Project Notes

October 31, 2025

1 Introduction

The core questions are:

1. ***How well do we need to measure*** *CBC parameters to infer the CBC population properties at a given* ***target redshift*** *with sufficient precision to learn conclusively if and how they differ from what we observe at redshift* ~ 0*?*

2. *What is the "****target redshift****"?*

2 Statistical tests

To begin with, we can start answering the following question: *What kind of observations do we need to conclude that the merger rate density in two mass-redshift bins is different?*

The idea is to use a statistical test to directly compare the number of events in two mass-redshift bins.

2.1 Likelihood Ratio Test (LRT)

- The statistic in each bin is poissonian distributed: $\mathrm{Pois}(N, \lambda) = \dfrac{\lambda^N \exp(-\lambda)}{N!}$.

- My likelihood is $\mathcal{L}(\lambda_1, \lambda_2) = \mathrm{Pois}(\lambda_1) \times \mathrm{Pois}(\lambda_2)$, where $\lambda_1 = N_1$ and $\lambda_2 = N_2$.

- **Define** $\alpha = \dfrac{N_1}{N_2} = \dfrac{\lambda_1}{\lambda_2}$: $\mathcal{L}(\alpha, \lambda_2) = \dfrac{(\alpha\lambda_2)^{N_1}}{N_1!} \exp(-\alpha\lambda_2) \dfrac{\lambda_2^{N_2}}{N_2!} \exp(-\lambda_2)$.

- Denote $\Theta = \{\alpha, \lambda_2\}$ the parameter space. The likelihood-ratio test statistic for the null hypothesis $H_0 : \theta \in \Theta_0$ is given by: $\lambda_{\mathrm{LR}} = -2\ln\left[\frac{\sup_{\theta\in\Theta_0}\mathcal{L}(\theta)}{\sup_{\theta\in\Theta}\mathcal{L}(\theta)}\right]$, where $\Theta_0 = \{\alpha = \alpha_0, \lambda_2\}$. λ_{LR} is asymptotically χ^2 distributed.

- It turns out that

$$\lambda_{\mathrm{LR}} = 2N_2\left[\alpha\ln\left(\frac{\alpha}{\alpha_0}\right) - (1+\alpha)\ln\left(\frac{1+\alpha}{1+\alpha_0}\right)\right] \qquad (1)$$ {eq:lambdaLR}

- In our case we have two parameters, one of which is fixed $\Longrightarrow$ 1 degree of freedom. If p-value $= P(\chi_1^2 > \lambda_{\mathrm{LR}}) <$ c.l. $\Longrightarrow \alpha \neq \alpha_0$

- The equation for λ_{LR} can be rewritten as:

$$\xi\alpha_0^{\overline{1+\alpha}^{\,\alpha}} - \alpha_0 - 1 = 0 \qquad (2)$$ {eq:lambdaLR2}

with

$$\xi = \left[\exp\left(\frac{\lambda_{\mathrm{LR}}}{2N_2}\right)\left(1+\frac{1}{\alpha}\right)^{\alpha}(1+\alpha)\right]^{\overline{1+\alpha}^{\,1}} \qquad (3)$$ {eq:xi}

.

- If I want to be sure that I've measured $\alpha \neq \alpha_0$ I need my p-value smaller than some confidence level (c.l.). Suppose c.l. $= 0.05 \Longrightarrow \lambda_{\mathrm{LR}} \simeq 3.841$. Then I can solve the equation above for α, for a fixed α_0, as a function of N_2.

2.2 Binomial test

- Use the fact that if two independent variables X_1 and X_2 are poissonian distributed with λ_1 and λ_2, then the distribution of X_1 conditioned on $X_1 + X_2$ is a binomial distribution with $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$ and $k = X_1 + X_2$.

- **Define** $N_1/N_2 = \lambda_1/\lambda_2 = \alpha$.

- **Then** $p = \frac{\alpha}{1+\alpha}$ and $p(N_1|N_1+N_2) = \mathrm{Bin}(\frac{\alpha}{1+\alpha}, N_1+N_2)$.

- The cumulative probability is given by:

$$P(N_1 \geq N_*|N_1+N_2) = \frac{1}{(1+\alpha)^{N_1+N_2}}\sum_{k=N_*}^{N_1+N_2}\binom{N_1+N_2}{k}\alpha^k. \qquad (4)$$ {eq:binomial}

- This test has the advantage of being exact regardless of the number of measured events in each bin.

2.3 Bayesian approach

- The number of events inside each bin are poissonian distributed as $\mathrm{Pois}(N_1|\lambda_1)$ and $\mathrm{Pois}(N_2|\lambda_2)$. We are interested in the posterior distribution of the parameters λ_1 and λ_2.

- The likelihood is given by $\mathcal{L}(N_1, N_2|\lambda_1, \lambda_2) = \mathrm{Pois}(N_1|\lambda_1)\mathrm{Pois}(N_2|\lambda_2)$.

- We need to choose a prior for λ_1 and λ_2. A convenient choice is the conjugate prior of the Poisson distribution: the **Gamma distribution**:

  $$p(\lambda) = \mathrm{Gamma}(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda). \qquad (5)$$
  {eq:Gamma_dist}

  where Γ is the Gamma function.

- In general the **posterior** is given by $p(\lambda_1, \lambda_2|N_1, N_2) \propto \mathcal{L}(N_1, N_2|\lambda_1, \lambda_2)p(\lambda_1)p(\lambda_2)$, but using the fact that bins and priors are independent and thanks to conjugate priors we have that

  $$p(\lambda_1, \lambda_2|N_1, N_2) = p(\lambda_1|N_1)p(\lambda_2|N_2), \qquad (6)$$
  {eq:joint_poster

  where

  $$p(\lambda|N) = \mathrm{Gamma}(\lambda|\alpha + N, \beta + 1). \qquad (7)$$
  {eq:posterior_la

  For generic priors we might use MCMC sampling.

- A possibile choice for the parameters of the prior distribution might be the (uninformative) *Jeffreys prior*: $\alpha = 0.5$ and $\beta = 0$.

- Now we can get the **posterior** of the **rates** inside each bin using $R(\Delta m_i, \Delta z_j) = \dfrac{\lambda_{ij}}{\langle VT \rangle_{ij}}$, where λ_{ij} are samples from the posterior of the bin.

- Most importantly we can get the **posterior** of rates **ratios**: $\dfrac{R_1}{R_2} = \dfrac{\lambda_1}{\lambda_2} \dfrac{\dfrac{1}{\langle VT \rangle_1}}{\dfrac{1}{\langle VT \rangle_2}}$.

- Actually the distribution of the ratio of two Gamma distributed random variables is also analytical and is given by the generalized **beta prime distribution** (also known as inverted beta distribution or beta distribution of the second kind):

$$\beta'(x|\alpha,\beta,p,q) = \frac{p\left(\frac{x}{q}\right)^{\alpha p-1}\left(1+\left(\frac{x}{q}\right)^{p}\right)^{-\alpha-\beta}}{qB(\alpha,\beta)} \tag{8}$$ {eq:betaprime_di

where B is the Beta function. Specifically, if $X_k \sim \mathrm{Gamma}(\alpha_k,\beta_k)$ are independent, then $\frac{X_1}{X_2} \sim \beta'(\alpha_1,\alpha_2,1,\frac{\beta_2}{\beta_1})$. Moreover, it can be useful to know that if $X \sim \beta'(\alpha,\beta,p,q)$, then $kX \sim \beta'(\alpha,\beta,p,kq)$.

- What if we want the $\langle VT\rangle_2/\langle VT\rangle_1$ distribution for a given value of R_1/R_2? The answer is still analytical due to the fact that if $X \sim \beta'(\alpha_1,\alpha_2)$, then $\frac{1}{X} \sim \beta'(\alpha_2,\alpha_1)$. We can make use of that because $\frac{N_1}{N_2} \sim \beta'(\alpha_1,\alpha_2,1,\frac{\beta_2}{\beta_1}) = \beta'(\alpha_1,\alpha_2)$ if $\beta_2/\beta_1 = 1$ (which is our case).

2.3.1 How to compute $\langle VT\rangle$?

- From a given population model $p(\theta)$.
- **Data driven** method assuming $p_i(\theta) = \delta(\theta-\theta_i)$, where θ_i are observed events:

$$\langle VT\rangle = \frac{N_{\mathrm{obs}}}{\hat{R}} = \frac{N_{\mathrm{obs}}V_{\mathrm{bin}}T_{\mathrm{obs}}}{\sum_k^{N_{\mathrm{obs}}}\frac{1+z_k}{p_{\mathrm{det},k}}} \tag{9}$$ {eq:VT_point}

3 Connection with the rates

We want to connect the number of events in a bin with the merger rate density in that bin. We have two ways to do that:

- $$R(\Delta m_i,\Delta z_j) = N_{\mathrm{obs}}^{ij}/\langle VT\rangle_{ij}, \tag{10}$$ {eq:rate}

  with

$$\langle VT\rangle_{ij} = T_{\mathrm{obs}}\int_{\Delta m_i,\Delta z_j}\frac{1}{1+z}\frac{dV_c}{dz}p(\theta)p_{\mathrm{det}}(m,\theta,z)dmdzd\theta. \tag{11}$$ {eq:VT}

- $$R(\Delta m_i, \Delta z_j) = \frac{1}{\Delta V_j T_{\rm obs}^{ij}} \sum_k^{N_{ij}} \frac{1+z_k}{p_{\rm det}(m_k, \theta_k, z_k)} \tag{12}$$ {eq:rate2}

  where $\Delta V_j = \int_{\Delta z_j} dz \frac{dV_c}{dz}$. In this case we can also get the variance of the rate by squaring the argument of the sum in (12).