# CHANGE - POINT MODELS

Uwe Jensen and Constanze Lütkebohmert

*Institut für Angewandte Mathematik und Statistik, Universität Hohenheim, Stuttgart, Germany*

*Key words:* Change-point, sequentially observed data, detection problem, optimal stopping, regression models, hazard rate models

**Abstract**

Change-point models describe formally the problem to decide whether a stochastic process is homogeneous in some sense or not. We review and classify various change-point models with a focus on the more recent ones. In particular, we consider change-point detection models in which the decision is based on sequentially observed data. Both, discrete and continuous models are presented. Finally, we give an overview of some parametric and nonparametric regression and hazard rate models with change-points.

## 1. Introduction

Change-point models have originally been developed in connection with applications in quality control, where a change from the *in-control* to the *out-of-control* state has to be detected based on the available random observations. Up to now various change-point models have been suggested for a broad spectrum of applications like quality control, reliability, econometrics or medicine.

The general change-point problem can be described as follows: A random process indexed by time is observed and we want to investigate whether a change in the distribution of the random elements occurs. In other words we are interested in determining whether the observed stochastic process is homogeneous or not. Formally, in the discrete time case, let $X_1, X_2, \ldots$ denote a sequence of independent random variables, where the elements $X_1, \ldots, X_{\theta-1}$ have an identical distribution function $F_0$ and $X_\theta, X_{\theta+1}, \ldots$ are distributed according to $F_1$ and the change-point $\theta$ is unknown. Several statistical tests of the null hypothesis $F_0 = F_1$ against the alternative $F_0 \neq F_1$ for some $\theta > 1$ have been suggested. In addition, estimates for the change-point have been proposed and their properties have been investigated. A lot of work has been done in this research field and it is impossible to give an exhaustive overview. The Zentralblatt MATH returns almost 3000 hits, when entering the word "change-point".

Change-point problems can be classified in different ways. Approaches in the classical framework as in the Bayesian framework have been made. Also, there exist models in continuous time as well as in discrete time. Furthermore, the analysis of change-points can be partitioned in sequential and posteriori detection models (ex post analysis). And of course, the problem can be viewed at in a parametric or nonparametric context. Another characterization is whether only one change-point exists or more than one. The early work in change-point analysis is described in the survey article of Zacks (1982). Other comprehensive reviews are given in Bhattacharya (1994) and in the book of Brodsky & Darkhovsky (1993) for nonparametric models. For an overview of limit theorems in change-point problems we refer to Csörgő & Horváth (1997). Change-point methods in process control are extensively studied in eqr275. In this article we want to concentrate on a review of models and methods for sequentially observed data and for change-point in regression and hazard rate models. We contemplate methods used for sequentially observed data in discrete and continuous time in the first section. The second section presents an overview of change-points in regression and hazard rate models. Of course, the selection of models follows our personal view.

## 2. Detection of a change-point in sequentially observed data

The aim of so-called disorder or detection problems is to detect the change-point "as soon as possible" but avoiding too many false alarms. We distinguish discrete time and continuous time models.

*2.1. Discrete time models*

In the discrete case let $X_1, X_2, \dots$ be a sequence of independent random variables which are observed sequentially. The first $X_1, \dots, X_{\theta-1}$, $\theta > 1$ are distributed according to some known distribution $F_0$ while $X_\theta, X_{\theta+1}, \dots$ have some known distribution function $F_1 \neq F_0$. The change-point $\theta$ is unknown. The time of alarm (a change-point has occurred) is determined by a stopping rule which takes the random observations into account. Concerning the change-point there exist Bayesian and non-Bayesian approaches. One of the first non-Bayesian methods is the CUSUM-procedure proposed by Page (1954), which was further investigated by Lorden (1971) and Moustakides (1986).

A Bayesian formalization of the disorder problem goes back to Shiryaev (1963). He postulated that the change-point $\theta$ has a geometric a priori distribution with some parameter $p$ and considered the following risk function $R(\tau)$ for stopping at $\tau$: $R(\tau) = P(\tau < \theta) + cE(\tau - \theta)^+$. Here the penalty costs of a false alarm are normed to 1, and the costs for the delay of stopping after the change-point are $c$ per unit time. Now the Bayes stopping rule is to stop at the smallest $n$ for which the posterior probability $\Pi_n = P(\theta \leq n | X_1, ..., X_n)$ of a change up to $n$ is greater than some threshold $A$ for some $0 < A < 1$.

*2.2. Continuous time model*

Shiryaev (1963) was also one of the first to present a model in continuous time in a Bayesian framework: the change-point is assumed to be a random variable with some prior distribution. Shiryaev considered the following observation process

$$W_t = B_t + r(t - \theta)^+, \quad t \in [0, \infty),$$

where $B$ denotes a standard Brownian motion, $r$ is a known fixed constant and $\theta$ is an unknown (random) change-point, which is assumed to be independent of $B$ and to have a mixed exponential prior distribution: $P(\theta = 0) = p$ and $P(\theta > t) = (1 - p)e^{-\lambda t}$, $p \in [0, 1)$, $\lambda > 0$, $t \geq 0$. The stopping time $\tau$ with respect to the filtration generated by

$W$ should signal the change in the drift as soon as possible. The speed of detection is measured by the risk function $R(\tau)$, which is the same as in the discrete time case. A Bayes solution $\tau^*$ should minimize the risk

$$R(\tau^*) = \inf_{\tau} R(\tau).$$

The optimal stopping time $\tau^*$ can be determined by means of the posterior distribution $\Pi_t = P(\theta \leq t \,|\, \mathcal{F}_t^W)$, with $\mathcal{F}_t^W = \sigma\{W_s : s \leq t\}$. Then the optimal stopping time is

$$\tau^* = \inf\{t > 0 \,|\, \Pi_t \geq p^*\},$$

for some properly chosen $p^* \in [0, 1)$. Details about this approach can be found in Shiryaev (1978). An explicit expression for the optimal threshold $p^*$, and further ramifications can be found in Beibel (1994, 1996).

Another type of change-point problems has been studied in recent years, namely the Poisson disorder problem. Instead of considering a Wiener process with changing drift a Poisson process with changing intensity is observed. Then the problem is to determine a stopping time which signals a change of the intensity of the observed Poisson process. Formally, a point process $(T_n)$, $n \in \mathbb{N}$ and its corresponding counting process $N_t = \sum_{n=1}^{\infty} I_{\{T_n \leq t\}}$ are observed, where $I$ is the indicator function. At an unknown random time $\theta$ the intensity of $N$ switches from $\mu_0$ to $\mu_1 > \mu_0$. This means, that if $\theta$ is given, $N$ is a Poisson process with intensity $\mu_0$ up to $\theta$ and a Poisson process with intensity $\mu_1$ after $\theta$. Peskir & Shiryaev (2002) described the structure of the solution in the general case. They assumed that $\theta$ is a random variable independent of the Poisson process having a mixed exponential prior distribution with parameters $p$ and $\lambda$ as before. The objective is to stop as soon as possible after $\theta$. The associated risk function $R(\tau) = P(\tau < \theta) + cE(\tau - \theta)^+$ is given as before and the optimal stopping time in the case $\mu_1 > \mu_0$ is $\tau^* = \inf\{t \geq 0 \,|\, \Pi_t \geq B^*\}$, where

$$B^* = \begin{cases} \frac{\lambda}{\lambda + c}, & c \geq \mu_1 - \mu_0 - \lambda \\ \frac{\lambda(\mu_1 - c)}{\lambda\mu_1 + c\mu_0}, & \mu^* \leq c < \mu_1 - \mu_0 - \lambda \end{cases}$$

is the optimal stopping threshold and $\mu^* = \frac{\mu_0\mu_1(\mu_0 - \mu_1 - \lambda)}{\mu_0\mu_1 + (\mu_1 - \mu_0)(\lambda + \mu_0)}$ and $(\Pi_t)$, $t > 0$ is the posterior probability process $\Pi_t = P(\theta \leq t \,|\, \mathcal{F}_t^N)$, with respect to the filtration $\mathcal{F}_t^N = \sigma\{N_s : s \leq t\}$ and the prior distribution. This stopping time $\tau^*$ is obtained by the use of the same methodology as in the Wiener process case. Brown & Zacks (2006) extended this result for $\mu^* > c$ using Dynamic Programming.

A different approach was made by Herberts & Jensen (2004). They solved the detection problem by deriving a semimartingale representation of a gain process whose expectation corresponds to the risk process. In a Bayesian framework the change-point has an exponential prior with parameter $\lambda$. The gain process considered by Herberts & Jensen (2004) is a generalization of the one introduced by Shiryaev and is given by

$$Z_t = c_0 \min(t, \theta) + c_1(t - \theta)^+ - k_0(1 - I_{\{\theta \leq t\}}) - k_1 I_{\{\theta \leq t\}}$$

if the process is stopped at time $t$. There is a gain of $c_0$ per unit of time and costs $k_0$ for stopping the process before the change occurs and a gain $c_1$ per unit time and stopping costs $k_1$ afterwards. The problem of detecting the change-point is then formulated as an optimal stopping problem: Find a stopping time $\zeta$ with respect to the given observation filtration $\mathbb{F}$ such that $EZ_\zeta = \sup_\tau\{EZ_\tau : \tau \in C^{\mathbb{F}}\}$ where $(Z_t), t \in \mathbb{R}_+$ is the gain process and $C^{\mathbb{F}}$ is the set of admissible stopping times. What goes beyond the classical approaches is in particular, that their method can be used for different information schemes, i.e. for sequentially observed data, ex post decision after observing the point process up to a fixed time $t^*$ and a combination of both observation schemes. In the sequential case the observation filtration $\mathbb{F} = (\mathcal{F}_t)$ is the point process history $\mathcal{F}_t = \mathcal{F}_t^N$. In the ex post case the history of $N$ is always known for the whole observation period up to $t^*$; therefore, we have in this case $\mathcal{F}_t = \mathcal{F}_{t^*}^N$ for all $t,\ 0 \leq t \leq t^*$. In both cases the optimal stopping time can be determined explicitly. In the case of sequential observation it is given by

$$\zeta^{\mathbb{F}^N} = \inf\{t \in \mathbb{R}_+ : \hat{Y}_t^{\mathbb{F}^N} \geq z^*\},$$

where $\hat{Y}_t^{\mathbb{F}^N} = P(\theta \leq t | \mathcal{F}_t^N)$ can be calculated explicitly and $z^* = \frac{c_0+\lambda(k_0-k_1)}{c_0-c_1+\lambda(k_0-k_1)}$. In the ex post case the solution is a little bit more complex and can be found among other details and more references in Herberts & Jensen (2004).

In recent years not only homogeneous but also inhomogeneous Poisson process models have been studied. We refer to Dabye et al. (2003) and Ruggeri & Sivaganesan (2005).

## 3. Change-Points in Regression and Hazard Rate Models

### *3.1. Regression models*

In the literature two different types of change-point regression models can be found: On the one hand so called time-varying regression models, in which the model parameters change at some unknown point in time, and on the other hand two-phase regression models. Both models are presented briefly in the following.

In the time-varying model a change in the regression coefficients takes place from the early to the late observations of a sequence $(X_n)$, $n \in \mathbb{N}$. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent random vectors. Then the model in the random design is given by

$$Y_i = \begin{cases} \alpha_0 + \alpha_1 X_i + \epsilon_i, & i \leq \tau \\ \beta_0 + \beta_1 X_i + \epsilon_i, & i \geq \tau + 1 \end{cases}$$

where $(X_i)$ and $(\epsilon_i)$, $i = 1, \ldots, n$ are mutually independent iid sequences with $E(\epsilon_i) = 0$ and $E(\epsilon_i^2) = 1$ and $(\alpha_0, \alpha_1) \neq (\beta_0, \beta_1)$. If $1 \leq \tau \leq n-1$, then $\tau$ is a change-point. This design is called fixed, if the sequence $(X_i)$, $i = 1, \ldots, n$ is non-stochastic.

A two-phase regression model is a regression model with piecewise linear regression functions over two different domains of the design-variable. The random design of a two-phase regression model is given by

$$Y_i = (\alpha_0 + \alpha_1 X_i) I_{\{X_i \leq \tau\}} + (\beta_0 + \beta_1 X_i) I_{\{X_i > \tau\}} + \epsilon_i = m(X_i) + \epsilon_i. \tag{1}$$

The two-phase regression models can be classified further into a restricted and an unrestricted case. In the restricted case the regression function $f$ is continuous but not differentiable, whereas in the unrestricted case the regression function is discontinuous. The discontinuity can be expressed in form of a fixed jump size or a contiguous jump size, in which the jump size tends to zero as the sample size tends to infinity.

Hinkley (1971) was one of the first authors to investigate a maximum likelihood estimator of the point of intersection for the special case of two line segments under normally distributed errors. A generalization of his model with multiple change-points was considered by Feder (1975a,b), who investigated least squares estimates and showed, that these estimates are consistent under suitable identifiability assumptions and the asymptotic distributions of these estimates are obtained by "classical" methods.

Koul & Qian (2002), Koul et al. (2003) considered M-estimators in the unrestricted two-phase random design with a fixed jump size. The M-process corresponding to a function $\phi : \mathbb{R} \to [0, \infty)$ is defined as

$$M_n(\boldsymbol{\theta})) = \sum_{i=1}^{n} \phi(Y_i - m(X_i, \boldsymbol{\theta})),$$

where $m(X; \boldsymbol{\theta})$ is the linear regression function of model (1) and the M-estimator $\hat{\boldsymbol{\theta}}$ is given as the minimizer of the M-process:

$$M_n(\hat{\boldsymbol{\theta}}) = \inf_{\boldsymbol{\theta}} M_n(\boldsymbol{\theta}) \quad \text{a. s.}$$

They showed that the estimate of the jump point converges with rate $O_p(n^{-1})$, whereas the consistency rate of the coefficient parameters is $O_p(n^{-1/2})$. The normalized M-process is asymptotically equivalent to the sum of two processes. One is a quadratic form in the standardized coefficient parameter vector, the other is a jump point process in the change-point parameter. This result can be exploited to show weak convergence. The suitably standardized M-estimator of the change-point converges weakly to the minimizer of a compound Poisson process. The estimates of the regression coefficients are asymptotically normal and independent of the jump point M-estimator. This is remarkable because the results differ from the restricted and unrestricted contiguous non-random design cases.

All models considered above assumed a parametric setting. Of course, there exist various nonparametric models as well. Müller (1992) studied the following fixed design nonparametric regression model

$$Y_{in} = g(t_{in}) + \epsilon_{in}, \quad t_{in} \in [0,1],\ 1 \leq i \leq n,$$

where $Y_{in}$ are noisy measurements of the regression function $g$ taken at points $t_{in}$ and $\epsilon_{in} \sim \mathcal{N}(0, \sigma^2)$ are iid errors. The assumption is made that there is a change-point for the $\nu$th derivative $g^{(\nu)}$ at $\tau,\ 0 < \tau < 1$ in the following sense: There exists a function $f \in \mathcal{C}^{k+\nu}([0,1])$ with $\nu \geq 0$ and $k \geq 2$ an even integer, such that

$$g^{(\nu)}(t) = f^{(\nu)}(t) + \Delta_\nu I_{[\tau,1]}(t), \quad \Delta_\nu > 0,\ 0 \leq t \leq 1.$$

The case $\Delta_\nu < 0$ can be treated analogously. Now, the jump size at the possible change-point $\tau$ of the $\nu$th derivative of $g$ is given by

$$\Delta_\nu = g_+^{(\nu)}(\tau) - g_-^{(\nu)}(\tau),$$

where $g_+^{(\nu)}(x) = \lim_{y \downarrow x} g^{(\nu)}(y)$ and $g_-^{(\nu)}(x) = \lim_{y \uparrow x} g^{(\nu)}(y)$ are the one-sided limits of the derivative $g^{(\nu)}(x)$. Hence, the idea is to base the inference of the change-points on differences between the left and right sided estimates of $g^{(\nu)}(t)$, which can be done by suitably chosen one sided kernel estimates. The location of the maximum of these differences is a reasonable estimator of the location of the change-point. Let $\tau$ be an element of a closed interval $T \subset (0,1)$. Then the estimator is

$$\hat{\tau} = \inf\{\rho \in T : \hat{\Delta}_\nu(\rho) = \sup_{x \in T} \hat{\Delta}_\nu(x)\}.$$

In this setting Müller (1992) proved weak convergence of the estimator $\hat{\tau}$. Loader (1996) considered a similar nonparametric regression model in which the mean function may have a discontinuity at an unknown point. His estimate is similar in principle to that studied by Müller (1992). But since he imposed different conditions on the kernel $K$, his estimate has different properties. It is shown that the change-point estimate converges in probability with rate $O_P(n^{-1})$ and that it has the same asymptotic distribution as maximum likelihood estimates in parametric models.

The same convergence rate is attained by Müller & Song (1997) in a two-step estimation of the change-point in a nonparametric fixed design regression model with fixed jump size, whereas the convergence rate in the contiguous case is $O_P(n^{-1}\Delta_n^{-2})$, where $\Delta_n$ is a sequence of jump sizes which tends to zero.

Another important problem in modeling data is the question whether an unknown function, which cannot parametrically be specified, should be modeled as a globally smooth function or a smooth function with isolated change-points. Müller & Stadtmüller (1999) proposed statistics which provide relevant information for this decision.

### *3.2. Hazard rate models*

Hazard rate models often occur in medical follow up studies after major surgery. The simplest one with a change-point can be expressed as follows

$$\lambda(t) = \begin{cases} \lambda_1 & t \leq \tau \\ \lambda_2 & t > \tau \end{cases}, \; t \geq 0\,, \; \tau \geq 0$$

with constants $\lambda_1, \lambda_2 > 0$ and change-point $\tau$. A first attempt to estimate these three parameters was made by Anderson & Senthilselvan (1982). They investigated as a special case of this simple model an extended Cox model with $\lambda_1 = e^{\boldsymbol{\alpha}^T \boldsymbol{Z}}, \lambda_2 = e^{\boldsymbol{\gamma}^T \boldsymbol{Z}}$, where $\boldsymbol{\alpha}$ and

$\boldsymbol{\gamma}$ are parameter vectors and $\boldsymbol{Z}$ is a vector of covariates. The parameters are estimated by the conditional log-likelihood given the value of $\tau$ and then the baseline hazard $\lambda(t)$ is estimated by a penalized maximum likelihood method conditioning on the parameter estimates. Liang et al. (1990) proposed a slightly different Cox model

$$\lambda(t) = \lambda_0(t) \exp\{(\beta + \theta I_{\{t \leq \tau\}})Z + \boldsymbol{\gamma}^T \boldsymbol{X}\},$$

where $Z$ is a one-dimensional covariate which should be included in possibly different magnitudes over time, $\boldsymbol{X}$ is another confounding covariate vector and the change-point at an unknown time is given by $\tau$. They tested the hypothesis of $H_0 : \ \theta = 0$ by using a test statistic

$$M = \sup_{\tau \in [a,b]} S(\tau),$$

where $S$ is a function of the first two derivatives of the partial log likelihood function with respect to $\beta$ and $\gamma$.

A further Cox model is presented by Luo & Boyett (1997):

$$\lambda(t) = \lambda_0(t) \exp\{\beta I_{\{X \leq \theta\}} + \boldsymbol{\alpha}^T \boldsymbol{Z}\},$$

where a constant is added to a covariate beyond a threshold, which is characterized by a random variable X. They proved consistency of their partial MLE.

A Cox model for independent and identically distributed right censored survival times with a change-point according to the unknown threshold of a covariate was introduced by Pons (2003):

$$\lambda(t) = \lambda_0(t) \exp\{\boldsymbol{\alpha}^T \boldsymbol{Z}_1(t) + \boldsymbol{\beta}^T \boldsymbol{Z}_2(t) I_{\{Z_3 \leq \zeta\}} + \boldsymbol{\gamma}^T \boldsymbol{Z}_2(t) I_{\{Z_3 > \zeta\}}\}$$

In this model it is shown that the partial MLE of the change-point $\zeta$ is $n$-consistent, i.e. the convergence rate is $O_p(n^{-1})$. Such a convergence rate was also attained for the change-point in the unrestricted two-phase random linear regression design with a fixed jump size (see Koul & Qian (2002)). Furthermore, Pons (2003) proved that the estimates of the regression parameter vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ are $n^{1/2}$-consistent and that $n(\hat{\zeta}_n - \zeta)$ converges weakly to a random variable $\hat{\nu}_Q$ which is a maximizer of a certain jump process. The estimates of the regression parameters are asymptotically normal.

Gandy et al. (2005) investigated a similar model. But instead of a jump there is a smooth change at the change-point $\xi$:

$$\lambda_i(t, \boldsymbol{\theta}) = \lambda_0(t) R_i(t) \exp\left\{ {\boldsymbol{\beta}_1}^T \mathbf{Z}_{1i}(t) + \beta_2 Z_{2i} + \beta_3 (Z_{2i} - \xi)^+ \right\},$$

where $\boldsymbol{\theta} = (\xi, \boldsymbol{\beta}^T)^T$ with $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \beta_2, \beta_3)^T \in \mathcal{B} \subset \mathbb{R}^{p+2}$ is the vector of regression parameters and $R_i$ the at risk indicator. The change-point is denoted by $\xi$ lying in a closed interval of known parameters $[\xi_1, \xi_2]$. For brevity, $\tilde{\boldsymbol{Z}}_i(t; \xi) = \left( \mathbf{Z}_{1i}^T(t), Z_{2i}, (Z_{2i} - \xi)^+ \right)^T$ is considered resulting in $\lambda_i(t, \boldsymbol{\theta}) = \lambda_0(t) R_i(t) \exp\left\{ \boldsymbol{\beta}^T \tilde{\boldsymbol{Z}}_i(t; \xi) \right\}$. In this new extended Cox model $\boldsymbol{\theta}$ is estimated by the value $\hat{\boldsymbol{\theta}}_n$ that maximizes the logarithm of the partial likelihood

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^{\tau} \boldsymbol{\beta}^T \tilde{\boldsymbol{Z}}_i(t; \xi)\, dN_i(t) - \int_0^{\tau} \log \left( \sum_{i=1}^{n} R_i(t) \exp(\boldsymbol{\beta}^T \tilde{\boldsymbol{Z}}_i(t; \xi)) \right) d \left( \sum_{i=1}^{n} N_i(t) \right).$$

The maximization can be carried out in the following way:

For fixed $\xi$, let $\hat{\boldsymbol{\beta}}_n(\xi) = \operatorname{argmax}_{\boldsymbol{\beta} \in \mathcal{B}} \log L(\xi, \boldsymbol{\beta})$ and $\log L(\xi) = \log L(\xi, \hat{\boldsymbol{\beta}}_n(\xi))$. Then $\xi$ can be estimated by $\hat{\xi}_n$ satisfying

$$\hat{\xi}_n = \inf\{\xi \in [\xi_1, \xi_2] : \log L(\xi) = \sup_{\xi \in [\xi_1, \xi_2]} \log L(\xi)\}.$$

The partial maximum likelihood estimator of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}}_n = (\hat{\xi}_n, \hat{\boldsymbol{\beta}}_n)$, where $\hat{\boldsymbol{\beta}}_n = \hat{\boldsymbol{\beta}}_n(\hat{\xi}_n)$.

It is proved that the partial MLE of all parameters is $n^{1/2}$-consistent. In contrast to Pons' model $n^{1/2}(\xi - \xi_0)$ is asymptotically normal. The proof uses results of the theory of $M$-estimators as presented in Van der Vaart (1998). The model can be extended further by substituting a vector $\boldsymbol{Z}_2$ for $Z_2$, which may also be time-dependent, and a vector $\boldsymbol{\xi}$ for $\xi$ allowing more than only one change-point. The asymptotic results are still the same.

Other approaches relying on the Cox model have recently been suggested by Dupuy (2006). He considered a model with a change-point in both hazard and regression parameters.

Estimates of the change-point, hazard and regression parameters are proposed and shown to be consistent.

**Acknowledgment**

Financial support of the *Deutsche Forschungsgemeinschaft* through Research Unit 460 is gratefully acknowledged.

**Related Article**

eqr275

**References**

Anderson, J. & Senthilselvan, A. (1982). A two-step regression model for hazard functions. *Appl. Statist.* **31**, 44–51.

Beibel, M. (1994). Bayes problems in change-point models for the Wiener process. In *Change-point Problems* (eds. E. Carlstein, H. Müller & D. Siegmund), 1–6, Institute of Mathematical Statistics, Hayward.

Beibel, M. (1996). A note on Ritov's Bayes approach to the minimax property of the CUSUM-procedure. *Ann. Stat.* **24**, 1804–1812.

Bhattacharya, P. (1994). Some aspects of change-point analysis. In *Change-point Problems* (eds. E. Carlstein, H. Müller & D. Siegmund), 28–56, Institute of Mathematical Statistics, Hayward.

Brodsky, B. & Darkhovsky, B. (1993). *Nonparametric Methods in Change-Point Problems*. Kluwer Academic Publishers, Dordrecht.

Brown, M. & Zacks, S. (2006). A note on optimal stopping for possible change in the intensity of an ordinary Poisson process. *Stat. Probab. Lett.* **76**, 1417–1425.

Csörgő, M. & Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, Chichester.

Dabye, A., Farinetto, C. & Kutoyants, Y. (2003). On Bayesian estimators in misspecified change-point problems for Poisson process. *Stat. Probab. Lett.* **61**, 17–30.

Dupuy, J.-F. (2006). Estimation in a change-point hazard regression model. *Stat. Probab. Lett.* **76**, 182–190.

Feder, P. (1975a). On asymptotic distribution theory in segmented regression problems. *Ann. Stat.* **3**, 49–83.

Feder, P. (1975b). The log likelihood ratio in segmented regression. *Ann. Stat.* **3**, 84–97.

Gandy, A., Jensen, U. & Lütkebohmert, C. (2005). A Cox model with a change-point applied to an actuarial problem. *Braz. J. Probab. Stat.* **19**, 93–109.

Herberts, T. & Jensen, U. (2004). Optimal detection of a change point in a Poisson process for different observation schemes. *Scand. J. Stat.* **31**, 347–366.

Hinkley, D. (1971). Inference in two-phase regression. *J. Am. Stat. Assoc.* **66**, 736–743.

Koul, H. & Qian, L. (2002). Asymptotics of maximum likelihood estimator in a two-phase linear regression model. *J. Stat. Plann. Inference* **108**, 99–119.

Koul, H., Quian, L. & Surgailis, D. (2003). Asymptotics of M-estimators in two-phase linear regression models. *Stochastic Processes Appl.* **103**, 123–154.

Liang, K.-Y., Self, S. & Liu, X. (1990). The Cox proportional hazard model with change-point: An epidemiologie application. *Biometrics* **46**, 783–793.

Loader, C. (1996). Change point estimation using nonparametric regression. *Ann. Stat.* **24**, 1667–1678.

Lorden, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Stat.* **42**, 1897–1908.

Luo, X. & Boyett, J. (1997). Estimations of a threshold parameter in Cox regression. *Commun. Stat., Theory Methods* **26**, 2329–2346.

Moustakides, G. (1986). Optimal stopping times for detecting changes in distribution. *Ann. Stat.* **14**, 1379–1387.

Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Stat.* **20**, 737–761.

Müller, H.-G. & Song, K.-S. (1997). Two-stage change-point estimators in smooth regression. *Stat. Probab. Lett.* **34**, 323–335.

Müller, H.-G. & Stadtmüller, U. (1999). Discontinuous versus smooth regression. *Ann. Stat.* **27**, 299–337.

Page, E. (1954). Continuous inspection schemes. *Biometrika* **41**, 100–115.

Peskir, G. & Shiryaev, A. (2002). Solving the Poisson disorder problem. In *Advances in Finance and Stochastics* (eds. K. Sandmann & P. Schönbacher), 295–312, Springer, Berlin.

Pons, O. (2003). Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *Ann. Stat.* **31**, 442–463.

Ruggeri, F. & Sivaganesan, S. (2005). On modeling change points in non-homogeneous Poisson processes. *Stat. Inference Stoch. Process.* **8**, 311–329.

Shiryaev, A. (1963). On optimum methods in quickest detection problems. *Theor. Probab. Appl.* **8**, 22–46.

Shiryaev, A. (1978). *Optimal Stopping Rules*. Springer, New York.

Van der Vaart, A. (1998). *Asymptotic statistics*. Cambridge University Press, New York.

Zacks, S. (1982). Classical and Bayesian approaches to the change-point problem: Fixed sample and sequential procedures. *Stat. Anal. Donnees* **7**, 48–81.