\alpha_2)$. Let F_1, F_2, r be mutually independent. Assume there exists a σ-finite measure μ on $(\mathbb{R}, \mathscr{B})$ such that:

1) *α_1, α_2 are absolutely continuous w.r.t. μ,*
2) *μ has mass one at each atom of α_1, α_2.*

Then

$$p(r \mid \boldsymbol{x}) \propto \frac{1}{\alpha_1(\mathbb{R})^{[r]}} \prod_{i=1}^{s} \alpha_1'(x_i^*) \, (m_1(x_i^*) + 1)^{[n_1(x_i^*)-1]} \cdot \frac{1}{\alpha_2(\mathbb{R})^{[n-r]}} \prod_{i=1}^{t} \alpha_2'(x_i^{**}) \, (m_2(x_i^{**} + 1)^{[n_2(x_i^{**})-1]} \, p(r) \,, \tag{3}$$

where

the product over a void set is defined to be zero,
$a^{[n]} = a(a+1)\dots(a+n-1)$,
$\alpha_j'(\cdot)$ denotes the Radon-Nikodym derivative of α_j w.r.t. μ $(j = 1, 2)$,
x_i^* is the i-th distinct value of X in $x^{(r)} \equiv (x_1, \dots, x_r)$,
x_i^{**} is the ith distinct value of X in $x_{(n-r)} \equiv (x_{r+1}, \dots, x_n)$,
$n_1(x_i^*)$ is the number of times the value x_i^* occurs in $x^{(r)}$,
$n_2(x_i^{**})$ is the number of times the value x_i^{**} occurs in $x_{(n-r)}$,
$m_j(x) = \alpha_j'(x)$ if x is an atom of α_j, zero otherwise,
s and t are the numbers of distinct values in $x^{(r)}$, $x_{(n-r)}$ respectively.

Proof. By the properties of DP (and of mixtures of DP's) the likelihood of x_{k+1} given $r, x_1, \dots, x_k$ is

$\dfrac{\alpha_1'(x_{k+1})\,\mathrm{d}\mu}{\alpha_1(\mathbb{R}) + k}$	for $k + 1 \leqq r$, if the value of X_{k+1} has not occurred previously in $x_1, \dots, x_k$,
$\dfrac{m_1((x_{k+1}) + j)\,\mathrm{d}\mu}{\alpha_1(\mathbb{R}) + k}$	for $k + 1 \leqq r$ if the value of x_{k+1} has occurred previously j times in $x_1, \dots, x_k$,
$\dfrac{\alpha_2'(x_{k+1})\,\mathrm{d}\mu}{\alpha_2(\mathbb{R}) + k - r}$	for $k + 1 > r$ if the value of x_{k+1} has not occurred previously in $x_{r+1}, \dots, x_n$,
$\dfrac{m_2((x_{k+1}) + j)\,\mathrm{d}\mu}{\alpha_2(\mathbb{R}) + k - r}$	for $k + 1 > r$ if the value of x_{k+1} has occurred previously j times in $x_{r+1}, \dots, x_n$.

Hence the likelihood of $(x_1, \dots, x_n)$, given r, is

$$\frac{1}{\alpha_1(\mathbb{R})^{[r]}} \prod_{i=1}^{s} \alpha_1'(x_i^*) \, (m_1(x_i^*) + 1)^{[n_1(x_i^*)-1]} \cdot \frac{1}{\alpha_2(\mathbb{R})^{[n-r]}} \prod_{i=1}^{t} \alpha_2'(x_i^{**}) \, (m_2(x_i^{**}) + 1)^{[n_2(x_i^{**})-1]} \,.$$

Multiplication by the prior distribution and normalization gives the result. □

Remark 1. The above proposition is analogous to lemma 1 of Antoniak (1974), in which he gave the posterior density of the index of a mixture of DP's. Note that in our problem F_1, F_2 are not mixtutes of DP's (with index r): in fact F_1, F_2 and r are assumed independent.

Remark 2. If the observations of the sample are all distinct and α_1 and α_2 are absolutely continuous w.r.t. Lebesgue measure, then (3) becomes

$$(4) \qquad p(r \mid x) \propto \frac{1}{(\alpha_1(\mathbb{R}))^{[r]}} \prod_{i=1}^{r} \alpha_1'(x_i) \frac{1}{(\alpha_2(\mathbb{R}))^{[n-r]}} \prod_{i=r+1}^{n} \alpha_2'(x_i)\, p(r)\,.$$

Factors $1/(\alpha_1(\mathbb{R}))^{[r]}$ and $1/(\alpha_2(\mathbb{R}))^{[n-r]}$ make the expression (4) different from the one obtaibed in the model with $F_i(t) = \alpha_i(t)/\alpha_i(\mathbb{R})$ ($i = 1, 2,$) known, i.e.:

$$p(r \mid x) \propto \prod_{i=1}^{r} F_1'(x_i) \prod_{i=r+1}^{n} F_2'(x_i)\, p(r)$$

where $F_1'(\cdot)$ and $F_2'(\cdot)$ are the densities of F_1 and F_2, respectively, w.r.t. some suitable dominating measure.

In this respect c.p. model behaves unlike other nonparametric models in which the posterior distributions of the index parameter are the same for the parametric and the nonparametric model under the hypotheses of no ties and absolute continuity of α. (See e.g. Cifarelli, Muliere, and Scarsini (1981) and Diaconis and Freedman (1982)).

Remark 3. If $\alpha_1(\mathbb{R})$ increases, *ceteris paribus*, then $p(r \mid x)$ moves towards little values of r. Conversely, if, *ceteris paribus*, $\alpha_2(\mathbb{R})$ increases, then $p(r \mid x)$ moves towards large values of r. This fact may be justified as follows: if $\alpha_1(\mathbb{R})$ increases, then the form of F_1 becomes more precisely known, so that it becomes more difficult for the sample data to be generated by F_1 and therefore it becomes more probable that they are generated by F_2 (less precisely specified). Analogously for α_2.

Remark 4. Suppose x_i is an atom of α_2 but not of α_1. In expression (3) $\alpha_1'(x_h^*)$ is zero for $x_h^* = x_i$ so that $p(r \mid x) = 0$ for $r \geqq i$. In other words, the probability that x_i is selected by F_1 is zero, while the probability that it is selected by F_2 is one.

3. INFERENCE ABOUT THE DISTRIBUTION FUNCTIONS

We now consider inference about F_1 and F_2. Properties of DP give the following posterior distributions for F_1 and F_2:

$$F_1 \mid r, X \in \mathscr{D}(\alpha_1(\cdot) + \sum_{i=1}^{r} \delta_{X_i}),$$

$$F_1 \mid X \in \sum_{r=0}^{n} \mathscr{D}(\alpha_1(\cdot) + \sum_{i=1}^{r} \delta_{X_i})\, p(r \mid X)$$

where δ_x is the measure that concentrates mass one at x.

Analogously, *mutatis mutandis*, for F_2.

If we choose a squared-loss function L, weighted according to some finite measure W on $\mathbb{R}$ (see Ferguson (1973))

$$L(F, F^*) = \int_{\mathbb{R}} (F(t) - F^*(t))^2 \, \mathrm{d}W(t),$$

we obtain the following Bayes estimate of F_1, given r and x

$$F_1^*(t \mid r, x) = \frac{\alpha_1(\mathbb{R})}{\alpha_1(\mathbb{R}) + r} F_{1,0}(t) + \frac{r}{\alpha_1(\mathbb{R}) + r} F_{1,r}(t)$$

where $F_{1,0}(t) = \alpha_1((-\infty, t])/\alpha_1(\mathbb{R})$ and $F_{1,r}(t) = \frac{1}{r} \sum_{i=1}^{r} \delta_{x_i}$ is the empirical d.f. of $x_1, \ldots, x_r$.

Therefore

$$F_1^*(t \mid x) = \sum_{r=0}^{n} F_1^*(t \mid r, x)\, p(r \mid x)$$

$$= \alpha_1(\mathbb{R})\, F_{1,0} q_0 + \sum_{i=1}^{n} \delta_{x_i} q_i$$

where

$$q_i = \sum_{r=i}^{n} \frac{1}{\alpha_1(\mathbb{R}) + r}\, p(r \mid x), \quad i = 0, \ldots, n.$$

Evidently $q_i \geqq q_{i+1}$, i.e., the weight of the observations decreases from one to another $F_2^*(t \mid x)$ will have an analogous structure, but the weight of the observations will be increasing.

If we define

$$\mu_1 = \int x \, \mathrm{d}F_1(x)$$

and assume a quadratic loss function, Bayes estimate of μ_1 given r and x will be

$$\mu_{1|r}^* = \frac{\alpha_1(\mathbb{R})}{\alpha_1(\mathbb{R}) + r} \mu_{1,0} + \frac{r}{\alpha_1(\mathbb{R}) + r} \frac{1}{r} \sum_{i=1}^{r} x_i$$

where

$$\mu_{1,0} = \int x \, \mathrm{d}\alpha_1(x)/\alpha_1(\mathbb{R}).$$

The unconditional Bayes estimate is

$$\mu_1^* = \sum_{r=0}^{n} \mu_{1|r}^*\, p(r \mid x) = \alpha_1(\mathbb{R})\, \mu_{1,0} q_0 + \sum_{i=1}^{n} x_i q_i$$

where q_i are as before.

Analogously for μ_2.

References

C. E. Antoniak (1974): Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist. 2*, 1152—1174.

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

D. M. Cifarelli, P. Muliere, and *M. Scarsini* (1981): Il modello lineare nell'approccio bayesiano nonparametrico. Research report N. 15, Istituto Matematico G. Castelnuovo, Roma.

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

P. Diaconis and *D. Freedman* (1982) Bayes rules for location problems, in *Statistical Decision Theory and Related Topics III*, (ed. by S. S. Gupta and J. O. Berger) vol. I, 315—327, Academic Press, New York.

T. S. Ferguson (1973): A Bayesian analysis of some nonparametric problems. *Ann. Statist. 1*, 209—230.

A. N. Pettit (1981): Posterior probabilities for a change-point using ranks, *Biometrika 68*, 443—450

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

A. F. M. Smith (1977): A Bayesian analysis of some time-varying models. In *Recent Developments in Statistics* (ed. by J. R. Barra et al.), 257—267, North-Holland, Amsterdam.

A. F. M. Smith (1980): Change-point problems: approaches and applications. *Trab. Estadist. 31*, 83—98.

Souhrn

PROBLĚMY BODU ZMĚNY: BAYESOVSKY NEPARAMETRICKÝ PŘÍSTUP

PIETRO MULIERE, MARCO SCARSINI

Problém bodu změny v posloupnosti náhodných veličin je studován z bayesovského hlediska při neparametrických hypotézách. Vychází se z Fergusonova-Dirichletova apriorního rozložení a odvozují se aposteriorní rozložení bodu změny a neznámých distribučních funkcí.

Author's addresses: Prof. *Pietro Muliere*, Istituto di Scienze Economiche e Statistiche, Università di Pavia, Via San Felice 5, 27100 Pavia; Prof. *Marco Scarsini*, Istituto di Matematica Finanziaria, Università di Parma Via Kennedy 4/B, 43100 Parma, Italy.