Оценивание плотности вероятности методами функции-ядра при исследовании звездных скоплений

А. Ф. Селезнев

Коуровская астрономическая обсерватория УрФУ

- Поверхностная плотность числа звезд
- Пространственная плотность
- Фазовая плотность
- Функция распределения скоростей
- Функция светимости
- Функция масс
- …

Для радиальных профилей поверхностной и пространственной плотности:

$$\mu(r) = \frac{2\pi r}{N} F(r) \qquad \nu(r) = \frac{4\pi r^2}{N} f(r)$$

Непараметрические методы оценивания плотности вероятности:

- Гистограмма
- Наивный метод (naive estimator)
- Метод функции-ядра (kernel estimator)
- Метод ближайшего соседа (nearest neighbor method)
- Метод функции-ядра с переменным (адаптивным) ядром
- Разложение в ряд по ортогональным функциям
- Метод максимального правдоподобия с использованием штрафных функций (maximum penalized likelihood estimator)
- Метод оценки с помощью весовой функции

Silverman B.W. “Density estimation for statistics and data analysis” Chapman & Hall, London, 1986

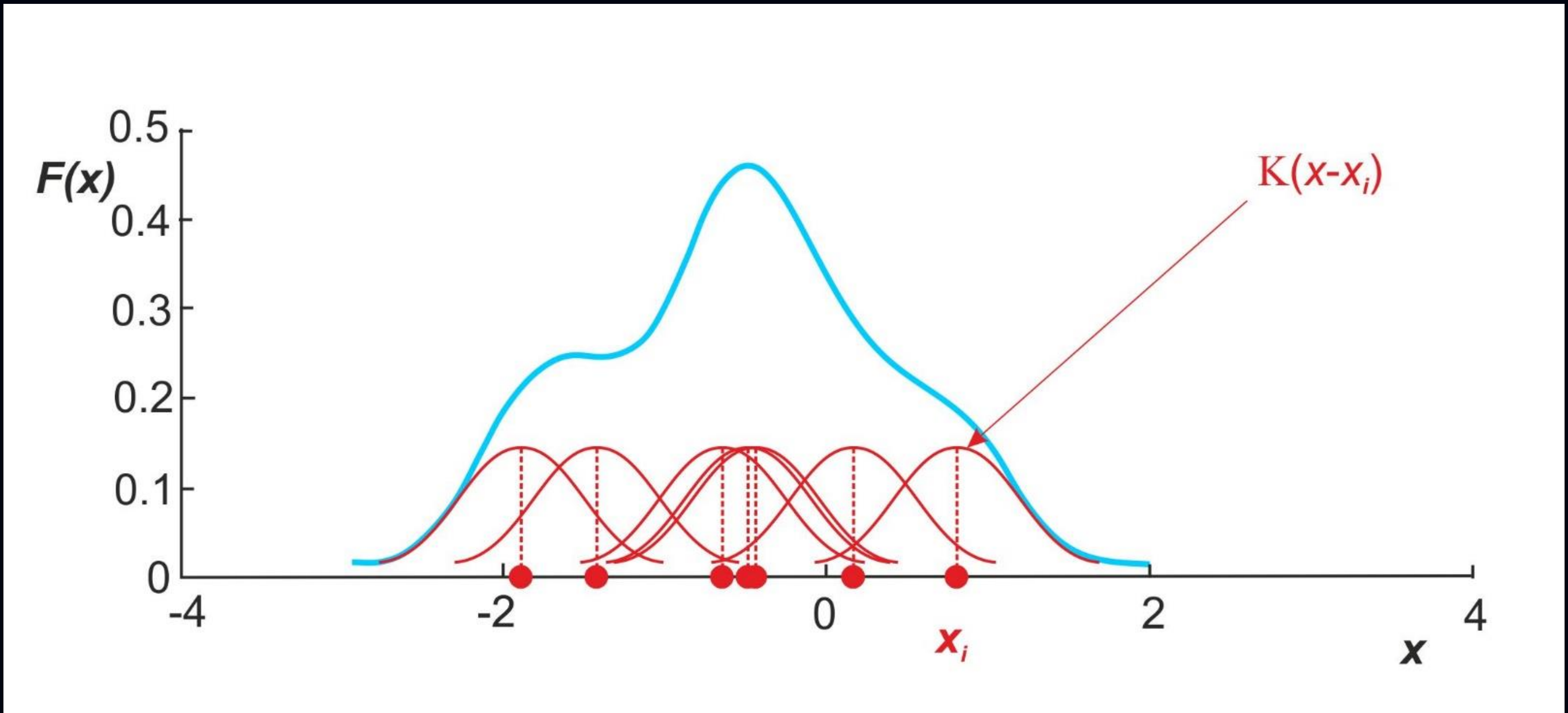
Гистограмма

- Не является непрерывной функцией
- Зависимость от выбора начальной точки
- Ошибка интервала[1]
- При определении дисперсии используется статистика Пуассона ($\sigma \sim \sqrt{N}$) [2,3].

$$\sigma \sim N^{\alpha}, \alpha = 0.89 \pm 0.04$$

1. Холопов П.Н. “Звездные скопления”, Наука, Москва, 1981
2. Василевский А.Е. “Методы звездной статистики”, Изд-во УрГУ, Свердловск, 1985
3. Danilov V.M., Matkin N.V., Pylskaya O.P., 1985, Sov. Astron., 29, 621

Метод функции-ядра (kernel estimator)



Результирующая оценка плотности вероятности наследует свойства функции-ядра (в том числе непрерывность и дифференцируемость).

Kernels

- Epanechnikov (quadratic) kernel

$$K(x) = \begin{cases} \frac{3}{4h}\left(1 - \frac{x^2}{h^2}\right), |x| \le h \\ 0 \qquad\qquad\qquad , |x| > h \end{cases}$$

- Quartic kernel

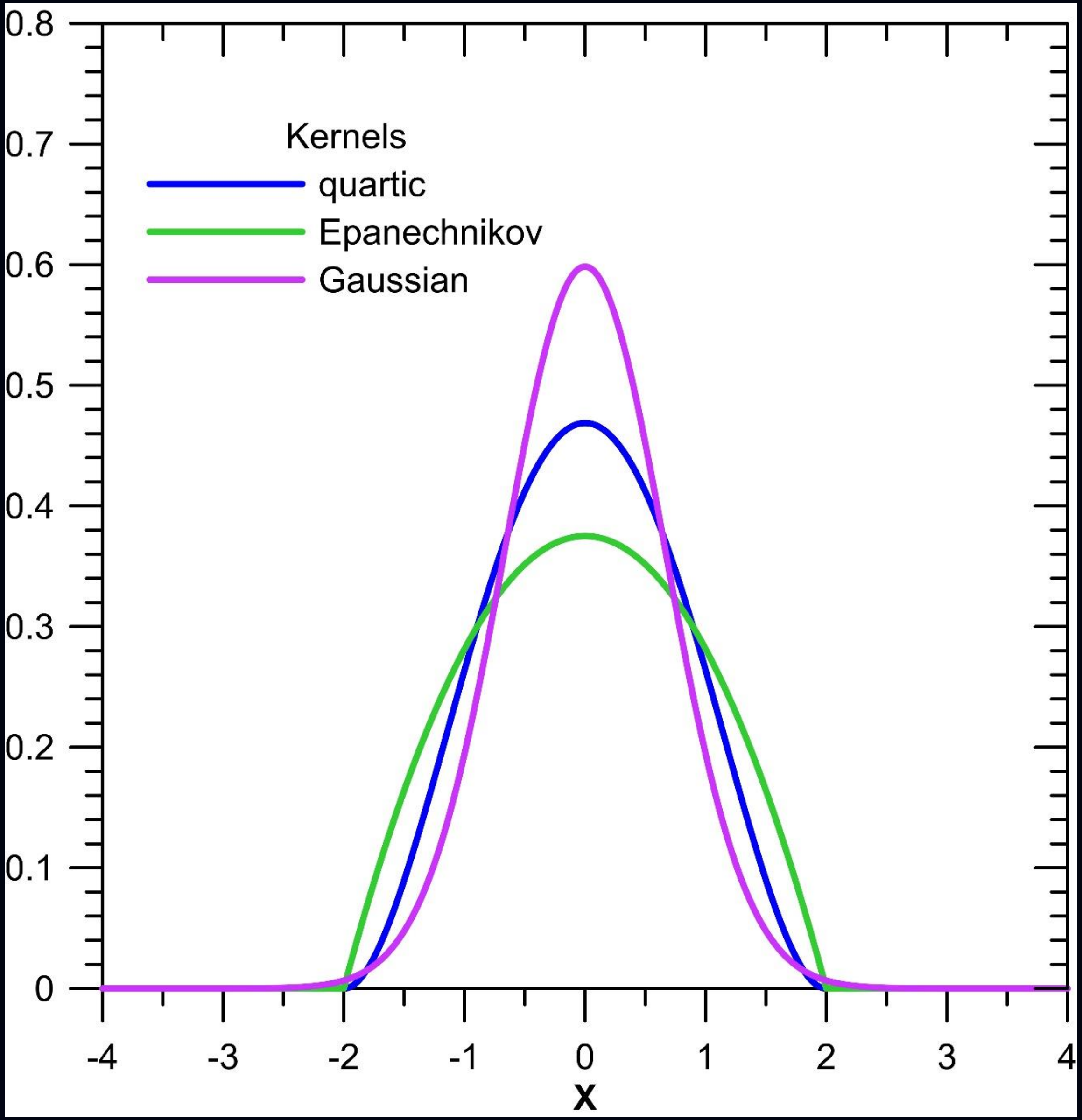
$$K(x) = \begin{cases} \frac{15}{16h}\left(1 - \frac{x^2}{h^2}\right)^2, |x| \le h \\ 0 \qquad\qquad\qquad , |x| > h \end{cases}$$

- Gaussian kernel

$$K(x)=\frac{1}{h\sqrt{2\pi}}\,e^{-\frac{x^2}{2h^2}}$$

В первых двух случаях h – полуширина ядра, в третьем - полуширина приближенно равна $3h$.

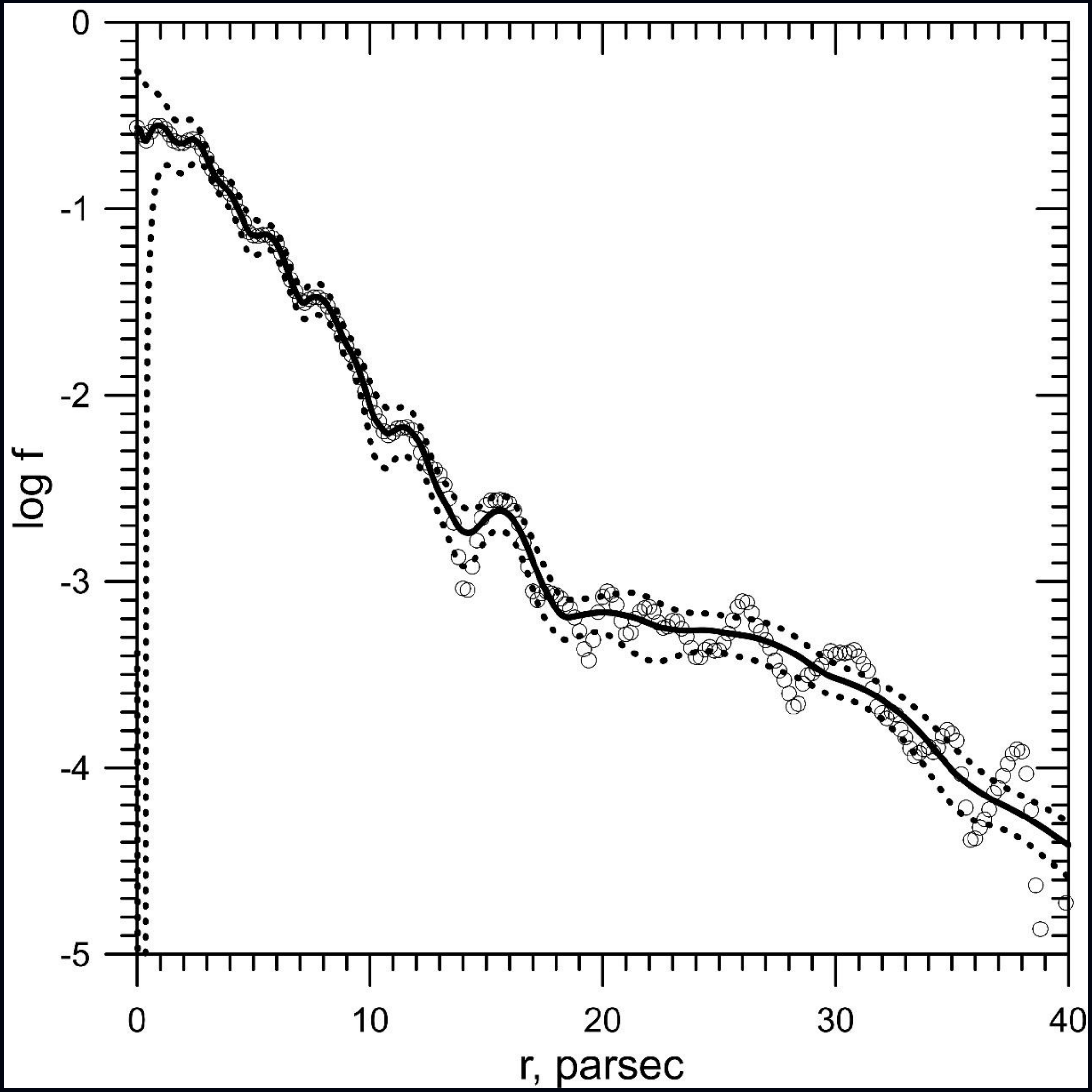
Kernels


Kernels
quartic
Epanechnikov
Gaussian
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-4
-3
-2
-1
0
1
2
3
4
X

Метод адаптивного ядра

- Полуширина ядра зависит от плотности.
- На первом шаге получается пилотная оценка – с фиксированным значением полуширины.
- На втором шаге, полуширина ядра в адаптивном методе обратно пропорциональна квадратному корню из пилотного значения плотности.
- Дает значительно лучший результат в случаях, когда отсутствует фон (например, при обработке результатов численного моделирования), при наличии протяженных слабых крыльев распределения.

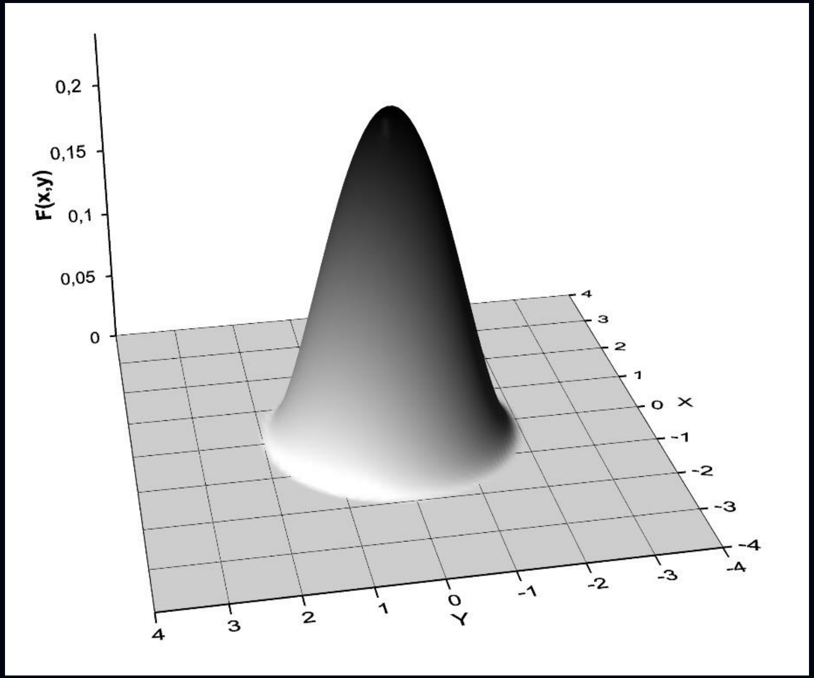
Метод адаптивного ядра



Обобщение для случая нескольких измерений

$$\hat{f}(\vec{x}) = \frac{1}{h^d} \sum_{i=1}^{n} K\left\{\frac{1}{h}\left(\vec{x} - \vec{X}_i\right)\right\}$$

$$\Delta F = \begin{cases} \dfrac{3}{\pi h^2}\left(1 - \dfrac{\rho^2}{h^2}\right)^2, |\rho| \le h \\ 0 \qquad\qquad\qquad\quad, |\rho| > h \end{cases}$$

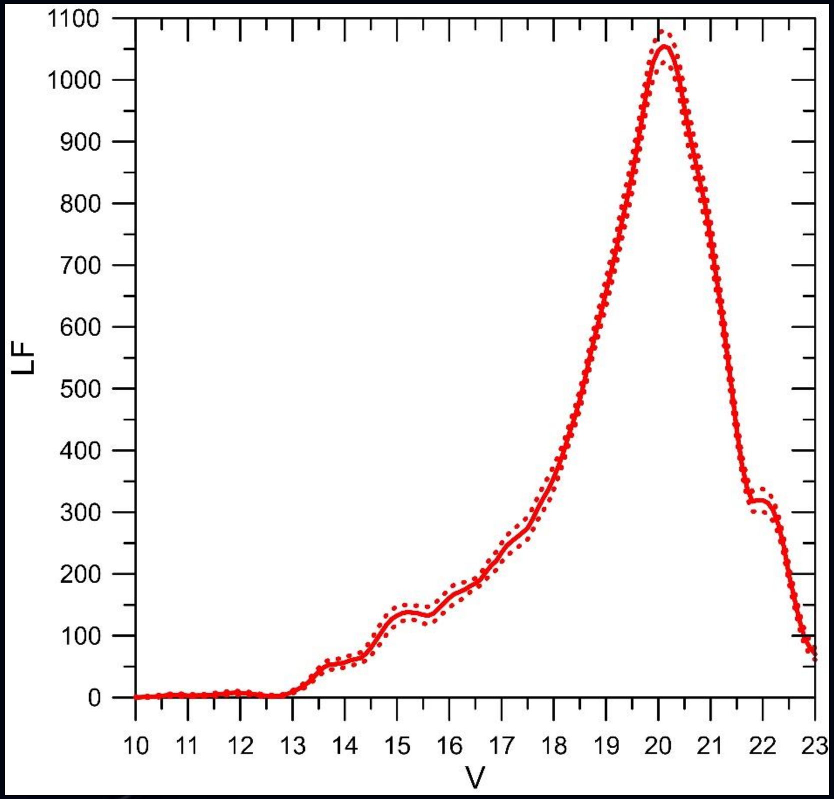


Обобщение для случая нескольких измерений

Размерность	Необходимый размер выборки
1	4
2	19
3	67
4	223
5	768
6	2790
7	10700

Размер выборки (с точностью до 3 значащих цифр), необходимый для достижения относительной среднеквадратичной ошибки 0.1 в точке x=0 при оценивании стандартного многомерного нормального распределения с помощью гауссового ядра и значением полуширины, которое минимизирует среднеквадратичную ошибку при x=0.

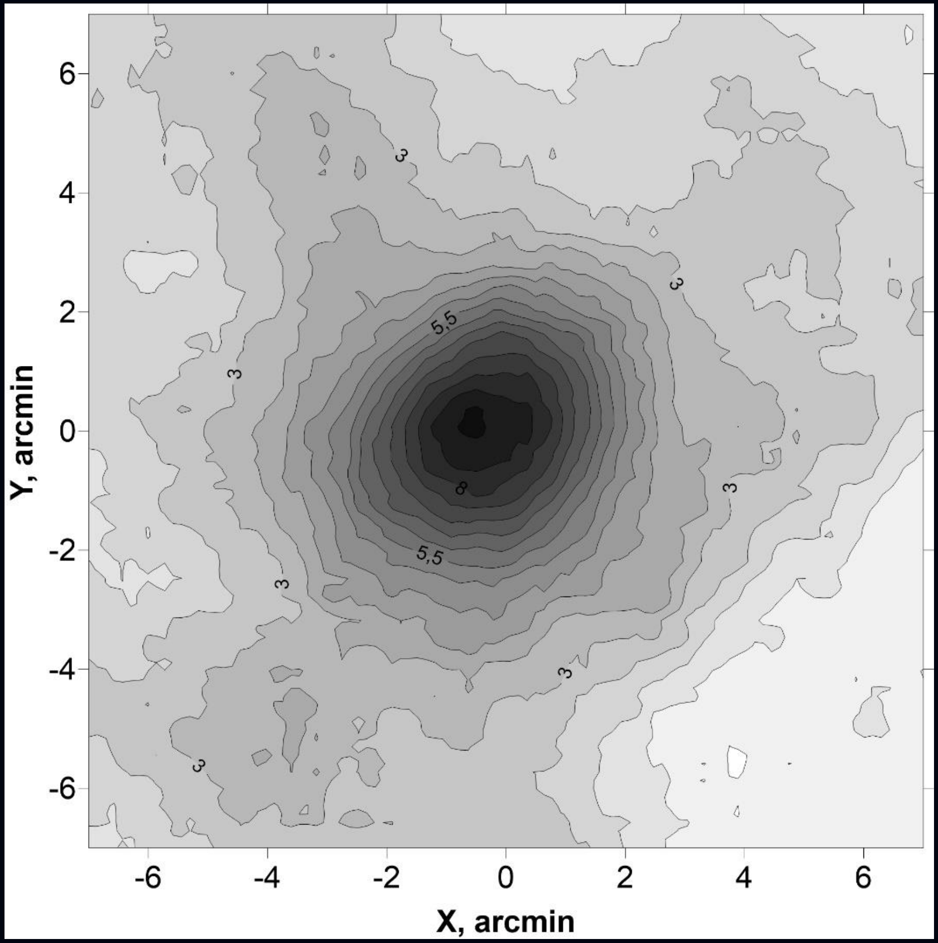
Построение доверительного интервала – smoothed bootstrap algorithm



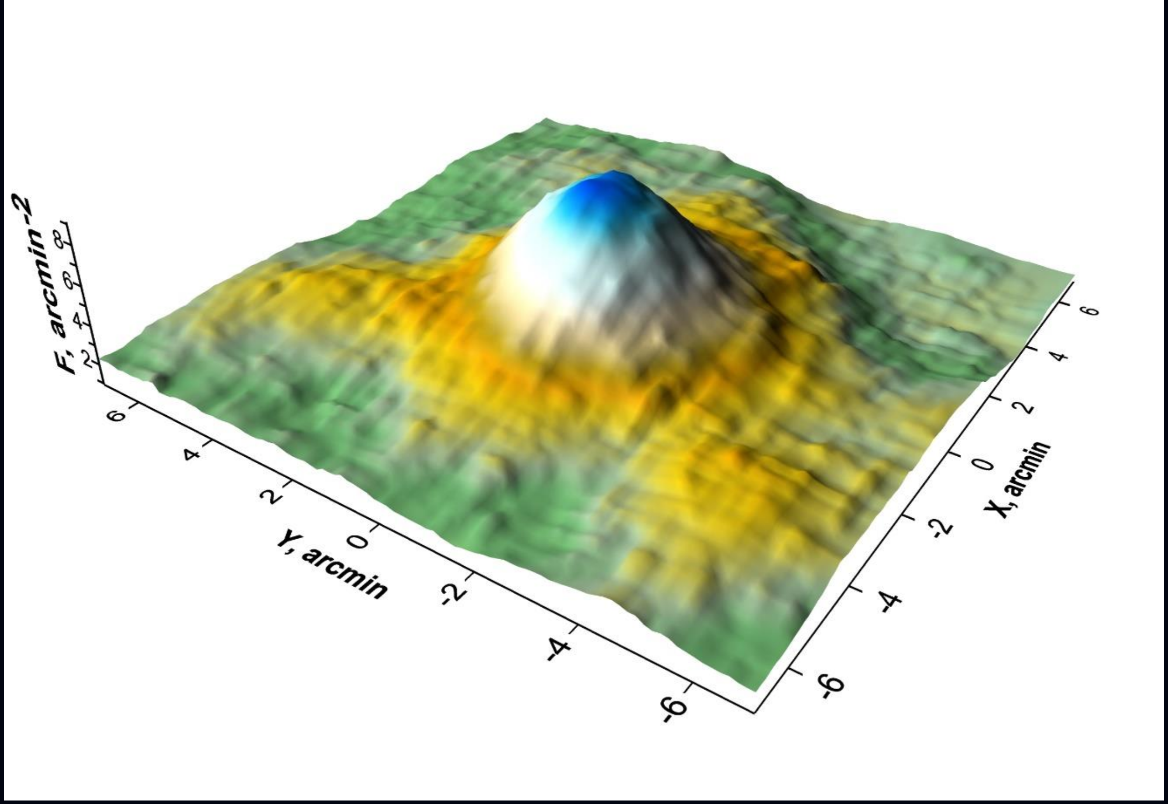
Для построения доверительного интервала были использованы 20 вторичных выборок (объем вторичной выборки должен совпадать с объемом исходной выборки). При «набрасывании» вторичной выборки используется метод Неймана. Исходная оценка аппроксимируется кубическим сплайном.

Двумерные карты распределения поверхностной плотности

Значения плотности рассчитываются на равномерной сетке координат в картинной плоскости. После этого строятся изолинии плотности или проекция 3-мерной поверхности.



Двумерные карты распределения поверхностной плотности



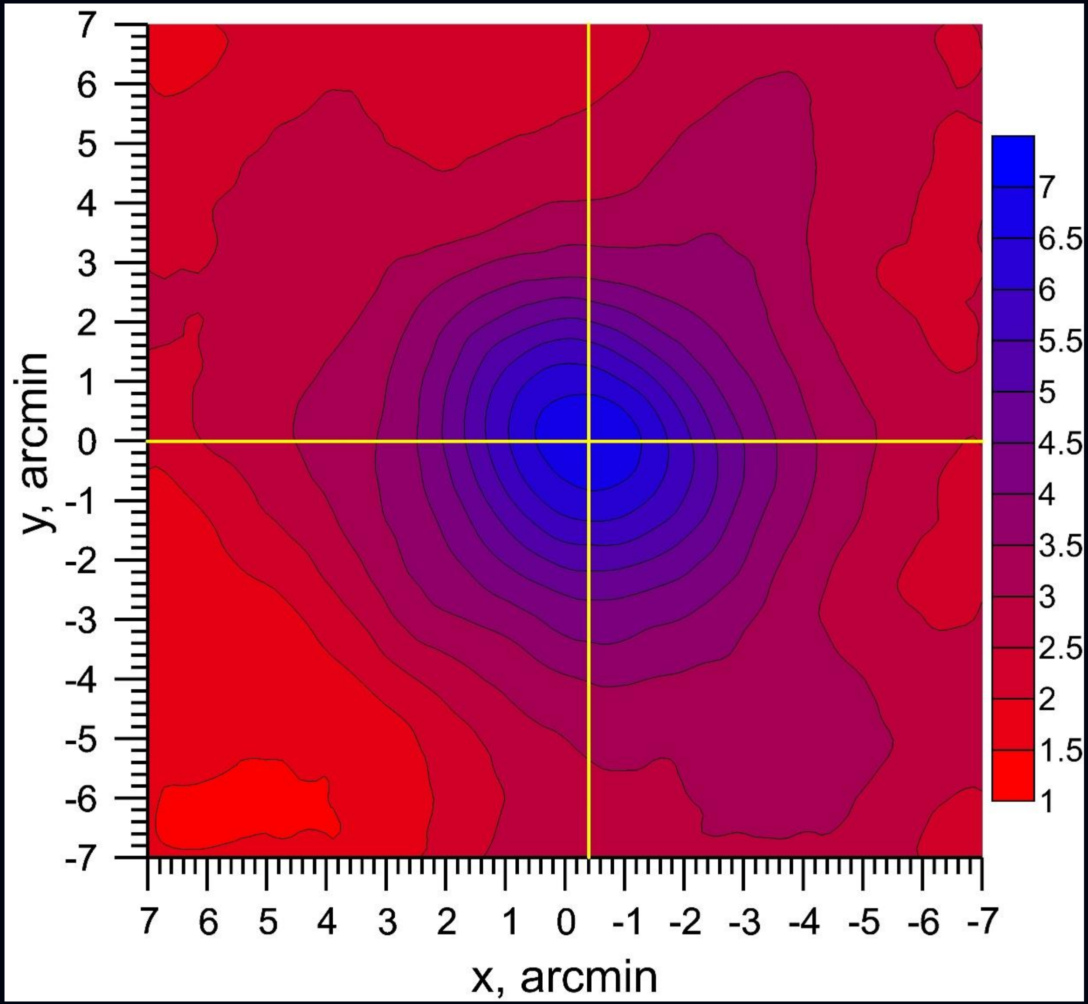
Необходимо отступать от границ области на величину h.

Двумерные карты распределения поверхностной плотности

- Общее представление об изучаемой области
- Предварительная оценка размеров скопления
- Выбор областей сравнения (например, для построения функции светимости скопления)
- Определение координат центра скопления
- Изучение структуры скопления (например, исследование эллиптичности)
- …

Определение координат центра скопления

NGC 4337



Исследование эллиптичности распределения звезд разных населений в ω Cen

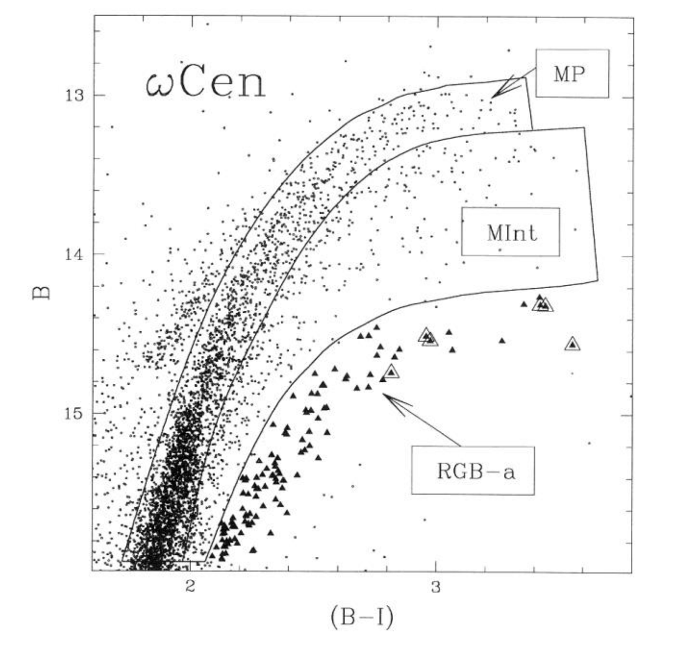
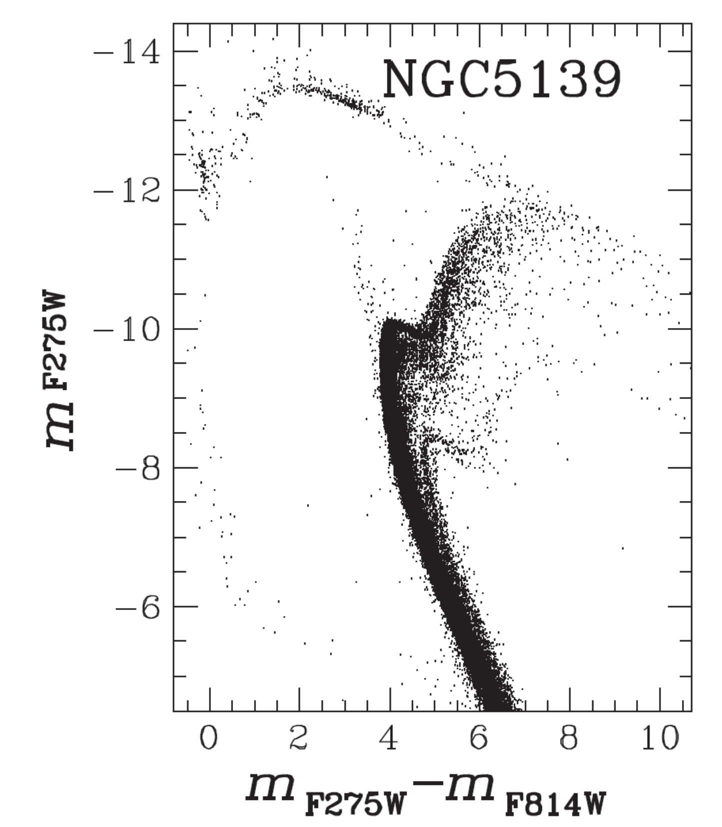
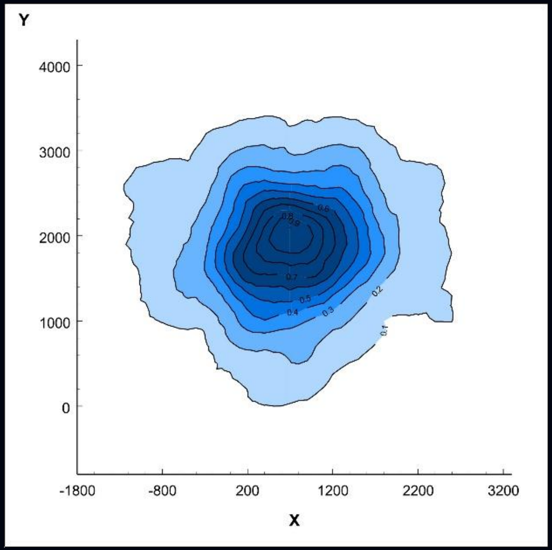


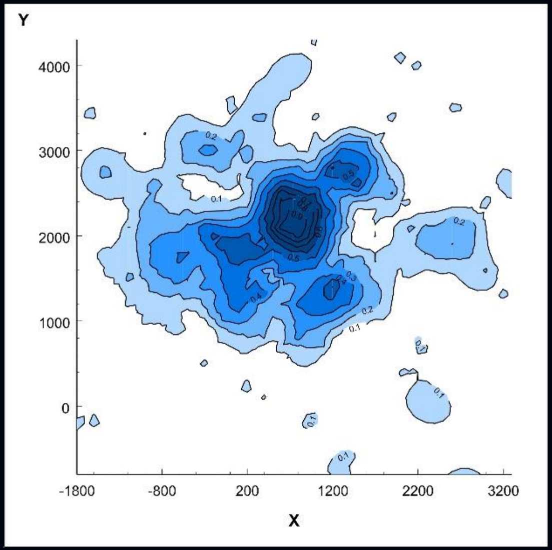
Figure 3.7: The definition of the three RGB sub-populations in ω Cen, as decribed in Section 3.3. This catalogation of the sub-populations in ω Cen constitutes the basis of the analysis presented in the following Chapters.



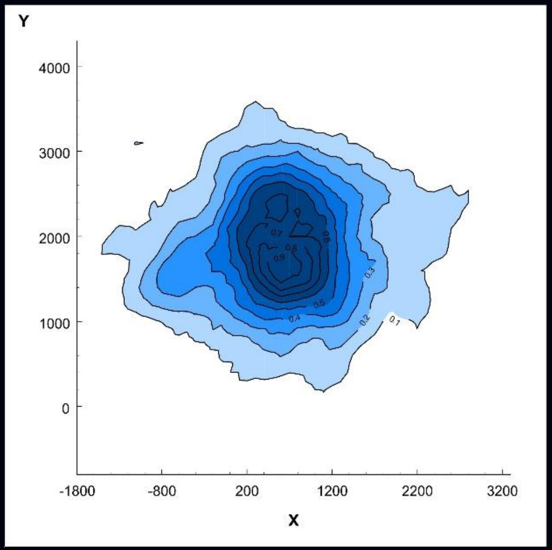
Исследование эллиптичности распределения звезд разных населений в ω Cen



metal-poor

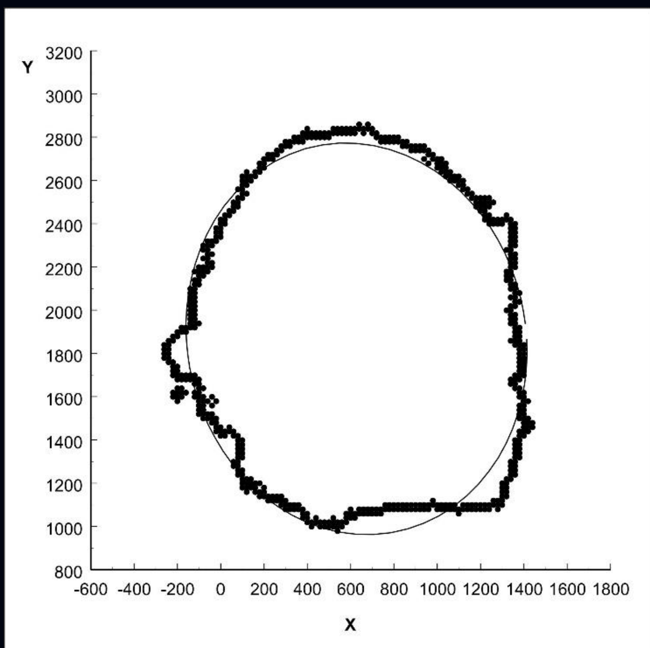


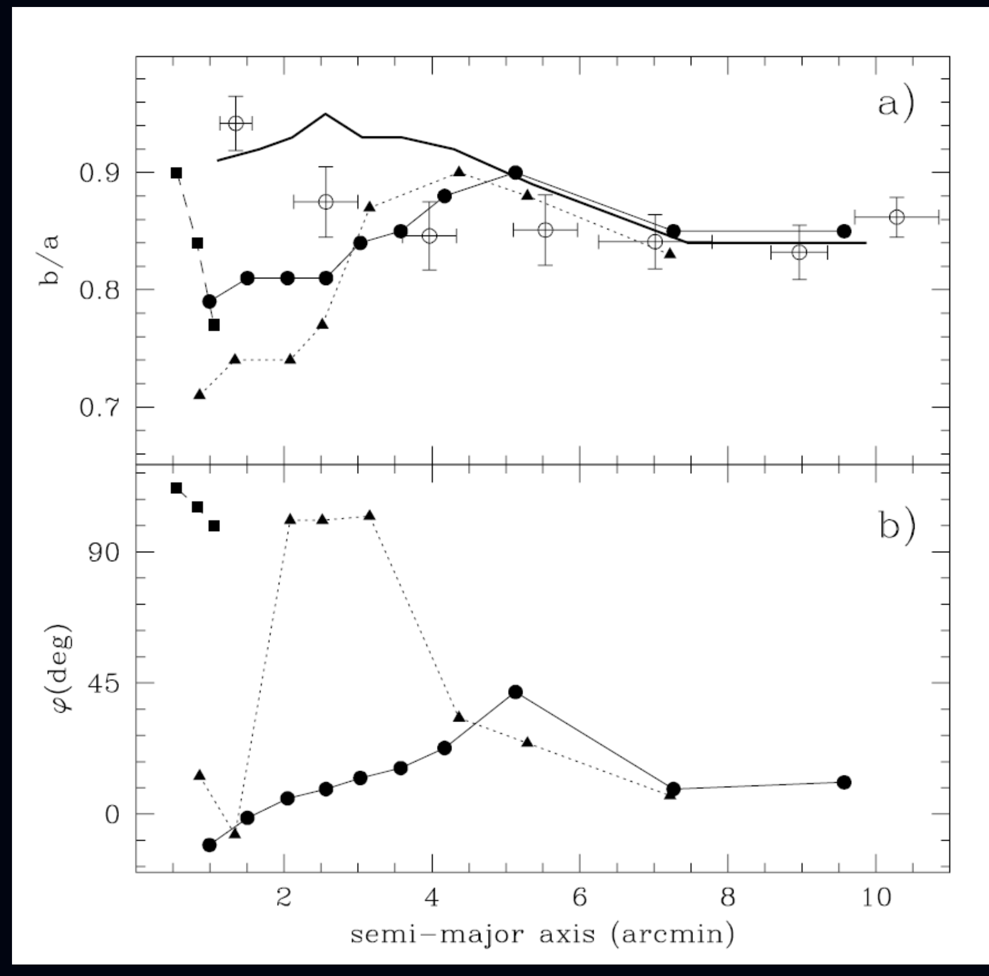
metal-rich



intermediate

Исследование эллиптичности распределения звезд разных населений в ω Cen





Профили поверхностной и пространственной плотности

Merritt D., Tremblay B., 1994, AJ, 108, 514:

- Показано, что одномерный метод функции-ядра не подходит, и требуется использование двумерного метода.
- Получены формулы для функции-ядра для случая построения радиального профиля поверхностной плотности.
- Исследована работа обоих методов (функции-ядра и максимального правдоподобия) по восстановлению профилей плотности в трех важных случаях распределений (Пламмера, де Вокулера, Кинга-Мичи).

Профили поверхностной и пространственной плотности

- Показано, что использование «оптимальной» ширины ядра, определяемой при минимизации интегральной среднеквадратичной ошибки, приводит к неудовлетворительному результату. Дается рекомендация подходить к выбору полуширины ядра h эмпирически - получать набор оценок при разных значениях h и выбирать наилучший вариант.

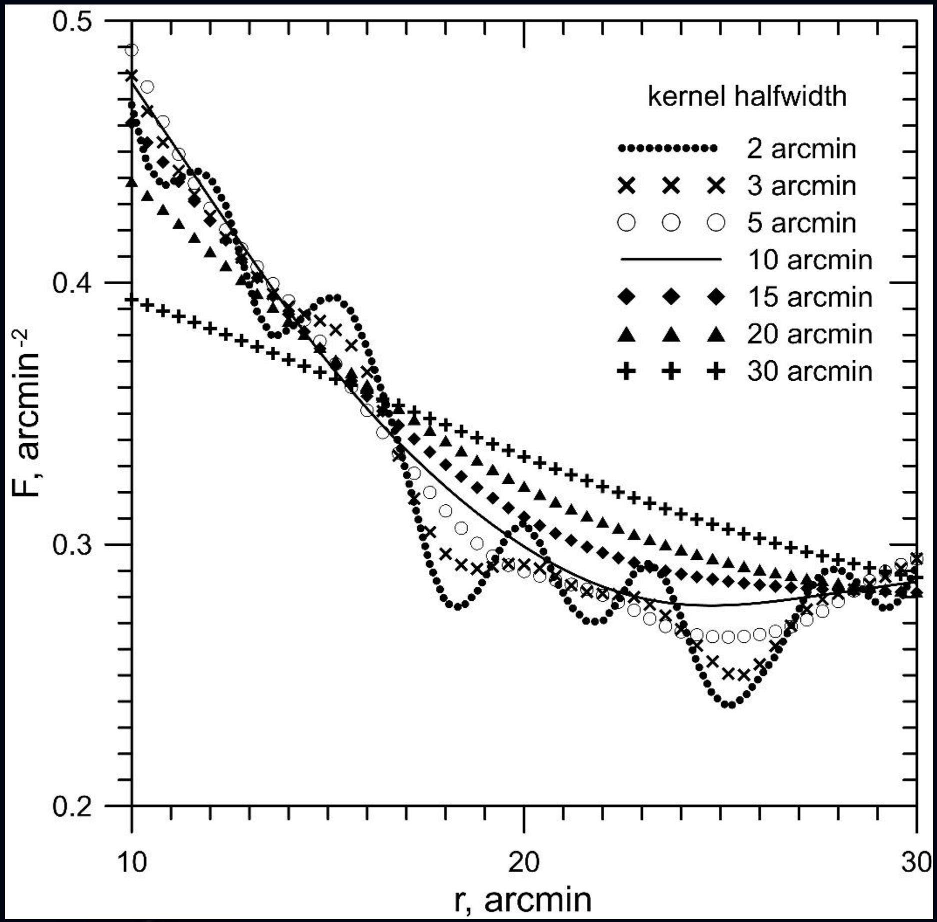
‘…simply looking at plots produced using several different values of the smoothing parameter, and accepting the one that is as smooth as possible without being obviously biased – that is, the smoothest curve that closely follows the mean trend defined by curves computed with much smaller smoothing parameter'

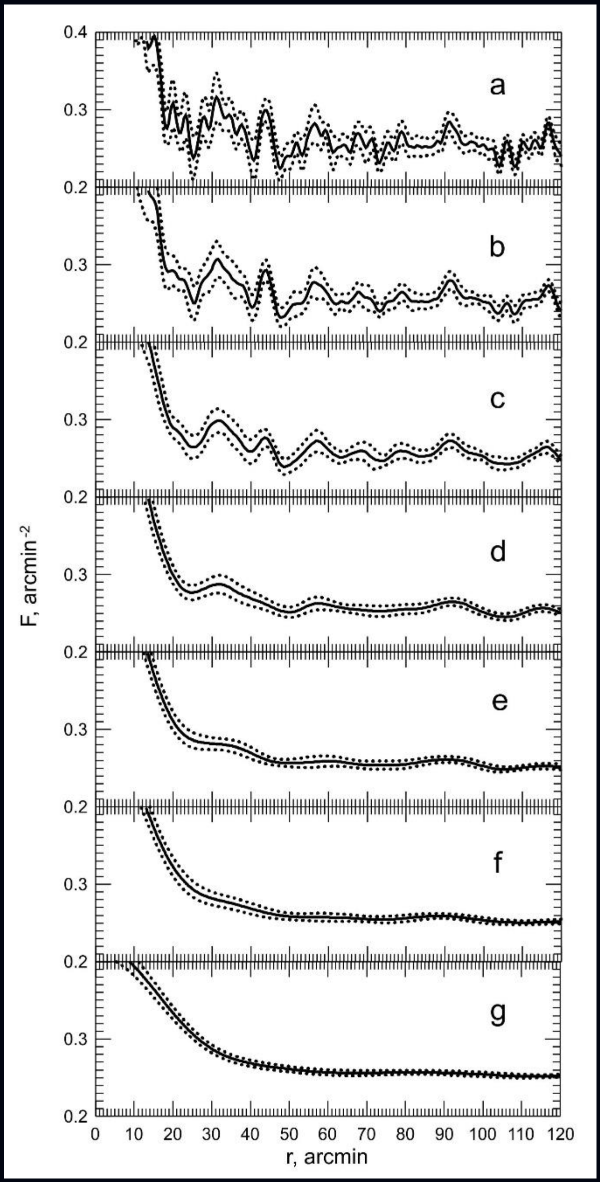
Профили поверхностной и пространственной плотности

Seleznev A.F., 2016, MNRAS, 456, 3757:

- Получены формулы для построения профиля пространственной плотности в случае, когда известны координаты звезд (x,y,z)
- Профили поверхностной плотности до разных предельных величин построены для 7 рассеянных звездных скоплений

Выбор оптимальной полуширины ядра

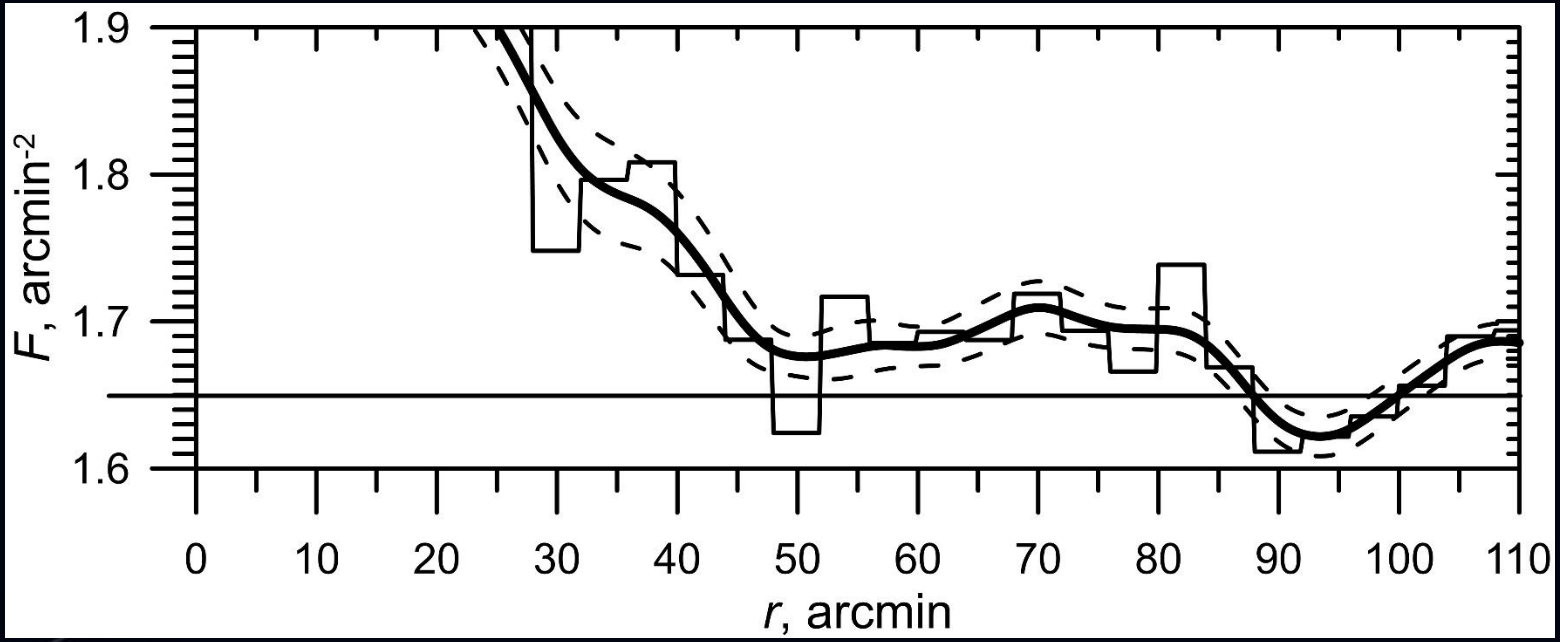




NGC 2287, 2MASS PSC, $J_{lim}=13^m$

Оценка числа звезд и массы звездного скопления

$$N=2\pi \int_0^R F(r)rdr$$



NGC 2516, 2MASS PSC, $J_{lim}=16^m$

Переход к пространственной плотности

$$f(r) = \frac{1}{\pi}\int_0^{\sqrt{R_c^2 - r^2}} S\left(\sqrt{r^2 + z^2}\right) dz$$

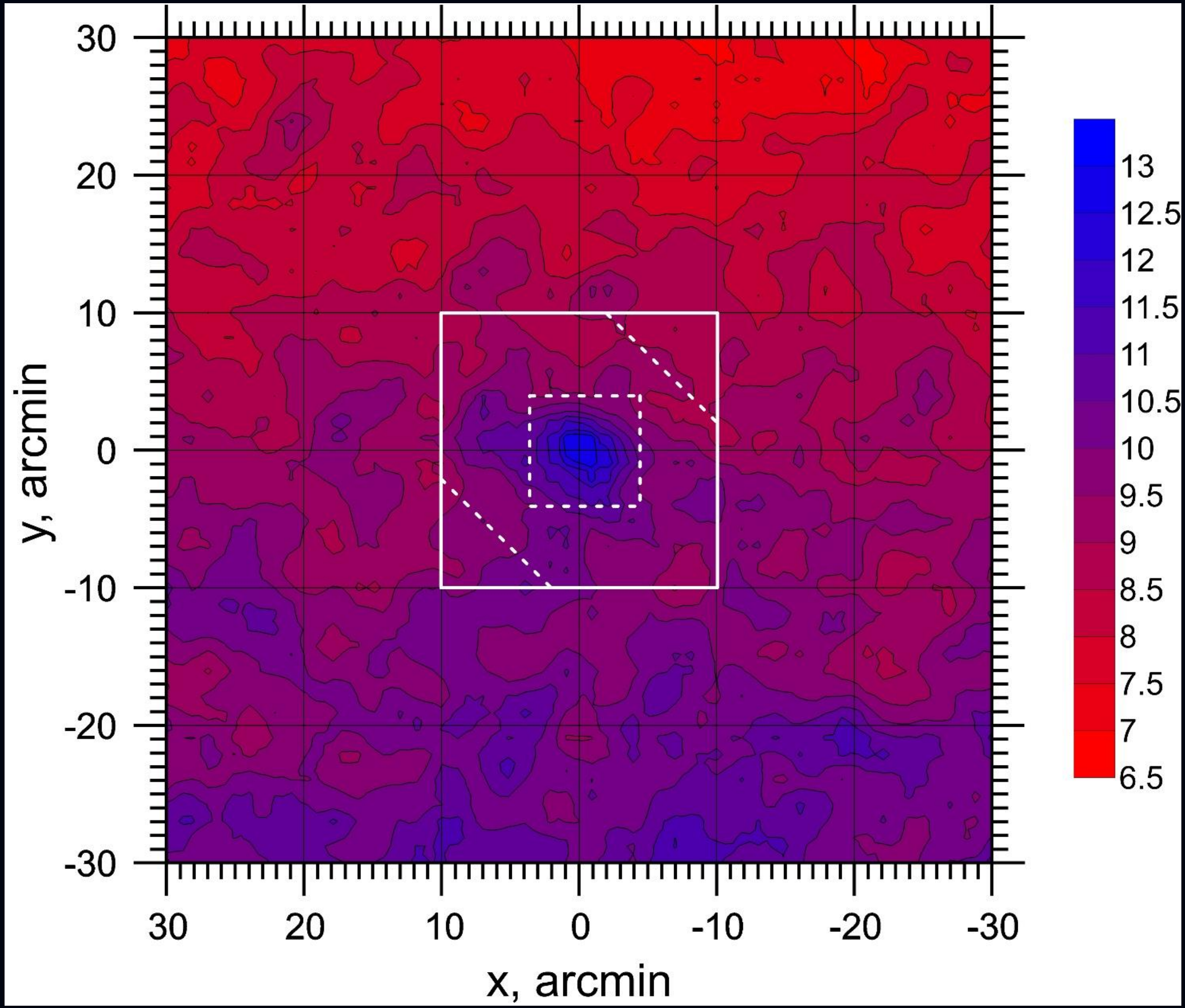
$$S(r) = -\frac{1}{r}\frac{dF(r)}{dr}$$

Пространственная плотность может быть использована, например, для моделирования скопления методом Монте-Карло с целью определения его структурных параметров:

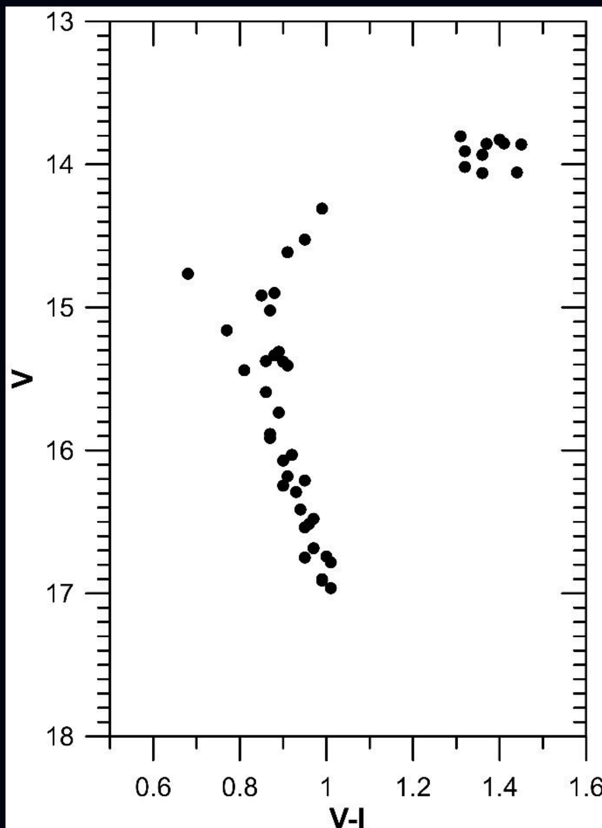
$$\bar{R} = \left\langle 1/r_{ij} \right\rangle^{-1} \qquad R_u = \langle 1/r_i \rangle^{-1} \qquad \langle r^2 \rangle$$

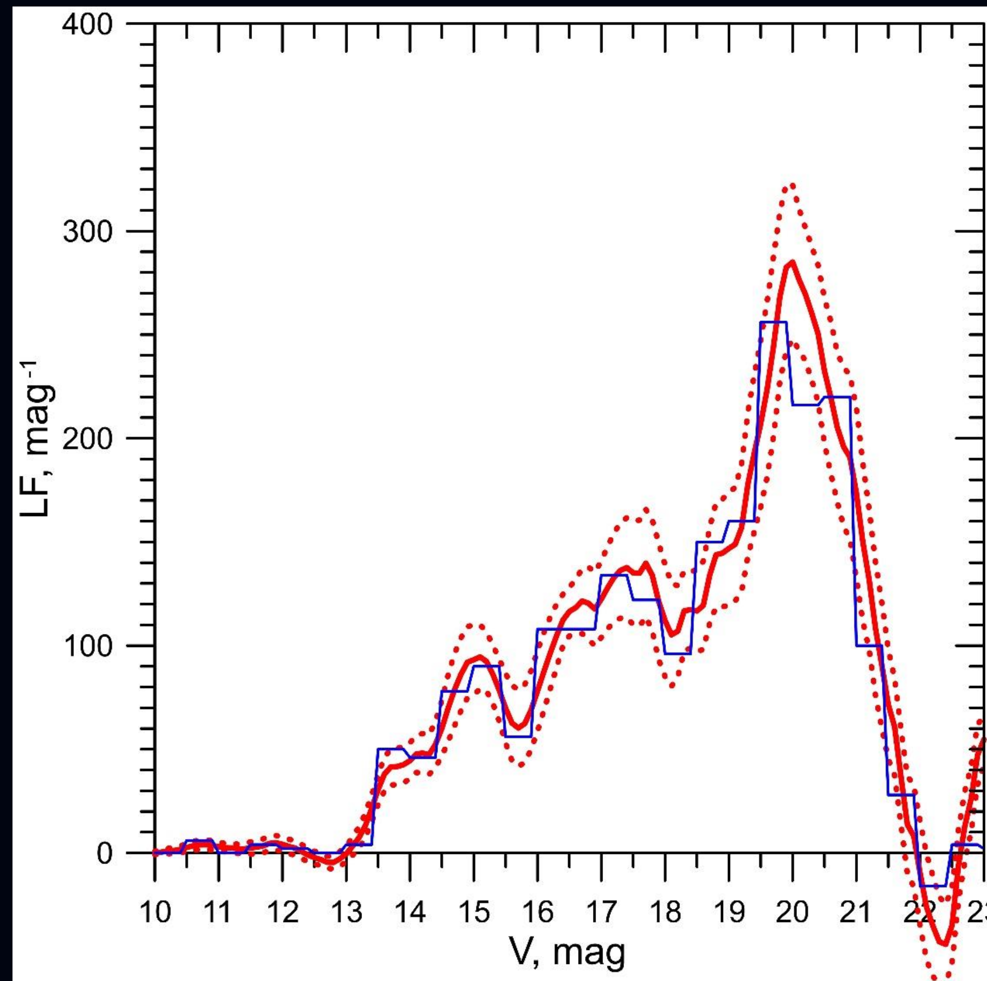
Функция светимости

NGC 4337, 2MASS PSC, карта области для выбора площадок сравнения.



Функция светимости





NGC 4337, функция светимости звезд скопления

Функция масс

$$\psi(m)=\frac{dn}{dm}$$

$$\int_{m_1}^{m_2}\psi(m)\mathrm{dm}=\mathrm{N}$$

$$\int_{m_1}^{m_2}\psi(m)\mathrm{mdm}=\mathfrak{M}$$

$$\varphi(V)=\frac{dn}{dV}$$

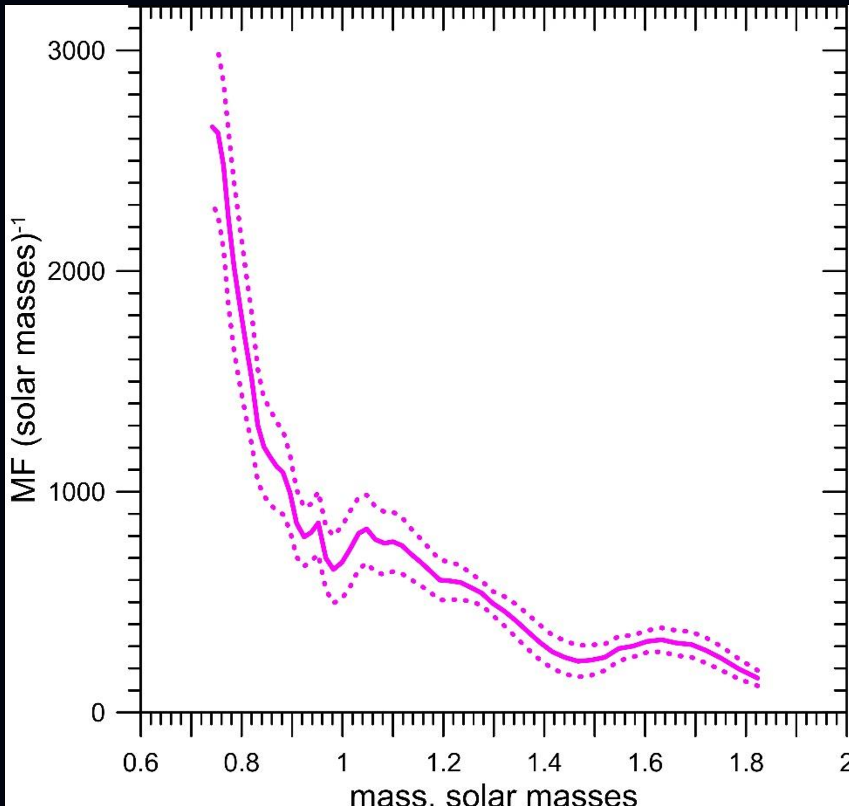
$$\int_{V_1}^{V_2}\varphi(V)\mathrm{dV}=\mathrm{N}$$

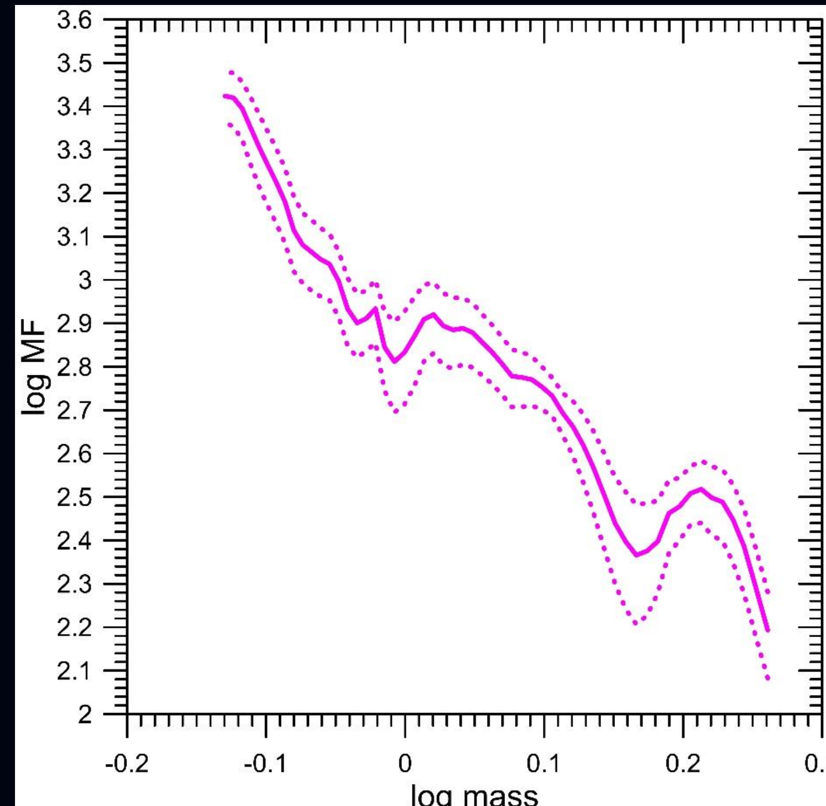
$$\mathrm{m}=\mathrm{m}(V)$$

$$\psi(m)=\frac{\varphi(V)}{m_V'}$$

(m(V) – соотношение масса-светимость, Padova suite of models, Bressan A. et al., 2012, MNRAS, 427, 127)

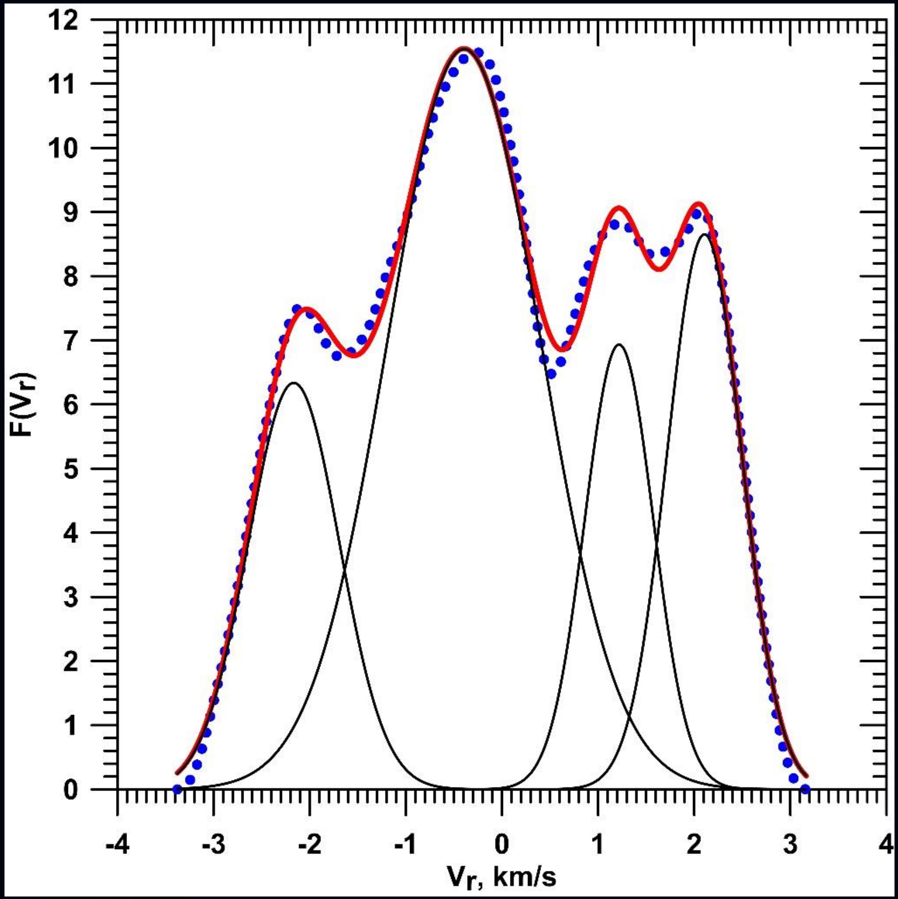
Функция масс

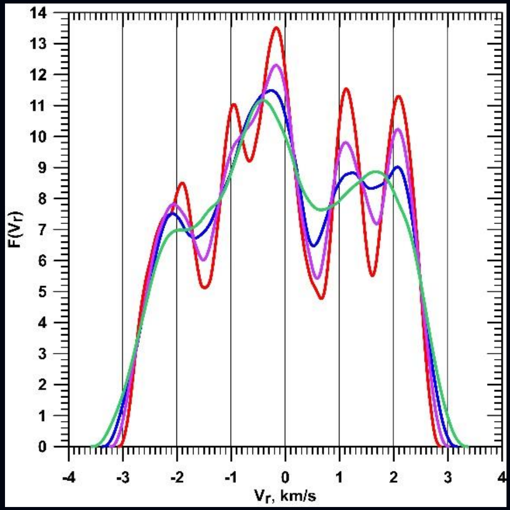


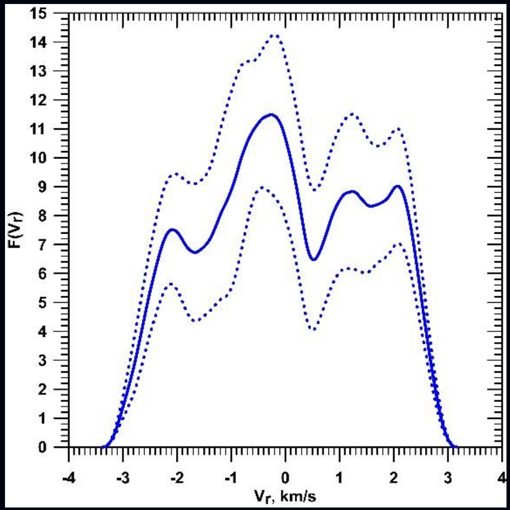


Наклон логарифмической функции масс -2.57±0.10

Функция распределения скоростей

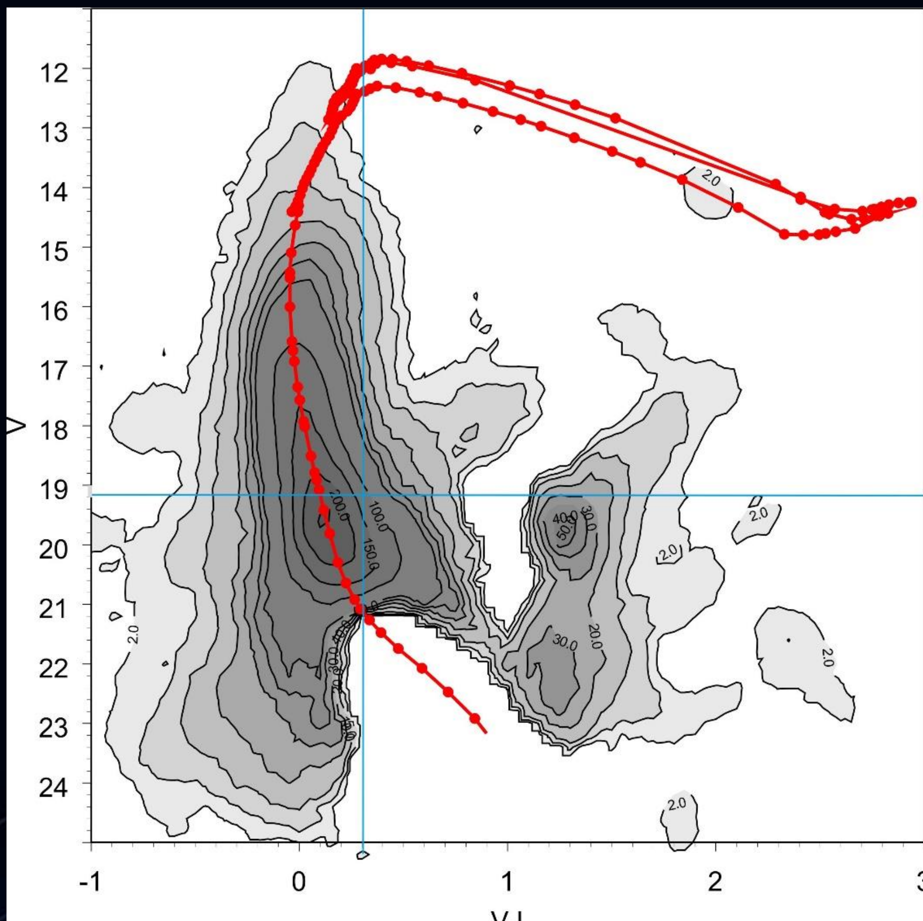


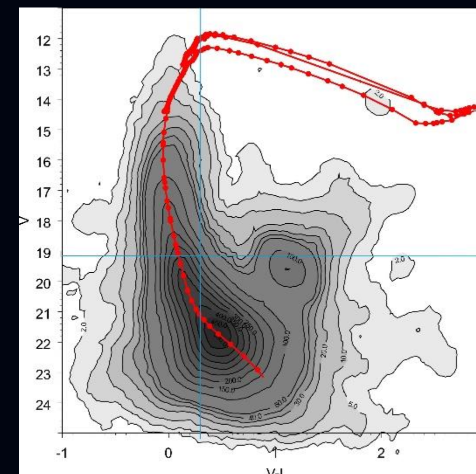


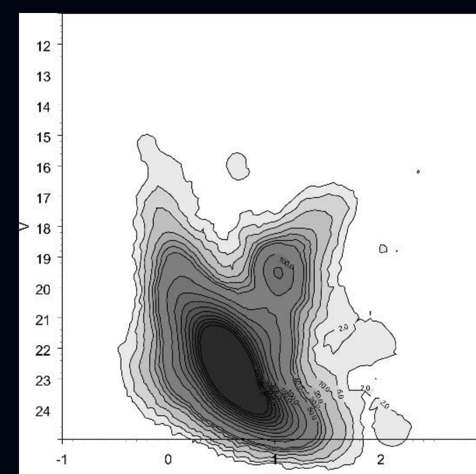


NGC 4337, вероятные члены скопления, отобранные по лучевым скоростям

Диаграммы звездная величина – показатель цвета (диаграмма Хесса)







NGC 2100

Выводы

- Метод функции-ядра может быть эффективно использован при решении многих задач звездной астрономии, везде, где необходимо оценивать и использовать функции распределения.
- Получаемые оценки функций распределения являются непрерывными и дифференцируемыми функциями, что очень важно для их использования.
- Написано большое количество программ для различных вариантов использования метода функции-ядра (Fortran), которые я с удовольствием предоставлю вместе с подробными инструкциями.

Спасибо за внимание!

Anton.Seleznev@urfu.ru