On the Bayesian Sequential Change-Point Detection

Gholamhossein Gholami

Department of Mathematics, Faculty of Sciences, Urmia University, Iran

Abstract. The problems of sequential change-point have several important applications, including quality control, failure detection in industrial, finance and signal detection. We discuss a Bayesian approach in the context of statistical process control: at an unknown time τ, the process behavior changes and the distribution of the data changes from p_0 to p_1. Two cases are considered: (i) p_0 and p_1 are fully known, (ii) p_0 and p_1 belong to the same family of distributions with some unknown parameters $\theta_1 \neq \theta_2$. We present a maximum a posteriori estimate of the change-point which, for the case (i) can be computed in a sequential manner. In addition, we propose the use of the Shiryaev's loss function. Under this assumption, we define a Bayesian stopping rule. For the Poisson distribution and in the two cases (i) and (ii), we obtain results for the conjugate prior.

Keywords. Bayesian sopping rule, Change-point detection, Maximum a posteriori estimation, Sequential Bayesian analysis, Shiryaev's loss function.

Gholamhossein Gholami (gh.gholami@urmia.ac.ir)

1   Introduction

The analysis of change-point describes sudden localized changes typically occurring in economics, medicine, and physical sciences. An example of application of change-point analysis is the Statistical process control (SPC). The SPC consists of methods for understanding, monitoring, and improving process performance, for instance, the performance of a production line of a soft-drink over time.

Suppose that, for a given experiment (or process), a sample of independent sequential observations $\{y_t, t \geq 1\}$ is available. When the process is *under-control,* the distribution of the process measurements equals a specific distribution $p_0(y)$. If the measurement distribution changes after an unknown time point, then the process becomes *out-of control* and the following observations are then derived from a different distribution $p_1(y)$. The major objective of the analysis is to detect such a change as soon as possible while keeping false alarms of being in out-of control as infrequent as possible so that the process can be stopped and the causes of the shift can be checked out in a timely fashion.

When $p_0(y)$ and $p_1(y)$ are completely specified, the problem is well known and has been studied under a variety of criteria. Some well-known methods to deal with this problem are Shewharts control charts, Moving Average control charts, Pages CUSUM procedure, and the Shiryaev-Roberts procedure (Lorden, 1971). However, in practice, the assumption of completely known under-control $p_0(y)$ and out-of control $p_1(y)$ distributions are not realistic. Then, a more flexible formulation may assume that these distributions belong to some known families with unknown parameters. Therefore, the change-point problem detection procedure should deal with estimating these unknown parameters as well.

In most applications, we are concerned with shifts in the mean parameter from μ_0 to $\mu_1 > \mu_0$ of the measurement distribution. These parameters may be known or unknown. A standard change-point detection procedure needs to specify a required rate of false alarms when $\mu_t = \mu_0$, to minimize detection delay if $\mu_t \geq \mu_1$ and to estimate these parameters (Lai, 1995; Lorden, 1971; Pollak, 1987; Siegmund and Venkatraman, 1995). These mean parameters could be a function of some explanatory variables. If we let $\mu_0 = f_0(\boldsymbol{x})$ and $\mu_1 = f_1(\boldsymbol{x})$, the problem is addressed as the change-point problem in regression analysis (Gholami, 2010). It is ideal if we could optimize all possible false alarm rates and all possible detection delays. Unfortunately, due to the lack of an attractive definition of optimality in the literature for this problem, this cannot be done (Mei, 2003).

In this paper, we discuss a Bayesian approach in the context of statistical process

control. We suppose that, at an unknown time τ, the process behavior changes and the distribution of the data changes from p_0 to p_1. As we do not assume any prior information for the change-point location, we use a discrete uniform distribution for τ. We consider two cases: (i) p_0 and p_1 are fully known, (ii) p_0 and p_1 belong to the same family of distributions with some unknown parameters $\theta_1 \neq \theta_2$. For the case (i), we present a maximum a posteriori estimate of the change-point which can be computed in a sequential manner. Furthermore, we propose the use of the Shiryaev's loss function which assumes that we lose c for each observation taken after τ, and lose 1 for a false alarm. Under this loss function, we define a Bayesian stopping rule. For the Poisson distribution and, in two cases (i) and (ii), we obtain results for the conjugate prior.

2   Problem Statement

The sampling distribution of the first t observations is given by

$$
p_t(y_{1:t}|\tau) = \begin{cases} \displaystyle\prod_{i=1}^{\tau} p_0(y_i) \prod_{i=\tau+1}^{t} p_1(y_i) & \text{if} \quad 1 \leq \tau \leq t-1 \\[2em] \displaystyle\prod_{i=1}^{t} p_0(y_i) & \text{if} \qquad \tau = t\,, \end{cases} \tag{2.1}
$$

where the discrete unknown parameter τ indicates a *change-point* in the sample. A related problem is to test sequentially whether the experiment is under control, i.e., for each t we want to test the null hypothesis of no change,

$$
H_0 : p_t(y_{1:t}|\tau) = \prod_{i=1}^{t} p_0(y_i),
$$

versus the alternative of a change-point,

$$
H_1 : \{p_t(y_{1:t}|\tau), 1 \leq \tau \leq t-1\}\,.
$$

As a consequence, we stop the experiment at the first value of t for which H_0 is rejected.

From a Bayesian point of view, we need to solve a testing problem which is formulated as a model selection problem between

$$\mathcal{M}_1 \quad : \quad p_t(y_{1:t}) \tag{2.2}$$

$$\mathcal{M}_2 \quad : \quad \{p_t(y_{1:t}|\tau), \pi_t(\tau)\}.$$

3 Maximum a Posteriori Estimation of Change-Point

By considering a discrete uniform distribution at times t for the change-point, i.e.,

$$\pi_t(\tau) = \frac{1}{t},$$

the posterior distribution will be

$$\pi_t(\tau|y_{1:t}) \quad \propto \quad \begin{cases} \pi_t(\tau)\displaystyle\prod_{i=1}^{\tau} p_0(y_i) \prod_{i=\tau+1}^{t} p_1(y_i) & \text{if} \quad 1 \le \tau \le t-1 \\ \\ \pi_t(\tau)\displaystyle\prod_{i=1}^{t} p_0(y_i) & \text{if} \quad \tau = t, \end{cases} \tag{3.1}$$

and the sequential relationship

$$\pi_{t+1}(\tau|y_{1:t+1}) \quad \propto \quad \begin{cases} \pi_t(\tau|y_{1:t})p_1(y_{t+1}) & \text{if} \quad 1 \le \tau \le t \\ \\ \pi_t(t|y_{1:t})p_0(y_{t+1}) & \text{if} \quad \tau \ge t+1. \end{cases} \tag{3.2}$$

holds between the posterior at times t and $t+1$.

Note that (3.1) is not in a sequential manner. Then, $\tau = t$ means that the change-point location is the last data (or there is no change-point), while (3.2) is written in a sequential way and $\tau \ge t+1$ means that the change-point will be at time $t+1$ or in a time after that. The maximum a posteriori (MAP) estimator is estimated as

$$\hat{\tau}_{t+1} \quad = \quad \begin{cases} \hat{\tau}_t & \text{if} \quad \pi_t(\hat{\tau}_t|y_{1:t})p_1(y_{t+1}) \ge \pi_t(t|y_{1:t})p_0(y_{t+1}) \\ \\ t+1 & \text{if} \quad \text{Otherwise,} \end{cases} \tag{3.3}$$

when $\hat{\tau}_1 = 1$.

4 Bayesian Stopping Rule

We consider the loss function

$$
L_t(\tilde{\tau}_t, \tau) \;=\; \begin{cases} 0 & \tilde{\tau}_t = \tau \\ c(\tilde{\tau}_t - \tau) & \tilde{\tau}_t > \tau \\ 1 & \tilde{\tau}_t < \tau, \end{cases}
$$

which is called Shiryaev's loss function (Lai, 1995), to penalize wrong decisions. This loss function naturally gives more weight to the situation that we have exceeded the true change-point. This cost is proportional to the delay in detecting the change-point, and $c > 0$ is a tuning parameter which should be chosen before performing the analysis.

At time t, our estimation for change-point is a value that minimizes the Bayes risk, i.e.,

$$
\bar{\tau}_t = \arg\ \min \mathbb{E}_{\pi_t(.|y_{1:t})}[L_t(\tilde{\tau}_t, .)], \tag{4.1}
$$

when

$$
\mathbb{E}_{\pi_t(.|y_{1:t})}[L_t(\tilde{\tau}_t, \tau)] = \sum_{i=1}^{\tilde{\tau}_t - 1} c(\tilde{\tau}_t - i)\pi_t(\tau = i|y_{1:t}) + \mathbb{P}_{\pi_t}(\tau \geq \tilde{\tau}_t + 1). \tag{4.2}
$$

The Bayesian stopping rule is defined by

$$
t^\star = \min\{t : \bar{\tau}_t < t\}, \tag{4.3}
$$

which solves the model selection problem mentioned in (2.2), i.e., at time $t^\star$, we select $\mathcal{M}_2$ meaning that a change-point has occurred.

5 The Parametric Change-Point Problem

Suppose that $p_0(y)$ and $p_1(y)$ are densities belonging to a parametric family which is denoted by $p(y|\theta_1)$ and $p(y|\theta_2)$, respectively, where θ_1 and θ_2 are unknown points of the parameter space Θ. Then, the sampling distribution of the first t observation is

$$
p_t(y_{1:t}|\tau, \theta_1, \theta_2) = \begin{cases} \displaystyle\prod_{i=1}^{\tau} p(y_i|\theta_1) \prod_{i=\tau+1}^{t} p(y_i|\theta_2) & \text{if} \quad 1 \leq \tau \leq t-1 \\[2em] \displaystyle\prod_{i=1}^{t} p(y_i|\theta_1) & \text{if} \quad \tau = t. \end{cases} \tag{5.1}
$$

In this case, for some prior distributions $\pi(\theta_1)$ and $\pi(\theta_2)$, the derived results in Section 1 hold with

$$
p_t(y_{1:t}|\tau) = \begin{cases} \int \prod_{i=1}^{\tau} p(y_i|\theta_1) \prod_{i=\tau+1}^{t} p(y_i|\theta_2)\pi(\theta_1)\pi(\theta_2)d\theta_1 d\theta_2 & \text{if} \quad 1 \le \tau \le t-1 \\[2em] \int \prod_{i=1}^{t} p(y_i|\theta_1)\pi(\theta_1)d\theta_1 & \text{if} \qquad \tau = t. \end{cases}
\tag{5.2}
$$

At time t, the posterior distribution of θ_1 and θ_2 is

$$
\pi_t(\theta_1, \theta_2|y_{1:t}) = \frac{\sum_{\tau=1}^{t} p_t(y_{1:t}|\tau, \theta_1, \theta_2)\pi_t(\tau)\pi(\theta_1, \theta_2)}{\sum_{\tau=1}^{t} \int p_t(y_{1:t}|\tau, \theta_1, \theta_2)\pi_t(\tau)\pi(\theta_1, \theta_2)d\theta_1 \theta_2}
$$

$$
= \frac{\sum_{\tau=1}^{t} p_t(y_{1:t}|\tau, \theta_1, \theta_2)\pi(\theta_1, \theta_2)}{\sum_{\tau=1}^{t} \int p_t(y_{1:t}|\tau, \theta_1, \theta_2)\pi(\theta_1, \theta_2)d\theta_1 \theta_2}.
\tag{5.3}
$$

6   Prior Elicitation

For a change-point analysis, choosing the prior distribution on the unknown parameters is a delicate consideration. The classical change-point literature reveals a preference for uniform priors on τ. Obviously, before seeing the data, every value for τ, apart from initial conditions and endpoints, is equally plausible. Therefore, at time t, the change-point prior will be $\pi_t(\tau) = \frac{1}{t}\mathbb{I}_{\{1,...,t\}}(\tau)$.

In the context of quality control Dehghan Monfared and Meshkani (2010) used EWMA[1] charts structure to construct an informative prior distribution for the change-point. In this context, one does not expect that the change happens immediately then the prior should give less weights to the data when the procedure starts to monitor. Note that these methods were developed for continuous data and could not be used for discrete distributions.

_____

[1]Exponentially Weighted Moving Average

For the sampling densities parameters θ_1 and θ_2, either conjugate priors or vague priors can be used. Note that the conjugate priors need to asses values for the hyper-parameters. Therefore, some sort of subjective input or empirical Bayes estimation is necessary. Furthermore, these hyper-parameters can be fixed in accord with the available information, which is the case in the control theory, for instance. Note that when we say the procedure is under-control, it implies that there is at least some information about the range of the parameters. It is possible to consider a hierarchical model structure where we introduce hyper-priors on the hyper-parameters. This also enables the data to solidate the choice of the prior. It is possible to choose the hyper-priors to be improper in many situations.

On the contrary, when the system is out-of control and our aim is to detect this situation as soon as possible, we do not expect to have enough data to asses the value of hyper-parameters in the case of conjugate priors. Note that the amount of available data for estimating θ_2 is controlled by the loss function. Using more data and consequently driving better estimation for θ_2 leads to long delays in reporting the occurrence of the change-point. Therefore, the priors we wish to use are the reference priors which have minimum effects on the posterior.

Girón *et al.* (2007) propose the use of *intrinsic priors* for the unknown parameters. They are derived from the arithmetic intrinsic Bayes factor (Berger and Pericchi, 1996) plus an asymptotic argument (Moreno *et al.*, 1998). While these priors do use the available information about the nature of underling process, they are not easy to compute for the models and we prefer to use an alternative prior such as hierarchical or conjugate priors.

7   The Poisson Case

Let us assume that the sampling density is a Poisson distribution. At time t, the likelihood function is

$$p_t(y_{1:t}|\tau, \lambda_1, \lambda_2) = \begin{cases} \displaystyle\prod_{i=1}^{\tau} \frac{e^{-\lambda_1}\lambda_1^{y_i}}{y_i!} \prod_{i=\tau+1}^{t} \frac{e^{-\lambda_2}\lambda_2^{y_i}}{y_i!} & \text{if} \quad 1 \leq \tau \leq t-1 \\[2em] \displaystyle\prod_{i=1}^{t} \frac{e^{-\lambda_1}\lambda_1^{y_i}}{y_i!} & \text{if} \qquad \tau = t. \end{cases} \tag{7.1}$$

If we consider a conjugate prior on the parameters

$$\lambda_1 \sim \mathcal{G}(\alpha_1, \beta_1), \quad \lambda_2 \sim \mathcal{G}(\alpha_2, \beta_2),$$

and a uniform prior on $\{1, \ldots, t\}$ for the change-point, by integrating out the parameters, the posterior distribution of τ will be

$$\pi_t(\tau|y_{1:t}) \quad \propto \quad \frac{\Gamma\left(\tau\bar{y}_1 + \alpha_1\right)}{(\tau + \beta_1)^{\tau\bar{y}_1 + \alpha_1}} \frac{\Gamma\left((t-\tau)\bar{y}_2 + \alpha_2\right)}{(t - \tau + \beta_2)^{(t-\tau)\bar{y}_2 + \alpha_2}} = \pi_t^*(\tau|y_{1:t}) \tag{7.2}$$

$$= \quad \frac{\pi_t^*(\tau|y_{1:t})}{\displaystyle\sum_{\tau=1}^{t} \pi_t^*(\tau|y_{1:t})},$$

where $\bar{y}_1 = \sum_{i=1}^{\tau} y_i/\tau$ and $\bar{y}_2 = \sum_{i=\tau+1}^{t} y_i/(t-\tau)$.

For given τ and t, the posterior distributions will be

$$\pi(\lambda_1|y_{1:\tau}) = \mathscr{G}(\lambda_1; \tau\bar{y}_1 + \alpha_1, \beta_1 + \tau)$$

and

$$\pi(\lambda_2|y_{\tau+1:t}) = \mathscr{G}(\lambda_2; (t-\tau)\bar{y}_2 + \alpha_2, \beta_2 + t - \tau);$$

then,

$$\hat{\lambda}_{1,t} = \frac{\tau\bar{y}_1 + \alpha_1}{\beta_1 + \tau} \qquad \text{and} \qquad \hat{\lambda}_{2,t} = \frac{(t-\tau)\bar{y}_2 + \alpha_2}{\beta_2 + t - \tau}.$$

For assessing the performance of the method, we can construct MSE and HPD intervals. For each t, MSE is defined as

$$MSE_t = \mathbb{E}\left(\hat{\tau}_t - \tau\right)^2 \approx \frac{1}{t}\sum_{j=1}^{t}(j - \tau)^2, \tag{7.3}$$

and a $100(1-\alpha)\%$ HPD interval for τ_t is defined by

$$C_t = \{\tau \mid \pi_t(\tau \mid y_{1:t}) \geq k\}, \tag{7.4}$$

where k is the largest value such that

$$\sum_{C_t} \pi_t(\tau \mid y_{1:t}) = 1 - \alpha.$$

If we consider the Jeffreys prior for the parameters, i.e., $\pi(\lambda) \propto \lambda^{-1/2}$, the posterior of τ will be

$$\pi_t(\tau|y_{1:t}) \propto \frac{\Gamma\left(\tau\bar{y}_1 + 0.5\right)}{\tau^{\tau\bar{y}_1 + 0.5}} \frac{\Gamma\left((t-\tau)\bar{y}_2 + 0.5\right)}{(t-\tau)^{(t-\tau)\bar{y}_2 + 0.5}}, \tag{7.5}$$

which is an special case of (7.2) for $\alpha_1 = \alpha_2 = 0.5$ and $\beta_1 = \beta_2 = 0$.
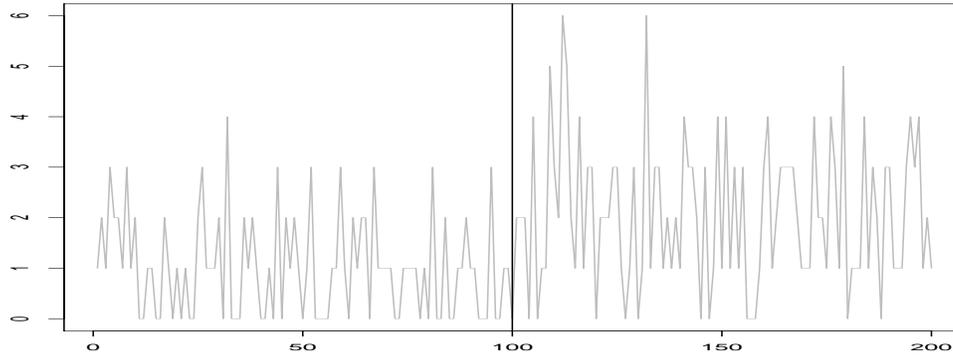
Figure 1: BDAEK. Simulated data from two Poisson distribution with parameters equal to 1 and 2. The change-point is located at $t = 100$.

7.1 Examples

7.1.1 Both Distributions Are Entirely Known (BDAEK)

We generate $T = 200$ observations form two Poisson distributions when $\lambda_1 = 1$, $\lambda_2 = 2$ and $\tau = 100$. Figure 2 shows the posterior distribution of change-point at different times. As it was expected, for $t < \tau$ (meaning that no change has occurred), the posterior mode is located at $t \geq \tau$. As we are receiving more observations and approaching the real change-point, the posterior starts to have its mode around the real change-point value. Furthermore, when the change occurs ($t \geq \tau$), the posterior mode is still located around the change-point and does not move, so that, for $t \geq 110$, the same posterior is approximately obtained.

Figure 3 exhibits the MAP estimate at different times. As it is observed, the MAP estimation is a non-decreasing function of time. The estimator has an increasing trend till time $t = 97$ and, after that, it is constant and takes 97 as the estimation of τ. Figure 4 gives the MAP estimates for different values of c, which is a penalization factor, at different times. As mentioned before, this parameter is a tuning parameter and should be fixed before the experiment. As expected, for its larger values, we can detect change-point rapidly which is almost a wrong alarm of detecting. In this figure, the solid line corresponds to $c = 0$ which does not penalize the delay in detecting the change-point and it is an extreme choice. Then, we never detect the change-point even when we stop the monitoring of the system (here when $t = T = 200$). Another extreme choice
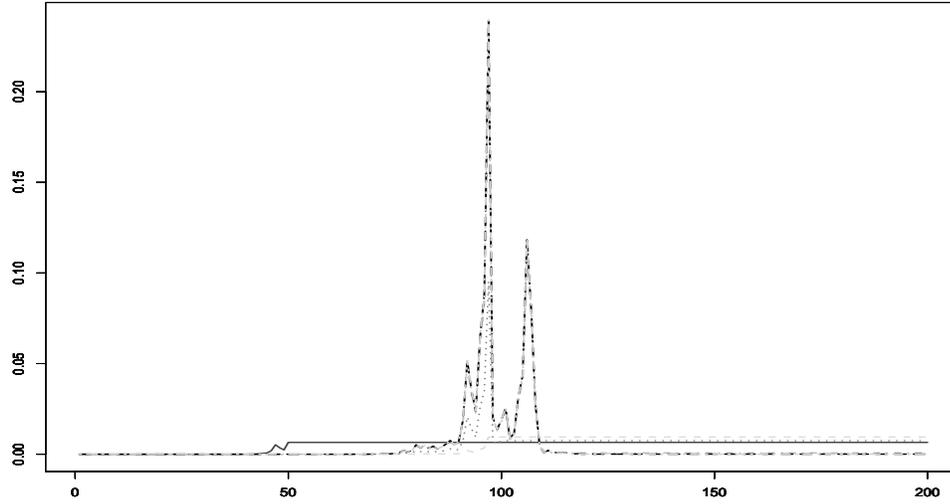
Figure 2: BDAEK. Posterior distribution at different times, solid line $t = 50$, dashed line $t = 97$, dotted line $t = 100$, dot-dashed line $t = 110$, long-dashed line $t = 120$ and two-dashed line $t = 150$.

is $c = 300$. For this, value the algorithm reports the presence of a change-point as the system starts to produce observations. It means that only the first observation belongs to the assumed distribution. For our data, it seems that $0 < c \leq 0.2$ is a reasonable choice.

Figure 5 shows the behavior of the stopping time rule for different values of $0 \leq c \leq 0.2$. As shown in the figure, for larger values of c, we detect the change-point earlier. For the data at hand, it seems that the best value for c is 0.034.

7.1.2 Second Distribution is Not Known (SDNK)

In this case, we assume that there is at least some information about the underlying procedure when it is out-of control. Furthermore, we assume that this information is of the form of the process mean λ_2^0. Then,

$$\mathbb{E}(Y) = \mathbb{E}[\mathbb{E}(Y|\lambda_2, \alpha_2, \beta_2)] = \mathbb{E}(\lambda_2|\alpha_2, \beta_2) = \frac{\alpha_2}{\beta_2} ,$$
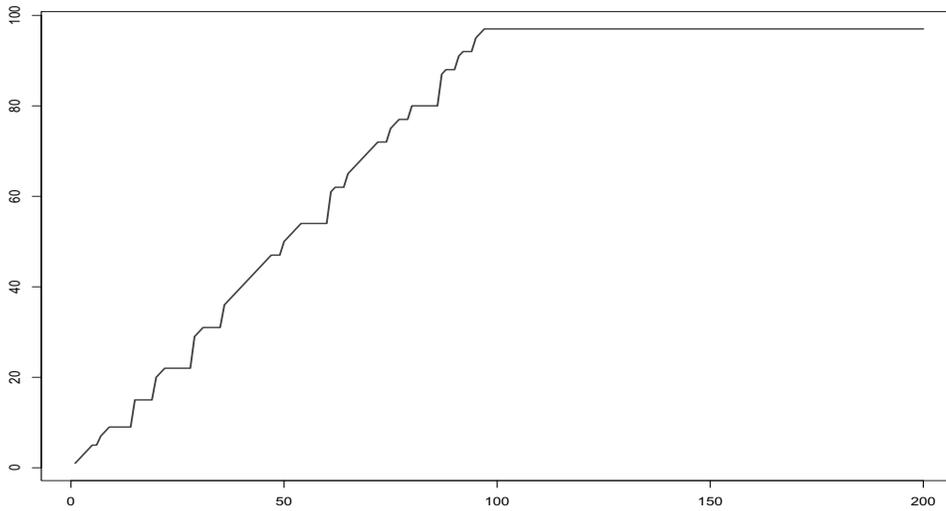
Figure 3: BDAEK. MAP estimation at $t = 1, \ldots, T$.

indicating that, for a conjugate prior, we have chosen $\alpha_2 = \lambda_2^0$ and $\beta_2 = 1$.

By letting $\lambda_2^0 = \bar{y}_{\tau+1:T}$, we run the same algorithm. The results are reported in Figures 7, 8, 9 and 10. By comparing these results with the BDAEK case, we observe that there is no considerable difference in the results.

The same results hold when the first distribution is not entirely known as well. We considered the conjugate prior on the parameter and assumed that the same type of information is available for the first distribution. Figure 11 shows the results in this case.

To measure the performance of the estimators, we generated $m = 100$ different samples from the same Poisson distribution and computed the estimations for these samples. In the sequel, we averaged the MSEs over them. Furthermore, for the Bayesian stopping time estimations, we derived the estimation for each sample. Figure 6 shows the results. For $t < \tau$, this figure shows that the MSEs increase and almost after that start to decrease. This was expectable as, before $t < \tau$, our MAP estimation was $\hat{\tau}_t = t + 1$ which might be far from the real $\tau = 100$. However, when we collect enough data ($t > \tau$), the estimations should be close to τ and so the MSEs start to decrease. This figure also shows that the Bayesian stopping rule estimations are scattered around the
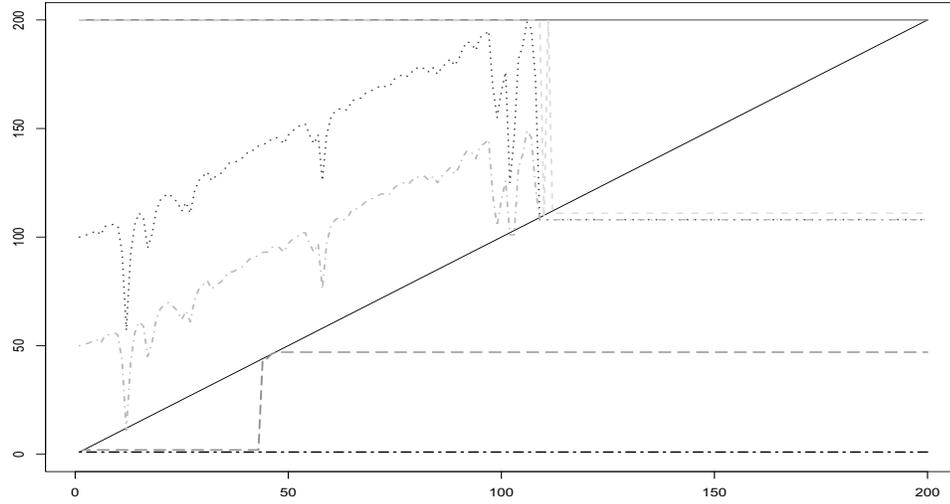
Figure 4: BDAEK. Bayesian estimation $\bar{\tau}_t$ for different values of c. Solid line $c = 0$, dashed line $c = 0.001$, dotted line $c = 0.01$, dot-dashed line $c = 0.02$, long-dashed line $c = 1$ and two-dashed line $c = 300$.

real change-point value.

Table 1: BDNKA: Averaged 0.95% HPD intervals for different values of t

t:	101	105	110	120	130	140	150	200
LB:	86.52	86.60	85.61	85.58	85.55	85.34	85.48	85.47
UB:	109.94	110.00	109.97	109.96	109.92	109.87	109.94	109.88

For our simulated data, the mean and standard deviation are 100.82 and 14.95, respectively. Table 1 reports the averaged 0.95% HPD intervals for different values of t. This table shows that, when $t > \tau$ (after having enough data), the HPD intervals tend to be almost constant which is in agreement with the posterior graphs from Figure 11. We gave these results only for this case because one can consider this case as the most difficult case to estimate.
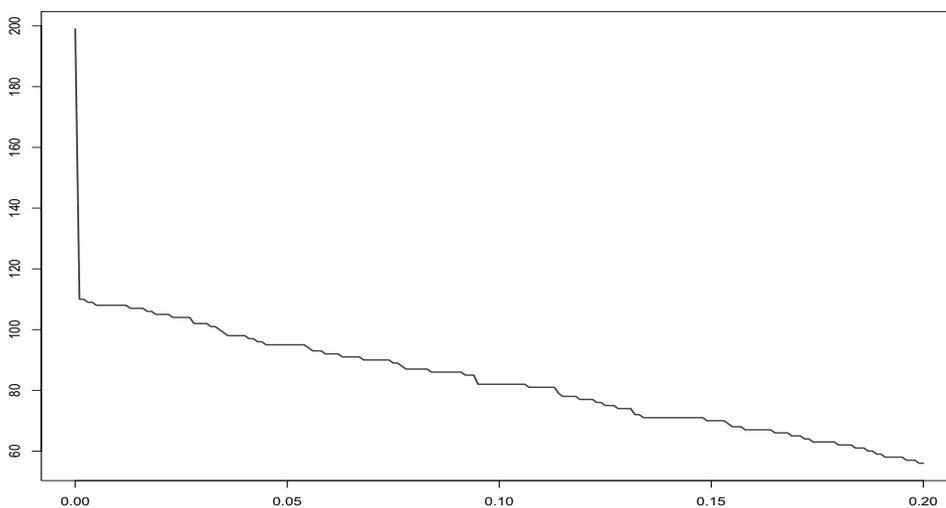
Figure 5: BDAEK. Stopping time for different values of *c*.

8   Conclusion

In this paper, we studied the single change-point problem. We gave a simple sequential algorithm to detect the change-point. We showed that, in the presence of some information about the data producing the system, our method is able to detect the change-point in a very efficient way.

Acknowledgements

I would like to thank the Editor and two anonymous referees for their helpful comments and considerations which improved the paper.

References

Berger, J. O. and Pericchi, L. R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction, *Journal of the American Statistical Association*, **91**(433),109–122.

Dehghan Monfared, M. E. and Meshkani, M. R. (2010). Bayesian Estimation of the Change Point Using EWMA Control Chart, *Journal of Statistical Theory and its Application*, **9**(2), 185–196.

Gholami, G. (2010). On the Bayesian Change-Point Problem in Regrassion Analysis, *Journal of Statistical Theory and its Application*, **9**(1), 9–27.

Girón, F. J., Moreno, E., and Casella, G. (2007). Objective Bayesian Analysis of Multiple Changepoints for Linear Models, In Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., and West, M., editors, *Bayesian Statistics*, **8**, 227–252. Oxford University Press.

Lai, T. L. (1995). Sequential Change Point Detection in Quality Control and Dynamical Systems, *Journal of the Royal Statistical Society*, Series B, **57**(4), 613–658.

Lorden, G. (1971). Procedure for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 1897–1908.

Mei, Y. (2003). *Asymptotically Optimal Methods for Sequential Change-Point Detection*. PhD thesis, California Institute of Technology.

Moreno, E., Bertolino, F., and Racugno, W. (1998). An Intrinsic Limiting Procedure for Model Selection and Hypotheses Testing, *Journal of the American Statistical Association*, **93**(444), 1451–1460.

Pollak, M. (1987). Average run lengths of an optimal method of detecting a change in distribution, *The Annals of Statistics*, **15**, 749–779.

Ruggieri, E., Antonellis, M. (2016). An exact approach to Bayesian sequential change point detection, *Computational Statistics Data Analysis*, **97**, 71–86.

Siegmund, D. and Venkatraman, E. (1995), Using the generalized likelihood ratio statistics for sequential detection of a change-point, *The Annals of Statistics*, **23**, 255–271.
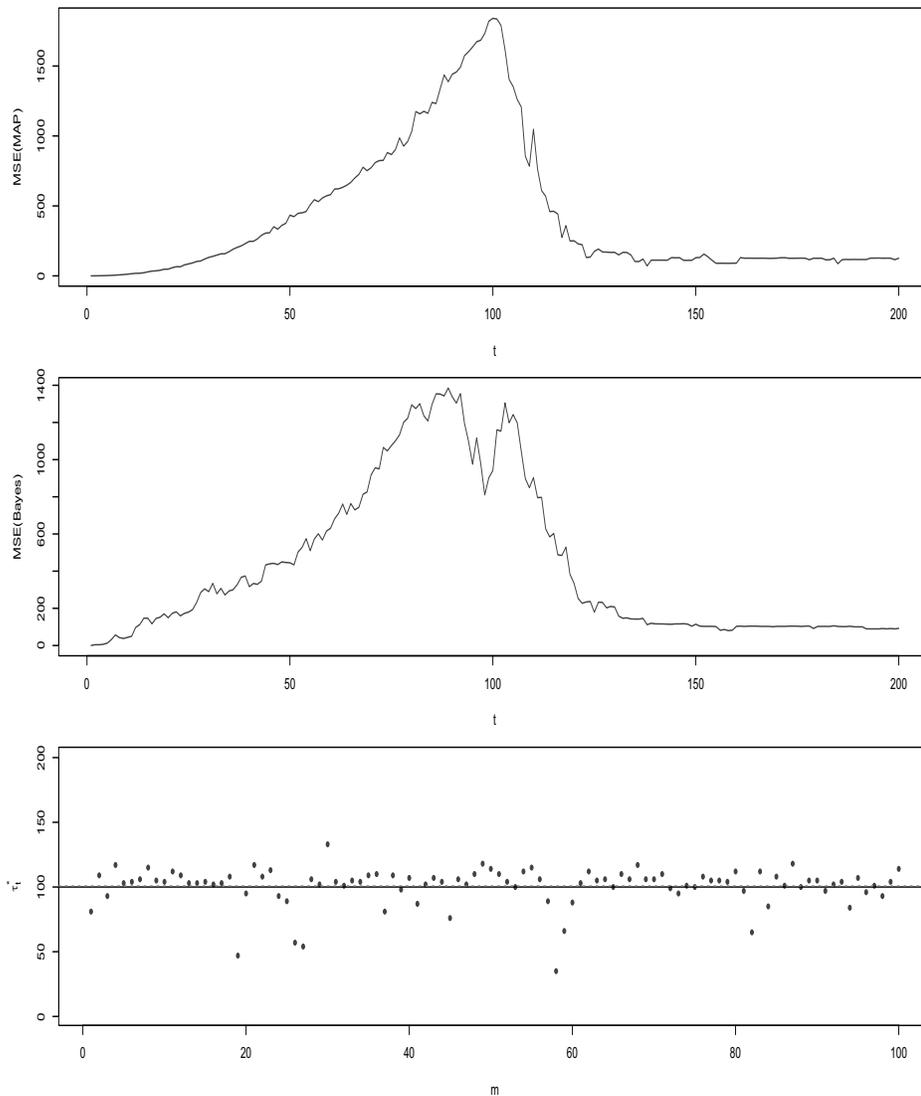
Figure 6: BDNK: The Averaged MSEs of MAP (*top*) and Bayes estimations (*mid*). The (*bottom*) shows the stopping time for different samples along with their average (*dashed line*) and the real change-point value (*solid line*).
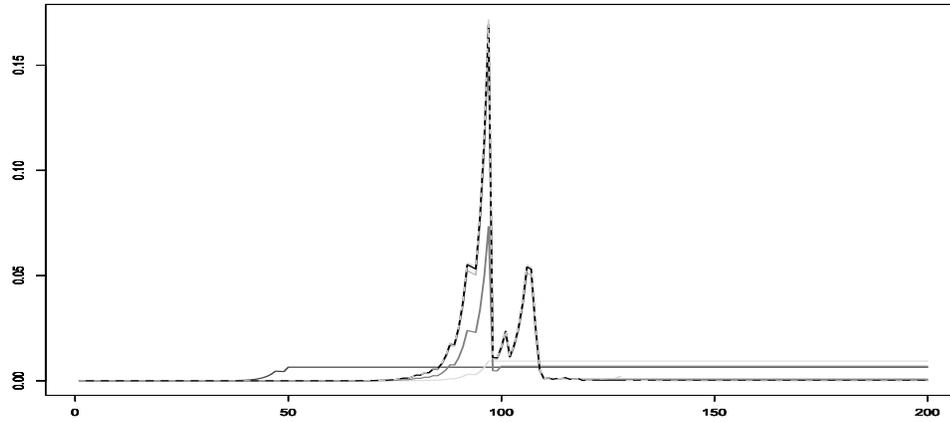
Figure 7: SDNK. Posterior distribution at different times, solid line $t = 50$, dashed line $t = 97$, dotted line $t = 100$, dot-dashed line $t = 110$, long-dashed line $t = 120$ and two-dashed line $t = 150$.
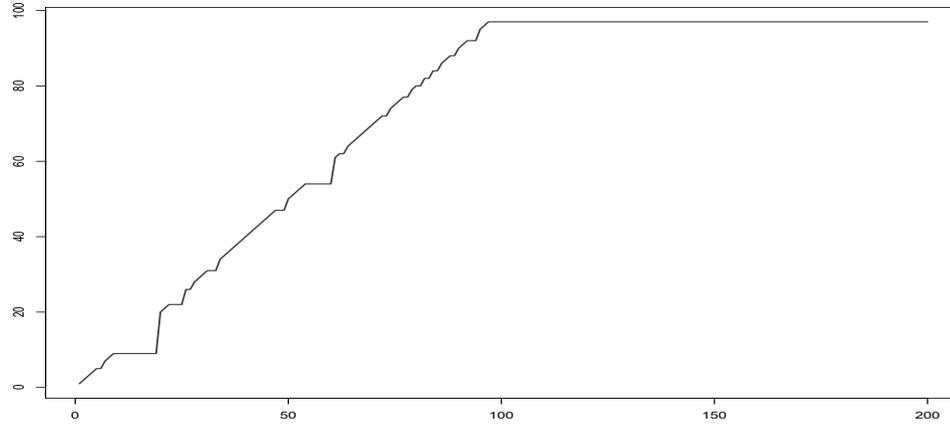
Figure 8: SDNK. MAP estimation at $t = 1, \ldots, T$.
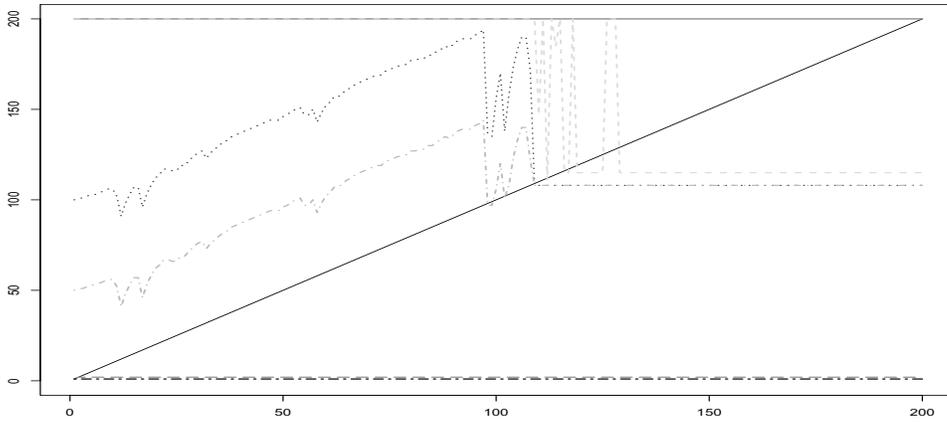
Figure 9: SDNK. Bayesian estimation $\bar{\tau}_t$ for different values of c. Solid line $c = 0$, dashed line $c = 0.001$, dotted line $c = 0.01$, dot-dashed line $c = 0.02$, long-dashed line $c = 1$ and two-dashed line $c = 300$.
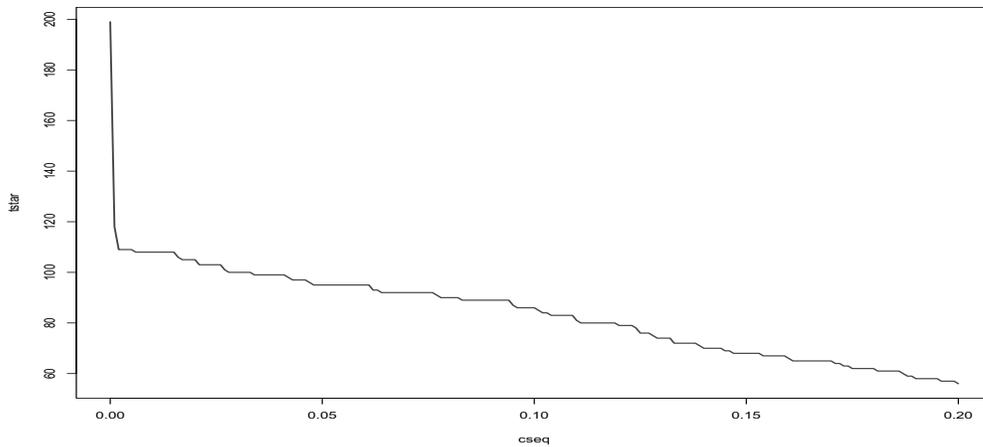
Figure 10: SDNK. Stopping time for different values of c.
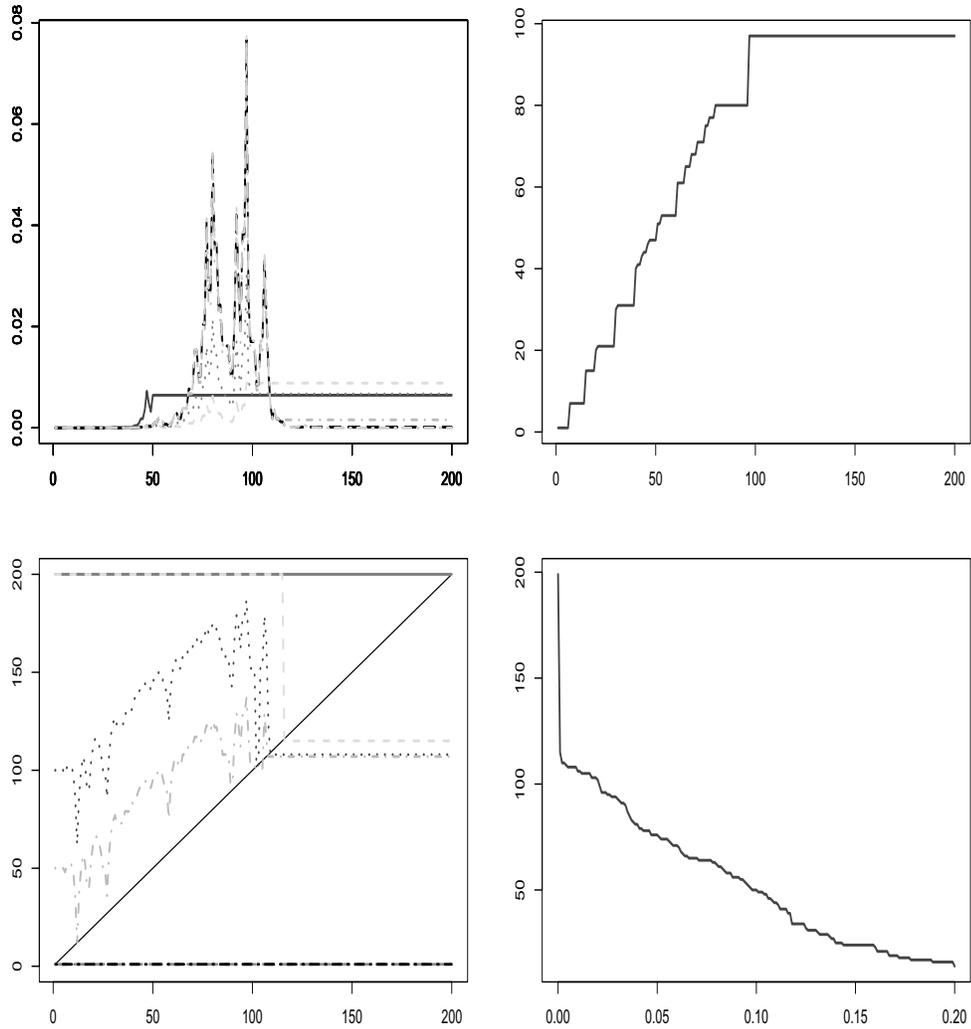
Figure 11: BDNK: Posterior distribution (*top left*), MAP estimation (*top right*), estimation of change-point for different values of *c* at different times (*bottom left*) and stopping time for different values of *c* (*bottom right*).