Interactive comment on "Instability and change detection in exponential families and generalized linear models, with a study of Atlantic tropical storms" *by* Y. Lu and S. Chatterjee

Y. Lu and S. Chatterjee

chatterjee@stat.umn.edu

Received and published: 26 July 2014

[11pt,letterpaper]article

Response to comments by referees
on
Instability and change detection in exponential families
and generalized linear models, with a study of Atlantic
tropical storms
Ying Lu and Snigdhansu Chatterjee

26 July 2014

We thank both reviewers for their very careful reading of our paper and their detailed and insightful comments. Below is our responses to their comments, itemized in the same order in which these comments were presented. We consider the comments of Referee#1 first, followed by the comments of Referee #2.

It is clear to us, while going over the comments of both referees and preparing our responses to them, that we need to make a major revision of our paper, which we are in the process of doing. The details are laid out in the following pages.

Referee #1 has two major concerns: (1) the assumption that all parameters other than the change-point are known, and (2) the lack of theory or simulations.

We understand and appreciate the referee's major concerns, and we believe we can address them. First, we assumed the parameters are known because of two main reasons: they made the mathematical formulation simpler and reduced the technical details considerably, and the traditional application domain of statistical process control where CUSUM and related techniques originated treat such parameters as known constants. Additionally, we may also consider the fact that under standard conditions the rate of convergence for the estimated change point to the true change point (if there is one) is faster than those of parameter estimates, consequently the asymptotics of parameter estimators can be fully de-linked from the asymptotics of the stability detection hypothesis test.

Nevertheless, the issue of parameter estimation (as opposed to considering them as known) is a very important one, and we now have the requisite methodology to incorporate the case of unknown but estimated parameters in our framework. **We will create a new subsection, in order to address this specific issue of the nature of the likelihood ratio stability test, when parameters are estimated.** We will make several comments in other parts of the manuscript, to address the differences between assuming parameters to be known, and estimating them. We hope that this process of simply adding in the "unknown but estimated parameters" cases would help remove this major concern.

In regard to the second major concern, we did not include any proof or simulation results in the first draft of the paper to save space. We will fully address this concern of the referee also, by doing exactly as the referee suggested, *i.e.*, by making them publicly available. We will also include a short sketch of one of the proofs in the actual paper to present the main idea.

<center>C351</center>

We respond to the specific points of the referee below:

- **Comment:** *The assumption that all the parameters other than τ are known seems unrealistic. In practice, the parameters of the exponential family distributions are typically unknown and are replaced by the corresponding estimates (such as MLEs). Similarly, in generalized linear models, the coefficients β are unknown and need to be replaced by suitable estimates (such as the estimates from estimating equation). The authors claimed that such extension is easy but did not provide any details or discussions. Based on my own experience, I believe that this is a nontrivial and important extension. Overall, the current setting seems rather narrow and may not be useful in real applications.*

  Response: We understand and appreciate the referee's concerns about this. As noted earlier, in the revision, we will include a new subsection specifically to address the case when parameters are estimated rather than presumed known. We will also include comments and discussion in various other sections of the paper, to address the estimated parameter case. Thus, we will directly address this comment, by building-in the parameter estimation issue into the paper.

- **Comment:** *The key theoretical results (e.g. Theorem 3.1, Theorem 3.3 and Theorem 4.1) in the paper are presented without providing any mathematical arguments. The absence of the technical details bothers me (the authors should at least make them publicly available). I also notice that some of the theoretical results are not rigorously presented. For example, in Theorem 3.3, it is unclear to me that whether the convergence holds in probability or almost surely and whether the convergence is uniform for all n.*

  Response: We will do as the referee suggests, that is, make the mathematical proofs available as part of an extended version of this paper in a publicly accessible domain. We had taken out the proof in order to keep the paper at a reasonable length, but putting the proofs back for a publicly available version is

<center>C352</center>

not a problem. We will also take care to make the statements and proofs of the mathematical results rigorous, in particular the Theorem that the referee mentions.

- **Comment:** *I am confused with the choice of L and the rationale behind it. Also I wonder how the value of ARL_0 can be determined in practice. It would be better to provide some discussions on this point.*

  Response: We will add a discussion on the rationale behind L, and on how ARL_0 is chosen. Essentially, L is the standard critical value that is used to compare the test statistic with in a hypothesis testing problem, and ARL_0 is related to the probability of Type-1 error. Thus, these quantities are versions of quantities that arise in the standard hypothesis testing protocol. We use L and ARL_0 in place of a standard critical value and level of a test just because of the historic relation of the problem being addressed in this paper (hypothesis testing for stability) to the literature on process monitoring and change detection. We will elaborate on this in the discussion that we will add in the paper in response to this comment. We thank the referee for bringing up this important issue.

- **Comment:** *The extension to the generalized linear model seems useless. The proposed test is infeasible as the coefficients β are in fact unknown.*

  Response: We understand the referee's concern. As mentioned earlier, in the revision, we will explicitly add-in the case where parameters are estimated. We will include a mathematical result to address the case when the parameter β is estimated.

- **Comment:** *In the data analysis, the unknown parameters are replaced by their estimates. I doubt that the proposed method is valid if the estimation effect is not taken into account.*

  Response: We understand the referee's concern. The validity of the proposed methods when the parameters are estimated requires a proof, which will be

C353

added in the longer, publicly available version of the paper that contains proofs. We naturally require additional technical conditions, and some involved algebra that was omitted from the first version of this paper. We will report these in the publicly available version. The reason the proposed techniques work is briefly as follows: using some amount of algebra and properties of exponential family distributions, we can show that the parameter estimates converge to the true unknown parameter values asymptotically. Since the sample sizes we consider are not particularly small, the difference between working with estimated parameters and true parameter values is not very large in many cases. Since the hypothesis testing for stability detection is a computationally challenging problem, we need to use numeric tools to obtain L and various properties of the test, both when parameters are known and when they are estimated. The latter case corresponds to parametric bootstrap, which is an effective tool in this situation.

- **Comment:** *If the underlying process is stable, how δ (which measures the magnitude of change) can actually be estimated as its true value is zero (the authors set $\delta = c\hat{\sigma}$ in the data analysis, which is ad hoc).*

  Response: The referee has raised an important point, which will require some additional discussion that we will inset in the revised manuscript. We describe δ in terms of multiples of σ since that is traditionally used, and since it makes sense to describe the distance between the null and alternative scenarios in terms of "units of standard deviation". Second, in samples of finite sizes, the only scenario where we get reasonable power in hypothesis tests is when the two hypotheses are sufficiently apart. Also 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, we decided to test hypotheses that are a reasonable number of standard deviation units away from each other. These observations hold even when parameters are estimated. We will include a discussion on this rather important issue in the revision.

C354

We thank the referee for observing that title of our paper is good and understandable for a wide audience, and the topic is interesting.

We thank the referee for bringing to our attention that we may need to discuss what is innovative in this paper. The new and innovative part of our paper is the fact that unlike existing methods for *detection* or *identification* of a change point, our formulation is that of a hypothesis test. This implies that 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, (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. In the revised manuscript, we will remark on these features.

We also note the referee's observations on Figures 1 and 2, and will strive to improve them as needed.

We respond to the specific comments of the referee in the itemized list below.

- **Comment:** *(a) With regard to the application of the EF-CUSUM on the Atlantic tropical storm, the authors apply the Poisson-distribution in the analysis, but they do not "prove" or show that the time series follows a Poisson-distribution. Why not include a figure which shows the frequency distribution of the time series they are focusing on and the authors can as well include a test which shows that the observations follow a Poisson-distribution or process – or at least show that the phenomenon is not normally distributed. Later in the analysis they argue that they detect changes in the parameter. Why not make a figure which show how lambda (the Poisson-lambda) varies over time and also plot the CUSUM.*

  Response: We thank the referee for this important observation. In the revised

manuscript, we will include the desired plots. We agree with the referee that these plots would improve the paper.

- **Comment:** *The authors could have described the time series according to whether the processes are autoregressive (AR) or moving average (MA).*

  Response: We understand the referee's concern that the data series may be correlated over time. Note however, we are dealing with *count* data, for which the standard Gaussian autoregression or moving average model would not be appropriate, partially for the same kind of reasons addressed in our manuscript. In particular, the variance and mean are related for a Poisson distribution. We noted however, that that we may consider a time-series version of a log-linear model. Considering that, we studied the autocorrelation and the partial autocorrelation function of the logarithms of the hurricane counts. These plots indicated a lack of dependency, justifying the data analysis.

- **Comment:** *(c) The authors do not discuss what kind of changes the time series are undergoing. Are we talking about an overall trend, short-term trends, short-term or long-term shifts in the mean and variances? Figure 2 shows clearly that volatility in the central pressure probably changes "significantly" short after the 1950s. The volatility is probably also changing for the maximum sustainable wind and the number of hurricanes. There is no discussion about these things which are important in a modeling context.*

  Response: This is an important observation. In addition to the referee's observations, we also note that the quality of data has not been uniform over time. The properties of the Poisson distribution imply that a change in the both the mean and variance. We will include a detailed discussion on the lines mentioned by the referee in the revision.

- **Comment:** *d) The method applied requires selection of average run length. The authors have chosen a fixed number 200. To me the choice of run-length is not*

clear.

Response: Both the referees have commented on this, and in the revised manuscript we will explain the process of selecting the average run length in detail. Note that the average run length is related to the probability of Type-1 error, and this governs the value that it is set at, similar to choosing the level of a test.

- **Comment:** *(e) The method applied requires estimate of the coefficients characterizing the distribution of the variable. In the empirical analysis the authors use the Poisson-distribution and base the lambda on the first observations, i.e. observation from 1850 to 1900 for the longest series and 1950 to 1970 for the short series. The estimates function as benchmarks. My impression is that these choices are ad-hoc and not supported by scientific arguments. They impose restrictions on the model/analysis. What will happen if they choose other time intervals as benchmarks?*

  Response: We appreciate this concern. However, we have considered other segments of time and have performed several checks around the different tuning parameter choices we have made.The results are largely invariant to such choices. We will discuss this point in the revised manuscript.

- **Comment:** *(f) Please look at table 6 and 7: the constants c = 1/4, 1/2 and 1 are unclear for me the role they play in 42 the analysis. I suggest the authors describe more clearly how they are included in the equations and the role they play.*

  Response: These constants quantify the degree of instability allowed by the alternative hypothesis, where we hypothesize a mean shift of $c\hat{\sigma}$ units at time τ. The tables the referee refers to show that the change point is detected and is significant for various choices of alternative hypothesis. This suggests that there

is strong possibility that the detected change is not an artefact of the method used here, but a feature of the data. We will discuss this in the manuscript.

- **Comment:** *(g) In the abstract the authors say: "We derive the related likelihood ratio test statistic". In the analysis of change-points I cannot see any statistical tests of the potential breakpoints the authors argue that they have identified. Where are the critical values and the estimates? I suggest they show more clearly how critical test-values are derived and applied.*

  Response: The critical value is reflected in L, for example as discussed in Theorem 3.1. We note the referee's concern, and will clarify these details in the revision.

- **Comment:** *(h) Please, look at pp. 392-383, Example 3.0.1, points 1 to 5: The authors derive the CUSUM statistic conditioned on the properties of the variance-covariance matrix. According to what I can see the authors do not use these derivations in the empirical part of the analysis. I'm therefore not certain what role or implication these deductions on pp 382-383 have in the paper.*

  Response: We understand the referee's point. The part of the manuscript that the referee points out is an example of the very Theorem 3.1 that precedes it. One important special case is the multivariate normal distribution where that theorem is applicable, and where some additional simplification is possible. The example lays out the different cases of the multivariate normal distribution and how Theorem 3.1 applies to these cases. It is true that we do not illustrate this important special case due to space limitations, but we feel that the results themselves are of independent interest, and may be useful to other researchers.

- **Comment:** *(i) As far as I understand, the method derived in the paper is based on the condition that the variables applied in the analysis independently distributed which means that the observations are not correlated over time. Application of the Poisson-distribution requires that the observations are not correlated in time*

(Please, see first and second line under section 3 "Distributional stability in exponential families", pp 377-378. In the empirical analysis the authors do not say anything about the whether the observations are correlated or not.

Response: We appreciate the referee's concern. However, as mentioned in an earlier response above, there is no evidence that the observations are dependent. We will remark on this in the revision. Also note that our method has a natural extension to time series and other dependent data with potential change points, for which a likelihood can be written and computed.

- **Comment:** *(j) The authors do not discuss or present in the concluding/summary section any critical remarks on the method. What is the strength and what are the weak parts of the method which challenge the validity?*

  Response: The referee has raised an important point here, and we thank them for bringing this to our notice. We will include appropriate comments on this in the revised manuscript.