A statistical analysis of trends for warm and cold spells by means of counts

Jesper Rydén

Department of Mathematics, Uppsala University
jesper@math.uu.se

Résumé. Nous nous intéressons à l'analyse des valeur extrêmes dans la climatologie ; plus précisément, les périodes de températures exceptionnellement chauds ou froid, c'est-à-dire, des *hot spells* (en cas des temperatures élevées). Ici, on regarde le nombre anuel des spells en Uppsala, Suéde. Ces nombres sont vraiment des nombres entiers. Donc, la théorie et la méthodologie statistique de la régression Poisson est approprié. On examine un trend possible pour la période 1840–2012. Le trend pour des spells chauds est trouvé positif, plus large que celui des chauds froids, et statistiquement significatif ($\alpha = 0.05$). La méthodologie peut être utilisée pour l'analyse des autres indicateurs climatiques.

Mots-clés. Régression Poisson, trend, température

Abstract. Statistical studies of extremes are of interest in the climatic sciences; for instance, trends of periods of unusually warm or cold weather, which could be labelled warm and cold spells, respectively. We study the yearly number of spells in Uppsala, Sweden, which from a data-analytic point of view truly are counts, and employ theory and methods from the field of regression models for counts. A possible trend for the period 1840–2012 was investigated. The trend for warm spells is positive and demonstrated to be larger in magnitude compared to the one for cold spells, and is found statistically significant ($\alpha = 0.05$). The methodology could be extended to analyse other climate indicators.

Keywords. Poisson regression, trend, temperature

1 Introduction

One of many problems in studies related to climate change is the closer study of so-called *heat waves* and *warm spells*, and related phenomena. These could be somewhat loosely described as periods of unusually hot weather. Corresponding notions for cold weather is also of interest, where applicable.

In this paper, we study measurements of extreme temperatures from Uppsala, Sweden. The quantity studied will be a climate indicator: the yearly number of warm spells and cold spells, respectively. A possible trend over time will be investigated, starting from 1840 until present. From a data-analytic point of view, a warm spell is a *run*, i.e. a period of consecutive days where the maximum is above a specified high value.

Often in the the climate literature, trends are investigated by non-parametric slope estimators, see e.g. Perkins and Alexander (2013). In this paper, we regard the yearly observations according to their true nature — counts — and hence face a time series of counts, sampled at a yearly basis. Generalised linear models and their extensions are then natural candidates for modelling.

The aim of this presentation is in principle two-fold: firstly, to analyse trends for the Uppsala series; secondly, to investigate the plausible use of count models for trend analysis of possibly other climatic indicators.

2 Data and definitions

2.1 Background and description

In this article, temperature measurements from Uppsala, Sweden (59°5'N, 17°4'E), are studied. A thorough description of this data set (historical accounts and data analysis) is found in Bergström and Moberg (2002). As spells are the subject ot the investigation, extreme observations are of interest.

In this paper we focus on periods of unusually warm or cold temperatures, what could be labelled warm and cold spells, respectively. Moreover, in the literature, the notion of heat wave is often reserved for periods of great severity, e.g. causing deaths among people. For typical Swedish conditions, we therefore prefer the notion of warm spells.

We have chosen here to use the definition by the Swedish Meteorological and Hydrological Institute (SMHI) when analysing data over the period 1840-2012. The definition is that the maximum temperature should exceed 25°C for at least five consecutive days, thus an absolute threshold. For the periods of cold spells, no clear definition seems to exist. We chose to use a percentile-based threshold; the 5% quantile of daily minimum temperatures for the standard reference period 1961-1990, which turns out to be -13.4°C. Periods longer than five days, fulfilling the criteria, were identified and extracted from data. For each year, the number of warm and cold spells, respectively, was calculated, and two time series formed. These are presented in Figure 1.

2.2 Analysis of distributions

The random quantity under study is the early number of warm and- cold spells, respectively. Histograms are shown in Figure 2. Reasonable distributions for counts can be found as members in the so-called Katz family of distributions, where the Poisson and negative binomial distribution are important examples (Winkelmann, 2008).

For the two data sets, χ^2 tests were carried out, testing for Poisson distribution or negative binomial distribution. These resulted in p-values 0.35 and 0.41 for warm spells; $1.2 \cdot 10^{-6}$ and 0.20 for cold spells. Hence, the conclusion can be drawn to reject the hypothesis about Poisson distribution for cold spells.

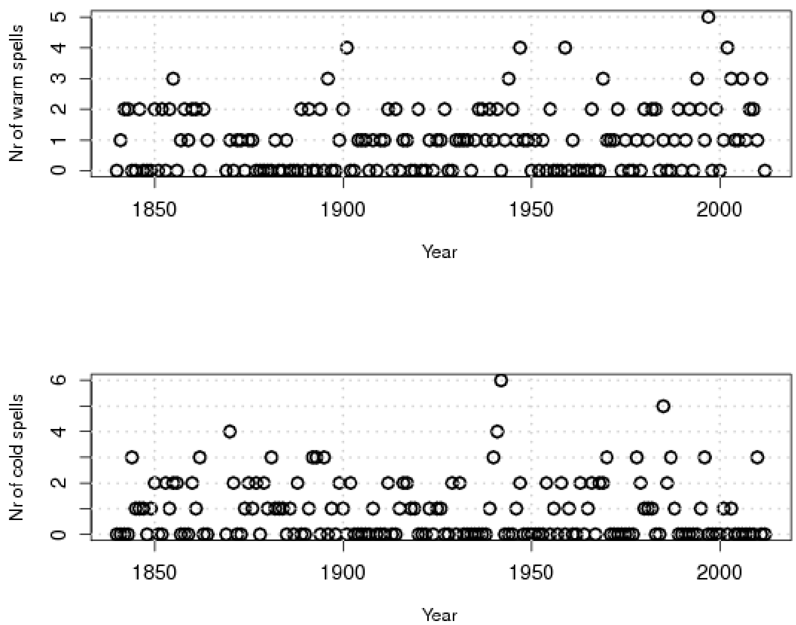

Figure 1: Time series. Top: Yearly number of warm spells. Bottom: Yearly number of cold spells.

3 Statistical modelling

The intention is to model the climate index as described above as a dependent variable in a regression model where time is the covariate. A crucial assumption for use of the various regression models of counts is lack of independence in the explanatory variable. Since data in our case are in fact time series, dependence might in fact be an issue. Regression models for counts will be fitted to the two data sets (warm and cold spells, respectively). it turns out, from visualisations and the Ljung–Box test that independent data in the time series is a reasonable assumption, see Rydén (2014).

A typical path to follow when modelling counts is to start with simpler types of models, e.g. Poisson regression. Common problems with that simplest approach are *overdispersion* (mean and variance not equal, as expected for a Poisson distribution) or excess of zeros (as might be noted from Figure 2, right panel). All models in the sequel are estimated by Maximum Likelihood, as implemented in the software R.

Given the value of the covariate x, time, the conditional mean of the response variable y is postulated by the models as

$$\mathsf{E}[y|x] = \exp(\beta_0 + \beta_1 x).$$

The coefficient β_1 is here interpreted that as the value of x increases one unit, the response

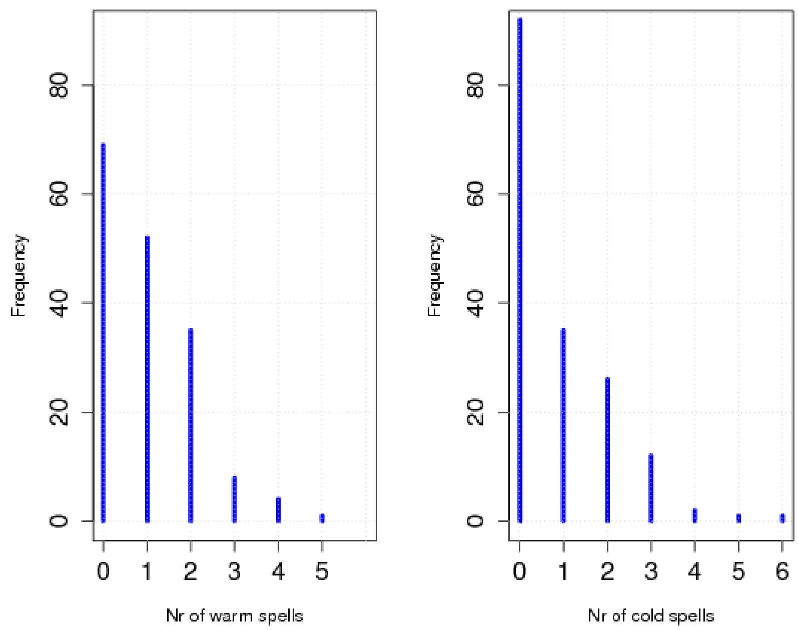

Figure 2: Histograms. Left: Yearly number of warm spells. Right: Yearly number of cold spells.

(i.e. number of yearly spells) increases with e^{β_1}. Moreover, the sign of β_1 is related to the nature of the trend: $\beta_1 < 0$ would mean a negative trend, i.e. a decreasing number of spells while $\beta_1 > 0$ means a positive trend.

Yearly number of warm spells

As a first step, a Poisson regression model was fitted. However, the resulting model suffered slightly from overdispersion (mean 0.99, variance 1.12). A test for overdispersion by Cameron and Trivedi (1990) was applied to investigate further (null hypothesis, equidispersion) and resulted in the p-value $p = 0.21$; hence, the null hypothesis about equidispersion is not rejected. The Poisson regression yields a significant result for the time covariate ($p = 0.019$) and a slightly positive estimated trend over time ($\hat{\beta}_1 = 0.0037$). The natural alternative to try out next, is the negative binomial distribution. The time covariate is then again significant ($p = 0.026$), with estimated coefficient $\hat{\beta}_1 = 0.0037$.

Yearly number of cold spells

Recall Figure 2, right panel. Although the χ^2 test in the previous section rejected the Poisson distribution, we first assume Poisson distribution for the response. However, the model suffers from overdispersion. The mean is 0.84 and the variance 1.31, and the hypothesis about equidispersion is here rejected (as expected): the test for overdispersion gives the p-value 0.0026. Guided by the histogram in Figure 2, a zero-inflated Poisson model is fitted, in which the inflation part is found significant. In the resulting model, the time covariate is not significant ($p = 0.28$), and the estimate of slope is $\hat{\beta}_1 = 0.0024$.

4 Discussion

In this paper, trends of the yearly number of warm and cold spells in Uppsala were examined for the time period 1840-2012. These observations were regarded as counts and statistical methodology developed for that particular data type applied. We found slightly positive changes, i.e. increasing yearly number, for both types of phenomena, where trends for warm spells, larger in magnitude compared to cold spells, were statistically significant. Note that the extremes of the data set have not been corrected for the urban heat island effect. Thus, the here analysed record includes some warming due to that effect.

Trends in extremes *per se* for long time series for the same geographical region have been analysed by e.g. Guttorp and Xu (2011), investigating a series from Stockholm, Sweden, and comparing with model simulations; moreover, by Rydén (2011), for the Uppsala series analysed in the present paper. An increasing trend in minimum temperatures was present.

For further details of this study, see Rydén (2014).

Bibliography

[1] Bergström, H. and Moberg, A., (2002), Daily air temperature and pressure series for Uppsala (1722-1998), *Climatic Change*, 53, 213–252.
[2] Cameron, A.C. and Trivedi, P.K., (1990), Regression-based tests for overdispersion in the Poisson model, *Journal of Econometrics*, 46, 347364.
[3] Guttorp, P. and Xu, J., (2011), Climate change, trends in extremes, and model assessment for a long temperature time series from Sweden, *Environmetrics*, 22, 456–463.
[4] Perkins, S.E. and Alexander L.V., (2013), On the measurement of heat waves, *Journal of Climate*, 26, 4500–4517.
[5] Rydén, J. (2011), Statistical analysis of temperature extremes in long-time series from Uppsala, *Theoretical and Applied Climatology*, 105, 193–197.
[6] Rydén, J. (2014), A statistical analysis of trends for warm and cold spells in Uppsala by means of counts, *Geografiska Annaler: Series A, Physical Geography.*
doi: 10.1111/geoa.12083
[7] Winkelmann, R., (2008), *Econometric Analysis of Count Data*, 5th edition. Springer-Verlag.