Statistics for Astronomers
Solutions to Homework #3

Prof. Sundar Srinivasan

October 25, 2020

1. ***Notational confusion*: I apologise that the parameters for the lower and upper limits for the mass function were denoted m_1 and m_2. To avoid confusion with the data points m_i, they are henceforth referred to as M_1 and M_2.**

   (a) The initial mass function

   $$p_{_M}(m) = C\ m^{-\alpha} \qquad \alpha > 1,\ M_1 \leq m \leq M_2$$

   is nothing but a probability density function for the number of stars. Thus, the number of stars with masses between m and $m + dm$ is given by $Cm^{-\alpha}dm$.

   As it is a PDF, it should be normalised (*i.e.*, it should integrate to unity). We can use this normalisation constraint to compute the value of C in terms of the parameters α, M_1, and M_2:

   $$\int\limits_{M_1}^{M_2} C\ m^{-\alpha}\ dm = \frac{C}{\alpha - 1}\Big(M_1^{1-\alpha} - M_2^{1-\alpha}\Big) = \frac{C}{\beta}\Big[M_1^{-\beta} - M_2^{-\beta}\Big] = 1,\ \text{with } \boldsymbol{\beta \equiv \alpha - 1 > 0}$$

   Therefore,

   $$C = \frac{\beta}{M_1^{-\beta} - M_2^{-\beta}} \tag{1}$$

   Note that C depends on the parameters! We must be careful not to ignore it when differentiating the likelihood.

   (b) The likelihood that a star observed at random has a mass in the range m_i and $m_i + dm$ is Cm_i^{α}. As the observations are independent, the combined likelihood of obtaining a set of masses $\{m_1, m_2, \cdots, m_N\}$ is

   $$\mathscr{L} = \prod_{i=1}^{N} Cm_i^{-\alpha} \Longrightarrow \ln \mathscr{L} = N \ln C - \alpha\ \sum_{i=1}^{N} \ln m_i \tag{2}$$

   Using Equations (1) and (2), we first compute the partial derivatives of $\ln C$ with respect to each of the parameters (recall: since $\beta = \alpha - 1$, it is convenient to find $\widehat{\beta}$, from which $\widehat{\alpha}$ can be readily computed):

$$\frac{\partial}{\partial M_1}\ln C = \frac{1}{C}\;\frac{\beta\ M_1^{-\alpha}}{M_1^{1-\alpha} - M_2^{1-\alpha}}$$
$$\frac{\partial}{\partial M_2}\ln C = -\frac{1}{C}\;\frac{\beta\ M_2^{-\alpha}}{M_1^{1-\alpha} - M_2^{1-\alpha}}$$
$$\frac{\partial}{\partial \beta}\ln C = \frac{1}{\beta} + \ln M_1 - \frac{\ln\left(\frac{M_2}{M_1}\right)}{\left(\frac{M_2}{M_1}\right)^{\beta} - 1}$$

We then use these relations to evaluate the partial derivatives of $\ln \mathscr{L}$:

$$\frac{\partial}{\partial M_1}\ln \mathscr{L} = N\;\frac{\partial}{\partial M_1}\ln C = -\frac{N}{C}\;\frac{(1-\alpha)\ M_1^{-\alpha}}{M_1^{1-\alpha} - M_2^{1-\alpha}} \tag{3}$$
$$\frac{\partial}{\partial M_2}\ln \mathscr{L} = N\;\frac{\partial}{\partial M_2}\ln C = \frac{N}{C}\;\frac{(1-\alpha)\ M_2^{-\alpha}}{M_1^{1-\alpha} - M_2^{1-\alpha}} \tag{4}$$
$$\frac{\partial}{\partial \beta}\ln \mathscr{L} = N\left[\frac{1}{\beta} + \ln M_1 - \frac{\ln\left(\frac{M_2}{M_1}\right)}{\left(\frac{M_2}{M_1}\right)^{\beta} - 1} - \overline{\ln m}\right], \tag{5}$$

where $\overline{\ln m}$ is the sample mean of $\ln m_i$.

The standard procedure to find the MLE values requires us to set the derivatives (Equations (3) and (4)) to zero. This method does not give us a meaningful solution for this particular problem. Instead, we will have to investigate the functional dependence of the likelihood on M_1 and M_2 from Equations (1) and (2).

The denominator of Equation (1) increases if either M_1 decreases or M_2 increases, increasing $\ln C$ and therefore $\ln \mathscr{L}$. The likelihood therefore achieves its largest value if the smallest (largest) possible data value is used as an estimate for M_1 (M_2). Accordingly, we have

$$\widehat{M_1} = \min(m_i) = m_{(1)}; \qquad \widehat{M_2} = \max(m_i) = m_{(N)}$$

That is, the ML estimates for the lower and upper mass limits of the IMF are the smallest and largest masses in the data set. $\boldsymbol{m_{(i)}}$ refers to the $\boldsymbol{i}^{\text{th}}$ order statistic.

(c) Having found $\widehat{M_1}$ and $\widehat{M_2}$, to find $\widehat{\alpha}$, we can use the standard method and equate Equation (5) to zero. The problem also gives us the value of the sample mean of $\ln m_i$. Equation (5) to zero gives us a non-linear equation for which $\widehat{\beta}$ are the roots. We can use a root-finder algorithm to solve for this exponent. The script provided here uses the `scipy.optimize.root_scalar` package to get $\widehat{\beta} \approx 1.24$, or $\widehat{\alpha} \approx 2.24$.

2. Since supernova explosions are independent events, we can assume Poisson statistics.

In solving problems involving the Poisson distribution, remember that the Poisson rate parameter λ <u>must be dimensionless</u>. Always use this as a sanity check!

Each galaxy is observed for a different time t_i, which means the expected number of supernova events (= the Poisson rate parameter!) is different for each galaxy.

The Poisson rate parameter λ_i for each galaxy is the product of the explosion rate p (dimensions: time^{-1}) and the exposure time t_i (dimensions: time). The total likelihood is therefore

$$\mathscr{L} = \prod_{i=1}^{N} \frac{\lambda_i^{n_i} e^{-\lambda}}{n_i!} = \prod_{i=1}^{N} \frac{(pt_i)^{n_i} e^{-pt_i}}{n_i!} \Longrightarrow \ln \mathscr{L} = \text{constant} + \sum_{i=1}^{N} (n_i \ln p - pt_i)$$

We set the first derivative of the log-likelihood to zero to compute the MLE for p:

$$\left(\frac{\partial \ln \mathscr{L}}{\partial p}\right)_{\widehat{p}} = \sum_{i=1}^{N} \left(\frac{n_i}{\widehat{p}} - t_i\right) = 0 \Longrightarrow \widehat{p} = \frac{\sum_{i=1}^{N} n_i}{\sum_{i=1}^{N} t_i}$$

(b) The Expected Fisher information is

$$\mathcal{I}(p) \equiv -\mathbb{E}\left[\frac{\partial^2 \ln \mathscr{L}}{\partial p^2}\right] = \mathbb{E}\left[\frac{1}{p^2} \sum_{i=1}^{N} n_i\right] = \frac{1}{p^2} \sum_{i=1}^{N} \mathbb{E}[n_i]$$

For Poisson statistics, $\mathbb{E}[n_i] = \lambda_i = pt_i$; therefore, the Cramér-Rao Lower Bound on the variance is

$$\text{CRLB} \equiv \frac{1}{\mathcal{I}(p)} = \frac{p}{\sum_{i=1}^{N} t_i}$$

Since the supernovae in different galaxies are independent occurrences, we can also define the total number of events $N = \sum_{i=1}^{N} n_i$ and the total observing time $T = \sum_{i=1}^{N} t_i$. In terms of these quantities,

$$\widehat{p} = \frac{N}{T} \text{ and } \text{CRLB}(\widehat{p}) = \frac{\widehat{p}}{T} = \frac{N}{T^2}$$

In other words, the standard error on $\widehat{p}$ is $\frac{\sqrt{N}}{T}$. This is consistent with the standard deviation on the total number of events N being $\sqrt{N}$ as a result of Poisson statistics.