Memo

19M00530 Takebe Wataru

July 3, 2020

Contents

Part I
Determination of alpha

1 Likelihood of α

First, we consider the normalized merger probability $R(m_1, m_2)$ such as

$$R(m_1, m_2) = Cf(m_1)f(m_2)(m_1 + m_2)^{\alpha} \int_0^{\infty} p_{det}(m_1, m_2, z)dz \tag{1}$$

where $f(m)$ is a function depending on the model of forming PBH, and C is some constant.
Let us make the grid cells in the $m_1 - m_2$ plane, each integer i is labeling each cell ; $m_{1,i} - dm < m_1 < m_{1,i} + dm, m_{2,i} - dm < m_2 < m_{2,i} + dm$, and the number of all cells are $K.(i = 1, 2, ..., K)$ Then, we can state the probability that each n_i events $((m_{1,1}, m_{2,1}), (m_{1,2}, m_{2,2}), ..., (m_{1,n_i}, m_{2,n_i})()$ are entering the each ith cell $(m_{1,i} - dm < m_{1,l} < m_{1,i} + dm), (m_{2,i} - dm < m_{2,l} < m_{2,i} + dm)$ like this,

$$P_m(\{n_i\}) = \frac{N!}{n_1! \cdots n_K!} p_1^{n_1} \cdots p_K^{n_K} \tag{2}$$

(this is a multinomial distribution.)
where

$$p_i = \int_{m_{1,i}-dm}^{m_{1,i}+dm} \int_{m_{2,i}-dm}^{m_{2,i}+dm} R(m_1, m_2) dm_1 dm_2 \tag{3}$$

$$N = \sum_{i=1}^{K} n_i \tag{4}$$

$$\sum_{i=1}^{K} p_i = 1 \tag{5}$$

Next, let us describe the probability that an given data is $\{n_i\}$. That is,

$$P(\{n_i\}|\alpha) = \frac{N!}{n_1! \cdots n_K!} p_1^{n_1} \cdots p_K^{n_K} \tag{6}$$

We can know the best α from this probability by using Bayes's theorem.

$$P(\alpha|\{n_i\}) = \frac{P_{pri}(\alpha)P(\{n_i\}|\alpha)}{P_{pri}(\{n_i\})} \tag{7}$$

If there is no reason or data to decide $P_{pri}(\alpha)$, we take it as uniform distribution;$P_{pri}(\alpha) = const$
We can see the examples in Next section.

2 χ^2 value and Hypothesis testing

In this section, we do Hypothesis testing to know which distribution R our data is following.Let us define the χ^2 value which is depending on actual data and model we consider. We define χ^2 value as below,

$$\chi^2(\{n_i\}) = \sum_{i=1}^{K} \frac{(n_i - Np_i)^2}{Np_i} \tag{8}$$

where

$$N = \sum_{i=1}^{K} n_i \tag{9}$$

Using this χ^2 value, we can know the plausibility of the model distribution which determines p_i.
Actually, after calculating this probability,

$$P_{\chi^2 > \chi^2_{\{n_i\}}} = \sum_{d_1=0}^{\infty} \cdots \sum_{d_K=0}^{\infty} \frac{D!}{d_1! \cdots d_K!} p_1^{d_1} \cdots p_K^{d_K} \Theta\left(\chi^2\left(\{d_i\}\right) - \chi^2\left(\{n_i\}\right)\right), \tag{10}$$

where

$$D = \sum_{i=1}^{K} d_i, \tag{11}$$

we can find plausibility of the model from this probability. However, calculation of this probability is too heavy to do numerical calculation. Actually, there is a way to know $P_{\chi^2 > \chi^2_{\{n_i\}}}$. The χ^2 value (see eq.(8)) follows χ^2 statistics that has $(K-1)$ degrees of freedom, when we can approximate the binomial distributions in eq.(2) with normal distributions. We will explain this later.(sorry, not today.)
In such case, we can use the chi-square distribution which is well known, then can save the calculation cost. We can see the examples in Next section.

3 Examples

In this section, we can see the examples of above framework. More specifically,

1. Generating the $100(=N)$ sample events following a particular distribution $R_0(m_1, m_2)$.(see eq.(1)) (Then we get $\{n_i\}$.)
2. Taking the model distribution $R(m_1, m_2)$ which can be different from true distribution $R_0(m_1, m_2)$. (but it should be excluded from our framework.) Then we could know $\{p_i\}$ from eq.(3).
3. Calculating χ^2 value (from eq.(8)).
4. We can get the probability of $\chi^2 > \chi^2_{\{n_i\}}$ from χ^2 distribution, and decide to exclude the distribution $R(m_1, m_2)$ or not.
5. Iterating $2 \sim 4$ by varying α.

Then we could determine the α by using $P\left(\alpha|\{n_i\}\right)$ (from eq.(7)).