

# 开发用于下一代 Ia 型超新星调查的超快似然距离推断框架

Dylan KUHN, Marc BETOULE, on behalf of the LEMAÎTRE collaboration

Sorbonne Université, CRNS, Université de Paris, LPNHE, 75252 Paris Cedex 05, France

在这项工作中，我们提出了 EDRIS (法语为不完整超新星调查的距离估算器)，一个专门用于从 Ia 型超新星光变曲线参数中重建无偏宇宙距离的宇宙学推断框架。这一目标是通过在统计模型中直接包含数据截断来实现的，该模型负责标准化光度距离。这使我们能够通过最大化相应的似然性来构建单步距离估计，避免了调查检测限制可能引入的偏差。此外，我们预计当前的世界统计数据在未来几年将增加到  $O(10)$  倍。这为处理提供了新的挑战，因为宇宙学分析必须保持计算上的可控性。我们展示了在 EDRIS 中使用优化方法可以实现合理的  $O(N^2)$  时间复杂度，从而导致非常快的推断过程 (对 1500 颗超新星进行推断仅需  $O(10)$  秒)。

## 1 介绍

EDRIS 的主要目标是处理已知为 Malmquist 偏差的仪器偏差。每个调查中固有的可观察到的星等限制导致倾向于观测内在更亮的天体，因此对距离估计器<sup>7</sup> 造成负偏差。通常防止此估计量产生偏差的方法是运行广泛的模拟以计算一个偏差校正，并在之后将其传播到重建的距离上<sup>2,4</sup>。然而，这种方法不随数据数量的增长而很好地扩展，我们预计当前的世界统计数据将在未来几年增加十倍。为了解决偏差问题和扩展性问题，我们提出了一种基于截断似然最小化的创新方法。这使得可以在距离标准化模型中整合选择效应的建模。所介绍的方法是在一个名为勒梅特的大数据组合背景下实现的。勒梅特分析是一个端到端的宇宙学分析，使用三个未发布的 Ia 型超新星样本 (ZTF、SNLS 5y、HSC/Subaru)，总共有  $O(4000)$  颗超新星。所有数据都通过一个共同的数据管道从像素处理到宇宙学推断，最后一步由 EDRIS 完成。

## 2 马尔基斯蒂偏差的建模

描述 Ia 型超新星行为的常见模型是 Tripp 模型<sup>8</sup> (参见方程 1)。通常的标准过程涉及两个参数 (颜色和拉伸)，但该模型可以很容易地推广，以考虑任意数量的标准参数。符号 \* 表示我们需要引入的潜在参数，以便正确处理  $c$  和  $x_1$  测量受到的影响且不能被忽略的不确定性。

$$Y_i = \begin{pmatrix} m_i \\ x_{1,i} \\ c_i \end{pmatrix} = \begin{pmatrix} M + \mu_i(z, \theta) + \alpha x_{1,i}^* + \beta c_i^* \\ x_{1,i}^* \\ c_i^* \end{pmatrix} + \begin{pmatrix} \epsilon_i \\ 0 \\ 0 \end{pmatrix} \text{ with } \epsilon_i \sim \mathcal{N}(0, \sigma_{int}^2) \quad (1)$$

$m_i x_{1,i}$ ,  $c_i$  和  $\mu_i$  分别指超新星  $i$  的视星等、其伸展参数、其颜色参数和其距离模数。 $M$  是超新星的绝对星等，而  $\sigma_{int}$  是 Ia 型超新星总体的绝对星等离散度。 $(m_i, x_{1,i}, c_i)$  参数是相关的，我们用  $C$  表示完整的协方差矩阵。截断效应可以写成：

$$\begin{aligned}
Y_i^{obs} &= Y_i + \eta_i \text{ if } m_i \leq m_{lim} + \kappa_i \text{ with } \eta_i \sim \mathcal{N}(0, C_i) \text{ and } \kappa_i \sim \mathcal{N}(0, \sigma_d^2) \\
Y_i^{obs} &\text{ is unobserved otherwise}
\end{aligned} \tag{2}$$

其中,  $\eta_i$  是测量噪声,  $m_{lim}$  是调查的绝对星等, 而  $\kappa_i$  是由于观测条件变化引起的波动  $\sigma_d$ 。与上述模型相关的负对数似然函数是基于多变量正态分布的标准似然性得出的。新颖之处在于方程 3 中的求和。从贝叶斯定理出发, 我们推导出两个新的似然项。这两个都依赖于正态分布的累积分布函数和选择函数参数, 允许考虑数据截断<sup>6</sup>。

$$\begin{aligned}
\Gamma &= -\ln(|W|) + r^\dagger W r \\
&+ \sum_i 2 \ln \left( \Phi \left( \frac{m_{lim} - M - \mu_i - \alpha x_{1,i}^* - \beta c_i^*}{\sqrt{\sigma_{int}^2 + \sigma_d^2}} \right) \right) - 2 \ln \left( \Phi \left( \frac{m_{lim} - m_i^{obs}}{\sqrt{\sigma_d^2 + f(C_i)}} \right) \right)
\end{aligned} \tag{3}$$

使用  $\Phi(z) = \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{z}{\sqrt{2}} \right) \right)$ ,  $W = C^{-1}$  和  $r = \begin{pmatrix} m^{obs} \\ x_1^{obs} \\ c^{obs} \end{pmatrix} - \begin{pmatrix} M^* - \mu - \alpha x_1^* - \beta c^* \\ x_1^* \\ c^* \end{pmatrix}$ 。

所描述的马尔库斯偏倚建模等同于通过三个独立的 S 形曲线 (每个调查一个) 对选择函数进行建模。每个 S 形曲线由一个中心值 ( $m_{lim}$ ) 和一个宽度 ( $\sigma_d$ ) 来表征。

### 3 距离和宇宙学参数的估计

在统计学中, 方差的估计量 (包括我们在本例中的  $\sigma_{int}$ ) 是有偏的。因此, EDRIS 本质上是偏差的。为了量化这种效应, 我们进行了 100 次蒙特卡罗模拟。生成的 Ia 型超新星光曲线和红移是从带有天空调查<sup>a</sup> 的观测日志中产生的。对于这个分析, 我们考虑一个简化的案例, 只有一个标准化参数 (颜色)。这允许用盐<sup>3</sup> 的一维版本拟合光曲线。因此, 我们在保持问题复杂性的同时 (光曲线模型仍然经过训练), 使分析运行得更快。模拟参数列于表 1 中, 而结果则展示在图 1 中。左面板显示了重建距离与模拟输入之间的差值的平均数, 在红移中以对数形式分布在 30 个区间内。蓝色点对应多变量正态分布的经典最大似然估计器, 并显示出由于每个调查的选择效应而导致重建距离存在强烈的负偏差。橙色点对应包含选择效应建模的距离估计器, 显示在  $\Omega_m$  上没有显著的偏差 (见表 2)。

Table 1: 蒙特卡罗分析的模拟参数。

Survey	Redshift range	$\beta$	$M$	Cosmology	$m_{lim}$	$\sigma_d$
ZTF	[0.0, 0.2]	3.15	-19.0	Planck 2018 <sup>l</sup>	18.59	0.1
SNLS 5y	[0.0, 1.07]	3.15	-19.0	Planck 2018 <sup>l</sup>	24.38	0.25
HSC/Subaru	[0.0, 1.6]	3.15	-19.0	Planck 2018 <sup>l</sup>	25.2	0.07

### 4 计算的加速

由于观测的协方差矩阵依赖于估计的  $\sigma_{int}$  参数, 我们需要在似然最小化的每一步中对其进行求逆。然而, 这种依赖性简单到足以利用舒尔补技术<sup>5</sup>。第一步包括考虑  $C$  的块结构 (见方程 4) 并将  $r$  写为  $(r_1, r_2)$  以匹配此结构。

<sup>a</sup><https://skysurvey.readthedocs.io/en/latest/>

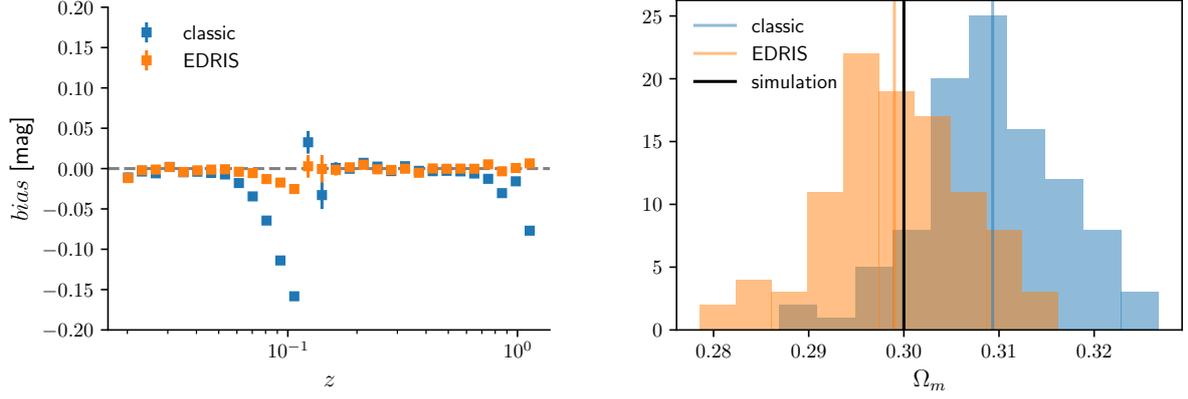


Figure 1: 左面板：重建的分箱距离模数对每个估计量的偏差。右面板：各个估计量重建的事物物质能密度参数直方图。

Table 2: 偏差和不确定度来自对两个估计量进行的蒙特卡洛模拟所得的  $\Omega_m$ 。

Estimator	Bias on $\Omega_m$	Mean of reconstructed $\Omega_m$	$\sigma(\Omega_m)$
classic	$0.0093 \pm 0.0007$	0.309	0.007
EDRIS	$-0.0010 \pm 0.0007$	0.299	0.007

$$W = \begin{pmatrix} C_{mm} + \sigma_{int}^2 I_N & C_1 \\ C_1^\dagger & C_2 \end{pmatrix}^{-1} \quad (4)$$

然后，我们可以计算并对角化（见公式 5）右下块  $C_2$  的舒尔补  $S = C_{mm} + \sigma_{int}^2 I_N - C_1 C_2^{-1} C_1^\dagger$  在  $C$  中。

$$S^{-1} = Q(\Lambda + \sigma_{int}^2 I_N)^{-1} Q^\dagger \quad (5)$$

似然性的前两项可以写成：

$$-\ln(|W|) = \sum_i \ln(\Lambda_i + \sigma_{int}^2) + \ln(|C_2|) \quad (6)$$

$$r^\dagger W r = r_1^\dagger S^{-1} r_1 - 2r_1^\dagger S^{-1} C_1 C_2^{-1} r_2 + r_2^\dagger C_2^{-1} r_2 + r_2^\dagger C_2^{-1} C_1^\dagger S^{-1} C_1 C_2^{-1} r_2 \quad (7)$$

最后一步包括预先计算方程 7 中的每个常数矩阵-矩阵乘积。因此，只剩下矩阵-向量乘积，这使得似然性计算能够从使用朴素行列式计算和矩阵求逆时的  $O(N^2)$  扩展到  $O(N^3)$ 。为了进一步优化，我们决定实现一个无海森矩阵最小化方法，使用 JAX<sup>b</sup>。总计算时间在图 2 中给出。我们能够达到在包括亮度标准化、估计  $\sigma_{int}$  和校正 Malmquist 偏差的情况下，对 1500 颗超新星的最小化时间为  $O(10)$  秒。

<sup>b</sup><https://jax.readthedocs.io/en/latest/>

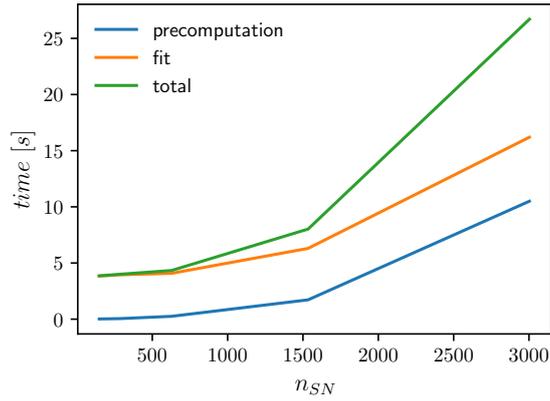


Figure 2: 似然最小化的时间缩放作为模拟超新星数量的函数。蓝色、橙色和绿色线条分别表示常数项的预计算、宇宙推断以及总时间。

### 参考文献

1. Aghanim, N., Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, et al. *Astronomy and Astrophysics*, vol. 641, September 2020, p. A6.
2. Betoule, M., et al. *Astronomy and Astrophysics*, vol. 568, August 2014, p. A22.
3. Hazenberg, François. PhD thesis, Sorbonne Université, 2019.
4. Kessler, Richard, et Dan Scolnic. *The Astrophysical Journal*, vol. 836, no 1, February 2017, p. 56.
5. Ouellette, Diane Valérie. *Linear Algebra and its Applications*, vol. 36, March 1981, p. 187-295.
6. Rubin, David, et al. *The Astrophysical Journal*, vol. 813, no 2, November 2015, p. 137.
7. Teerikorpi, P. *Astronomy and Astrophysics*, vol. 339, November 1998, p. 647-57.
8. Tripp, Robert. *Astronomy and Astrophysics*, vol. 331, March 1998, p. 815-20.