HIGHER CRITICISM STATISTIC: THEORY AND APPLICATIONS IN NON-GAUSSIAN DETECTION

J. JIN

Statistics Department, Purdue University, 150 N. University Street, West Lafayette, IN 47907, USA
E-mail: jinj@stat.purdue.edu

Higher Criticism is a statistic recently proposed by Donoho and Jin[5]. It has been shown to be effective in resolving a very subtle testing problem: whether n normal means are all zero versus a small fraction is nonzero. Higher Criticism is also useful for non-Gaussian detection in Cosmic Microwave Background (CMB) data. In this report, we review the theory developed in Donoho and Jin[5] and discuss the use of Higher Criticism for two settings: detecting the non-Gaussian component in a superposed image of CMB and cosmic strings (CS), and detecting non-Gaussianity in the WMAP first year data.

1. Introduction

The Cosmic Microwave Background (CMB) is the relic radiation emitted when the universe was about $380,000$ years old. It is an almost perfect black body at a temperature of ≈ 2.726 Kelvin. The Standard Inflation model predicts that temperature anisotropies of the CMB (i.e. small angular fluctuations of the temperature) are the imprint of the initial density perturbations which gave rise to the large scale galaxies we see today. The study of the CMB is expected to improve our understanding of the very early universe, and it is of great interest to cosmologists.

The standard Inflationary model predicts that temperature anisotropies in the CMB have a Gaussian distribution. However, many other models (e.g. multi-field inflation[2], super string and topological defects[6, 7, 10]) as well as secondary effects (inverse Compton scattering etc.) predict deviations from a Gaussian distribution. The goal of non-Gaussian detection is to disentangle different non-Gaussian sources from one another.

The wavelet transformation is a powerful approach for non-Gaussian detection, and many wavelet-based methods have been investigated (see page 3 in Jin[8] for references to these works). Particularly, it was shown in Aghanim *et al* and P. Viela *et al*[1, 12] that the excess kurtosis of the wavelet coefficients outperformed all other methods.

However, the effectiveness of a detection tool depends highly on the underlying non-Gaussianities: a detection tool can be sensitive to some types of non-Gaussianities, but totally immune to other types. It is thus of interest to introduce more statistical tools to this field, and to compare their strengths as well as weaknesses. Higher Criticism is one of these new tools.

2. Higher Criticism

Higher Criticism (HC) was first proposed in Donoho and Jin[5] for a multiple comparison setting, where it was shown to be effective in resolving a very subtle testing problem: whether n normal means are all zero versus a small fraction of them being nonzero. Higher Criticism can also be viewed as a goodness-of-fit measure, and a tool for non-Gaussian detection.

Consider a setting in which we have n observations $\{X_i\}_{i=1}^n$. The problem of non-Gaussian detection is to test the following hypothesis: H_0 : $X_i \stackrel{iid}{\sim} N(0,1)$, where for simplicity we assumed that the data are standardized. To implement Higher Criticism[5, 8, 3], we first obtain individual p-values: $p_i = P\{N(0,1) \geq X_i\}$, we then sort them in ascending order $p_{(1)} < p_{(2)} < \ldots < p_{(n)}$, and calculate the normalized z-scores:

$$HC_{n,i} = \sqrt{n} \cdot [|i/n - p_{(i)}|]/[\sqrt{p_{(i)}(1-p_{(i)})}].$$

The Higher Criticism statistic is then defined as $HC_n^* = \max_{\{1 \leq i \leq n\}} HC_{n,i}$.

The rationale behind the normalization is that, when the hypothesis H_0 is indeed true, then for almost all i (except when i is close to 1 or n), $HC_{n,i} \approx N(0,1)$, and moreover $HC_n^* \approx \sqrt{2 \log\log n}$. Thus a large HC_n^* value implies non-Gaussianity.

2.1. *Sensitive to Unusually Large Amount of Moderate Significances*

In a data set, the extreme value refers to the data point which is largest in absolute value. It is a well-known result in statistics that out of n samples from the standard Gaussian the extreme value $\approx \sqrt{2\log n}$. In contrast, moderate significances refer to the tiny portion of the data points that are slightly smaller (in absolute values) than the extreme value, e.g. data points $\approx \sqrt{\log n}$. The proportion of moderate significances is very small, e.g. $P\{N(0,1) \geq \sqrt{\log n}\}$, the proportion of samples $\geq \sqrt{\log n}$ approximately equals to $n^{-1/2}$.

In Donoho and Jin[5], the authors have considered a sparse normal mean problem: we have n observations from $X_i \sim N(\mu_i, 1)$, with all $\mu_i = 0$ except a possible tiny fraction ϵ_n of them satisfying $\mu_i = \mu_n$, where ϵ_n and μ_n depend on n but not on i. The goal is to test whether the sparse mean effect is present or not, or equivalently to test whether $\epsilon_n = 0$ or $\epsilon_n > 0$. They considered a range of (ϵ_n, μ_n) which concerns the situation of "very sparse signal with moderate significant amplitude": on one hand, ϵ_n is too small so that the sparse mean effect can't be detected by statistics based on moments (cumulants, kurtosis, etc.); on the other hand, as the signals are only of moderately significant, the sparse mean effect can't be detected by merely looking at the extreme values.

It was proved in Donoho and Jin[5] that the Higher Criticism statistic is optimally adaptive in detecting the sparse normal mean effect. Roughly put, for fixed ϵ_n, whenever μ_n is large enough so that it is possible to reliably tell that $\epsilon_n > 0$, the Higher Criticism statistic is able to do so.

We now take a heuristic approach for understanding the mechanism of Higher Criticism. The sparse mean effect can be thought of as the situation in which one has n samples from the standard Gaussian, and now you want to sneak in a bunch of μ_n by the following two steps: (a). randomly select a tiny portion of the samples, leave others untouched, and (b). add μ_n to each selected samples. The problem is then to tell whether such a process has occured or not. Higher Criticism works by picking a sequence of significance levels and asking whether there are too many samples found above each significance level. If the answers are all "no", the Higher Criticism claims Gaussian and nothing is found, but claims non-Gaussian otherwise. Higher Criticism uses the normalized z-score for deciding whether there are too many samples found above each significance level or not: $HC_{n,\alpha} = \sqrt{n}[\{\text{Fraction at Level } \alpha\} - \alpha]/\sqrt{\alpha(1-\alpha)}$; when all samples are truly from the standard Gaussian, $HC_{n,\alpha} \approx N(0,1)$ and should be relatively small, so a large $HC_{n,\alpha}$ implies non-Gaussianity. Thus Higher Criticis works across the full range of significance levels, looking for evidence against possible types of "sneak-in" we mentioned above.

We now come back to the sparse normal mean problem. The most strong evidence for the presence of the sparse mean effect is that when you look at the portion of data points of moderate significance, there are too many moderate significances than there would be if the null hypothesis is true (i.e. all samples are truly from the standard Gaussian). Higher Criticism immediately reports a very large normalized z-score and rejects the null. This property of Higher Criticism is sensitive to an unusually large amount of moderate significances.

2.2. *Useful for Locating nonGaussianity*

The previous section pointed out that Higher Criticism is useful for locating the non-Gaussianity. To illustrate this point, suppose Higher Criticism picks all levels from 0 to 1 with 1% increment. Suppose the answers at levels $\alpha = 10\%$ and $\alpha = 9\%$ are "yes", while those at other levels are "no". Then on the one hand we are told that too many samples are observed at Level 10%. On the other hand we are told that *not* too many samples are observed at Level 8%. We then conclude that there are too many samples that fall between Level 8% and Level 10%, and this *slice* of data is suspected of non-Gaussianity. Notice here that the extreme values don't have to be more "non-Gaussian".

3. Detecting Cosmic Strings

We have considered using the Higher Criticism statistic for detecting cosmic strings (CS). Refer to Jin[8] for detailed discussion of the following results.

We consider a setting in which we have a superposed image of a simulated map of CMB and CS: $Y = \sqrt{1-\lambda}CMB + \sqrt{\lambda}CS$, and we are interested in testing whether $\lambda = 0$ or not. The simulated map of

CS was kindly provided by F. Bouchet. Though the real map of the CMB exists (from WMAP), we used simulated CMB maps instead in order to ensure that there is no non-Gaussianity in the maps.

Since the pixel values of the simulated CMB are correlated, working in the frequency domain is more convenient than in the space domain. Let $\{X_i\}_{i=1}^n$ be the wavelet coefficients of Y, then $X_i = \sqrt{1-\lambda} z_i + \sqrt{\lambda} w_i$, where $z_i \stackrel{iid}{\sim} N(0,1)$ are the transform coefficients of CMB and $w_i \stackrel{iid}{\sim} W$ are the coefficients of the CS map. The distribution of W is unknown, but is symmetrical and has a heavy-tail. Without loss of generality, both $\{z_i\}_{i=1}^n$ and $\{w_i\}_{i=1}^n$ have been standardized with standard deviations equal to 1. The testing problem is then equivalent to testing a null hypothesis H_0 under which $X_i \stackrel{iid}{\sim} N(0,1)$ versus an alternative hypothesis $H_1^{(n)}$ under which $X_i = \sqrt{1-\lambda} z_i + \sqrt{\lambda} w_i$. We are interested in which pair (λ, W) do the two hypotheses asymptotically merge together so that no test can separate them, versus which pair of (λ, W) the two hypotheses asymptotically separate from each other. By saying asymptotically, we mean n tends to ∞.

Clearly, if we fix $\lambda > 0$, then when n gets larger and larger, the difference between the two hypotheses becomes increasingly large, and eventually it is trivial to tell one from another. Thus the interesting range for λ is that it tends to 0 as n tends to ∞, so we set $\lambda = \lambda_n = n^{-r}$, $0 < r < 1$. At the same time, motivated by the heavy-tailed behavior of W, we assume that the tail probability of W decays algebraically:

$$\lim_{x \to \infty} x^{\alpha} P\{|W| \geq x\} = C_{\alpha},$$

where C_{α} is a constant.

Intuitively, as λ_n is algebraically small, we expect that the majority of relatively smaller samples from W will not have much influence on testing. Instead, a tiny fraction of very large samples from W would play the decisive role. This turns out to be true, and there is a *threshold effect* for the testing problem. We call the curve $r = \rho^*(\alpha)$ in the α-r plane the *detection boundary*: if (r, α) falls below the detection boundary, then the null and the alternative hypothesis separate asymptotically; if (r, α) falls above the detection boundary, the null and the alternative merge asymptotically. It turns out that $\rho^*(\alpha) = 2/\alpha$ when $\alpha < 8$ and $1/4$ otherwise.

We now compare the asymptotical performance of the excess kurtosis and Higher Criticism. If $\alpha > 8$ or the 8-th moment of W exists, then the excess kurtosis is better than Higher Criticism. When (r, α) falls into the region that $\{(r, \alpha) : \alpha > 8, \frac{2}{\alpha} < r < 1/4\}$, then asymptotically the excess kurtosis has full power for detection, while the power of Higher Criticism tends to 0. If on the other hand $\alpha < 8$, then Higher Criticism is better than the excess kurtosis. When (r, α) falls into the region $\{(r, \alpha) : \alpha < 8, 1/4 < r < \frac{2}{\alpha}\}$, then asymptotically Higher Criticism will have full power, while the power of the excess kurtosis tends to 0.

The phenomenon can be explained as follows. Take $\alpha = 5$ for example. When you look at the data, before you notice any difference in the excess kurtosis, the largest sample from W is quite apparent, so detectors concentrated on the tail are more sensitive. However, when α ranges between 5 and ∞, the tail is gradually thinned out, and at some point, it will not tell you anything by merely looking at the data tail. You need to shift your attention to relatively smaller samples, or the bulk of the data, for which the excess kurtosis is more sensitive. It is interesting to study the α parameter corresponding to the tail behavior of W. Our study[8] supports the assumption that W has a power law tail: implementing the Hill estimator[9] gives $\alpha \approx 6.1$, where the standard error of this estimate approximately equals to 0.9.

Finally, the above result is highly asymptotical. It would be interesting to investigate the performances for moderately large n. Reports in this direction are included in Jin[8].

4. WMAP First Year Data

We have implemented Higher Criticism to analyze the WMAP first year data. The detailed study is in Cayon *et al*[3]. We work with the WMAP data from the LAMBDA website (*lambda.gsfc.nasa.gov*). We construct a weighted combination of released foreground cleaned Intensity Maps at bands Q, V, and W (refer to Cayon *et al*[3] for details). We then generated 5,000 Gaussian simulations of CMB maps (including observational constraints imposed by noise and beam profiles), and take the wavelet transforms for each simulated map as well as the WMAP map. Finally, we carry out the statistical analysis on the wavelet coefficients.

The wavelet basis we used is the Spherical Mexican Hat. The wavelet coefficients we obtained do not fit very well with *iid* Gaussian samples. One reason is that the wavelet basis is not orthogonal. Despite this, Higher Criticism can still be used as a criterion for non-Gaussianity: it can be thought that the larger the Higher Criticism value, the larger the deviation from Gaussianity. We thus take the approach in which we compare the Higher Criticism values of the $5,000$ simulated maps with that of the WMAP data, and claim non-Gaussianity if 99% of the simulated CMB maps have a smaller Higher Criticism value than that corresponding to the WMAP. It would be interesting to try the analysis with some orthogonal basis; we leave this for future study.

In addition to Higher Criticism, we have also implemented the excess kurtosis to the above setting. The Higher Criticism reports non-Gaussian detection at 99.46%. In comparison, the excess kurtosis is slightly better by reporting non-Gaussian detection at 99.7%.

However, Higher Criticism has more to offer. We pointed out earlier in the report that Higher Criticism can be used to automatically identify a tiny fraction of data as suspected of non-Gaussianity. We isolated 490 wavelet coefficients, at the scale of 5 degrees of the WMAP data. In detail, we set a threshold t_0 as the 1%-upper percentile of the $5,000$ Higher Criticism values (HC_n^*) based on simulated CMB maps. Then out of all wavelet coefficients of the WMAP, we select those with an associated normalized z-score ($HC_{n,i}$) larger than t_0.

Last, we map these 490 wavelet coefficients back to pixels in the WMAP. There are two ways to do the mapping. In the first way, we map each coefficient back to all pixels involving the coefficients, i.e. all pixels convoluted with the wavelet basis when calculating this coefficient. Notice that each of the coefficients naturally maps back to a cluster of pixels. It is interesting to note that the 490 pixels, and those correlated with them by the wavelet convolution, are at the cold spot found by Vielva et al.[13] and Cruz et al.[4]. In the second way, we map each coefficient back to only one pixel: the one at the center of the pixel-cluster mentioned above. This way of mapping has the advantage of visualization. By the second way, the selected 490 coefficients map back to a ring on the outer part of the cold spot. Interestingly, the "coldest" wavelet coefficient (i.e. largest in absolute value but is negative) maps back to a pixel in the center part of the cold spot, which is not in the ring. We clarify here that, both in this report and in Cayon *et al*[3], our result doesn't attempt to conclude that there is a ring structure in the WMAP map. Instead, the ring visualizes the position of pixels corresponding to the 490 moderately significant wavelet coefficients we extracted.

5. Conclusions

We introduced the Higher Criticism statistic for non-Gaussian detection. We have studied the application of Higher Criticism to the detection of cosmic strings and to the WMAP first year data. Higher Criticism is useful in applications by adding discussions to the field of non-Gaussian detection.

Acknowledgments

We would like to thank Louis Lyons, the anonymous referee, Laura Cayon, Jean-Luc Starck, and Anna Treaster for valuable discussions. We acknowledge the use of the Legacy Archive for Microwave Background Data Analysis (LAMBDA). Support for LAMBDA is provided by the NASA Office of Space Science. This work was partially supported by NSF Grant DMS-0505423.

References

1. N. Aghanim *et al.*, *Astronomy and Astrophysics* **406**, 797 (2003)
2. F. Benardeau and J. Uzan, *Phys. Rev. D* **66**, 103506 (2002).
3. L. Cayon *et al.*, *Mon. Not. Roy. Astron. Soc.* **362**, 826 (2005).
4. M. Cruz *et al.*, *Mon. Not. Roy. Astron. Soc.* **356**, 29 (2005).
5. D. Donoho and J. Jin, *Ann. Statist.* **32,3**, 962 (2004).
6. A. Gaugui *et al.*, *Astrophysical Journal* **430**, 447 (1994).
7. A. H. Jaffe, *Phys. Rev. D* **49**, 3893 (1994).
8. J. Jin *et al.*, *to appear in EURASIP Journal on Applied Signal Processing* , (2004).
9. B.M.Hill, *Ann. Statist.* **3**, 1163 (1975).
10. X. Luo, *Astrophysical Journal Letter* **427**, L71 (1994).
11. G.R. Shorack and J.A.Wellner, *Empirical Processes with Applications to Statistics, John Wiley & Sons* , (1986).
12. J. Starck *et al.*, *Astronomy and Astrophysics* **416**, 9 (2004).
13. P. Viela *et al.*, *Astrophysical Journal* **609**, 22 (2004).