

Inferring the velocity distribution of nearby stars from *Hipparcos* data

J. Bovy

Center for Cosmology and Particle Physics, New York University

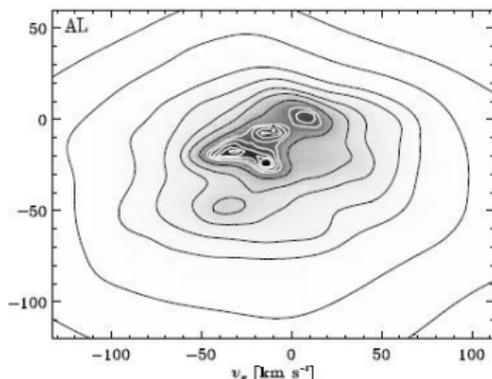
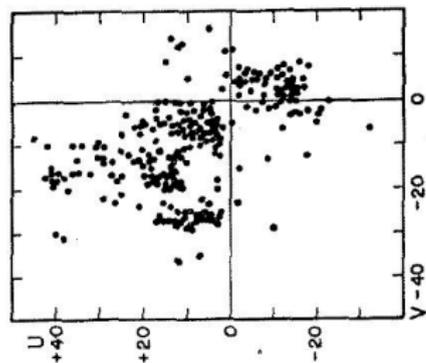
New York Workshop on Computer, Earth, and Space
Sciences, Feb. 2009

Density estimation in the presence of noisy, heterogeneous, and incomplete data

- ▶ Astronomical data sets: Low S/N, incomplete data . . . on a per observation basis
- ▶ Find *underlying* distribution, *not* observed distribution
- ▶ Extreme deconvolution: Each sample is drawn from a different distribution
- ▶ But CS only gives us methods for noiseless data!! (or noise = constant)

Velocity distribution from *Hipparcos* data

- ▶ *Hipparcos*: positions, proper motions and parallaxes for nearby stars
- ▶ Infer velocity distribution
→ streams
- ▶ Low S/N ($\sigma_\pi/\pi \sim 10\%$), incomplete data (no radial velocities).



O. Eggen 1965, W. Dehnen 1998

Modeling the distribution function

- ▶ Distribution = Sum over K Gaussians
- ▶ Fit for amplitudes, means, and covariances
- ▶ Observations: Noisy projections of true values
 - ▶ $\mathbf{w}_j = \mathbf{R}_j \mathbf{v}_j + \text{noise}$
 - ▶ Gaussian noise
 - ▶ incomplete data \equiv noisy data $\rightarrow \mathbf{R}_j$

Objective function for optimization

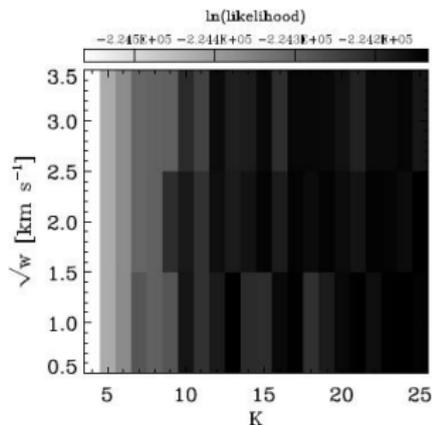
- ▶ Likelihood of the model = \prod_i probability of the data point given the model
- ▶ $P(\text{data point} \mid \text{model}) = (\sum \text{Gaussians}) * (\text{noise})$
- ▶ ...optimize
- ▶ And we're done!

Objective function for optimization

- ▶ Likelihood of the model = \prod_i probability of the data point given the model
- ▶ $P(\text{data point} \mid \text{model}) = (\sum \text{Gaussians}) * (\text{noise})$
- ▶ ... optimize
- ▶ And we're done!

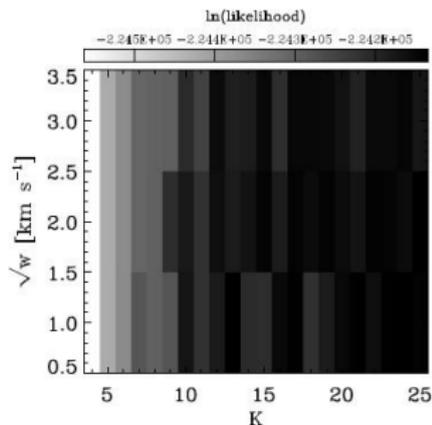
Not so fast! Complications

- ▶ Optimization is hard
 - ▶ Generic optimizer: constraints on amplitudes, covariances
 - ▶ Expectation-Maximization: Deals naturally with incomplete data, but slow
- ▶ Prior information: use conjugate priors to
 - ▶ Regularize covariances
 - ▶ Regularize amplitudes
 - ▶ Works well with EM
- ▶ Local maxima: use “split-and-merge” extension



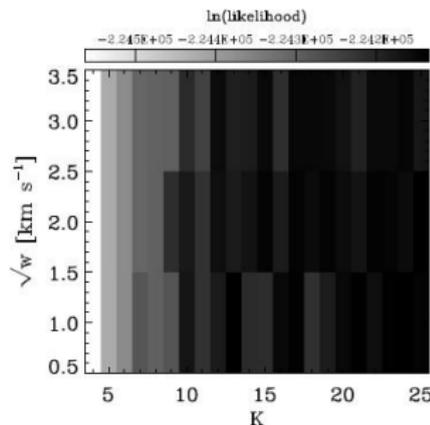
Not so fast! Complications

- ▶ Optimization is hard
 - ▶ Generic optimizer: constraints on amplitudes, covariances
 - ▶ Expectation-Maximization: Deals naturally with incomplete data, but slow
- ▶ Prior information: use conjugate priors to
 - ▶ Regularize covariances
 - ▶ Regularize amplitudes
 - ▶ Works well with EM
- ▶ Local maxima: use “split-and-merge” extension



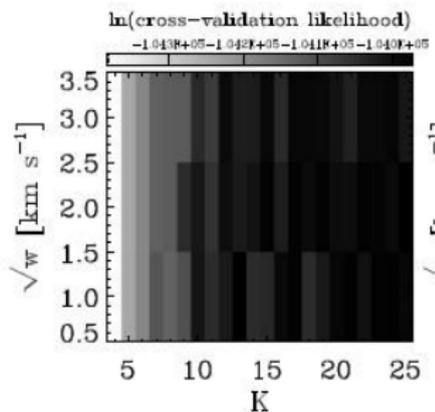
Not so fast! Complications

- ▶ Optimization is hard
 - ▶ Generic optimizer: constraints on amplitudes, covariances
 - ▶ Expectation-Maximization: Deals naturally with incomplete data, but slow
- ▶ Prior information: use conjugate priors to
 - ▶ Regularize covariances
 - ▶ Regularize amplitudes
 - ▶ Works well with EM
- ▶ Local maxima: use “split-and-merge” extension



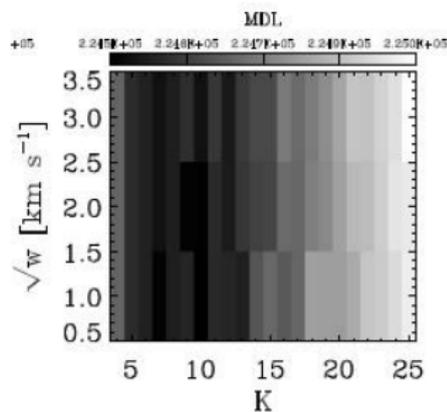
Model selection: Set # of Gaussians K (+hyperparameters)

- ▶ Cross-validation: slow, impractical
- ▶ Minimum coding inference: best model has shortest message length
- ▶ Probability of an external data set: e.g., radial velocities



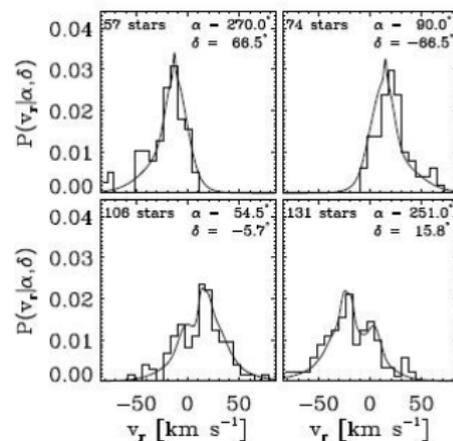
Model selection: Set # of Gaussians K (+hyperparameters)

- ▶ Cross-validation: slow, impractical
- ▶ Minimum coding inference: best model has shortest message length
- ▶ Probability of an external data set: e.g., radial velocities



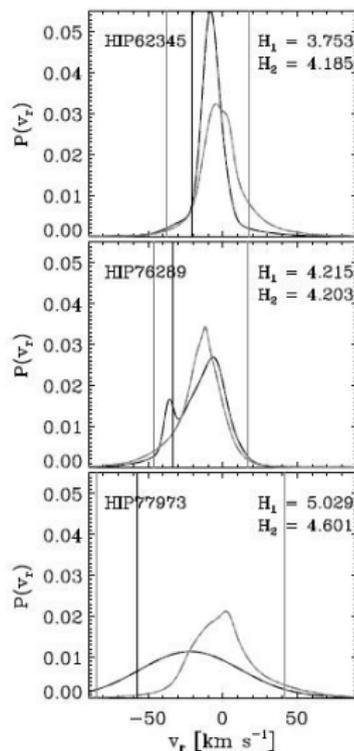
Model selection: Set # of Gaussians K (+hyperparameters)

- ▶ Cross-validation: slow, impractical
- ▶ Minimum coding inference: best model has shortest message length
- ▶ Probability of an external data set: e.g., radial velocities



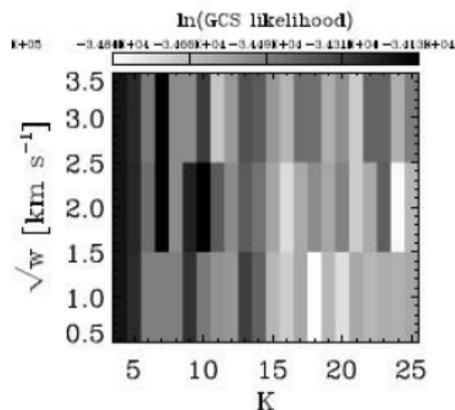
Model selection: Set # of Gaussians K (+hyperparameters)

- ▶ Cross-validation: slow, impractical
- ▶ Minimum coding inference: best model has shortest message length
- ▶ Probability of an external data set: e.g., radial velocities

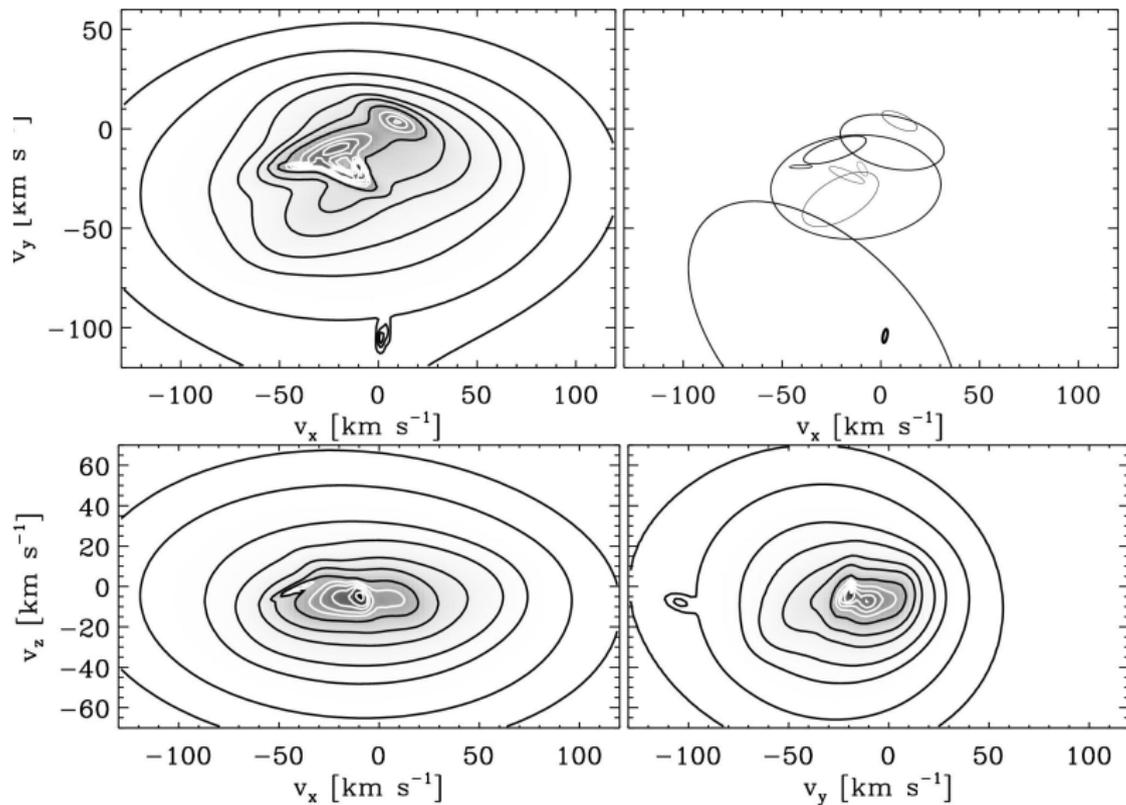


Model selection: Set # of Gaussians K (+hyperparameters)

- ▶ Cross-validation: slow, impractical
- ▶ Minimum coding inference: best model has shortest message length
- ▶ Probability of an external data set: e.g., radial velocities



Preliminary Results



Summary

- ▶ New technique for deconvolving an observed distribution function consisting of
 1. justified, scalar objective function
 2. Stable optimizer
 3. Model selection recipe
- ▶ Establish the statistical significance of the “moving groups” in the velocity distribution
- ▶ Joint work with: David Hogg (NYU), Sam Roweis (Toronto)

Summary

- ▶ New technique for deconvolving an observed distribution function consisting of
 1. justified, scalar objective function
 2. Stable optimizer
 3. Model selection recipe
- ▶ Establish the statistical significance of the “moving groups” in the velocity distribution
- ▶ Joint work with: David Hogg (NYU), Sam Roweis (Toronto)

Summary

- ▶ New technique for deconvolving an observed distribution function consisting of
 1. justified, scalar objective function
 2. Stable optimizer
 3. Model selection recipe
- ▶ Establish the statistical significance of the “moving groups” in the velocity distribution
- ▶ Joint work with: David Hogg (NYU), Sam Roweis (Toronto)