A nonparametric method to evaluate significance of events in search for gravitational waves with false discovery rate

Hirotaka Yuzurihara
Institute for Cosmic Ray Research, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8582, Japan
E-mail: yuzu@icrr.u-tokyo.ac.jp

Shuhei Mano
The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

Hideyuki Tagoshi
Institute for Cosmic Ray Research, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8582, Japan

Abstract. During recent observation runs by LIGO and Virgo, tens of gravitational wave events have been observed. The gravitational wave network of LIGO/Virgo/KAGRA is planning longer observation runs with improved sensitivity in the future. In this situation, it is important to evaluate the significance of each event with reliable methods since we expect, as the observation period becomes longer and as the sensitivity of the detector improves, we expect to detect more false positive events with a given threshold for the false alarm rate. In this paper, we introduce \( q \)-value which was first introduced by Storey\[1\], and is interpreted as False Discovery Rate (FDR). We introduce an algorithm to evaluate FDR in a non-parametric way, and apply it to publicly available results of the analysis of the advanced LIGO O1 data, and find that we can obtain a better significance for GW151012 which has the lowest significance in the O1 events. For other events, we obtain consistent results already obtained by LIGO and Virgo.

1. Motivation
After the first observation of gravitational waves, GW150914, in 2015, 10 gravitational wave events from binary black hole coalescences have been observed. The gravitational waves from the coalescence of binary neutron stars were also observed \[2\], which was observed coincidently observed with a gamma-ray telescope. Electromagnetic follow-up observations were successful, and an optical counterpart was discovered. This event opened the era of gravitational wave astronomy.

In compact binary coalescence searches, we search for gravitational wave signals by maximizing the detection statistic over the template bank in a short time window. We obtain one trigger from each time window (for more details, see \[4\]). Accordingly, for a given threshold,
as the observing time and the template bank becomes larger, the probability that we obtain false triggers by noise (false alarm probability) becomes larger. This is called the family wise error. When the candidate events are selected only by the threshold from a large amount of the data, we need to care about this problem. There are methods to control the false alarm probability as we require. The Bonferroni correction is one of the methods (see Chapter 9 of [5]). However, these methods generally reduce the detection probability while we can control the false alarm probability.

The false discovery rate (FDR) can treat the family wise error [6, 1]. In this proceeding, we present a consistent procedure to assess the significance of candidate events by using FDR. We propose a procedure to evaluate the FDR of each event by extending the procedure proposed by Storey and Tibshirani [1]. We apply this procedure to the results of the analysis of O1 data set computed by Nitz et al. [7], and re-evaluate the significance of each candidate event. The main advantage of our procedure is that our procedure is completely nonparametric, namely, we do not assume any parametric model behind data.

2. False discovery rate

The false discovery rate is based on the p-value of the event which is defined by the detection statistic \( \rho \) and the background distribution as

\[
\hat{p}_{\text{trigger}}(\rho) = \frac{n_{\text{bg}}(\rho)}{N_{\text{bg}}},
\]

where \( N_{\text{bg}} \) is the total number of noise events in the background distribution and \( n_{\text{bg}}(\rho) \) is the number of the noise events whose detection statistics are larger than \( \rho \) in the background distribution.

Benjamini and Hochberg [6] introduce the false discovery rate, which is defined as the expected value of \( F/S \), that is, \( E(F/S, S > 0) \), where \( F \) and \( T \) are the numbers of noise and signal events called significant by a statistical test, respectively, and \( S = F + T \) is the total number of events called significant. Although \( S \) is a known number in the statistical test, \( F \) and \( T \) are unknown numbers. To estimate FDR from the data, Storey and Tibshirani [1] proposed the \( q \)-value for a particular event. The definition of the \( q \)-value is the minimum FDR that can be attained when calling the event significant, namely,

\[
q_i := \min_{u \geq p_i} \text{FDR}(u),
\]

where threshold \( u \) where \( 0 < u \leq 1 \) and \( \text{FDR}(u) = E\left[\frac{F(u)}{S(u)}, S(u) > 0\right] \). Here, \( F(u) \) is the number of the noise events which p-value is smaller than or equals to the threshold \( u \), and \( S(u) \) is the total number of events which p-value is smaller than or equals to the threshold \( u \).

3. Algorithm to evaluate false discovery rate

In the search for compact binary coalescence, we propose the following algorithm to evaluate FDR of each event by extending the procedure proposed by Storey and Tibshirani [1]. Let \( m \) to be the number of false alarm rates which are less than some value. Assume p-values in the region around and larger than \( \hat{p}(m) \) are noises. The procedure is

(i) Compute estimated p-values as \( \hat{p}_i = (\text{false alarm rate of } i\text{-th event}) \times (\text{time window width}) \), where \( i = 1, \ldots, m \).
(ii) Let \( \hat{p}_1(1) \leq \hat{p}_2(2) \leq \cdots \leq \hat{p}_m(m) \) be the ordered p-values.
(iii) Set \( \pi_0 = 1 \) or \( \pi_0 = \hat{f}(\hat{p}_m) \), where \( \hat{f}(\lambda) \) is the natural cubic spline curve with three degree of freedom fitted to \( f(\lambda) = \frac{n_{\text{obs}} - \#(\hat{p} < \lambda)}{n_{\text{obs}}(\hat{p}_m(\lambda) - \lambda)} \), where \( 0 < \lambda < \hat{p}_m \).
(iv) Set \( \hat{q}(m) = \hat{\pi}_0 n_{\text{obs}} \hat{p}(m)/m. \)

(v) For \( i = m-1, m-2, \ldots, 1, \) compute

\[
\hat{q}(i) = \min \left( \frac{\hat{\pi}_0 n_{\text{obs}} \hat{p}(i)}{i}, \hat{q}(i+1) \right).
\]

(vi) The estimated \( q \)-values for the \( i \)-th most significant event is \( \hat{q}(i) \).

4. Result

We introduced an algorithm to evaluate \( q \)-value in a non-parametric way, and applied it to the results of the analysis of LIGO O1 data \([7]\) which is publically available at https://github.com/gwastro/1-ogc. Table 1 shows the result of \( q \)-values from the \( \text{bbh} \) data set. The top three events should be called significant, if we call events with \( q \)-value smaller than 0.05 significant. These results are consistent with the result in the recent catalog of gravitational waves \([7, 8]\). We also find that the \( q \)-value of the third event (GW151012) is \( 9.83 \times 10^{-5} \). If we compare it with \( 1 - P_{\text{astro}} = 0.024 \), we can say that GW151012 is significant with a better reliability. The detailed method and other results of our analysis are described in a separate paper \([9]\).

Table 1. Estimated \( p \)-values and \( q \)-values of the events of the \( \text{bbh} \) data set. Events are sorted by a false alarm rate and the top 6 events are shown. The inverse false alarm rates are obtained from the data set. \( p \)-values are computed by (1). \( q \)-values are computed by our algorithm. \( P_{\text{astro}} \) are the probability of the astrophysical origin (a complement of a false discovery rate) which are obtained from \([7]\).

| Designation (false alarm rate)\(^{-1} \) [yr] | \( p \)-value \( \times 10^{-13} \) | \( q \)-value \( \times 10^{-6} \) | \( 1 - P_{\text{astro}} \) |
|-----------------------------------------------|--------------------------------|--------------------------------|----------------|
| 150914+09:50:45UTC > 6.55 \times 10^4 | <4.84 \times 10^{-13} | <1.11 \times 10^{-6} | - |
| 151226+03:38:53UTC > 5.91 \times 10^4 | <5.36 \times 10^{-13} | <1.11 \times 10^{-6} | - |
| 151012+09:54:43UTC 447 | 7.10 \times 10^{-11} | 9.83 \times 10^{-5} | 0.024 |
| 160103+05:48:36UTC 0.396 | 8.00 \times 10^{-8} | 0.0831 | 0.939 |
| 151213+00:12:20UTC 0.309 | 1.03 \times 10^{-7} | 0.0853 | 0.953 |
| 151216+18:49:30UTC 0.106 | 2.98 \times 10^{-7} | 0.207 | 0.983 |

References

[1] Storey J.D., Tibshirani R. Statistical significance for genomewide studies. Proc. Natul. Acad. Sci. USA 100, 9440–9445 (2003)
[2] Abbott B.P. et al. (LIGO Scientific Collaboration, Virgo Collaboration). GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. Phys. Rev. Lett. 119, 161101 (2017)
[3] LIGO Scientific Collaboration et al. Multi-messenger observations of a binary neutron star merger. Astrophys. J. Lett. 848, L12 (59pp) (2017)
[4] Usman S.A. et al. The PyCBC search for gravitational waves from compact binary coalescence. Class. Quantum Grav. 33, 215004 (2016)
[5] Lehmann E.L., Romano J.P. Testing Statistical Hypotheses. 3rd. ed. Springer (2008)
[6] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B 57, 289–300 (1995)
[7] Nitz A.H. et al. I-OGC: The first open gravitational-wave catalog of binary mergers from analysis of public advanced LIGO data. arXiv:1811.01921v1
[8] The LIGO Scientific Collaboration, the Virgo Collaboration. GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs. arXiv: 1811.12907v2
[9] Yuzurihara H, Mano S, Tagoshi H. A nonparametric method to assess significance of events in search for gravitational waves with false discovery rate. arXiv:1907.00379