

Применение критерия согласия Пирсона

Руководитель: И. И. Никифоров

Выполнил: П. Л. Соболев

Задачи

- Оптимизировать параметры моделей нормального и бимодального распределений металличностей по представленным данным, используя метод наибольшего правдоподобия;
- Применить критерий согласия Пирсона к каждой комбинации моделей, размеров ячеек и данных.

Ход выполнения и результаты

Модельные функции:

$$\varphi_{\text{normal}}(f; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(f-\mu)^2}{2\sigma^2}}, \quad (1)$$

$$\varphi_{\text{bimodal}}(f; \mu_1, \sigma_1, \mu_2, \sigma_2, c) = \frac{c}{\sqrt{2\pi}\sigma_1} e^{-\frac{(f-\mu_1)^2}{2\sigma_1^2}} + \frac{1-c}{\sqrt{2\pi}\sigma_2} e^{-\frac{(f-\mu_2)^2}{2\sigma_2^2}}. \quad (2)$$

Для этих двух модельных распределений находим статистику

$$\chi_q^2 = \sum_{j=1}^J \frac{(N_j - Np_j)^2}{Np_j} \quad (3)$$

и уровень значимости

$$\alpha_q = P(X \geq \chi_q^2) \text{ при } X \sim \chi^2(k), \quad k = J - 1. \quad (4)$$

Здесь N — объём выборки, J — число ячеек (разрядов), N_j — число объектов, попавших в ячейку с индексом j , а p_j — вероятность попадания случайной величины f в эту ячейку:

$$p_j = \int_{\hat{f}_j}^{\hat{f}_{j+1}} \varphi_{\text{mod}}(f) df. \quad (5)$$

Для нормального распределения имеем

$$p_j = \int_{\hat{f}_j}^{\hat{f}_{j+1}} \varphi_{\text{normal}}(f; \mu, \sigma) df = \frac{1}{2} \operatorname{erf} \left(\frac{f-\mu}{\sqrt{2}\sigma} \right) \Big|_{\hat{f}_j}^{\hat{f}_{j+1}}, \quad (6)$$

где $\operatorname{erf}(z)$ — функция ошибок:

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt. \quad (7)$$

Результаты получены с помощью скрипта, написанного на языке программирования [Julia](#). Код расположен в GitLab репозитории [Stellar Astronomy Special Workshop S10-2022](#). Для воспроизведения результатов следуй инструкциям в файле `README.md`.

Таблица 1: Точечные оценки параметров моделей

Параметр	all.dat	wo_p1.dat	wo_p12.dat	wo_p123.dat
μ	-1.26717	-1.30727	-1.27778	-1.30282
σ	0.56236	0.55809	0.56445	0.56652
μ_1	-1.5162	-1.52075	-1.48453	-1.52882
σ_1	0.39482	0.40862	0.42066	0.40132
μ_2	-0.51945	-0.49915	-0.44657	-0.47336
σ_2	0.22909	0.18125	0.15908	0.17974
c	0.75016	0.79104	0.80081	0.78588

Таблица 2: Результаты применения критерия согласия Пирсона при $\Delta\hat{f} = 0.1$

Переменная	all.dat	wo_p1.dat	wo_p12.dat	wo_p123.dat
N	152	132	108	85
J	24	24	24	24
k	23	23	23	23
$\chi^2_{q, \text{normal}}$	30.44766	30.11229	33.11314	29.90528
$\alpha_{q, \text{normal}}$	0.13698	0.14621	0.07912	0.15214
$\chi^2_{q, \text{bimodal}}$	14.92319	11.37252	11.69287	12.74771
$\alpha_{q, \text{bimodal}}$	0.89737	0.97906	0.97494	0.95725

Таблица 3: Результаты применения критерия согласия Пирсона при $\Delta\hat{f} = 0.2$

Переменная	all.dat	wo_p1.dat	wo_p12.dat	wo_p123.dat
N	152	132	108	85
J	12	12	12	12
k	11	11	11	11
$\chi^2_{q, \text{normal}}$	14.04021	15.01636	13.95291	10.95251
$\alpha_{q, \text{normal}}$	0.23078	0.18175	0.23561	0.44725
$\chi^2_{q, \text{bimodal}}$	2.46976	1.88955	1.6971	3.57468
$\alpha_{q, \text{bimodal}}$	0.99605	0.99885	0.99931	0.98075

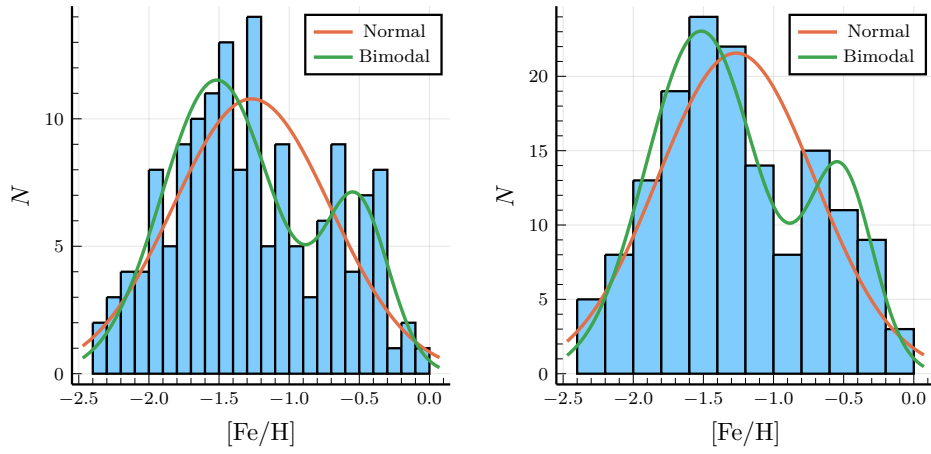


Рис. 1: Гистограммы данных из файла `all.dat`
и оптимизированные модельные функции

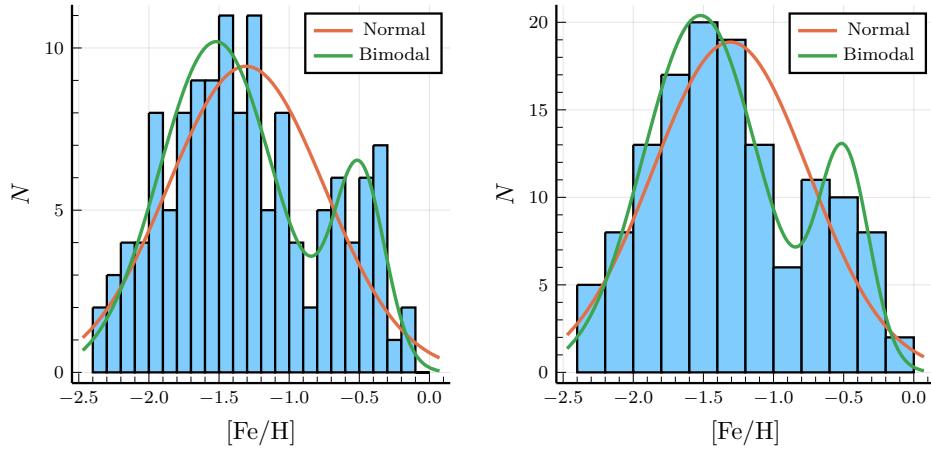


Рис. 2: Гистограммы данных из файла `wo_p1.dat`
и оптимизированные модельные функции

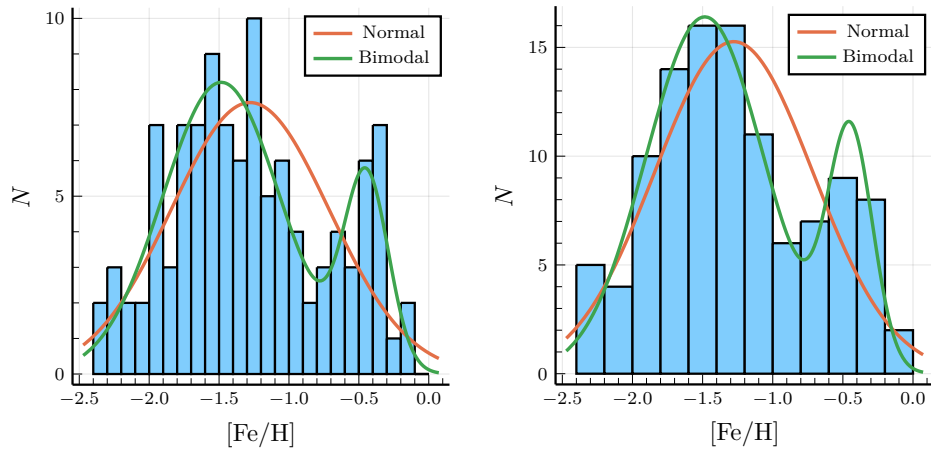


Рис. 3: Гистограммы данных из файла `wo_p12.dat` и оптимизированные модельные функции

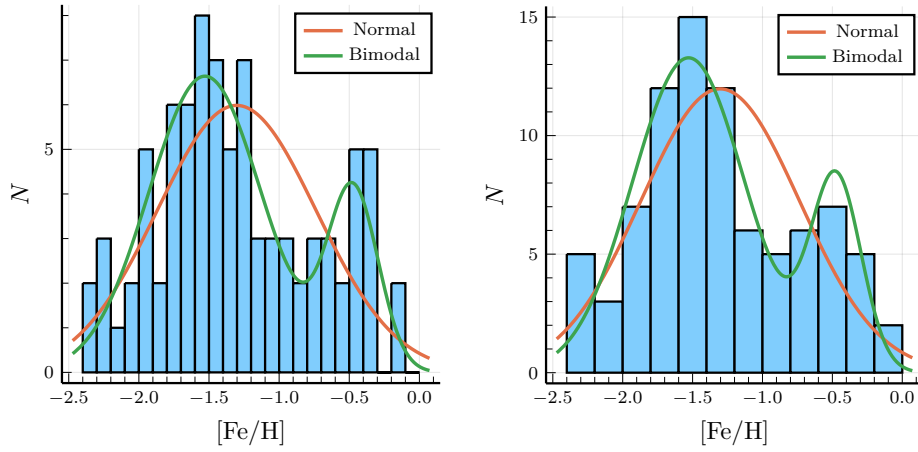


Рис. 4: Гистограммы данных из файла `wo_p123.dat` и оптимизированные модельные функции

Таким образом, нуль-гипотеза об описании данных модельным распределением принимается как статистически значимая ($\alpha_q \geq 0.05$) во всех комбинациях исходных данных, размеров ячеек и моделей. Однако по значениям α_q и представленным графикам наглядно видно, что бимодальное распределение лучше описывает исходные данные.