Proceedings of the World Congress on Engineering 2016 Vol I
WCE 2016, June 29 - July 1, 2016, London, U.K.

Problem of Change-Point Detection for the Poisson Observations

Dariush Ghorbanzadeh, Philippe Durand, Luan Jaupi

***Abstract*—Inference is considered about the point in a sequence of random variables at which the probability distribution changes. In particular, we consider the case a Poisson distribution with changing in parameter.**

Keywords: Change-point, likelihood estimator, Poisson distribution.

I. Introduction

The applications of change-point analysis are huge: we find relevant literature in many fields including: Biology, Chemistry, Environmental Sciences and Climate Change, Engineering, Econometrics, Medicine, Behavioral Sciences and also in Political Science, Finance, Image Analysis and Security.
When the observations reveal a phenomenon of change point, it is useful to determine firstly the relevant models and also to specify where the break occurs. The aim of this work is to present a theoretical point of view different types of point of change and to demonstrate the use of these models in an application. The application concerns Poisson observations. we apply the results to the data in [5] and [8], the data represent the number of haemolytic uraemic syndrome (HUS) cases in Birmingham and Newcastle from 1970-1989.

II. Statistical model

The observed data are a sequence of random variables $X_1, \ldots, X_n$ that may change law after the first k observations. The time k (called the time of change-point), strictly between 1 and n such that the observations $X_1, \ldots, X_k$ are independent and identically distributed according to a distribution P_{θ_1} and observations $X_{k+1}, \ldots, X_n$ are independent and identically distributed according to a distribution P_{θ_2}.
The model parameters are estimated from the likelihood function (the maximum likelihood method). The parameter estimates θ_1 and θ_2 will depend on the time k. In other words, each time k we have a probability, thus estimate the parameter k, we consider the function likelihood maximum.

III. Model for the Poisson observations

We consider the following model:

$$\begin{cases} X_1, \ldots, X_k \sim \mathcal{P}(\lambda) \\ X_{k+1}, \ldots, X_n \sim \mathcal{P}(c\lambda) \end{cases} \tag{1}$$

where $\lambda > 0$ and $c > 0$.
In the following, we will note

$$S_k = \sum_{i=1}^{k} X_i \,,\; S_k^\star = \sum_{i=k+1}^{n} X_i \,,\; \overline{X}_k = \frac{S_k}{k} \,,\; \overline{X_k^\star} = \frac{S_k^\star}{n-k} \;.$$

The likelihood function is given by

$$\begin{aligned} L(x, \lambda, c, k) &= \prod_{i=1}^{k} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \prod_{i=k+1}^{n} e^{-c\lambda} \frac{(c\lambda)^{x_i}}{x_i!} \\ &= \left(\prod_{i=1}^{n} \frac{1}{x_i!}\right) e^{-k\lambda-(n-k)c\lambda} \lambda^{s_k+s_k^\star} c^{s_k^\star} \end{aligned} \tag{2}$$

The log-Likelihood function is given by

$$\begin{aligned} &\ell(x, \lambda, c, k) = \log\big(L(x, \lambda, c, k)\big) \\ &= \log\big(\prod_{i=1}^{n} \tfrac{1}{x_i!}\big) - k\lambda - (n-k)c\lambda \\ &+ (s_k + s_k^\star)\log\lambda + s_k^\star \log c \end{aligned} \tag{3}$$

The maximum likelihood estimators $\hat{\lambda}$ and $\hat{c}$ is then that value of λ and c that maximises ℓ defined in the equation (3). They are determined by taking the partial derivative of ℓ with respect to λ and c and setting it to zero. We have

$$\begin{cases} \frac{\partial}{\partial \lambda}\ell(x, \lambda, c, k) = -k - (n-k)c + \frac{s_k+s_k^\star}{\lambda} \\ \frac{\partial}{\partial c}\ell(x, \lambda, c, k) = -(n-k)\lambda + \frac{s_k^\star}{c} \end{cases} \tag{4}$$

From the equation (4), we deduce the maximum likelihood estimators $\hat{\lambda}$ and $\hat{c}$,

$$\hat{\lambda} = \overline{X}_k \;,\; \hat{c} = \frac{\overline{X_k^\star}}{\overline{X}_k} \tag{5}$$

By injecting $\hat{\lambda}$ and $\hat{c}$ in equation (3), we obtain

$$\begin{aligned} &R(k) = \ell(x, \hat{\lambda}, \hat{c}, k) \\ &= \log\big(\prod_{i=1}^{n} \tfrac{1}{x_i!}\big) - s_n + s_k \log(\overline{x}_k) + s_k^\star \log(\overline{X_k^\star}) \end{aligned} \tag{6}$$

where $s_n = s_k + s_k^\star$.

The maximum likelihood estimators $\hat{k}$ is defined by,

$$\hat{k} = \max_{1<k<n-1} R(k). \tag{7}$$

IV. Simulation results

In order to study the performance of the method, we simulated a sample of size 500 and the values set $\lambda = 5$, $c = 2$ and $k = 150$. The following figure show the results obtained.

Dariush Ghorbanzadeh, Philippe Durand, Luan Jaupi are with the Department of Mathematics (Ingenirie mathmatique: IMATH), Conservatoire National des Arts et Métiers, 292 rue Saint martin, 75141 Paris FRANCE e-mail: philippe.durand@lecnam.net, leon.jaupi@lecnam.net, dariush.ghorbanzadeh@lecnam.net

ISBN: 978-988-19253-0-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

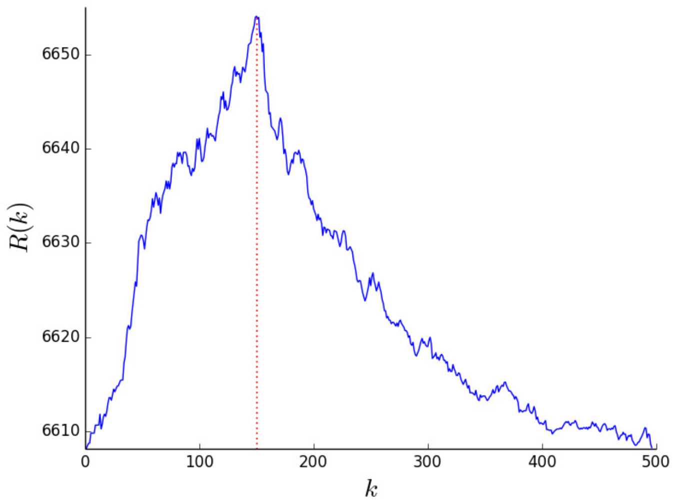

Fig. 1. Graph of $R(k)$, with the values estimator $\hat{\lambda} = 5.198675$, $\hat{c} = 1.458383$ and $\hat{k} = 151$.

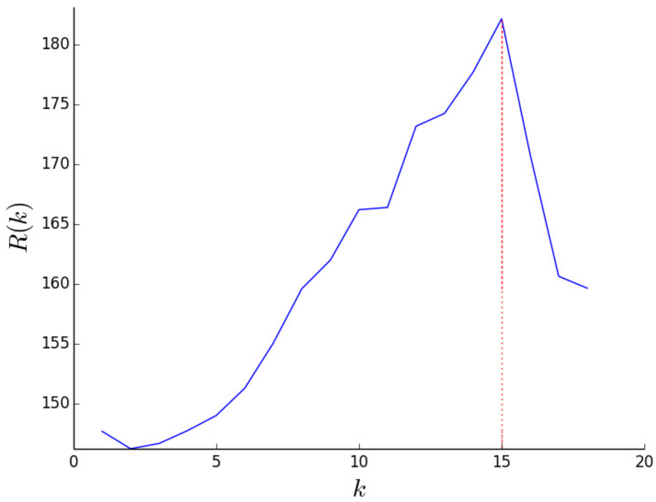

Fig. 2. Graph of $R(k)$for the Birmingham data, with $\hat{\lambda} = 2.066667$ and $\hat{c} = 6.096774$.

TABLE I
COUNTS OF CASES OF HUS AT BIRMINGHAM AND NEWCASTLE.

Year	Count at Birming-ham	Count at New-cas-tle	Year	Count at Birm-ingham	Count at New-castle
1970	1	6	1980	1	4
1971	5	1	1981	7	0
1972	3	0	1982	11	4
1973	2	0	1983	4	3
1974	2	2	1984	7	3
1975	1	0	1985	10	13
1976	0	1	1986	16	14
1977	0	8	1987	16	8
1978	2	4	1988	9	9
1979	1	1	1989	15	19

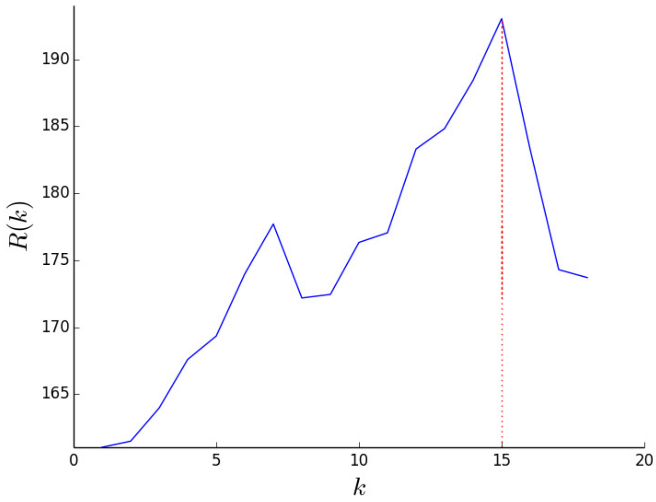

Fig. 3. Graph of $R(k)$for the Newcastle data, with $\hat{\lambda} = 2.466667$ and $\hat{c} = 5.108108$.

V. APPLICATIONS

In this section we apply the previous results to the data used by [5] and [8]. The data represent the number of cases of diarrhoea-associated haemolytic uraemic syndrome (HUS) for the years 1970-1989, in two cities Birmingham and Newcastle within the United Kingdom.

The following table displays the annual number of cases of HUS collected in Birmingham and Newcastle of England, respectively, from 1970 to 1989.

The following figures show $R(k)$, for the Birmingham data and Newcastle data ,based on k, the maximum is reached at the point $\hat{k} = 15$ (year 1984).

VI. CONCLUSION

The method used in this work is very simple to implement and particularly easy to program. In the application part, we obtained the same results as [5] and [8] who used another change point detection method.

REFERENCES

[1] D. Ghorbanzadeh & N.Point (2008). Modles probabilistes de dtection de rupture dans le comportement. Sciences pour le gnie civil.LCPC . 119-123.
[2] D. Ghorbanzadeh & R. Lounes (1998). A Bayes procedure for detecting change in exponential random observations. Studen, 2 2.
[3] D. Ghorbanzadeh (1996). Detection of random change point in one-parameter exponential families. Appl. Math. and Compu.,Vol.77, 2 & 3, 167-177.
[4] E.Gombay (2003). Sequential Change-Point Detection and Estimation. Sequential Analysis. Volume 22, Issue 3, 203-222.
[5] R. Henderson & J. N. S. Matthews (1993). An Investigation of Changepoints in the Annual number of Cases of Haemolytic Uraemic Syndrome. Appl. Statist, 42, 3. 461-471.
[6] Y.Lan & M.Banerjee & G. Michailidis (2009). Change-point estimation under adaptive sampling. Ann. Statist. Vol. 37, N 4, 1752-1791.

[7] J. D. Storey & J. E. Taylor & D.Siegmund (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. Journal of the Royal Statistical Society: Series B, Vol. 66, Issue 1, 187-205.
[8] R. W. West & R. T. Ogden (1997). Continuous-time estimation of a change-point in a Poisson process. Journ.Statist. Comput. Simul, 56, 293-302.

ISBN: 978-988-19253-0-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)