Estimation of background distribution in gravitational wave search

Hirotaka Yuzurihara (ICRR, UTokyo)
Hideyuki Tagoshi (ICRR, UTokyo)
Shuhei Mano (ISM)

The 5th KAGRA International Workshop
Feb. 15th, 2019
Perugia

Contents

- Motivation
  - gravitational wave events
  - false alarm rate from background distribution

- Alternative methods to estimate background distribution
  - generalized extreme value distribution
  - extrapolate the distribution by fitting
  - permutation test

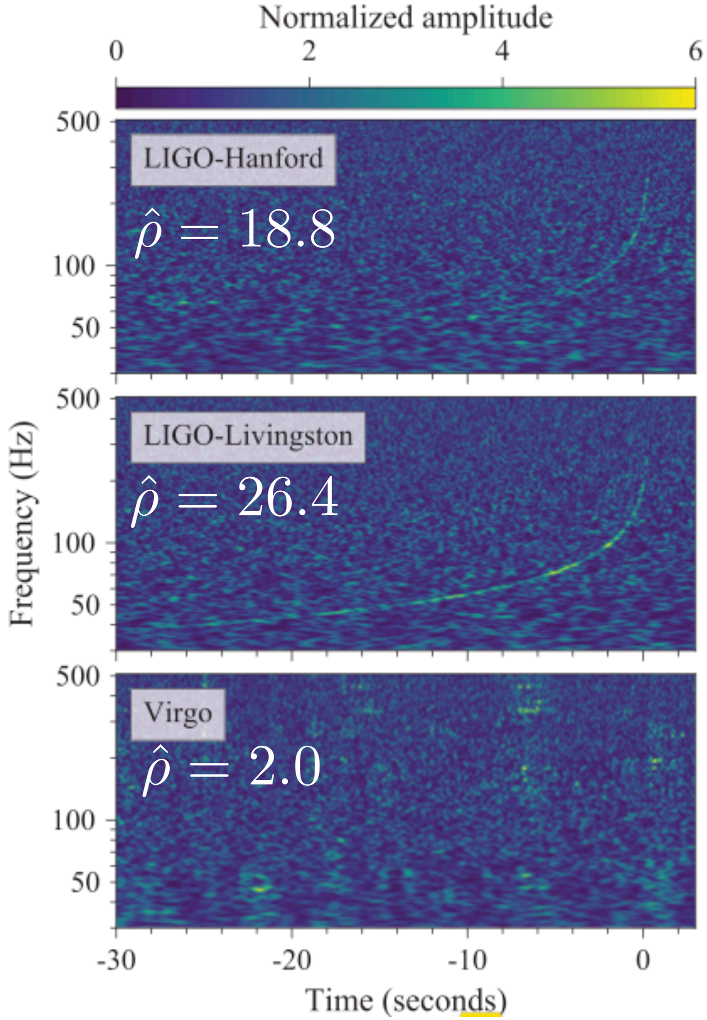
- Simulation
  - generate simulation Gauss noise data
  - analyze the data and estimate background distribution
- Summary

First detection of neutron star binary event GW170817

[PRL 119, 161101 (2017)]

- advanced LIGO and advanced Virgo detected the GW from binary neutron star at the same time.
  - 17 August 2017 12:41:04 UTC
  - total SNR : 32.4
  - distance : ~40Mpc

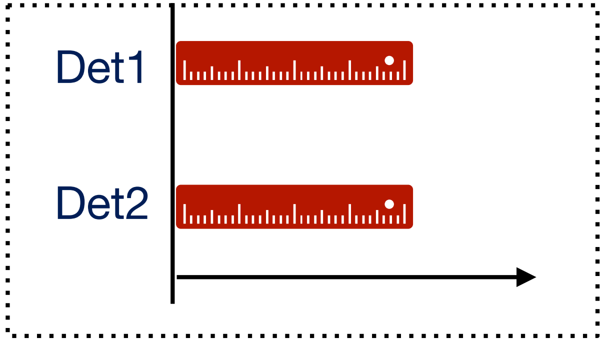
- The localization is restricted within 31 deg^2.
  → Follow-up observations
  → Short Gamma ray burst
  →Gravitational wave astronomy



- To claim the detection of the gravitational wave, we need to
  - compare with the background distribution of the detection statistics ρ
  - quantify how often the detection statistics ρ appear from the noise data (=false alarm rate).

Conventional method : time shift method

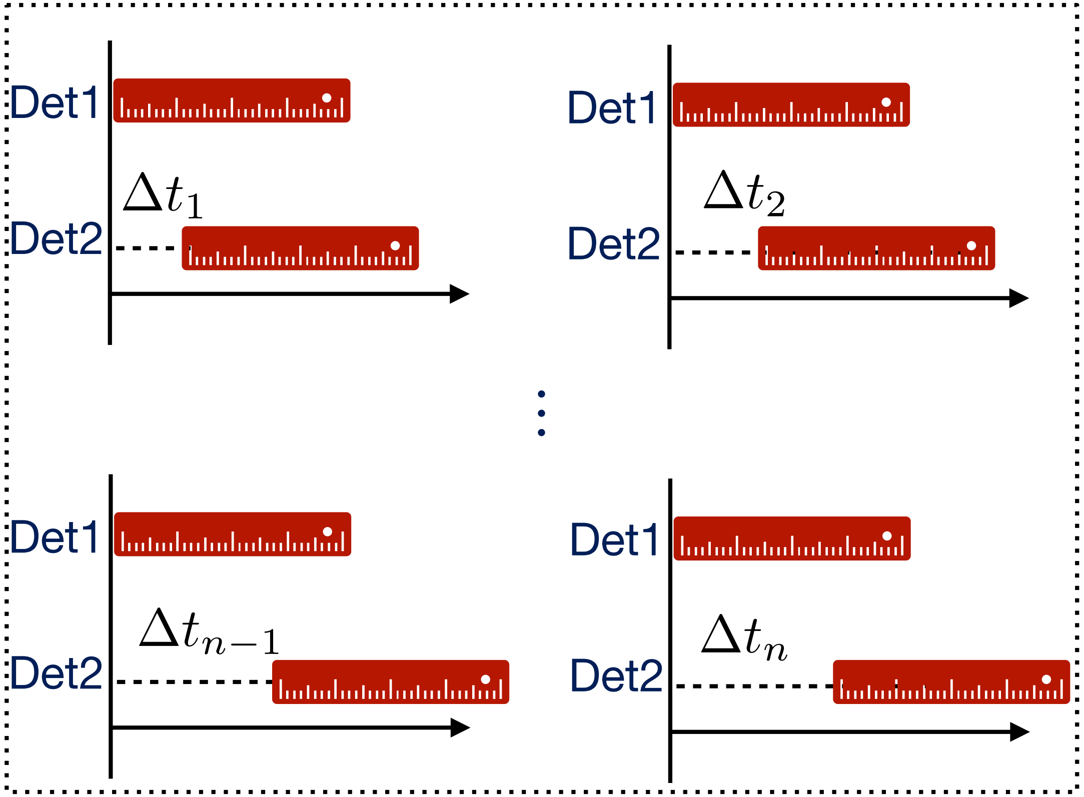
In pyCBC pipeline, time shift method is used to estimate background distribution



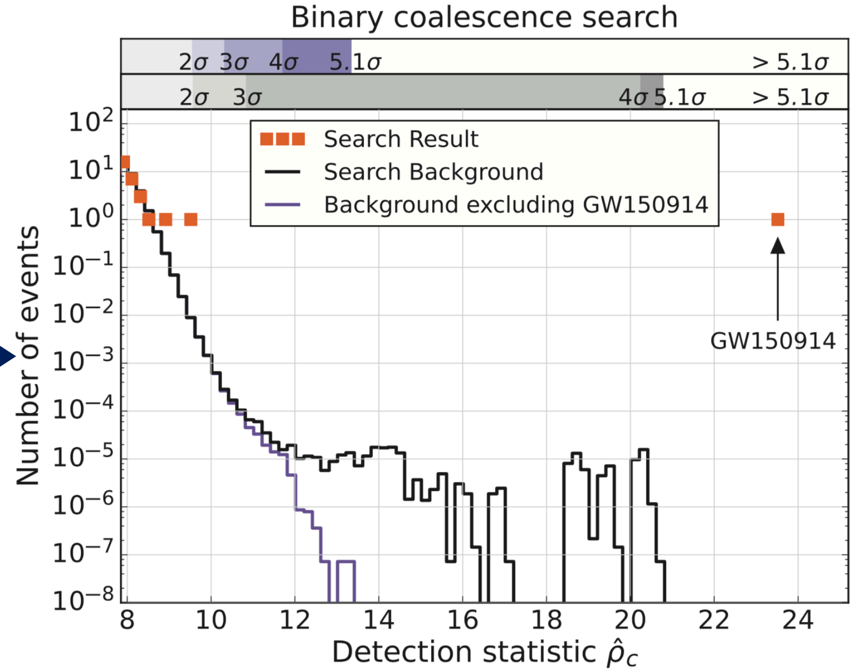
Procedure

1. Analyze data without time offset and find the trigger event
2. Analyze the data with time offset and
   repeat analysis with different Δt

pyCBC pipeline needs 5.2 days data to estimate background distribution.



[PRL 116, 061102 (2016)]



Weakness of the time shift method

- Although time shift method guarantee all the data don't have coincident signal, there are weaknesses;
  - We need at least two detectors to do time shift.
    → We can't use time shift method for single detector data.
  - We need huge computations to get close to low false alarm rate.
  - It is possible the noise condition changes during the data for time shift.

- We investigate two methods to estimate background distribution.
  - Fit the distribution by generalized extreme value distribution and extrapolate the distribution
  - Permutation test

Contents

- Motivation

  - gravitational wave events

  - false alarm rate from background distribution

- **Alternative methods to estimate background distribution**

  - generalized extreme value distribution

  - extrapolate the distribution by fitting

  - permutation test

- Simulation

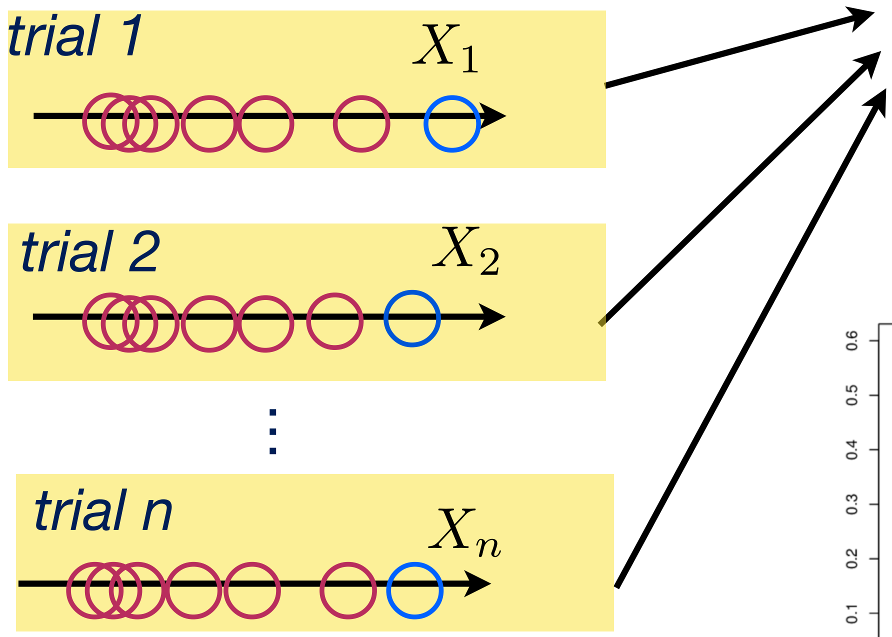
  - generate simulation Gauss noise data

  - analyze the data and estimate background distribution

- Summary

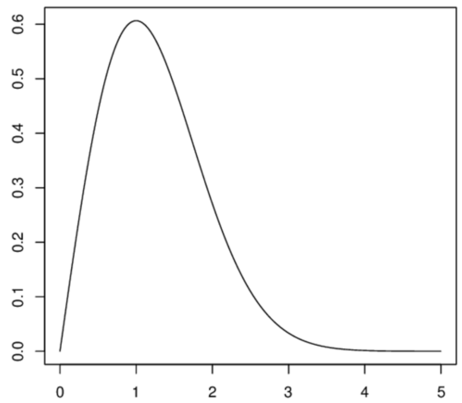
Generalized extreme value distribution

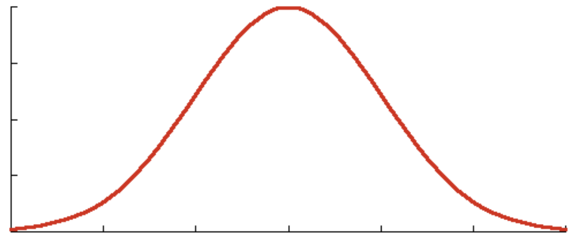
General example

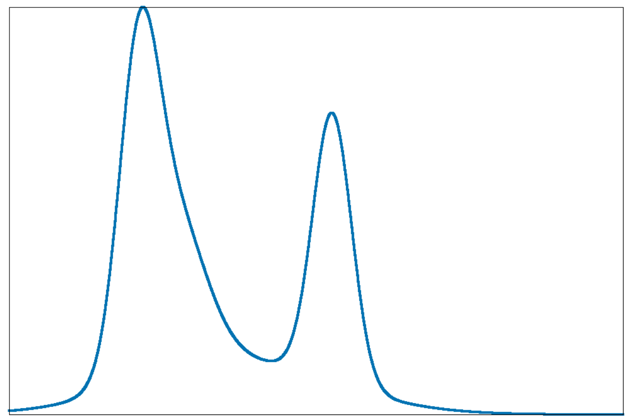


$X_1, X_2, \cdots, X_n$

are maximum value in each trial.
They follow the generalized extreme value (GEV) distribution,
regardless of the distribution of the original data.





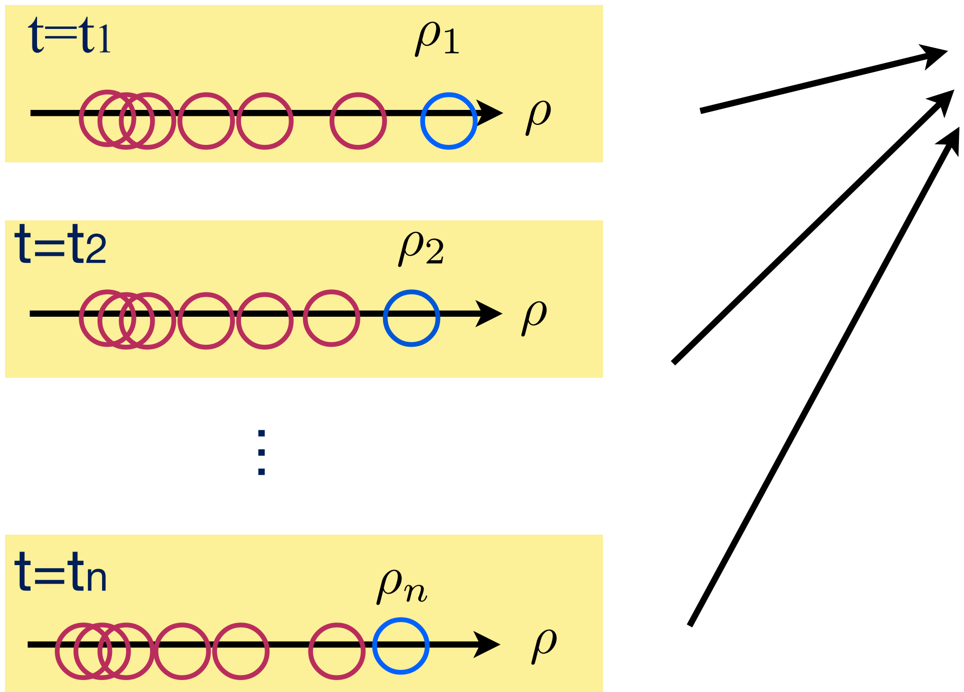


- GEV distribution is used to estimate
  - maximum of amount of rainfall in one year

    → How high dam we should build?
  - maximum of horse racing refund in one year

Generalized extreme value distribution

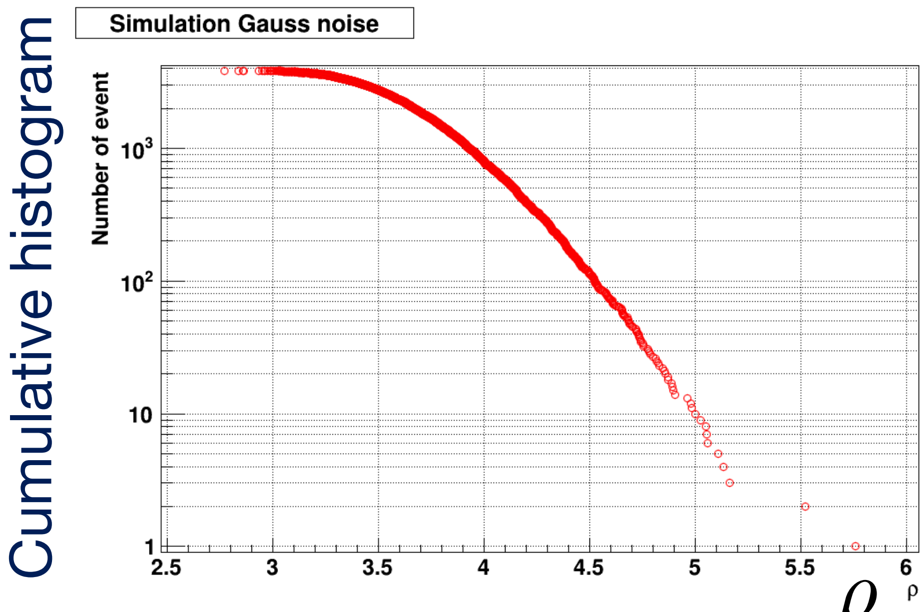
<u>Example in CBC search</u>



$\rho_1, \rho_2, \cdots, \rho_n$

are maximum value calculated by each template and each time window. They follow the generalized extreme value(GEV) distribution, **regardless of the distribution of the original data**.

- Maximized over short time window
- Maximized over templates



Estimation of background distribution

- We fit the cumulative distribution of ρ by using Generalized extreme value (GEV)distribution;
- Fitting function :

$$f_{GEV}(x, A, \theta, \mu, \gamma) = A\left(1 - \exp\left\{-\left[1 + \gamma\left(\frac{x-\mu}{\theta}\right)\right]^{-1/\gamma}\right\}\right)$$

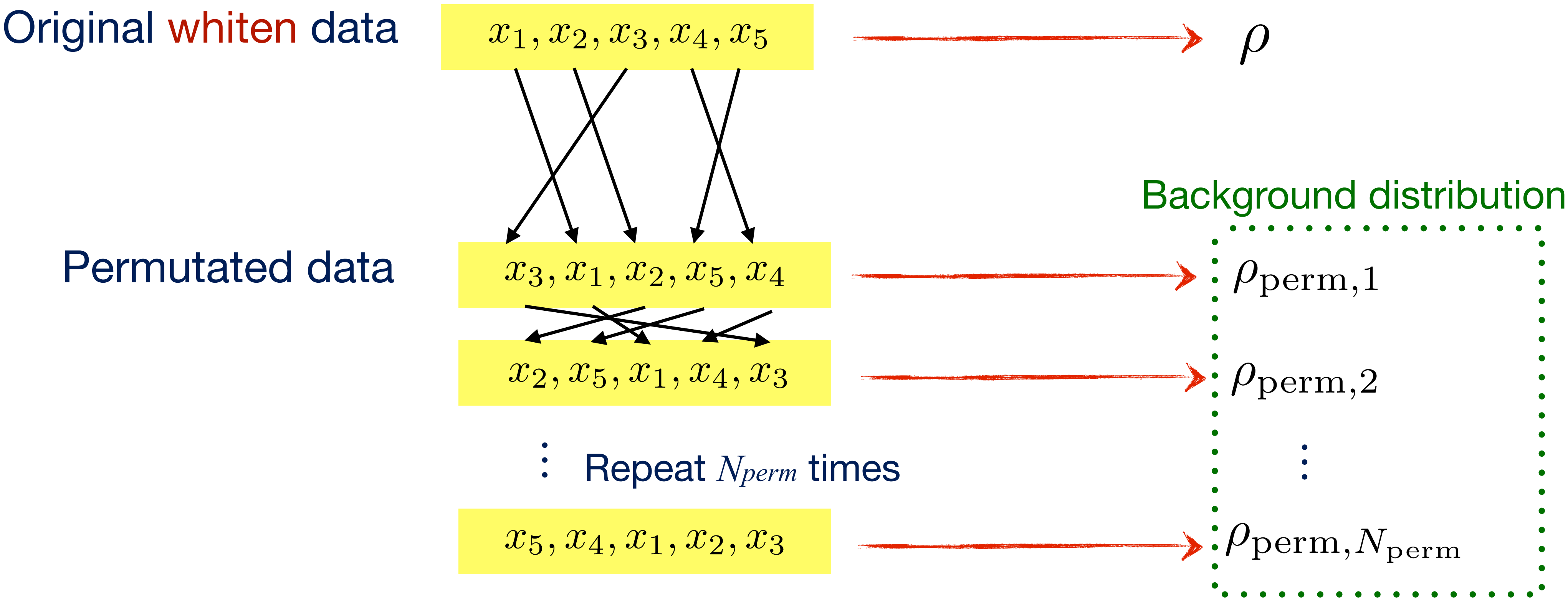
- Because we extrapolate cumulative distribution, we can calculate false alarm rate as $f_{GEV}(\rho, A, \theta, \mu, \gamma)$.

- Advantages of extrapolating by GEV distribution
  1. We can apply the single detector data.
  2. We don’t need additional calculation
  3. We don’t need long data.

Procedure of Permutation test

- Here we consider a simple example (n=5).

→ : Calculate the statistics (matched filter)



- We permutate the index of the data and analyzing the pemutated data.
  We can estimate the background distribution and the false alarm rate.

[Fisher, R.A. (1935) The design of experiments, Section 21.]
[Lehmann, E.L. and Romano, J.P. (2005) Testing statistical hypotheses]

Example of permutation test

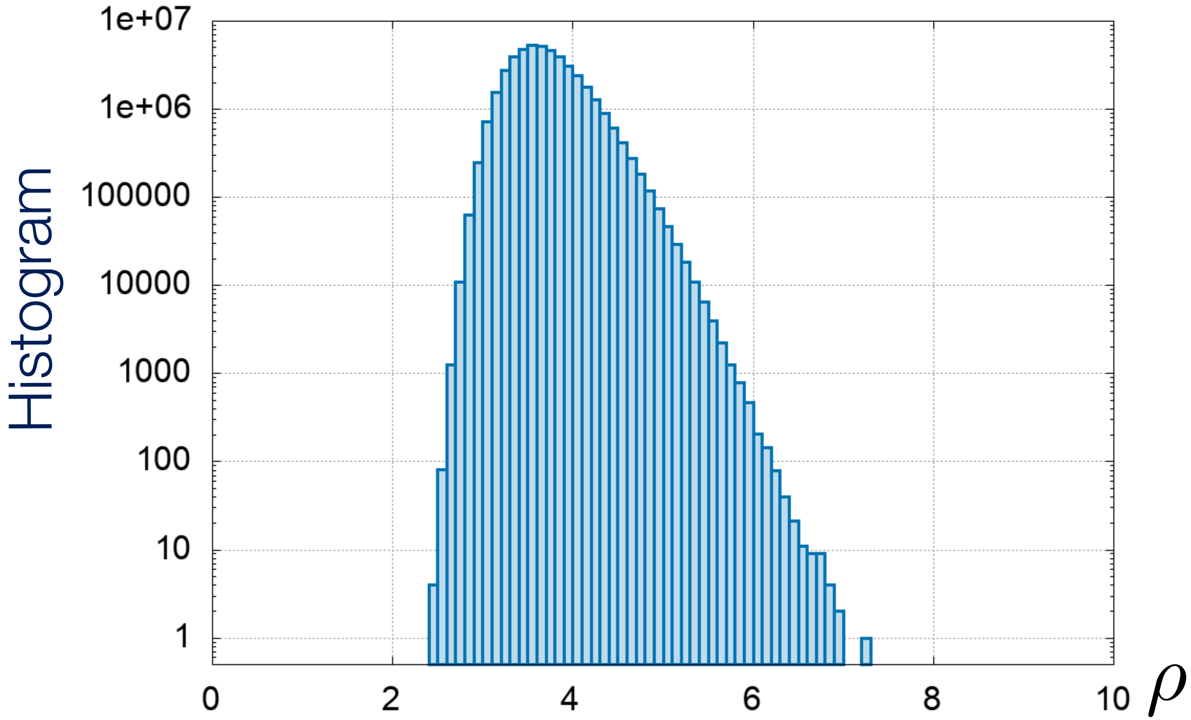
Background distribution

$$\rho_{\mathrm{perm},1}$$

$$\rho_{\mathrm{perm},2}$$

$$\vdots$$

$$\rho_{\mathrm{perm},N_{\mathrm{perm}}}$$



- By counting the number of the event,
we can evaluate how often the detection statistics ρ appear
from the noise data.

$$\text{false alarm rate} = \frac{N(\rho \leq \rho_{\mathrm{perm}})}{N_{\mathrm{perm}}}$$

- Advantages of permutation test
  1. We don’t need long data.

Comparison of methods to estimate background distribution

	Can apply to single detector?	Huge computational cost?	We need long data?
Time shift	No. At lease two detectors	Yes.	Yes. In LSC case, 5.2 days data.
Extrapolate by GEV	Yes	No. No additional analysis	No. No additional data
Permutation test	Yes	Yes.	No. No additional data

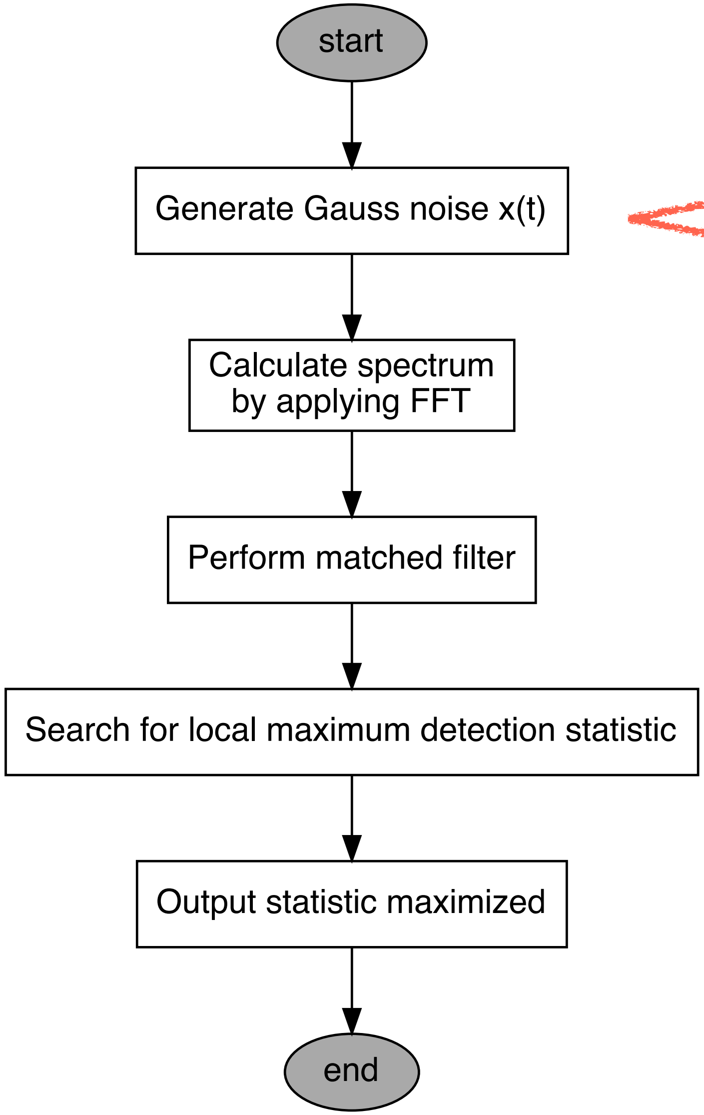
Contents

- Motivation
  - gravitational wave events
  - false alarm rate from background distribution

- Alternative methods to estimate background distribution
  - generalized extreme value distribution
  - extrapolate the distribution by fitting
  - permutation test

- **Simulation**
  - **generate simulation Gauss noise data**
  - **analyze the data and estimate background distribution**
- Summary

Flow of simulation to apply GEV fitting (1)

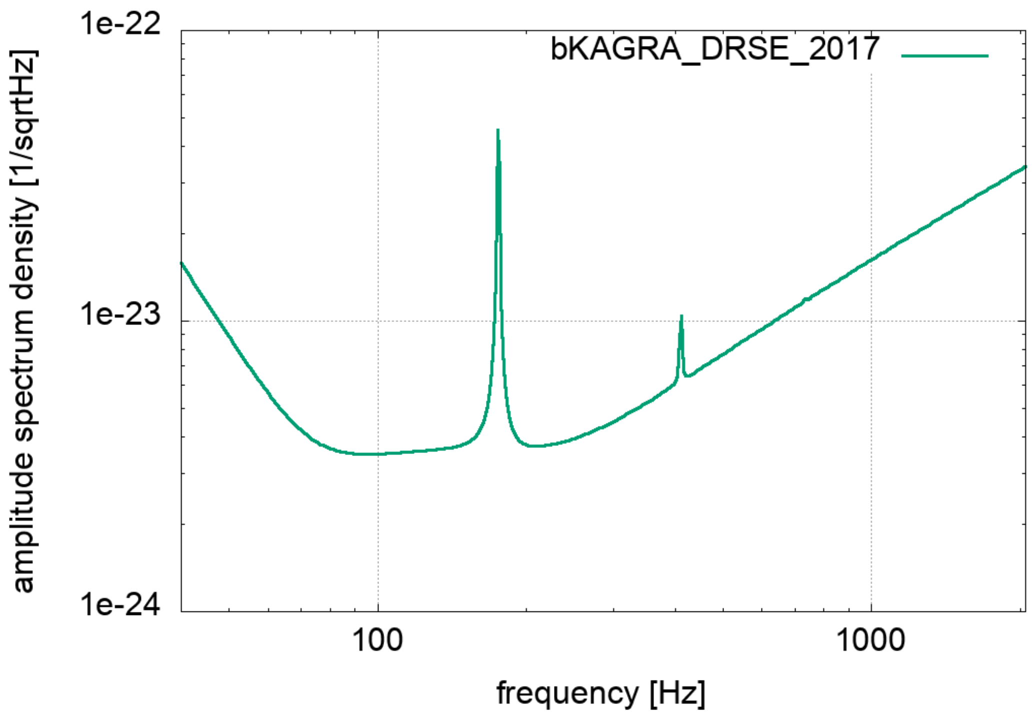


Set-up of simulation

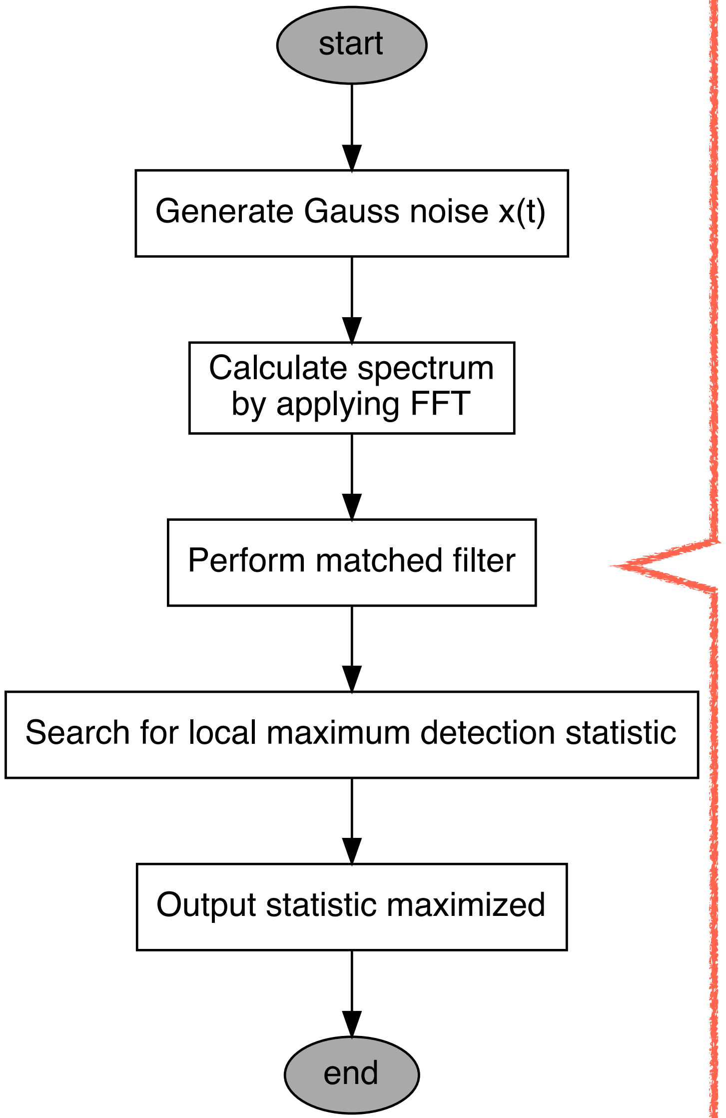
- sampling rate = 4096Hz
- duration = 512 second
- bKAGRA design sensitivity

https://gwdoc.icrr.u-tokyo.ac.jp/cgi-bin/private/DocDB/ShowDocument?docid=7038

- data with only noise



Flow of simulation to apply GEV fitting (2)



Matched filter

$$(x, u) = 4\mathrm{Re} \int_{f_{\min}}^{f_{\max}} \frac{\tilde{u}^*(f)\tilde{x}(f)}{S_{\mathrm{n}}(|f|)} df$$

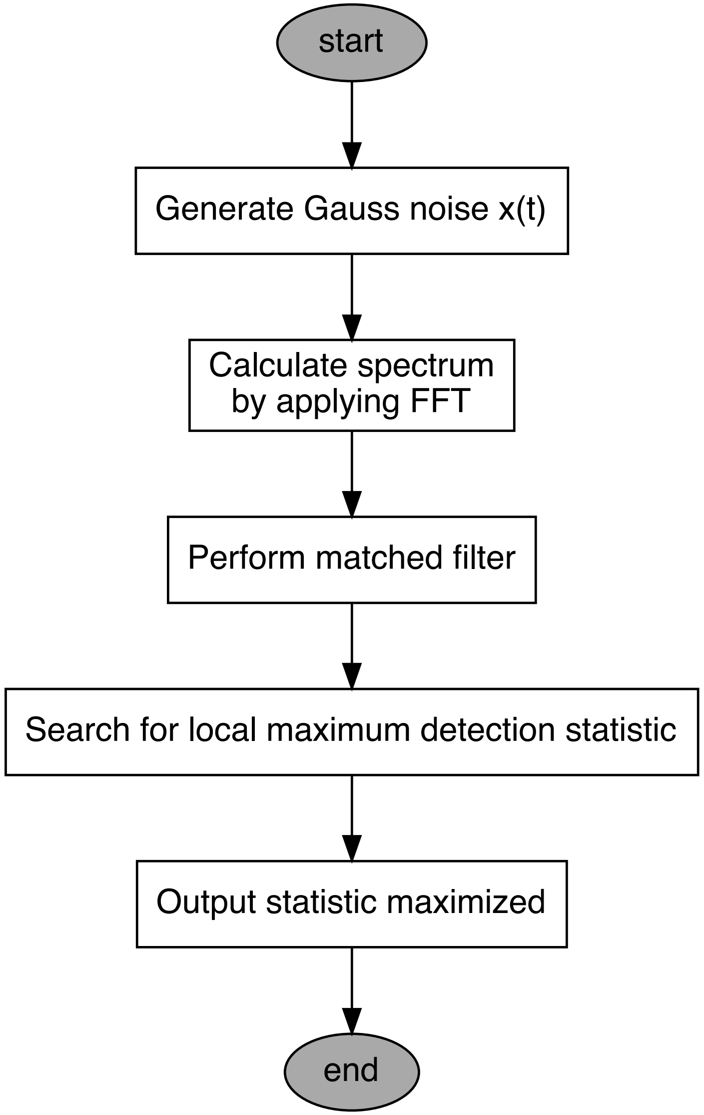
$$\boxed{\rho(t)} = \sqrt{(x, u_c)^2 + (x, u_s)^2}$$

$u_{\mathrm{c}}, u_{\mathrm{s}}$: two orthogonal and normalized waveform (template)
We can maximize quantity over unknown phase.

$S_{\mathrm{n}}(f)$: one-sided power spectrum density
For simplicity, we used design sensitivity.

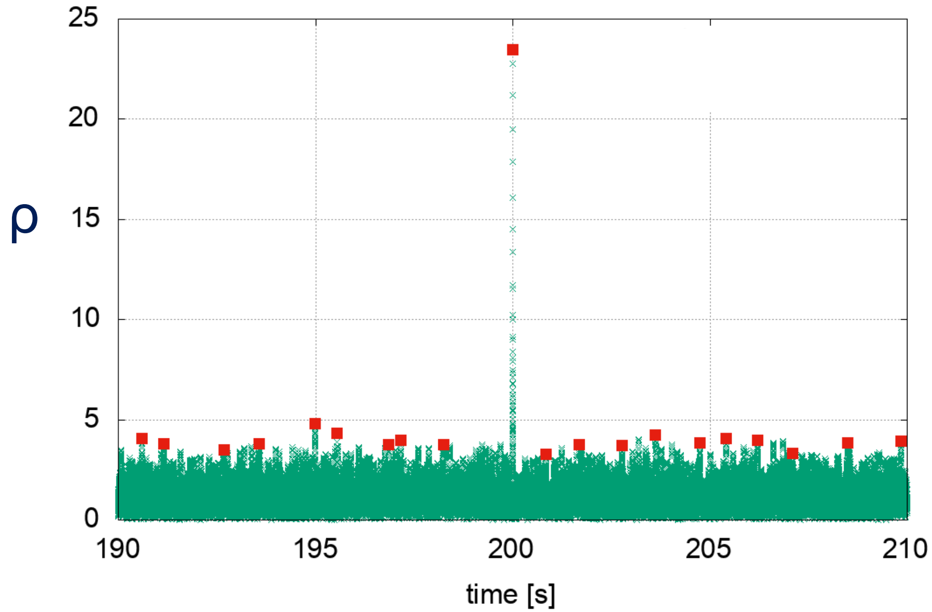
- waveform : TaylorF2
- We set $f_{\min} = 40\mathrm{Hz}$.
- Now we don't calculate chi^2 statistics, because the data consist of Gauss noise.

Flow of simulation to apply GEV fitting (3)



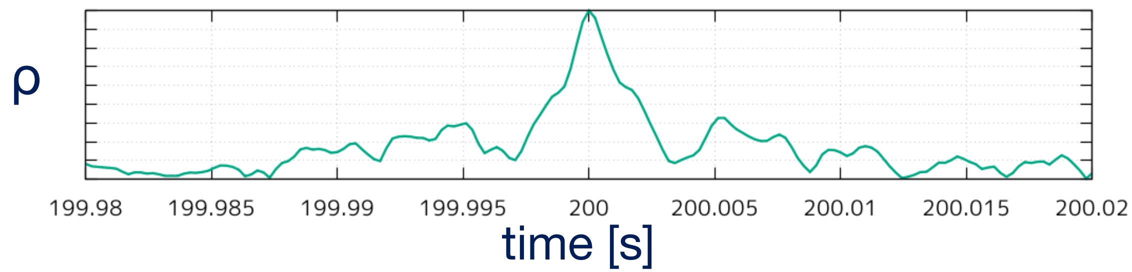
- We search for local maximum of ρ over 1 s.

=> ρ is expected to follow GEV.

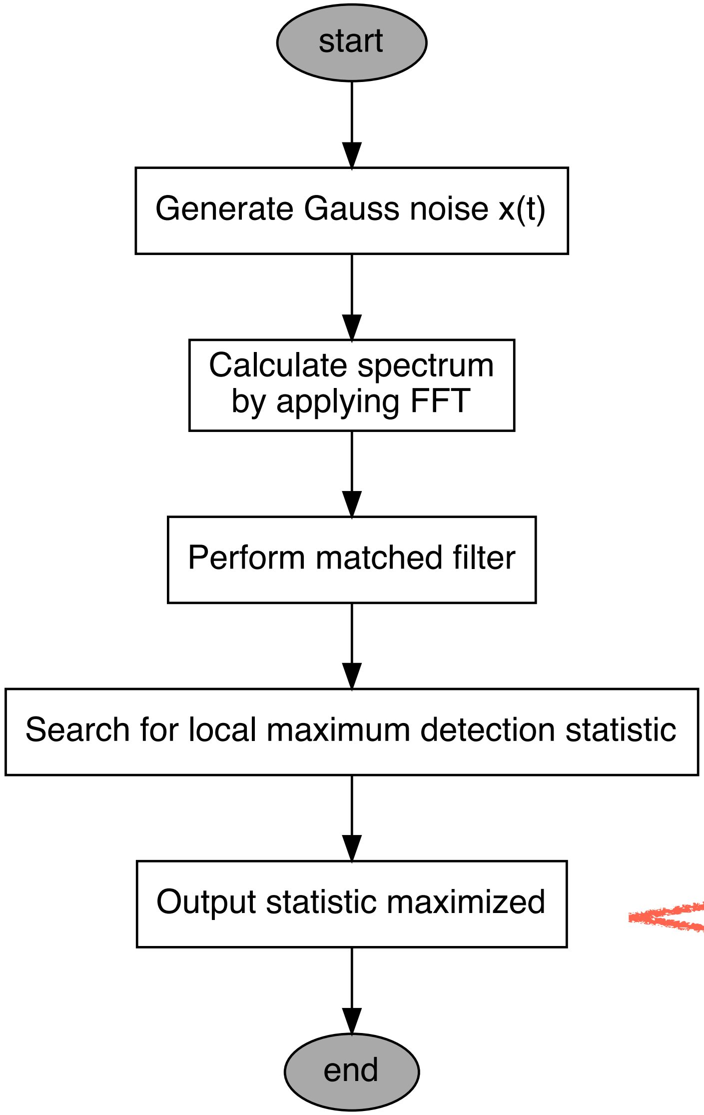


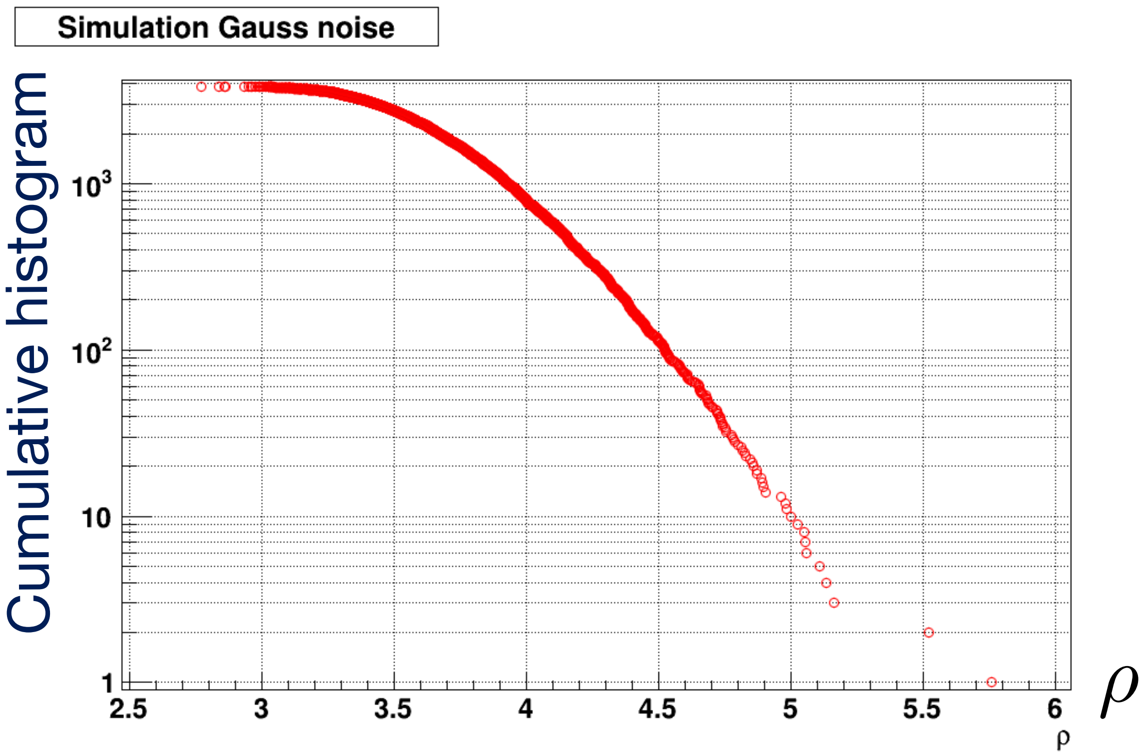
- The clustering is needed,

because the effect of signal can be spread.



Flow of simulation to apply GEV fitting (4)



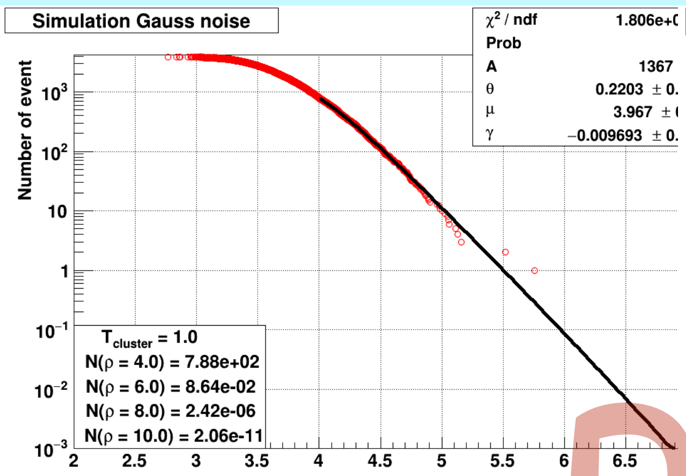


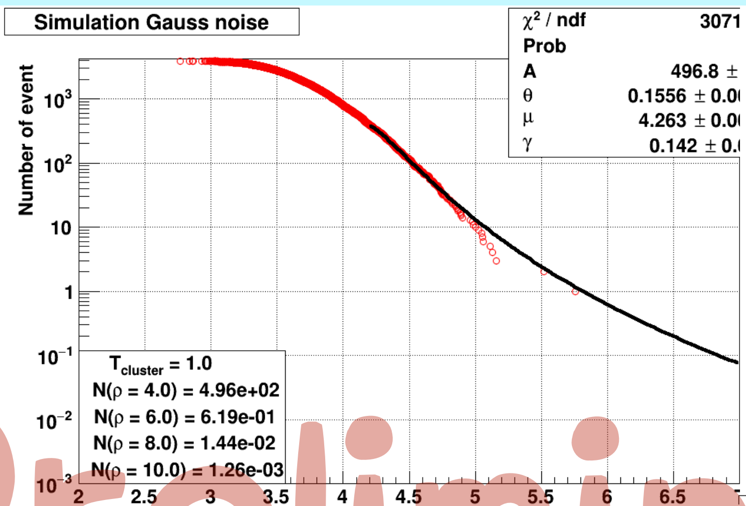
- We obtain the distribution of ρ.

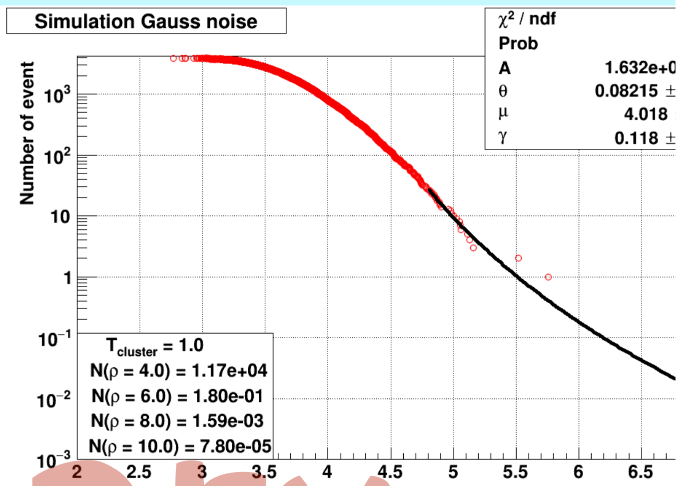
=> Fit the distribution and extrapolate

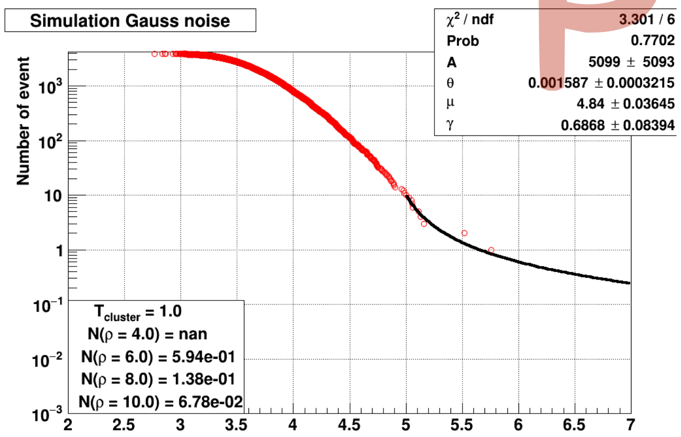
$$f_{GEV}(x, A, \theta, \mu, \gamma) = A\left(1 - \exp\left\{-\left[1 + \gamma\left(\frac{x-\mu}{\theta}\right)\right]^{-1/\gamma}\right\}\right)$$

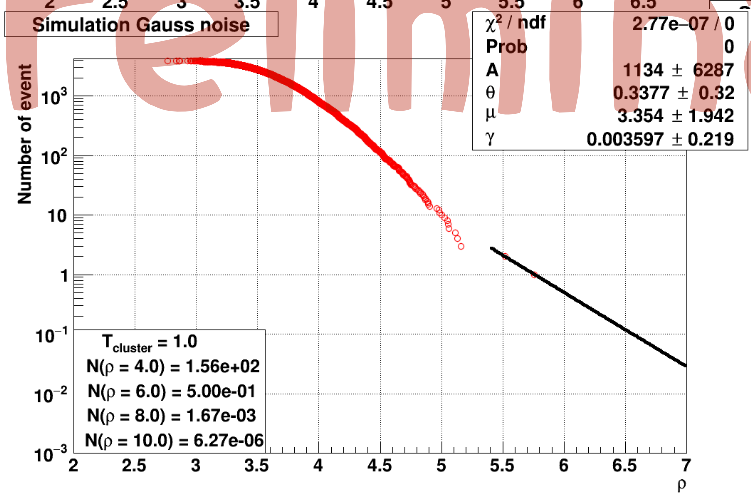
Result of GEV fitting

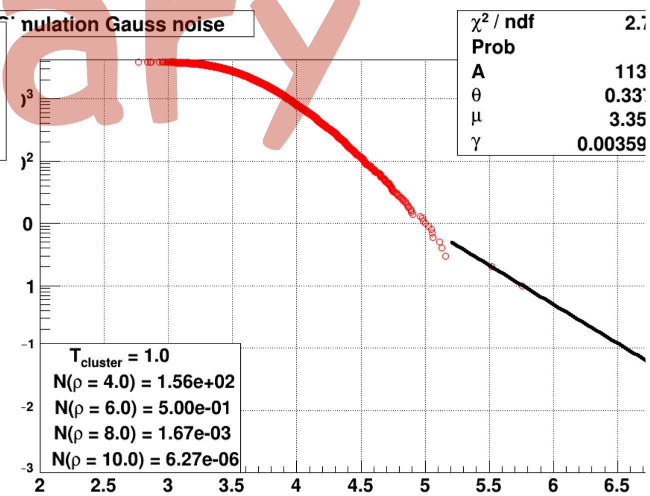


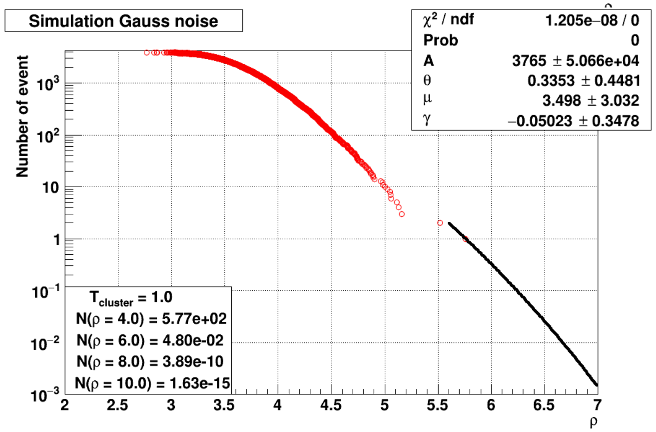






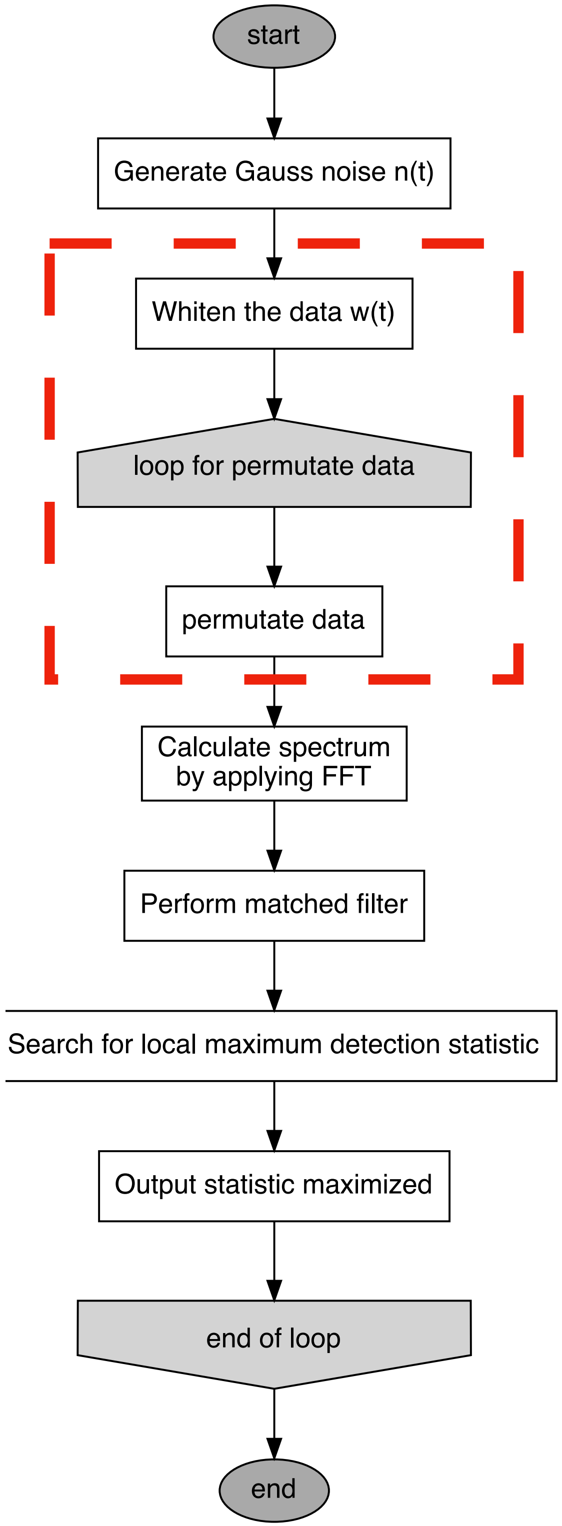






- The ranges for fitting are different.
- Under discussion about results.
- To justify this method, we perform the simulation for permutation test.

Flow of permutation test



- Almost same as previous simulation.
  The parts in red box are different.

- We need whitening data before permutating the data, because data is colored.

- From one permutation test, we obtain 3840 s of $\rho(t)$.

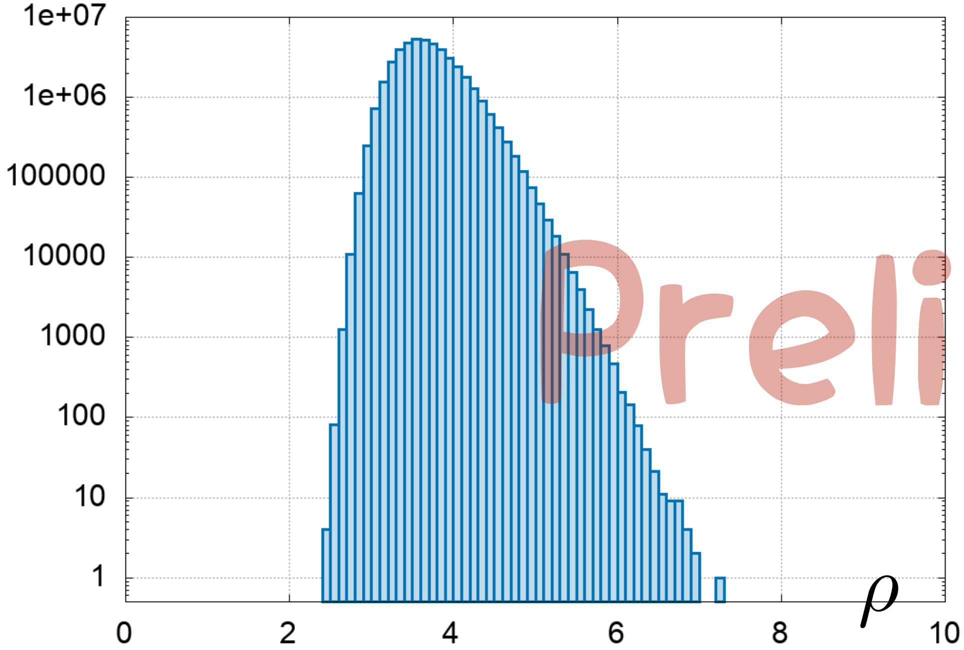
=> From 100 permutation tests, 10hours of $\rho(t)$
=> From 10000 permutation tests, 44 days of $\rho(t)$

- Latest LSC catalog paper (arXiv:1811.12907) requires a FAR less than 1 per 30 days to claim the GW event.

Result of permutation test

Permutation test

Histogram

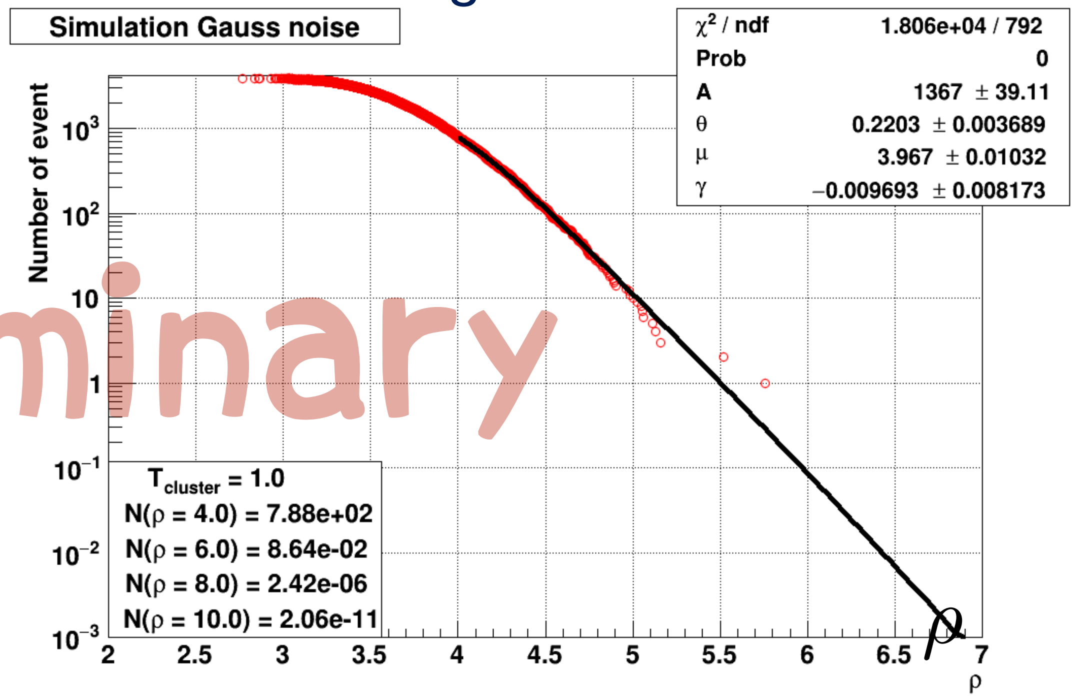


$$N(\rho = 4.0) = 7.073 \times 10^2$$

$$N(\rho = 6.0) = 4.570 \times 10^{-2}$$

GEV fitting

Cumulative histogram



$$N(\rho = 4.0) = 7.88 \times 10^2$$

$$N(\rho = 6.0) = 8.64 \times 10^{-2}$$

Summary

- We are investigating alternative methods to estimate background distribution
  - extrapolate by using generalized extreme value distribution
  - permutation test

- We are performing the simulation to estimate background distribution.

- In the future, we will apply these method to the real data and test the reliability.