Testing for unimodality in mortality trends

Rueda[1], C., Ugarte[2,3] M.D. and Militino[2], A.F.
[1]*IMUVA. Instituto de Matemáticas. Universidad de Valladolid, SPAIN*
[2]*Departamento de Estadística e Investigación Operativa, Universidad Pública de Navarra, SPAIN*
Corresponding author e-mail[3]: lola@unavarra.es

Abstract

In some diseases it is well-known that a unimodal mortality pattern exists. A clear example in developed countries is breast cancer, where mortality increased sharply until the nineties and then decreased. This clear unimodal pattern is not necessarily generalizable to all regions within a country. In this work, we develop statistical tests to check if this unimodality persists within regions using order restricted inference. The same methodology will be also used to provide change-points within regions as well as confidence intervals. Results will be illustrated using age-specific breast cancer mortality data from Spain in the period 1975-2005.

Keywords: breast cancer trend patterns, order restricted inference, change-points

1 Introduction

In an experiment with n response levels (time points) it is common for the response variable (mortality rate) to increase up to a certain unknown point and then decrease. This up-then-down response pattern is usually called an umbrella or unimodal ordering. In many applications of regression in different fields, the response and the explanatory variable exhibit a unimodal pattern. In particular, unimodal regression often arises in the analysis of epidemiological data. For example, when studying breast cancer mortality data in Spain and other developed countries (Ugarte et al., 2010). The main point of interest in many of these applications, and in particular in our case, is that of locating the peak, that in practice represents a change of the mortality time trend of a specific disease, and also testing the null hypothesis of a simpler increasing pattern. This paper deals with the analysis of breast cancer mortality rates but the methodology derived is also useful to analyze data from a variety of applications in many fields whenever the functional relationship between the response and the explanatory variable is unimodal.

There exist several parametric approaches to estimate a unimodal relationship, and the corresponding peak, such as fitting a quadratic regression or fitting a piece-wise linear regression with two segments. This latter method is known as segmented regression (see for example, Muggeo, 2003) and it is often use in epidemiological contexts. On the other side, a nonparametric approach could also be used to estimate the peak ranging from a simple choice of selecting the maximum of a smoothed version of the data, to other more sophisticated approaches as the one using splines recently proposed in the literature by Köllmann et al.(2012). Both the parametric and the nonparametric solutions have important drawbacks. The parametric approaches require a careful choice of the form of the function giving biased estimates of the location of the peak under an erroneous specification of the function. Meanwhile, the nonparametric approach is not simple to implement and requires user-specified choices, such as bandwidth, or smoothing parameters and the number and location of knots. Besides, in the application at hand, the variability is high as the mortality rates refer to small areas. This feature must be taken into account to select the most appropriate methodology in order to derive useful and reliable conclusions. In this paper we have adopted a semiparametric regression approach where the mortality rates at different time points are estimated assuming independent Poisson distributions with the log-means exhibiting a unimodal pattern. A likelihood approach of inference is used to obtain the MLE of the unimodal restricted means and thus, the maximum. The optimization problem using the likelihood is solved using isotonic regression (Robertson et al., 1988). An algorithm for carrying out isotonic regression has been implemented in the R package Iso (Turner, 2011). The main advantages of the proposed approach is that is simple and likelihood-based, it provides an estimate of the location of the peak without assuming a specific form of the regression function, it will give us the possibility to derive a test for the unimodal assumption and we will be able to compute AIC measures. As far as we know, although the problem of testing monotonicity has received some attention in the recent past, the problem of considering unimodality as alternative hypothesis is new in the literature. The gain in power when the alternative is reduced can be decisive in cases like the one above mentioned about change point detection. In this paper a conditional test for testing monotonicity against unimodality is derived in a regression context. The proposed test is illutrated under Poisson assumptions for the application at hand. However, it is also possible to be used in general exponential family models including the normal case. Several authors have deal with related problems using isotonic regression for monotone regression models or for other shape constraints.

2 The model and the likelihood based statistics

The model is defined as follows: let $y = (y_1, ..., y_n)'$ be the response vector. In particular, when dealing with Spanish regions, $n = 17$, and y_i is the number of deaths in region i assumed to be Poisson distributed with mean $\mu_i = n_i r_i$, where r_i is the unknown rate of mortality and n_i is the population at risk. Namely,

$$y_i \sim P(\mu_i), \quad \mu_i = n_i r_i.$$

We also assume a unimodal pattern of mortality rates which is formulated mathematically as:

$$r = (r_1, ..., r_n)' \in U,$$

where $U = \{x \in \Re^n / \exists m, 1 \leq m \leq n, x_1 \leq \ldots \leq x_m \geq x_{m+1} \geq \ldots \geq x_n\}$ is a convex region in $\Re^n$ defining the unimodal restriction.

The approach is called semiparametric because, on one hand, the number of parameters of the model depends on the sample size and tends to infinity, and, on the other hand, given n, the way we do inferences is parametric.

The MLEs, $\widehat{r}_i$, are the solution to the optimization problem

$$\min_{r \in U} \left(\sum_{i=1}^{n} y_i \log(r_i) + y_i \log(n_i) - \log(y_i!) - r_i n_i \right). \tag{1}$$

The solution to this problem is the isotonic regression of $y_i/n_i, i = 1, \ldots, n$ with weights $w_i = e_i$ (see Robertson et al., 1988), pp 38). $\widehat{r}$ is also defined as $P_W(v/U)$, the projection using the metric of the matrix $W = diag((w_1, ..., w_n)$ of the vector $v = y/n = (y_i/n_i, ..., y_n/n_n)'$ in the convex region U.

On the other hand, the monotone is formulated mathematically, using the convex cones M in $\Re^n$, as follows

$$M = \{x \in \Re^n / x_1 \leq \ldots \leq x_n\} \quad \text{and} \quad M \subset U.$$

The testing problem we are dealing with is

$$H_m : \mu \in M \qquad H_u : \mu \in U.$$

The LR (Likelihood Ratio) statistic for these hypothesis is given by

$$T = -2l(\widehat{r}_m) - 2l(\widehat{r}_u), \tag{2}$$

where $l()$ is the loglikelihood.

3 Results

In this section the unimodality test is illustrated using breast cancer mortality data from Spain in the period (1975-2005). It is known that in Spain as well as in other developed countries, breast cancer mortality increased sharply until the nineties and then decreased. Here, we are interested in testing if this unimodality pattern persists in all the Autonomous Regions of the country. This problem is of interest as Autonomous Regions have their own health system and habits, and then, some disparities may be found. Isotonic regression will be also used to find break points and confidence intervals for the regions where unimodality persists. However, as the common procedure to estimate break points in epidemiology is using the well-known joinpoint regression model (see for example, Statistical Research and Applications Branch, National Cancer Institute, 2009), in this section, a comparison between isotonic regression and joinpoint regression will be made first, to show the reasons why isotonic regression will be used in this context. The comparison will be made analyzing real data and in a simulation study under the scenario of having a unimodality pattern (a single break point) imitating the real mortality data from Spain. A similar procedure is used by Turner and Wollan (1997). These authors demonstrate that the isotonic estimator of the mode is consistent, simple to understand, and performs satisfactorily in practice. In this work we show that the isotonic estimator is working well in the epidemiological context considered here. However, the joinpoint regression method could be seriously biased and it is not robust because when we modify the studied period a bit, the estimator provides a different mode, in contrast to the isotonic estimator. With respect to results concerning the monotonicity test, all the Autonomous Regions is Spain except one, show a unimodal pattern.

Acknowledgements

The authors gratefully acknowledge the financial support of the Spanish of Science and Education, project MTM2012-37129 (Rueda) and project MTM2011-22664 jointly sponsored with Feder grants (Ugarte and Militino).

References

[1] Statistical Research and Applications Branch, National Cancer Institute (2009). Joinpoint Regression Program, Version 3.4.

[2] Köllmann, C., Bornkamp, B. and Ickstadt, K. (2012). Unimodal regression using Bernstein-Schoenberg-splines and penalties. Technical Report. Universiy of Dortmund.

[3] Muggeo, V.M.R.(2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, 22, pp. 3055-3071.

[4] Robertson, T., Wright, F.T., Dykstra, R.L. (1988). *Order Restricted Statistical Inference*. Wiley.

[5] Turner, R. (2013). *Iso: Functions to Perform Isotonic Regression*. R package version 0.0-12. `http://CRAN.R-project.org/package=Iso`.

[6] Turner, T.R. and Wollan, P.C. (1997). Locating a maximum usin isotonic regression. *Computational Statistics and Data Analysis*, 25, pp. 305-320.

[7] Ugarte, M.D. and Goicoa, T. and Etxeberria, J. and Militino, A.F. and Pollán, M.(2010). Age-specific spatio-temporal patterns of female breast cancer mortality in Spain (1975–2005). *Annals of Epidemiology*, 20, pp. 906-916.