Nonlinear Processes
in Geophysics

Instability and change detection in exponential families and generalized linear models, with a study of Atlantic tropical storms

Y. Lu[1] and S. Chatterjee[2]

[1]Amazon.com, Inc., Seattle, Washington, USA
[2]University of Minnesota, School of Statistics, Minneapolis, Minnesota, USA

Correspondence to: S. Chatterjee (chatterjee@stat.umn.edu)

Abstract. Exponential family statistical distributions, including the well-known normal, binomial, Poisson, and exponential distributions, are overwhelmingly used in data analysis. In the presence of covariates, an exponential family distributional assumption for the response random variables results in a generalized linear model. However, it is rarely ensured that the parameters of the assumed distributions are stable through the entire duration of the data collection process. A failure of stability leads to nonsmoothness and nonlinearity in the physical processes that result in the data. In this paper, we propose testing for stability of parameters of exponential family distributions and generalized linear models. A rejection of the hypothesis of stable parameters leads to change detection. We derive the related likelihood ratio test statistic. We compare the performance of this test statistic to the popular normal distributional assumption dependent cumulative sum (Gaussian CUSUM) statistic in change detection problems. We study Atlantic tropical storms using the techniques developed here, so to understand whether the nature of these tropical storms has remained stable over the last few decades.

1 Introduction

One important way in which nonlinear structures may be present in data related to many physical and natural phenomena is by structural breaks and changes. Generally, elicitation of the time and nature of such breaks with statistical guarantees involves change detection techniques like the cumulative sum (CUSUM) or the exponentially weighted moving average (EWMA).

The standard framework for applying such change detection techniques requires assuming that the order in which the sampled observations arrive is known, with the question of interest being whether the data generating process has remained stable over time. The observations are assumed to follow a known Gaussian distribution, and are monitored for a potential change to a different, but still known, Gaussian distribution. Statistical guarantees are typically expressed in terms of expected run length, i.e., how long it takes on average for a true change to be detected, when there is a control for the expected length of time before false signaling occurs. These normality-based sequential monitoring and stability detection techniques originated from industrial process control (Page, 1954), although they have far ranging applications at the present time. Examples of such applications are in health care monitoring (Steiner et al., 1999), detection of genetic mutation (Krawczak et al., 1999), credit card and financial fraud detection (Bolton and Hand, 2002), insider trading in stock markets (Meulbroek, 1992), and detection of jamming attacks in wireless networks (Chen et al., 2007).

Note that in many modern applications, the assumption of normality is not tenable. In this paper, we discuss change detection in a general exponential family, and in regression models including generalized linear models like logistic regression and log-linear regression. We present several mathematical results concerning the different kinds of CUSUM statistics that may result, depending on the probabilistic structure under consideration, and whether certain parameters are estimated or assumed to be known. A natural question here is on the performance of the normality-based CUSUM statistic, when the probability models do not

satisfy the Gaussian assumptions. We study this issue, and present mathematical results, simulation studies, and discussions about when and how the Gaussian CUSUM may yield high quality results. Finally, we discuss properties of Atlantic tropical storms, and use the techniques developed in the rest of this paper to study structural changes in the fundamental physical properties for which we have data records for such storms.

In order to generalize the scope of statistical change detection tools, in this paper we propose a variant of the sequential industrial monitoring framework, by considering the stability of the data generation process as a problem of detecting the time of the distributional change; in other words, we conduct a hypothesis test, and under the null hypothesis, the data generation process remains stable through the entire sampling time $t = 1, \ldots, n$. Under the alternative hypothesis, the distribution of the individual observations remains stable up to an unknown point of time $\tau \leq n$ and then it changes to another distribution. With this hypothesis testing framework, we are in a position to (a) consider models with none, one, or more change points in the same statistical framework, (b) quantify uncertainty associated with any potential result using standard concepts of hypothesis tests like size, power, level of significance, or properties of the run length, and (c) extend the scope of the study beyond the traditional frameworks where the data either arrives sequentially, or there are sufficient observations before and after each change point. We may consider problems where some parameters are known for some duration of the process, while others are estimated. The sequential process monitoring statistics like CUSUM are obtained as a special case, so there is no loss of generality in using the hypothesis testing approach proposed here. Two of these generalizations, that of extension to any partitioning of the data and that of using multiple change times, can be easily visualized in this hypothesis testing framework, but we do not pursue them here for brevity. However, we briefly comment on these generalizations in Sect. 3 below. Also, our framework allows for cases where parameter values are unknown and estimated from data, but we present first our results for the known-parameter case for clarity, and restrict the discussion of the estimated parameter case in Sect. 3.3 below. We call the proposed testing procedure the exponential family CUSUM (or EF CUSUM in short), while the statistic obtained under the Gaussian framework is called normal CUSUM or Gaussian CUSUM.

Simulation studies show that in most situations the EF CUSUM method performs better than Gaussian CUSUM. The EF CUSUM has a shorter average run length, smaller variation of run length and shorter maximum run length compared with Gaussian CUSUM. Moreover, smaller shifts can be detected more quickly by EF CUSUM than by Gaussian CUSUM, which is a big advantage of using EF CUSUM. Under some circumstances the Gaussian CUSUM approximates the EF CUSUM well, we discuss this issue below. It is also important to note that whether the change point τ is at the be-

ginning, in the middle, or at the end, the EF CUSUM generally outperforms the Gaussian CUSUM, so the unknown parameter τ plays little role in our analysis. Finally, in the case of a large parameter shift, the EF CUSUM and the Gaussian CUSUM perform similarly. This is not unusual, and even visual and ad hoc techniques suffice for many cases of large changes.

We also extend our study to that of parameter change in the generalized linear model. In this context, Brown et al. (1975) and Jandhyala and MacNeill (1991) discussed general linear model, Lee et al. (2004), Chihwa and Ross (1995), and Ploberger et al. (1989) focused on detecting the linear model with different types of error terms. In this paper we propose methodology for detecting change in regression coefficients in the generalized linear model setting and the EF CUSUM scheme associated with it.

Our case study for illustrating our instability and change detection techniques is based on Atlantic tropical storm data. There are several studies in recent times on whether, and how, the properties of these storms have changed with climate change; see for example Robbins et al. (2011). Such storms can do immense harm to life and property, consequently a change in their patterns is of interest. Apart from being of current interest, the presence of some amount of evidence for change in the literature is helpful for evaluating whether our proposed methods can detect known instabilities. We study the yearly number of such storms, as well as the joint relationship between pressure and wind speed. We detect changes compatible with known facts. Interestingly, we find that although wind speeds and central pressure values of Atlantic tropical storms have changed, they have changed in sync, that is, their mutual relationship has remained stable over time. This lends credence to the idea that our methodology might be able to detect true changes and discard false signals well, since large-scale energy balance relationships (such as that between pressure and wind speed) are not expected to change.

Section 2 contains a brief literature review. Section 3 deals with EF CUSUM statistic derivation. Multivariate Gaussian CUSUM is discussed as well, with covariance matrix either singular or positive definite. A few examples are given as to how to derive CUSUM statistic, and Tables 1 and 2 are provided for the convenience of readers. Section 3.2 talks about change detection in the generalized linear model setting. Section 4 contains simulation studies. The data analysis for Atlantic tropical storms is provided in Sect. 5, followed by conclusions and discussion in Sect. 6.

2 Literature review

In this section we provide a partial list of techniques for change detection. As mentioned earlier, some of these originated in industrial quality context, and related methods include Shewhart charts (Shewhart, 1931), EWMA control

Table 1. Exponential family CUSUM: binomial, exponential, gamma and multivariate normal distributions.

Type of distribution	Density function	EF CUSUM based on
Binomial(n,p): $p \to p + \delta$	$\binom{n}{k} p^x (1-p)^{n-x}$	$x \log(\frac{p+\delta}{p}) + (N-x) \log(\frac{1-p-\delta}{1-p})$
Poisson(λ): $\lambda \to \lambda + \delta$	$\frac{\lambda^x e^{-\lambda}}{\lambda!}$	$x \log \frac{\lambda+\delta}{\lambda} - \delta$
Gamma(α,β): $\alpha \to \alpha + \delta_1, \beta \to \beta + \delta_2$	$\frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$	$\frac{\delta_2}{\beta(\beta+\delta_2)} x + \delta_1 \log \frac{x}{\beta+\delta_2} - \alpha \log \frac{\beta+\delta_2}{\beta} - \log \frac{\Gamma(\alpha+\delta_1)}{\Gamma(\alpha)}$
Multivariate normal: $N_p(\mu, \Sigma) \to N_p(\mu + \delta, \Sigma)$ Σ is positive definite	$\frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(x-\mu)' \Sigma^{-1} (x-\mu)\}$	$(x - \mu - \frac{1}{2}\delta)' \Sigma^{-1} \delta$

Table 2. CUSUM statistic for normal distribution: the first row is more general with both mean and variance change. The remaining three rows are special cases of the first one.

Distribution	CUSUM statistic
$N(\mu, \sigma_1^2) \to N(\mu + \delta_1, \sigma_2^2)$	$\log \sigma_1 + \frac{1}{2}\sigma_1^{-2}(x_i - \mu)^2 - \log \sigma_2 - \frac{1}{2}\sigma_2^{-2}(x_i - \mu - \delta_1)^2$
$N(\mu, \sigma^2) \to N(\mu + \delta, \sigma^2)$	$\sigma^{-2}(x_i - \mu - \frac{1}{2}\delta_1)\delta_1 \propto (x_i - \mu - \frac{1}{2}\delta_1)\delta_1$
$N(\mu, \sigma_1^2) \to N(\mu, \sigma_2^2)$	$\log(\sigma_2^{-1}\sigma_1) + \frac{1}{2}\sigma_1^{-2}\sigma_2^{-2}(\sigma_2^2 - \sigma_1^2)(x_i - \mu)^2$
$N(\theta, \theta^2) \to N(\theta + \delta_1, (\theta + \delta_1)^2)$	$\log((\theta + \delta_1)^{-1}\theta) + \frac{1}{2}\theta^{-2}(x_i - \theta)^2 - \frac{1}{2}(\theta + \delta_1)^{-2}(x_i - \theta - \delta_1)^2$

charts (Roberts, 1966), and CUSUM (Page, 1954). In the context of the CUSUM statistic, which originated from Page (1954, 1955), various optimality results are available in Lorden (1971), Khan (1979), Moustakides (1986), Ritov (1990), Pollak (1987), and Atienza et al. (2000) showing the versatility of this procedure.

The CUSUM technique has been extended to better suit practical needs, including Shu, Yeung and Jiang (2010) on adaptive CUSUM, Hawkins (1992) on robust average run length with Winsorization, Liu et al. (2006) on transformation of exponential data, and Yashchin (1993) on transforming serially correlated observations. In other directions, Lucas and Saccucci (1990) compared the average run length properties of EWMA with CUSUM, MacEachern, et al. (2007) developed robust CUSUM by modifying the likelihood function, Albers and Kallenberg (2009) proposed cummulative minimum (CUMIN) charts for grouped data and compared CUMIN with CUSUM and Shewhart charts, Chatterjee and Qiu (2009) proposed CUSUM control charts with control limits estimated using bootstrapping when the distribution was unknown, Steiner et al. (1999) used simultaneous CUSUM control charts to monitor correlated bivariate outcomes in the field of medical research, Crosier (1988) proposed vector CUSUM and Hotelling T^2-based CUSUM when dealing with multivariate case and compared

them to a Shewhart scheme, Lucas (1982) proposed Shewhart–CUSUM scheme to draw advantages of both methods for quick detection of mean change in the normal distribution setting, and Morais and Pacheco (2006) extended the approach to binomial data.

Some researchers have treated special cases in the EF CUSUM family, including Hawkins and Olwell (1997) on detecting known location and shape change in inverse gamma distribution, Hawkins and Zamba (2005) on change point detection in unknown mean and variance for normal distribution, Watkins et al. (2008) used negative binomial CUSUM to study outbreaks of Ross River virus disease and compared it to Early Aberration Reporting System CUSUM algorithms, Wu et al. (2008) studied large shifts in fraction non-conforming, and Lucas (1985) improved the Poisson CUSUM with fast initial response (FIR) and introduced the two-in-a-row rule to robust CUSUM. Healy (1987) discussed shift in mean and covariance for multivariate normal distribution using CUSUM, Alwan (2000) proposed transformation to normality to deal with EF CUSUM chart, Severo and Gama (2010) discussed using Kalman filter and CUSUM to detect residual mean and variance in the regression model, and Qiu and Hawkins (2001) used a rank-based CUSUM procedure to deal with multivariate measurements without normality assumption.

3 Distributional stability in exponential families

3.1 Known parameter case

Let the data be the random sample $\{X_1, \ldots, X_n\}$, where we know X_1 is observed first, then X_2 is observed, and so on. We assume that $X_1, \ldots, X_\tau$ are identically and independently distributed following an exponential family (EF) distribution with probability density or mass function given by

$$p(x; \theta, \phi) = \exp\left\{a(\phi)^{-1}(x\theta - b(\theta)) + c(x, \phi)\right\}.$$

Here the parameters are θ, which is of the same dimensionality as each of the data points, and ϕ.

We assume that $X_{\tau+1}, \ldots$ are identically and independently distributed from another EF distribution, with probability density function given by

$$p(x; \theta + \delta_1, \phi + \delta_2) = \exp\left\{a(\phi + \delta_2)^{-1}(x(\theta + \delta_1) \right.$$
$$\left. - b(\theta + \delta_1)) + c(x, \phi + \delta_2)\right\}.$$

Here τ is a fixed but unknown parameter denoting the time of change from one distribution to another, and $0 < \tau < \infty$. In the testing for distributional stability (TDS) framework we adopt in this paper, our interest is in testing the null hypothesis $H_0 : \tau \geq n$ against the alternative hypothesis $H_1 : \tau < n$. In keeping with the traditional process monitoring literature, we consider all parameter values, other than τ as known constants for now. Then in Sect. 3.3, we extend a selection of our results to the case where the parameters are estimated from the available data. Assuming some, or all, of these parameters as unknown requires additional technical conditions and assumptions.

Note that the time ordering of the observations is not an integral part to our methodology. Also, multiple change points may be allowed. For the former, we would assume that there is some permutation of the data, say $X_{\sigma_1}, \ldots, X_{\sigma_n}$ such that $X_{\sigma_1}, \ldots, X_{\sigma_\tau}$ are independent and identically distributed with some EF distribution with parameters θ and ϕ, while $X_{\sigma_{\tau+1}}, \ldots$ independent and identically distributed with the same distribution with a different set of parameter values. Also, multiple change points $\tau_1, \ldots, \tau_k$ can be easily accommodated in the above framework, and both the null and alternative hypothesis made more complex. In other words, we can extend our study to the case where, for some permutation of the indices, the data may be partitioned into k_0 segments under the null and k_1 segments under the alternative. Here, each segment of data is a set of independent, identically distributed exponential family random variables with its own distinct set of parameters. Our current problem may be thought of as the special case where $\sigma_i = i$ for $i = 1, \ldots, n$, $k_0 = 1$ and $k_1 = 2$. Extensions like those described above may lead to new approaches for solving several problems in applied statistics. However, in the interest of clarity of presentation, and to keep this paper at a reasonable length, we do not pursue such extensions here. Our method has a natural extension to time series and other dependent data with potential (unknown) change points, for which a likelihood can be written and computed, and an equivalent CUSUM testing framework can be established.

In our first result below, we obtain the test statistic for the hypothesis test described above. We adopt the convention that $\sum_{i=a}^{b} Y_i = 0$ whenever $a > b$, for any sequence of (possibly random) reals $\{Y_i\}$.

Theorem 3.1.

Let

$$
\begin{aligned}
Y_i &= a(\phi + \delta_2)^{-1}(X_i(\theta + \delta_1) - b(\theta + \delta_1)) \\
&\quad + c(X_i, \phi + \delta_2) - a(\phi)^{-1}(X_i\theta - b(\theta)) - c(X_i, \phi),
\end{aligned}
$$

for $i = 1, \ldots, n$, and further define $S_k = \sum_{i=1}^{k} Y_i$, adopting the convention that $S_0 = 0$.

The likelihood ratio test statistic for testing the null hypothesis $H_0 : \tau \geq n$ against the alternative hypothesis $H_1 : \tau < n$ is given by $T_n = S_n - \min_{0 \leq k < n} S_k$, and the null hypothesis is rejected if $T_n \geq L$ for a critical value L.

We omit the proof of this and several other theorems in the interest of brevity.

In general, the distribution of the test statistic T_n is intractable under both null and alternative hypothesis, consequently p value, power, and critical value L are difficult to find. Numeric methods are typically used to obtain these, and a parametric bootstrap is used when the distributional parameters are unknown and estimated. We discuss this issue in greater detail in Sect. 3.3.

The critical value L may be chosen by standard hypothesis testing protocol, by setting an upper bound α (significance level) to the probability of falsely rejecting the null hypothesis, i.e., type 1 error. However, in the framework of sequential process monitoring, the expected number of tests that may be performed before a false rejection is traditionally used as a control in place of the probability of a single test turning out to be a false rejection, and may be more meaningful in some applications. The former is called average run length (ARL) under the null hypothesis, denoted by ARL_0, and is related to the probability of type 1 error. A deeper discussion on this relation may be found in Li et al. (2013). Formally, the run length is $R = \inf\{n : S_n - \min_{0 \leq k < n} S_k = T_n \geq L\}$. The value of L is obtained by fixing the value of $\mathbb{E}R (= \mathrm{ARL})$ assuming $\tau = \infty$, at a pre-determined value ARL_0. In this paper we adopt the statistical process control-based approach of specifying control over false rejections using ARL_0. We set the value $\mathrm{ARL}_0 = 200$ for our examples and data analysis below. This implies a significance level of $\alpha = 0.005$ for a sequence of independent tests. More importantly, in our data sets of a few dozen observations, this implies that we are very unlikely to make a false rejection of the null hypothesis, since a hypothesis test for change at every single data point would still need an average of 200 observations for type 1 error to occur.

Note that the test statistic T_n may be written recursively as $T_n = \max\{0, T_{n-1} + Y_n\}$, with $T_0 = 0$. This form is reminiscent of the the celebrated CUSUM statistic. In view of this, we call T_n the EF CUSUM statistic. We obtain the classical CUSUM statistic as a special case as a corollary to Corollary 3.1 below. Note that $T_n \geq 0$ almost surely, hence a non-trivial test is obtained only when L is strictly positive. Our next result shows that this relation is fairly easy to ensure in practice.

Theorem 3.2. $\mathbb{E}_{\tau=\infty} R$ ($= \mathrm{ARL}_0$) ≥ 1 if and only if the critical value L is positive.

PROOF OF THEOREM 3.2

The necessity part: if $L \leq 0$, since $R = \inf\{n : S_n - \min_{0 \leq k < n} S_k \geq L\}$, we have $S_0 - \min_{0 \leq k < 0} S_k = 0 \geq L$ almost surely. Hence, we have $R = 0$ almost surely, and therefore $\mathbb{E}_{\tau=\infty}(R) = 0$, which is contradictory to $\mathrm{ARL}_0 \geq 1$.

The sufficiency part: If $L > 0$, then R cannot be zero because $S_0 - \min_{0 \leq k < 0} S_k = 0 < L$ almost surely; hence, R is at least 1 almost surely. Therefore $\mathrm{ARL}_0 \geq 1$.

$\square$

We now state some special cases of Theorem 3.1, which are of interest. Our first such result deals with the case where the observations are normally distributed. We use the notation $\overset{i.i.d.}{=}$ for independent and identically distributed.

Corollary 3.1. Suppose $X_1, \ldots, X_\tau \overset{i.i.d.}{=} N(\mu, \sigma_1^2)$ and $X_{\tau+1}, \ldots, X_n \overset{i.i.d.}{=} N(\mu + \delta_1, \sigma_2^2)$. For testing the null hypothesis $H_0 : \tau \geq n$ against the alternative $H_1 : 0 \leq \tau < n$, the likelihood ratio statistic is given by $C_n = S_n - \min_{0 \leq k < n} S_k$, where $S_k = \sum_{i=1}^{k} Y_i$ and $Y_i = \log(\sigma_1/\sigma_2) + \frac{1}{2}\sigma_1^{-2}(X_i - \mu)^2 - \log(\sigma_2) - \frac{1}{2}\sigma_2^{-2}(X_i - \mu - \delta_1)^2$.

In the very special case where $\sigma_1 = \sigma_2 = 1, \mu = 0$, we obtain $Y_i = (X_i - \delta/2)$, and hence obtain $S_n - \min_{0 \leq k < n} S_k = C_n = \max\{0, C_{n-1} + X_i - \delta/2\}$, with $C_0 = 0$. This expression is that of the classical Gaussian CUSUM, where the factor $\delta/2$ is often called the allowance constant.

The statistic C_n defined as $C_n = \max\{0, C_{n-1} + X_i - \delta/2\}$ (with $C_0 = 0$) is often used as a default statistic for change detection. Our result above shows that this statistic may also be obtained in a non-sequential framework; however, the assumption of normal distribution seems unavoidable. Since C_n is also used for change detection in non-normal data, it is of interest to know under what circumstance it may obtain reasonable accuracy and precision with change detection. Our next theorem describes the conditions under which using C_n as a statistic may be a reasonable procedure.

Theorem 3.3. Consider the framework of Theorem 3.1. Additionally, assume that the third derivative of $b(\cdot)$ at θ_0 is zero, i.e., $b'''(\theta_0) = 0$, that δ_1 is small and $\delta_2 = 0$.

Under these assumptions, the difference between the normality-based CUSUM C_n and the EF CUSUM T_n is as follows: $|C_n - T_n| = o_p(n\delta_1)$.

Example 3.1.1

Binomial change detection: in the case of binomial distribution with parameter p, the natural parameter is $\theta = \log((1-p)^{-1}p)$, and $b(\theta) = n\log(1 + \exp\{\theta\})$, ϕ is taken as a constant. Also $b'''(\theta) = (1 + \exp\{\theta\})^{-4}\{n\exp\{\theta\}(1 + \exp\{\theta\})(1 - \exp\{\theta\})\}$, $b'''(\theta_0) = 0$ if and only if $\theta_0 = 0$. In that case, $p = \frac{1}{2}$. To conclude, when $p = \frac{1}{2}$, a change from $p \to p + \delta_1$ using Gaussian CUSUM $\widetilde{y}$ and EF CUSUM y yield similar performance.

Corollary 3.2. For the same detection problem as above, under the condition of $b'''(\theta_0) = b''''(\theta_0) = 0, \delta_1$ is small and $\delta_2 = 0$, we get an even stronger result $|C_n - T_n| = o_p(n\delta_1^2)$.

Example 3.1.2

Change from $N_p(\mu, \Sigma_1)$ **to** $N_p(\mu + \delta, \Sigma_2)$.

The CUSUM for multivariate normal distribution is somewhat more complicated and therefore we divide this problem into the following cases based on the nature of the variance–covariance matrix. In all the cases listed below, the test statistic is $C_n = S_n - \min_{0 \leq k < n} S_k$, where $S_k = \sum_{i=1}^{k} Y_i$ and Y_i depends from one case to another. This result is a corollary of Theorem 3.1, but is of independent interest owing to the multitude of applications involving the normal distribution.

1. $\Sigma_1 = \Sigma_2 = \Sigma$, where Σ is positive definite. Based on the following density function: $f(x|\mu, \Sigma) = (2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\}$ it is straightforward to derive the CUSUM statistic based on $Y_i = (x_i - \mu - \frac{1}{2}\delta)'\Sigma^{-1}\delta$. If we let $p = 1$, we are back to the univariate normal situation.

2. $\Sigma_1 = \Sigma_2 = \Sigma$, where Σ is a singular.

   Assume rank $(\Sigma) = r, r < p$. There exists an orthogonal matrix Q_{p*p}, such that $Q\Sigma Q' = \Lambda$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)$, where $\lambda_i > 0, i = 1, 2, \ldots, r$. So $Z = QX \sim N_p(Q\mu, \Lambda)$. Let $P = (I_r 0_{r \times (p-r)})$, and $K = PZ \sim N_r(PQ\mu, \widetilde{\Lambda})$, where $\widetilde{\Lambda} == \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$. Thus the problem is reduced to a change of $N_r(PQ\mu, \widetilde{\Sigma})$ to $N_r(PQ(\mu + \delta), \widetilde{\Sigma})$, and we are back to case 1. The CUSUM statistic is based on $Y_i = (x_i - \mu - \frac{1}{2}\delta)'(PQ)'\widetilde{\Sigma}^{-1}PQ\delta$.

3. $\Sigma_1 \neq \Sigma_2$, where Σ_1, Σ_2 are both positive definite. Following previous discussion, the CUSUM statistic is based on $Y_i = \frac{1}{2}\log(|\Sigma_1|^{-1}|\Sigma_2|) + \frac{1}{2}(x_i - \mu - \delta)'\Sigma_2^{-1}(x_i - \mu - \delta) - \frac{1}{2}(x_i - \mu)'\Sigma_1^{-1}(x_i - \mu)$.

4. $\Sigma_1 \neq \Sigma_2$, where Σ_1, Σ_2 are both singular.

   Based on discussion of case 2, our CUSUM statistic is based on $Y_i = (\frac{r_2}{2} - \frac{r_1}{2})\log(2\pi) + \frac{1}{2}\log(|\widetilde{\Lambda_1}|^{-1}|\widetilde{\Lambda_2}|) - $

$\frac{1}{2}(P_1 Q_1 (x_i - \mu))' \widetilde{\Lambda_1}^{-1} (P_1 Q_1 (x_i - \mu)) + \frac{1}{2}(P_2 Q_2 (x_i - \mu - \delta))' \widetilde{\Lambda_2}^{-1} (P_2 Q_2 (x_i - \mu - \delta))$. Here P_1, Q_1, P_2, Q_2 are such that $P_1 Q_1 \Sigma_1 Q_1' P_1' = \widetilde{\Lambda_1}$, $P_2 Q_2 \Sigma_2 Q_2' P_2' = \widetilde{\Lambda_2}$, and rank $(\widetilde{\Lambda_1})$ = rank (Σ_1), rank$(\widetilde{\Lambda_2})$ = rank (Σ_2), $\widetilde{\Lambda_1}, \widetilde{\Lambda_2}$ are $r_1 \times r_1$ and $r_2 \times r_2$ diagonal matrix.

5. $\Sigma_1 \neq \Sigma_2$, where Σ_1 is positive definite, Σ_2 is singular. In this case we have $Y_i = \frac{r_2 - p}{2} \log(2\pi) + \frac{1}{2} \log(|\widetilde{\Lambda_1}|^{-1} |\widetilde{\Lambda_2}|) + \frac{1}{2}(P_2 Q_2 (x_i - \mu - \delta))' \widetilde{\Lambda_2}^{-1} (P_2 Q_2 (x_i - \mu - \delta)) - \frac{1}{2}(x_i - \mu) \Sigma_1^{-1} (x_i - \mu)$, where $P_2 Q_2 \Sigma_2 Q_2' P_2' = \widetilde{\Lambda_2}$, rank$(\widetilde{\Lambda_2})$ = rank (Σ_2), $\widetilde{\Lambda_2}$ is $r_2 \times r_2$ diagonal matrix.

3.2 Generalized linear model and CUSUM

In this section, we consider data of the form $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$. Here, the y_i's are the responses, and the $\mathbf{x}_i$'s are covariates that are considered to be fixed constant vectors. We assume that y_i's come from the distribution $p(y_i | \theta_i) = \exp\{a(\phi)^{-1}(y_i \theta_i - b(\theta_i)) + c(y_i, \phi)\}$, where $\theta_i = \mathbf{x}_i' \beta$ is the canonical parameter under stable distributional regime and $a(\phi) > 0$ is a dispersion parameter. Our main result below generalizes the main result of the previous section, and presents a change detection test statistic for generalized linear models:

Theorem 3.4. Assume that $(y_1, \mathbf{x}_1), \ldots, (y_\tau, \mathbf{x}_\tau)$, the true model is $\theta_i = \mathbf{x}_i' \beta$, and for $(y_{\tau+1}, \mathbf{x}_{\tau+1}), \ldots, (y_n, \mathbf{x}_n)$, the true model is $\theta_i = \mathbf{x}_i' (\beta + \delta)$, where β, δ is known. For the hypothesis testing $H_0 : \tau \geq n$ vs $H_1 : 0 \leq \tau < n$, if we denote $z_i = y_i \mathbf{x}_i' \delta - b(\mathbf{x}_i' (\beta + \delta)) + b(\mathbf{x}_i' \beta)$ and $S_k = \sum_{i=1}^{k} z_i$, then the test statistic is $S_n - \min_{0 \leq k < n} S_k$.

3.3 Estimated parameter cases

We now illustrate that the results presented above extend to the case where the parameters are unknown. For simplicity of presentation, we omit the scaling function $a(\phi)$ for the first two results below. We begin with the single parameter framework where $X_1, \ldots, X_{\tau_n}$ are independent and identically distributed with density

$$p(x; \theta_0) = \exp\{(x\theta_0 - b(\theta_0)) + c(x)\},$$

and $X_{\tau_n+1}, \ldots$ are i.i.d. with density

$$p(x; \theta_1) = \exp\{(x\theta_1 - b(\theta_1)) + c(x)\}.$$

We assume $\theta_1 \neq \theta_0$ throughout. We test the null hypothesis $H_0 : \tau_n \geq n$ against the alternative $H_1 : 0 \leq \tau_n < n$. Let us denote the maximum likelihood estimator for θ_0 based on $X_1, \ldots, X_n$ as $\hat{\theta}_{00}$; note that this is under the null hypothesis scenario. Also, under the alternative hypothesis scenario, the likelihood $L(\theta_0, \theta_1, \tau_n) = \prod_{i=1}^{\tau_n} p(X_i; \theta_0) \prod_{i=\tau_n+1}^{n} p(X_i; \theta_0)$ is maximized at $(\hat{\theta}_{10}, \hat{\theta}_{11}, \hat{\tau}_n)$. We have the following result:

Theorem 3.5. In the framework described above, the likelihood ratio test statistic is given by

$$
\begin{aligned}
T_{n1} = & (\hat{\theta}_{10} - \hat{\theta}_{00}) \sum_{i=1}^{\hat{\tau}_n} X_i + (\hat{\theta}_{11} - \hat{\theta}_{00}) \sum_{i=\hat{\tau}_n+1}^{n} X_i \\
& - \hat{\tau}_n b(\hat{\theta}_{10}) - (n - \hat{\tau}_n) b(\hat{\theta}_{11}) + n b(\hat{\theta}_{00}).
\end{aligned}
$$

Further, under either $\tau_n \geq n$ or $\tau_n/n \in (0, 1)$, the parametric bootstrap scheme may be used to estimate the distribution of T_{n1}, and consequently obtain a rejection region and p value of the above hypothesis test.

It may be noted, however, that the above test statistic can suffer from extremely low power, depending on the values of θ_0, θ_1, and τ_n. One reason for this performance deficiency is that θ_{00} is not a consistent estimator for θ_0 under the alternative hypothesis. In order to address this issue and improve the performance capabilities of our testing procedure, we propose a modification of the usual likelihood ratio test, whereby we use $\hat{\theta}_{10}$ as the estimator for θ_0, even under the null hypothesis. We have the following result:

Theorem 3.6. In the framework of Theorem 3.5, the profile likelihood ratio test statistic is

$$T_{n2} = (\hat{\theta}_{11} - \hat{\theta}_{00}) \sum_{i=\hat{\tau}_n+1}^{n} X_i - (n - \hat{\tau}_n)(b(\hat{\theta}_{11}) - b(\hat{\theta}_{00})).$$

Further, under either $\tau_n \geq n$ or $\tau_n/n \in (0, 1)$, the parametric bootstrap scheme may be used to estimate the distribution of T_{n1}, and consequently obtain a rejection region and p value of the above hypothesis test. Further, the power of this test tends to one when $\tau_n/n \in (0, 1)$. In addition, $(\hat{\theta}_{10}, \hat{\theta}_{11}, \hat{\tau}_n)$ converge in probability to $(\theta_0, \theta_1, \tau_n)$ under standard conditions.

The above test statistic can be obtained from the profile likelihood (for null and alternative), when θ_0 is replaced with $\hat{\theta}_{10}$. Another useful variant is the case where both θ_0 and θ_1 may be estimated from the full data, perhaps under some restrictions on the model. An example is where the the null distribution is $N(\theta_0, \sigma^2)$, and after τ_n it changes to $N(\theta_0 + c\sigma, \sigma^2)$ for some known constant c. This formulation is particularly useful for applications, where it may be of importance to detect only practically significant lack of stability of distributions, and not just statistically significant ones. In our simulation examples and the real data analysis below, we consider the above specification where we test for a change in mean in terms of c standard deviation units. We study results with $c = 1, 1/2, 1/4$ as potential cases of relatively easy, not easy, and hard change-detection scenarios. This framework is adopted in this paper since it makes sense to describe the distance between the null and alternative scenarios in terms of units of standard deviation. Also, in samples of finite sizes, the only scenario where we get reasonable power in hypothesis tests is when the two hypotheses are sufficiently apart. Additionally, for practical purposes, even if there is a change

but the change is minute and negligible, the hypotheses test may be redundant. Based on all these considerations, it is advisable to test hypotheses that are a reasonable number of standard deviation units away from each other.

There can be several other results relating to stability detection with estimated parameters, under various assumptions and technical conditions, which we will address in future work. We conclude this section with a result on stability detection when parameters are estimated in a generalized linear model.

Theorem 3.7. Assume that for $(y_1, \mathbf{x}_1), \ldots, (y_{\tau_n}, \mathbf{x}_{\tau_n})$, the true model is $\theta_{i0} = \mathbf{x}_i'\beta_0$, and for $(y_{\tau_n+1}, \mathbf{x}_{\tau_n+1}), \ldots, (y_n, \mathbf{x}_n)$, the true model is $\theta_{i1} = \mathbf{x}_i'\beta_1$. For the hypothesis testing $H_0 : \tau_n \geq n$ vs. $H_1 : 0 \leq \tau_n < n$, the test statistic is $T_{n3} = \sum_{i=1\hat{\tau}_n+1}^{n} a^{-1}(\hat{\phi}) \left\{ y_i \mathbf{x}_i'(\hat{\beta}_1 - \hat{\beta}_0) - b(\mathbf{x}_i'\beta_1) + b(\mathbf{x}_i'\beta_0) \right\}$.

We present below a sketch of the proof of the above result.

SKETCH OF PROOF OF THEOREM 3.7

The likelihood function under the alternative hypothesis is

$$
\begin{aligned}
&L_1(\beta_0, \beta_1, \tau_n, \phi) \\
=\ & \prod_{i=1}^{\tau_n} \exp\{a(\phi)^{-1}(y_i \mathbf{x}_i'\beta_0 - b(\mathbf{x}_i'\beta_0)) + c(y_i, \phi)\} \\
&\times \prod_{i=\tau_n+1}^{n} \exp\{a(\phi)^{-1}(y_i \mathbf{x}_i'\beta_1 - b(\mathbf{x_i}'\beta_1)) + c(y_i, \phi)\}.
\end{aligned}
$$

Suppose this function is maximized at $(\hat{\beta}_0, \hat{\beta}_1, \hat{\tau}_n, \hat{\phi})$. We evaluate the likelihood under the null hypothesis at $\hat{\beta}_0, \hat{\phi}$, and obtain the profile likelihood ratio as

$$
\begin{aligned}
\Lambda(\tau) \ =\ & \frac{L_1(\hat{\beta}_0, \hat{\beta}_1, \hat{\tau}_n, \hat{\phi})}{L_0(\hat{\beta}_0, \hat{\phi})} \\
=\ & \exp\left[\sum_{i=1\hat{\tau}_n+1}^{n} a^{-1}(\hat{\phi}) \left\{ y_i \mathbf{x}_i'(\hat{\beta}_1 - \hat{\beta}_0) \right. \right. \\
& \left. \left. \left[\sum_{i=\hat{\tau}_n+1}^{n} a^{-1}(\hat{\phi}) \left\{ -b(\mathbf{x}_i'\beta_1) + b(\mathbf{x}_i'\beta_0) \right\} \right] \right. \right].
\end{aligned}
$$

$\square$

Furthermore, in the generalized linear model case, the parametric bootstrap is a viable way of approximating the distribution of T_{n3}, and thus eliciting the properties of the test for stability.

4 Simulation study

In this section, we discuss a simulation study on the change of parameter(s) for binomial, exponential, gamma, and Poisson distributions, and compare the EF CUSUM statistic with the Gaussian CUSUM statistic, under the constraint that the mean and the standard deviation of both distributions are equal. Based on the exponential family density $f(x; \theta, \phi) = $ $\exp\{a(\phi)^{-1}(x\theta - b(\theta)) + c(x, \phi)\}$, it is easy to calculate $E(X) = b'(\theta)$, and var $(X) = b''(\theta)a(\phi)$. When there is change in parameter from θ to $\theta + \delta_1$ and from ϕ to $\phi + \delta_2$, we have $E(X) = b'(\theta + \delta_1)$ and var$(X) = b''(\theta + \delta_1)a(\phi + \delta_2)$. So the corresponding Gaussian assumption-based setting is a change from $N(b'(\theta), b''(\theta)a(\phi))$ to $N(b'(\theta + \delta_1), b''(\theta + \delta_1)a(\phi + \delta_2))$.

The simulation procedure can be described as follows: First, we control false alarms by carefully choosing L under the null distribution by fixing $\mathrm{ARL}_0 = 200$. Second, we compute $E((R - \tau)^+)$ under the alternative distribution. Let τ be the time of change. We simulate $x_1, \ldots, x_\tau \stackrel{i.i.d.}{=} f(x|\theta)$ and $x_{\tau+1}, \ldots, x_T \stackrel{i.i.d.}{=} f(x|\theta + \delta)$ for 2500 replications, where δ is known. For each $\tau = 0, 1, \ldots, 100$, use the L from the first step and compute R for the 2500 replications to get the mean, median, standard deviation, and maximum of $(R(\tau))$. We simultaneously carry out the same procedure for the Gaussian CUSUM case for comparison with the EF CUSUM.

From the simulation results in Fig. 1, one key finding is that in most cases, EF CUSUM statistic performs better than Gaussian CUSUM statistic except for one occasion when the underlying distribution is exponential distribution. Also note that for a small shift in parameter, exponential CUSUM has a considerable advantage over the Gaussian CUSUM, while for a large shift in parameter, EF CUSUM still works better than Gaussian CUSUM, but not significantly different.

We also discover that $E_1(R(\tau))$ does not vary a lot with τ changing from 0 to 100 for a particular distribution in the exponential family. Particularly, for τ close to 0 or close to 100, $E_1(R(\tau))$ is still quite stable. In addition, the median, standard deviation, and maximum of average run length tell the same story as the mean.

5 Tropical storm data analysis

We now discuss a case study of Atlantic tropical storms, for which we use HURDAT (hurricane database) data from the US National Hurricane Center. For each storm, the following information is recorded: date and time, tropical storm identity, tropical storm name, position in latitude and longitude, maximum sustained winds in knots, and central pressure in millibars.

We present our results from three studies on Atlantic tropical storms here. Each of these studies are carried out on two data sets: a longer series from 1851 to 2008 and a shorter series from 1951 to 2008. The expectation–maximization algorithm was used for missing data segments in the longer series when required, this problem does not arise in the shorter series.

First, we consider the problem of TDS for the yearly number of tropical storms between 1851 and 2008. This yearly data is modeled as Poisson($\hat{\mu}$), and a potential change to Poisson($\hat{\mu} + \delta$) is studied. We assume that any potential change point occurred after 1900, and use the data previ-
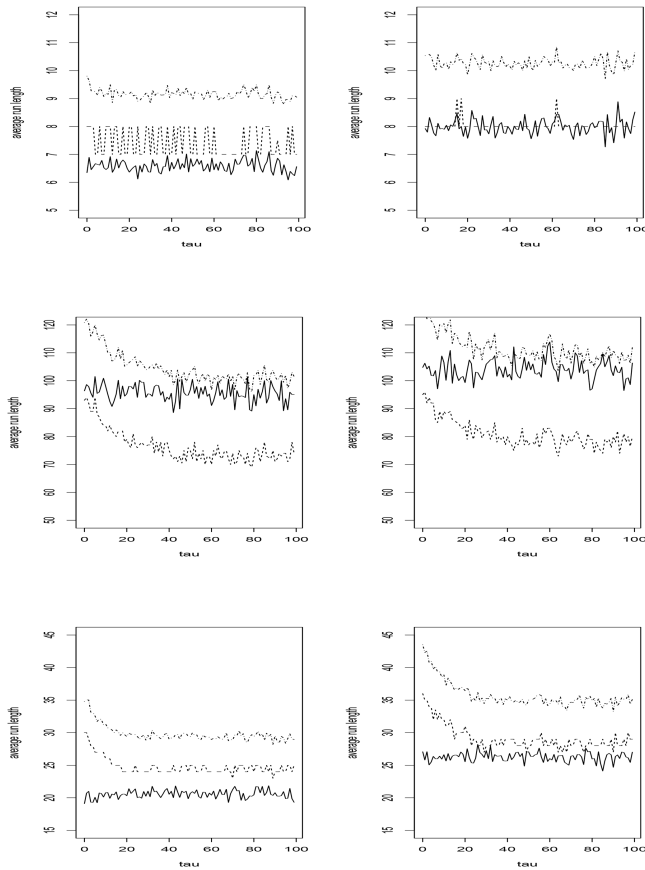
Figure 1. Performance comparison: EF CUSUM with Gaussian CUSUM. Dot-dash, dashed, and solid line stand for mean, median, and standard deviation. The top panel describes run length comparison from Binomial(15,0.95) to Binomial(15,0.90), the middle panel describes run length comparison from Poisson(3) to Poisson(3.1), the bottom panel describes run length comparison from Gamma(1,2) to Gamma(1.5,1.5). Due to length limitation of the graphs, we do not include the MAX line here.

Figure 2. The observed data and a Poisson fit for the number of tropical storms between 1951 and 2008.

ous to it for estimating parameters. We estimate $\hat{\mu} = 7.54$, and fix $\delta = c\hat{\sigma}$, where c is predetermined as $\frac{1}{4}$, $\frac{1}{2}$, and 1, and $\hat{\sigma} = 2.75$ is the estimated standard deviation. Note that $\sigma \approx \mu^{\frac{1}{2}}$ because for the Poisson distribution, the mean equals the variance. Then we create the Poisson CUSUM statistic as given in Table 1. We get L based on $E_0(R) = 200$, and search for the first n that makes $S_n - \min_{0 \le k < n} S_k \ge L$ with the tropical storm data.

In view of the fact that the data from the nineteenth century and the first half of the twentieth century may not be entirely reliable, we repeated the above analysis on detecting change for the Atlantic tropical storms from year 1951 to 2008. We assume that the potential change could only occur after 1970. For detecting potential change Poisson($\hat{\mu}$) to Poisson($\hat{\mu} + \delta$), we now have $\hat{\mu} = 9.8$, and $\delta = c\hat{\sigma}$, where c is predetermined as $\frac{1}{4}$, $\frac{1}{2}$, and 1, and $\hat{\sigma} = 2.97$. Note that in both analyses, the
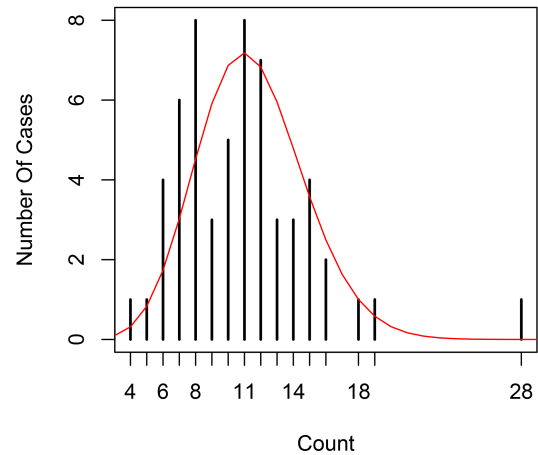
sample standard deviation is close to the sample mean, again verifying the Poisson model assumption.

In both of these analyses, our results are not particularly sensitive to the choice of the initial segment when no change is assumed to occur (i.e., until 1900 and 1970, respectively, in the first and second analysis described above). We also verified that the assumption that the number of tropical storms in a given year follows a Poisson distribution is reasonable. For example, a goodness-of-fit p value for testing Poisson distribution fit is 0.8, thus strongly rejecting that Poisson is a bad fit. Note Fig. 2 also has an observed and expected plot for the data between 1951 and 2008. We also explored the possibility that there may be a temporal pattern in the number of tropical storms over the years, but that was ruled out from autocorrelation and partial autocorrelation computations on both the original and logarithmic scales.

The second study has two parts. For the data from 1851 to 2008, we model the maximum sustained winds and maximum central pressure as N_2 ($\hat{\mu}$, $\hat{\Sigma}$), and study potential change to N_2 ($\hat{\mu} + \delta$, $\hat{\Sigma}$). We estimate the mean $\hat{\mu}$ and variance–covariance matrix $\hat{\Sigma}$ based on the first 50 observations. Here $\hat{\mu} = \begin{pmatrix} 104.8 \\ 982.99 \end{pmatrix}$, and $\hat{\Sigma} = \begin{pmatrix} \hat{\sigma_{11}} & \hat{\sigma_{12}} \\ \hat{\sigma_{21}} & \hat{\sigma_{22}} \end{pmatrix} = \begin{pmatrix} 199.96 & -20.66 \\ -20.66 & 367.56 \end{pmatrix}$. Let $\delta = \begin{pmatrix} c\hat{\sigma_{11}} \\ c\hat{\sigma_{22}} \end{pmatrix}$, where c is predetermined as $\frac{1}{4}$, $\frac{1}{2}$, and 1.

In a variation of the second study, we consider maximum sustained wind speed and minimum central pressure as $N_2(\hat{\mu}, \hat{\Sigma})$ and study potential change to $N_2(\hat{\mu} + \delta, \hat{\Sigma})$. Here $\hat{\mu} = \begin{pmatrix} 129.5 \\ 937.6 \end{pmatrix}$, and

$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma_{11}} & \hat{\sigma_{12}} \\ \hat{\sigma_{21}} & \hat{\sigma_{22}} \end{pmatrix} = \begin{pmatrix} 376.05 & -220.47 \\ -220.47 & 237.41 \end{pmatrix}$. Let $\delta = \begin{pmatrix} c\hat{\sigma_{11}} \\ c\hat{\sigma_{22}} \end{pmatrix}$, where c is predetermined as $\frac{1}{4}$, $\frac{1}{2}$, and 1.

Table 3. Atlantic tropical storm data from 1851 to 2008 are used to detect any mean change in tropical storm characteristics. Here c is the magnitude representing the number of standard deviation from the mean. The result shows that the number of tropical storm had a significant increase around 1933–1936, and strength of the tropical storm increased around 1923–1924.

Distribution	$c = \frac{1}{4}$	$c = \frac{1}{2}$	$c = 1$
Poisson	1936	1933	1933
Bivariate normal	1924	1923	1924

Table 4. Atlantic tropical storm data from 1951 to 2008 are used to detect any mean change in tropical storm characteristics. Here c is the magnitude representing the number of standard deviation from the mean. The result shows that the number of tropical storm had a significant increase around the year of 2000, and strength of the tropical storm has not changed.

Distribution	$c = \frac{1}{4}$	$c = \frac{1}{2}$	$c = 1$
Poisson	2001	2001	2000
Bivariate normal	2008	2008	2008

The results are summarized in Tables 3 and 4. We discover that the number of tropical storms had a significant increase around 1933–1936, and the strength of the tropical storms had a sharp increase around 1923–1924. This is consistent with the historical records. In history, the 1924 tropical storm Cuba was the earliest officially classified Category 5 Atlantic hurricane on the Saffir–Simpson scale, and it became the strongest hurricane on record to hit the country; furthermore, the 1928 Okeechobee hurricane was the second recorded hurricane to reach Category 5 status on the Saffir–Simpson scale in the Atlantic basin after the 1924 Cuba hurricane. The 1933 Atlantic tropical storm season was the second most active Atlantic tropical storm season on record with 21 storms, and the 1936 season was fairly active, with 17 tropical cyclones including a tropical depression. From the analysis of the shorter series, we detect that the period 2000–2001 saw an increase in the number of tropical storms. According to National Hurricane Center, the 2001 Atlantic tropical storm season produced 17 tropical storms. Notice that the results we obtain are consistent for $c = 1, 1/2, and 1/4$ which strongly suggests that the changes we see are not false discoveries. As a further corroborative check, we present a moving estimate of the average number of tropical storms in Fig. 3, which strongly suggests there is a change in the average around the fiftieth observation, i.e., in 2000 for the 1951–2008 data. Our results are similar to those obtained by Robbins et al. (2011) (see Sect. 5 therein), who notice changes in North Atlantic tropical storm patterns circa 1930 and 1995.

In the third study, we consider the relationship between the number of tropical storms Y, the maximum sustained winds

Figure 3. A moving average estimate of the average number of tropical storms between 1951 and 2008.

X_1, and maximum (minimum) central pressure for data between 1851 and 2008 (1951 and 2008) X_2. We model Y as Poisson(λ), where $\theta = \log \lambda$, $p(y, \theta) = \exp\{y\theta - e^\theta - \log y!\}$ and use the canonical link $\theta = (1, X)'\beta$.

For the 1851–2008 data, we take the first 50 observations, and get $\hat{\beta} = (-4.99, 0.01, 0.006)'$. We also estimate the bivariate mean and covariance as $\hat{\mu} = (104.8, 982.99)'$ and $\hat{\Sigma} = \begin{pmatrix} 199.96 & -20.66 \\ -20.66 & 367.56 \end{pmatrix}$. Second, we select $\delta = c\hat{\beta}$, where $c = \frac{1}{4}, \frac{1}{2}, 1$. Next we search for L, assuming $\mathrm{ARL}_0 = 200$. To implement this, we simulate the bivariate series X using $\hat{\mu}$ and $\hat{\Sigma}$. Based on equation $\log(\hat{\lambda}) = (1, X)'\hat{\beta}$, we get $\hat{\lambda}$, and we can simulate Y from Poisson ($\hat{\lambda}$). Construct the CUSUM statistic and the stopping rule $S_n - \min_{0 \le k < n} S_k \ge L$ to satisfy $\mathrm{ARL}_0 = 200$. Finally, we fit the stopping rule to the real data and discover the signal. Results shows that there is no significant change in terms of β, which means the way the maximum sustained winds and maximum central pressure of how a tropical storm relates to the number of tropical storms has not changed over the past 158 years.

For the 1951–2008 data, we take the first 20 observations, and get $\hat{\beta} = (3.08, 0.003, -0.0016)'$. We also estimate the bivariate mean and covariance as $\hat{\mu} = \begin{pmatrix} 129.5 \\ 937.6 \end{pmatrix}$, and $\hat{\Sigma} = \begin{pmatrix} 376.05 & -220.47 \\ -220.47 & 237.41 \end{pmatrix}$. Second, we select $\delta = c\hat{\beta}$, where $c = \frac{1}{4}, \frac{1}{2}, 1$. Results shows that there is no significant change in terms of β, which means the way how the maximum sustained winds and minimum central pressure of a tropical storm relate to the number of tropical storms has not changed over the past 58 years. Thus, the third part of our study shows broad physical relations between wind speeds and pressures have not changed, which is to be expected.

6 Conclusion and future work

The EF CUSUM generally performs better than the Gaussian CUSUM. In practice, in situations where the data do not follow normal distribution, we should consider the appropriate distribution for modeling the data and choose the corresponding CUSUM statistic to effectively detect the change in parameter(s) if there is any. Further details for the mathematical proofs, simulation studies, and our analysis of Atlantic tropical storms record are available from the authors.

In general, optimality results for our proposed methods should follow along the lines similar to those established by Moustakides (1986) and related works, but this requires a separate proof. There are other situations of interest in geophysical studies where an exponential family model may not be appropriate. Examples include extremes, cases where the parameter is a boundary point of the support of the random variable, and mixtures of distributions. Our future work will consist of stability detection for such cases.

The presence of temporal dependence is typically not problematic; furthermore, our likelihood-based schemes generalize easily to standard time series frameworks, although additional mathematical technicalities cannot be avoided. In addition, cases where the observations are not temporally ordered, or when there are multiple break points, need suitable generalizations and mathematical treatment. Note that there is a relationship between the number of structural breaks in the distribution of a data sequence, the size of such breaks, and the probabilities of true/false inference from hypothesis testing. Establishing the limits of our proposed methodology along these lines is work to be realized in the future.

It should be noted that the methodology discussed here may fail under several different scenarios. For example, when parameters of the distributions are unknown, there will be no reasonable way of obtaining the null or alternative distribution consistently if there are too few observations before or after any change point. This also suggests that the proposed method may not be able to adapt to situations where there are many change points, or when one or more changes in the parameters asymptotes to zero quickly. Although we consider exponential family distributions here which lends itself to several standard statistical techniques, our proposed tests may require modifications if other distributions are involved, and a parametric bootstrap is not guaranteed to produce consistent distributional approximations.

Edited by: D. Wang
Reviewed by: three anonymous referees

References

Albers, W. and Kallenberg, W. C. M.: CUMIN Charts, Metrika, 70, 111–130, 2009.

Alwan, L. C.: Designing an Effective EF-CUSUM Chart without the Use of Nomographs, Communic. Stat.-Theory Methods, 29, 2879–2893, 2000.

Atienza, O. O., Tang, L. C., and Ang, B. W.: A Uniform Most Powerful Cumulative Sum Scheme Based on Symmetry, J. Roy. Stat. Soc. Series D, 49, 209–217, 2000.

Bolton, R. J. and Hand, D. J.: Statistical fraud detection: A review, Stat. Sci., 17, 235–255, 2002.

Brown, R. L., Durbin, J., and Evans, J. M.: Techniques for Testing the Constancy of Regression Relationships over Time, J. Roy. Stat. Soc. B, 37, 149–192, 1975.

Chatterjee, S. and Qiu, P.: Distribution-Free Cumulative Sum Control Charts Using Bootstrap-Based Control Limits, The Ann. Appl. Stat., 3, 349–369, 2009.

Chen, W., Chen, D., Sun, G., and Zhang, Y.: Defending Against Jamming Attacks in Wireless Local Area Networks, Lect. Notes Comput. Sci., 4610, 519–528, 2007.

Chihwa, K. and Ross, S. L.: A Cusum Test in the Linear Regression Model with Serially Correlated Disturbances, Econom. Rev., 14, 331–346, 1995.

Crosier, R. B.: Multivariate Generalizations of Cumulative Sum Quality-Control Schemes, Technometrics, 30, 291–303, 1988.

Hawkins, D. M.: Evaluation of Average Run Lengths of Cumulative Sum Charts for an Arbitrary Data Distributions, Commun. Stat.-Simulat. Comput., 21, 1001–1020, 1992.

Hawkins, D. M. and Olwell, D. H.: Inverse Gaussian Cumulative Sum Control Charts for Location and Shape, J. Roy. Stat. Soc. D, 46, 323–335, 1997.

Hawkins, D. M. and Zamba, K. D.: A Change-point Model for a Shift in Variance, J. Qual. Technol., 37, 21–31, 2005.

Healy, J. D.: A Note on Multivariate CUSUM Procedures, Technometrics, 29, 409–412, 1987.

Jandhyala, V. K. and MacNeill, I. B.: Tests for Parameter Changes At Unknown Times in Linear Regression Models, J. Stat. Plann. Inference, 27, 291–316, 1991.

Khan, R. A.: A Sequence Detection Procedure and the Related Cusum Procedure, The Indian J. Stat. B, 40, 146–162, 1979.

Krawczak, M., Ball, E., Fenton, I., Stenson, P., Abeysinghe, S., Thomas, N., and Cooper, D. N.: Human Gene Mutation Database: A biomedical Information and Research Resource, Human Mutat., 15, 45–51, 1999.

Lee, S., Tokutsu, Y., and Maekawa, K.: The CUSUM Test for Parameter Change in Regression Models with ARCH Errors, J. Jap. Stat. Soc., 34, 173–188, 2004.

Li, Z., Qiu, P., Chatterjee, S., and Wang, Z.: Using P-Values To Design Statistical Process Control Charts, Stat. Papers, 54, 523–539, 2013.

Liu, J. Y., Xie, M., and Goh, T. N.: CUSUM Chart with Transformed Exponential Data. Communic. Stat.-Theory Methods, 35, 1829–1843, 2006.

Lorden, G.: Procedures for Reacting to a Change in Distribution, The Ann. Stat., 42, 1897–1908, 1971.

Lucas, J. M.: Counted Data CUSUM's, Technometrics, 27, 129–144, 1985.

Lucas, J. M.: Combined Shewhart-CUSUM Quality Control Schemes, J. Qual. Technol., 14, 51–59, 1982.

Lucas, J. M. and Saccucci, M. S.: Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements, Technometrics, 32, 1–12, 1990.

MacEachern, S. N., Rao, Y., and Wu, C.: A Robust-Likelihood Cumulative Sum Chart, J. Am. Stat. Assoc., 102, 1440–1447, 2007.

Meulbroek, L. K.: An empirical analysis of illegal insider trading, The J. Finance, 47, 1661–1699, 1992.

Morais, M. C. and Pacheco, A.: Combined CUSUM-Shewhart Schemes for Binomial Data, Econom. Qual. Control, 21, 43–57, 2006.

Moustakides, G. V.: Optimal Stopping Times for Detecting Changes in Distribution, The Annal. Stat., 14, 1379–1387, 1986.

Page, E. S.: Continuous inspection schemes. Biometrika, 41, 100–115, 1954.

Page, E. S.: A Test for a Change in a Parameter Occurring at an Unknown Point, Biometrika, 41, 100–115, 1955.

Pollak, M.: Optimal Detection of a Change in Distribution. The Annal. Stat., 13, 206–227, 1985.

Pollak, M.: Average Run Lengths of an Optimal Method of Detecting a Change in Distribution. The Annal. Stat., 15, 749–779, 1987.

Ploberger, W., Kramer, W., and Alt, R.: A Modification of the CUSUM Test in the Linear Regression model with Lagged Dependent Variables, Empirical Econom., 2, 65–75, 1989.

Qiu, P. and Hawkins, D.: A Rank Based Multivariate CUSUM Procedure, Technometrics, 43, 120–132, 2001.

Ritov, Y.: Decision Theoretic Optimality of the CUSUM Procedure, The Annal. Stat., 18, 1464–1469, 1990.

Robbins, M. W., Lund, R. B., Gallagher, C. M., and Lu, Q.: Change-points in the North Atlantic Tropical Cyclone Record, J. Am. Statist. Assoc, 106, 89–99, 2011.

Roberts, S. W.: A Comaprison of Some Control Chart Procedures, Technometrics, 8, 411–430, 1966.

Severo, M. and Gama, J.: Change Detection with Kalman Filter and CUSUM, Lecture Notes in Computer Science, 6202/2010, 148–162, 2010.

Shewhart, W. A.: Economic Control of Quality of Manufactured Product, Van Nostrand, Princeton, 1931.

Shu, L., Yeung, H. and Jiang, W.: An Adaptive CUSUM Procedure for Signaling Process Variance Changes of Unknown Sizes, J. Qual. Technol., 42, 69–85, 2010.

Steiner, S. H., Cook, R. J., and Farewell, V. T.: Monitoring Paired Binary Surgical Outcomes Using Cumulative Sum Chart, Stat. Medicine, 18, 69–86, 1999.

Watkins, R. E., Eagleson, S., Veenendaal, B., Wright, G., and Plant, A. J.: Applying Cusum-Based Methods for the Detection of outbreaks of Ross River virus disease in Western Australia, BMC Medical Informatics and Decision Making, 8, doi:10.1186/1472-6947-8-37, 2008.

Wu, Z., Jiao, J. and Liu, Y.: A binomial CUSUM Chart for Detecting Large Shifts in Fraction Non Conforming, J. Appl. Stat., 35, 1267–1276, 2008.

Yashchin, E.: Performance of CUSUM Control Schemes for Serially Correlated Observations, Technometrics, 35, 37–52, 1993.