Improved Astronomical Inferences via Nonparametric Density Estimation

Chad M. Schafer, InCA Group

www.incagroup.org

Department of Statistics

Carnegie Mellon University

Work Supported by NSF, NASA-AISR Grant

January 2010

The Core Collaborators

Susan M. Buchman

Peter E. Freeman

Ann B. Lee

Joseph W. Richards

Motivation

Theory predicts **the distribution of** observables as a function of cosmological parameters.

Motivation

For example,

$$\begin{aligned}
\Omega_m &= \text{total matter density} \\
\Omega_b &= \text{baryonic matter density} \\
\Omega_\Lambda &= \text{dark energy density} \\
H_0 &= \text{the Hubble parameter} \\
\tau &= \text{the optical depth} \\
n_s &= \text{spectral index of initial spectrum} \\
A &= \text{amplitude of initial spectrum}
\end{aligned}$$

parameterize the power spectrum of the CMB anisotropy.

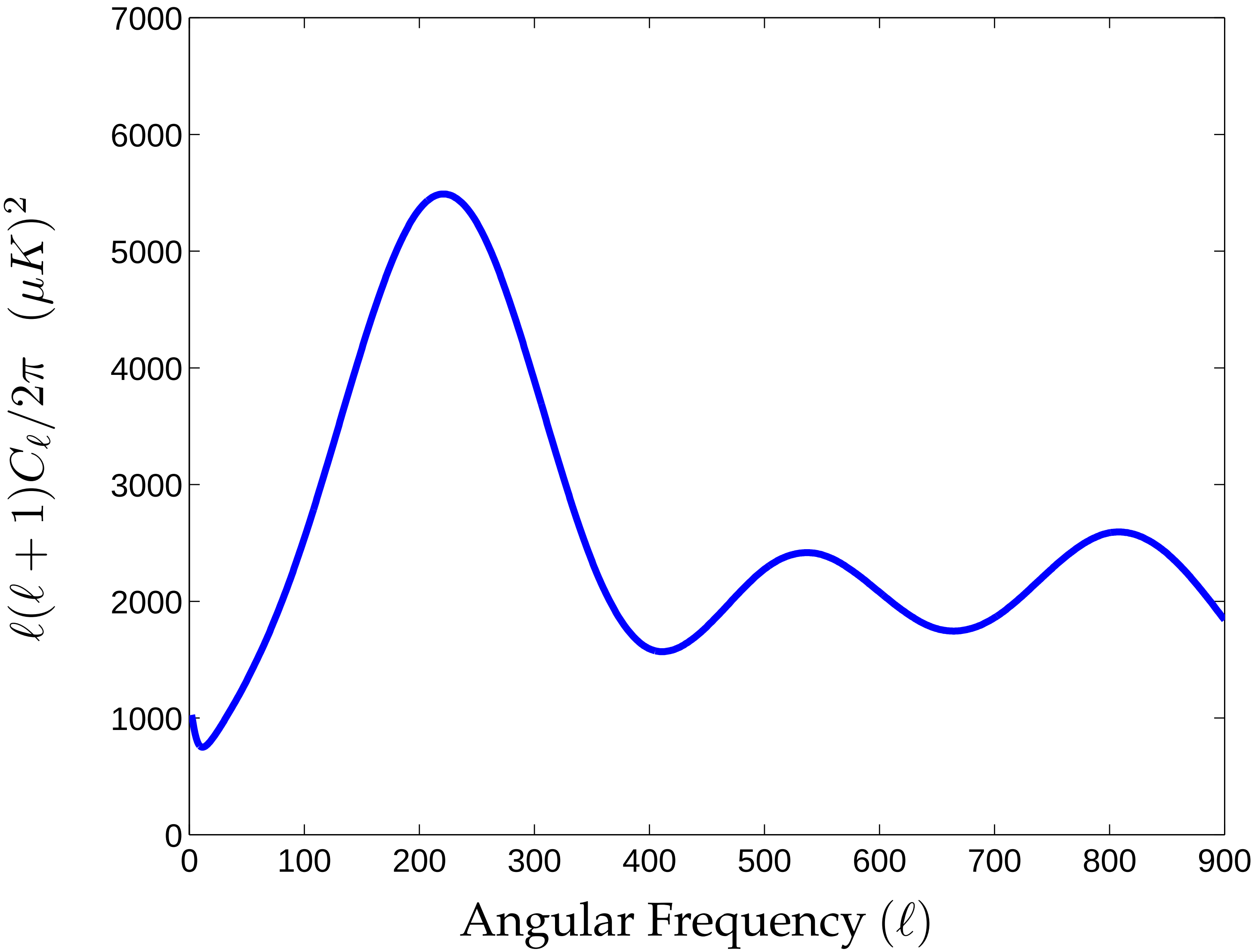
Motivation

For example,

Ω_m	$=$	total matter density	0.40
Ω_b	$=$	baryonic matter density	0.056
Ω_Λ	$=$	dark energy density	0.60
H_0	$=$	the Hubble parameter	64.6 km/s/Mpc
τ	$=$	the optical depth	0.075
n_s	$=$	spectral index of initial spectrum	0.99
A	$=$	amplitude of initial spectrum	0.79

parameterize the power spectrum of the CMB anisotropy.

Motivation



7000
6000
5000
4000
3000
2000
1000
0
0
100
200
300
400
500
600
700
800
900
$\ell(\ell+1)C_\ell/2\pi \quad (\mu K)^2$
Angular Frequency (ℓ)

Motivation

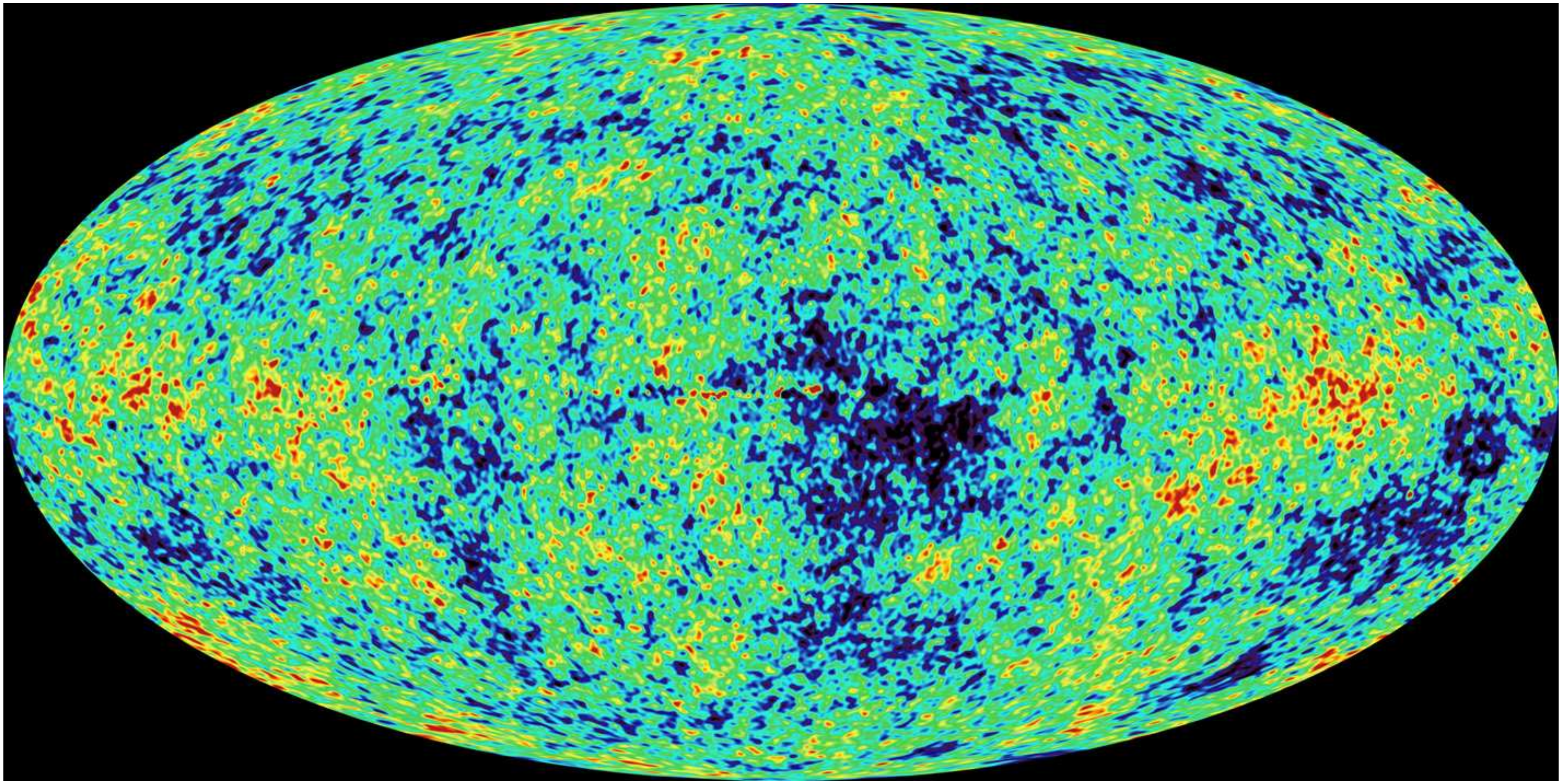


Image courtesy of WMAP Science Team.

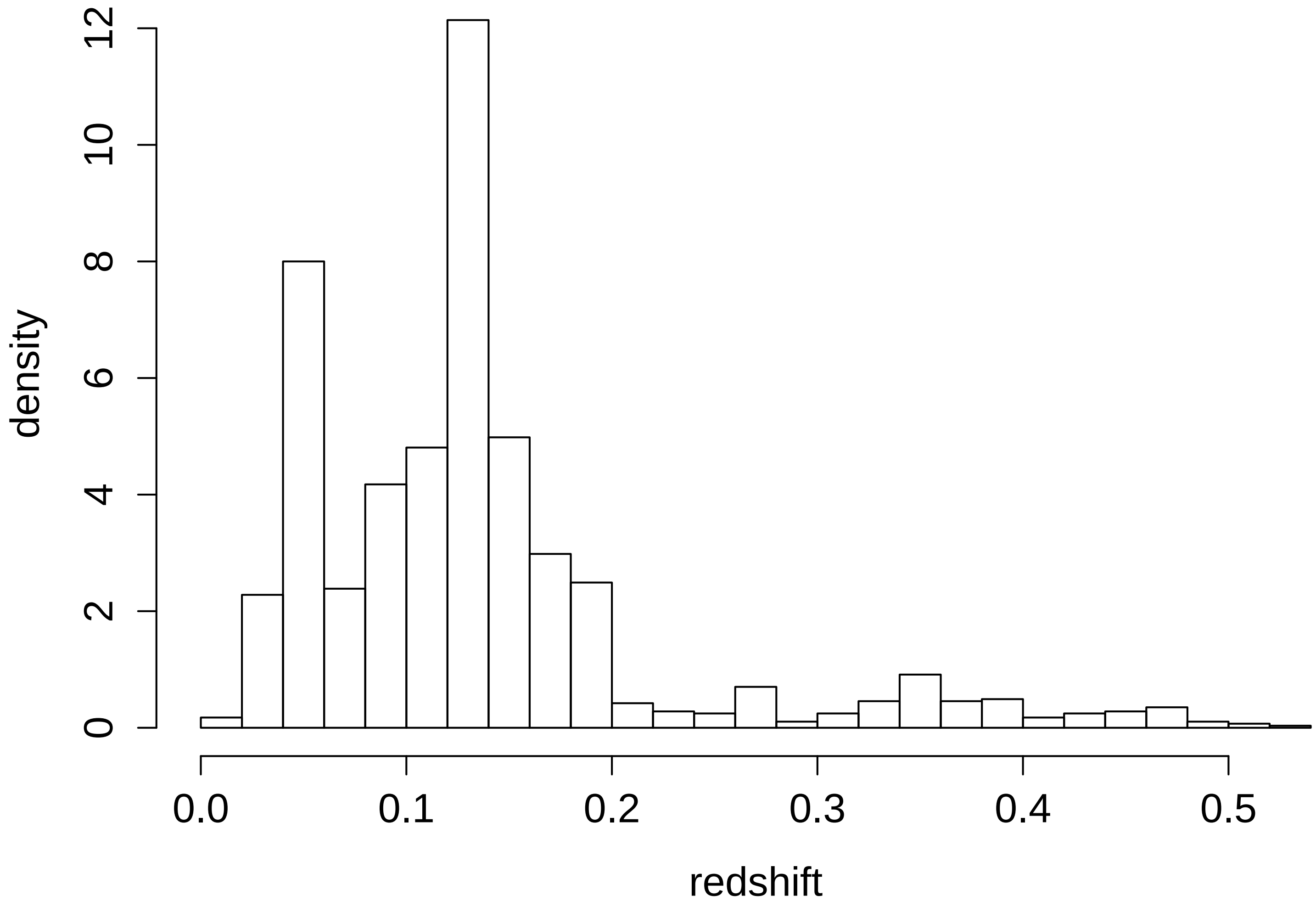
Motivation

The key role of Density estimation, i.e., estimating the distribution from which a sample of data were drawn

Assuming a parametric form is convenient, but often difficult to justify.

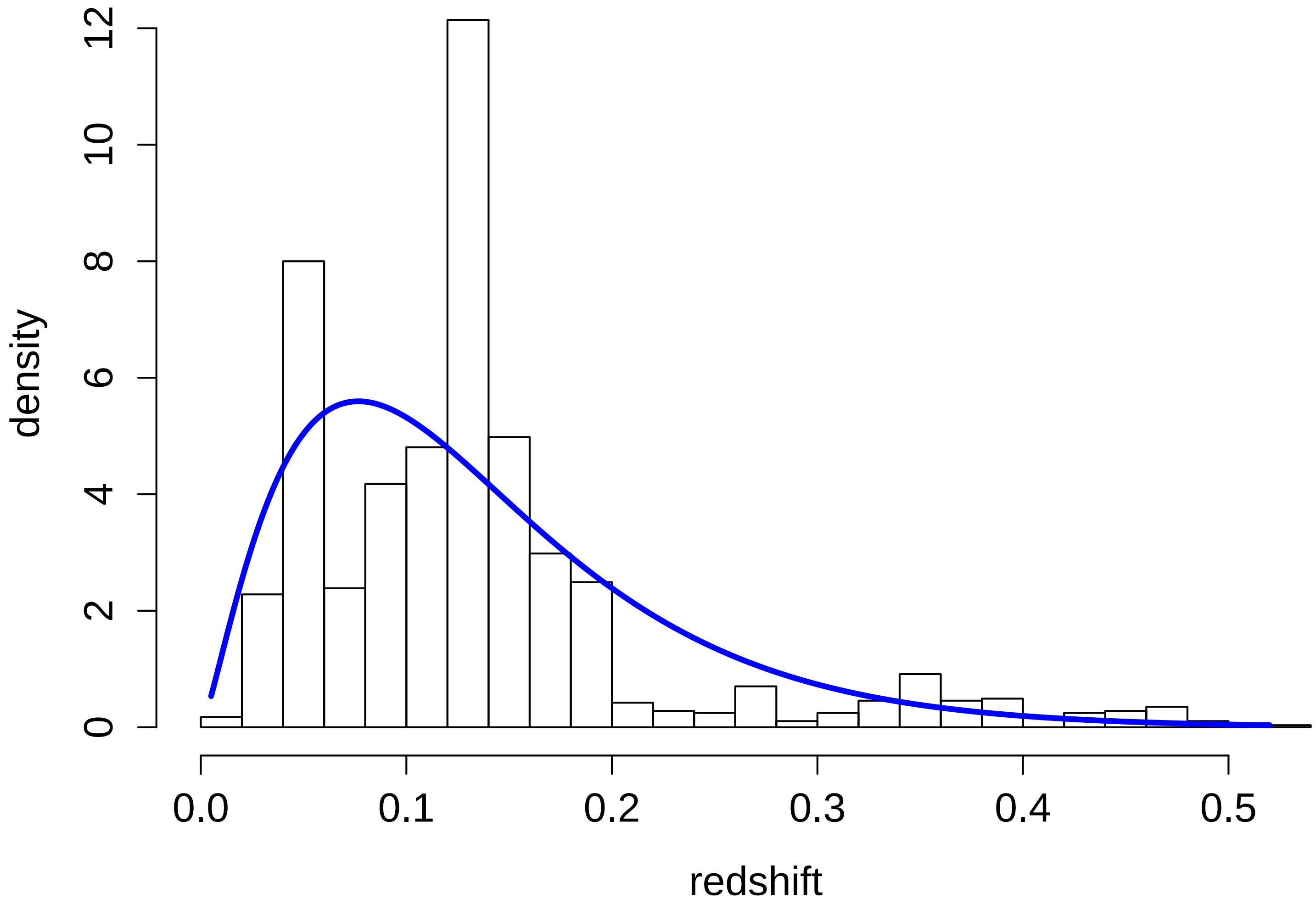
Nonparametric density estimation drops these restrictions

Nonparametric Density Estimation



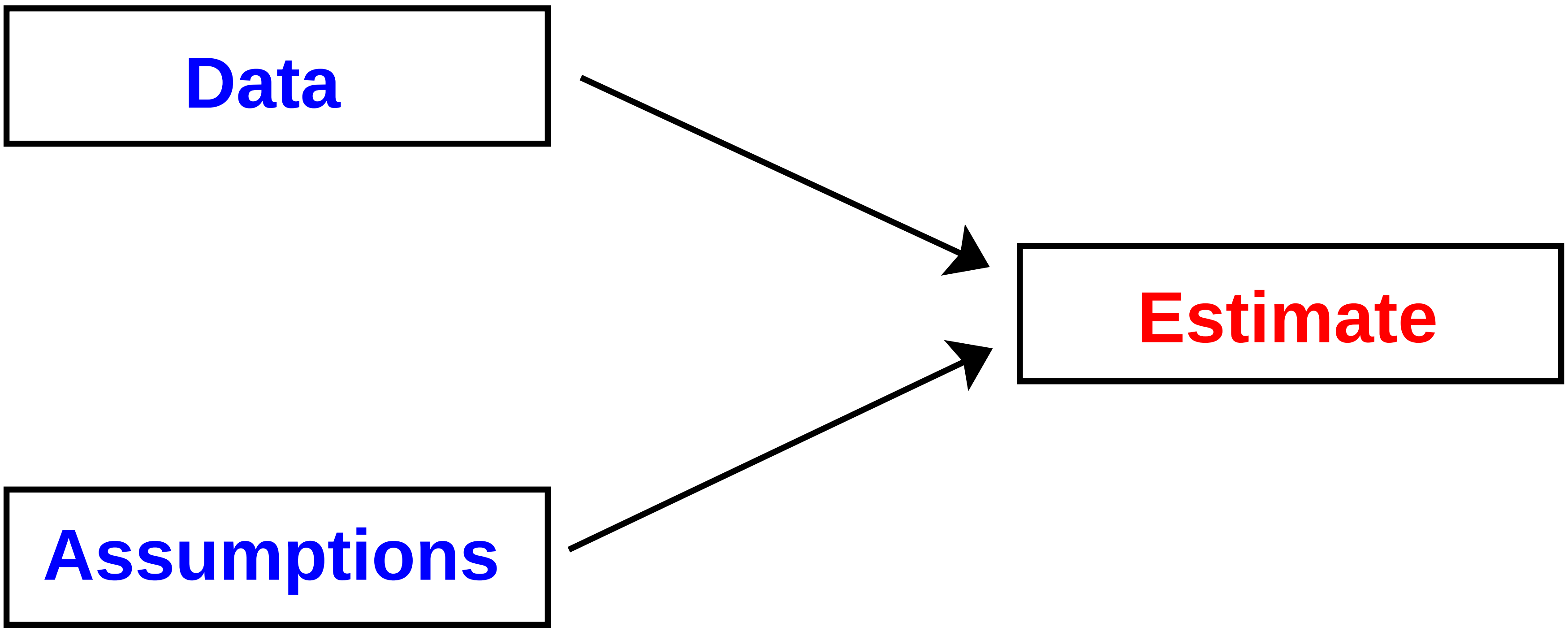
Histogram of 1,425 galaxy redshifts.

Nonparametric Density Estimation



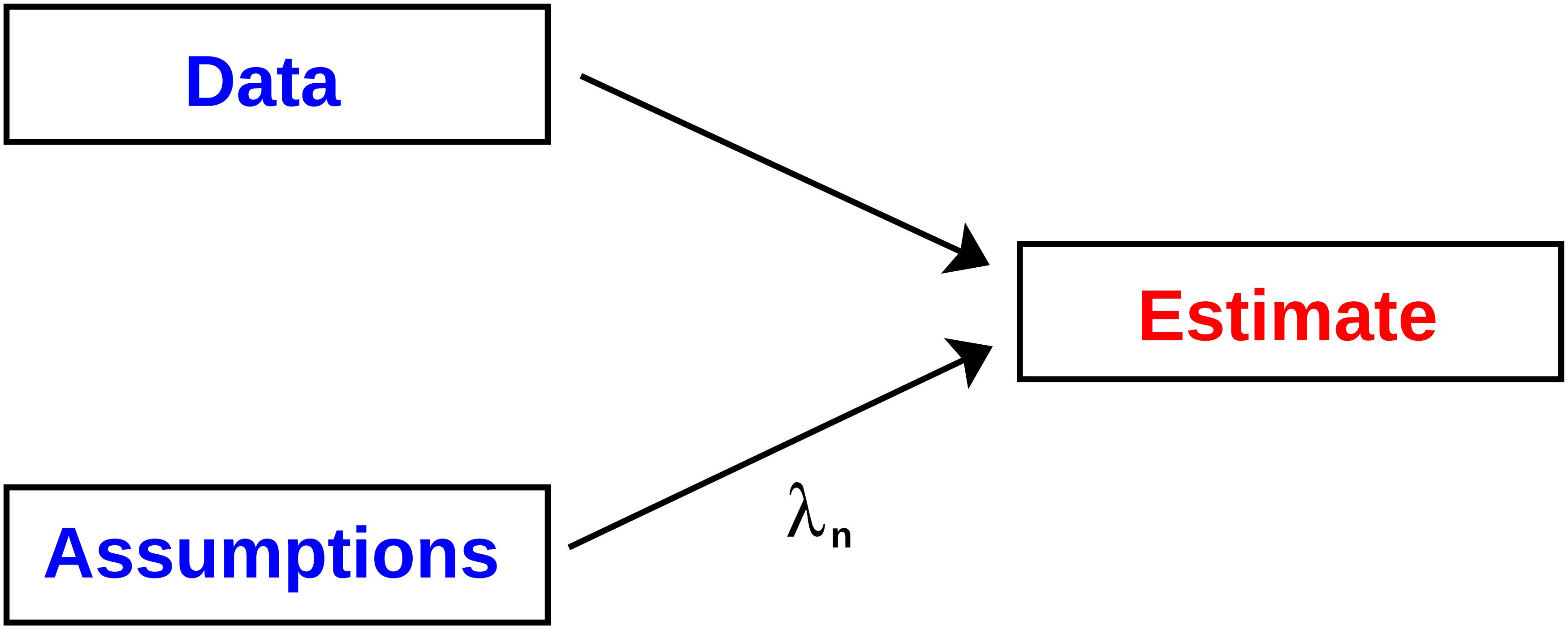
Compared with best fitting gamma distribution.

Nonparametric Density Estimation



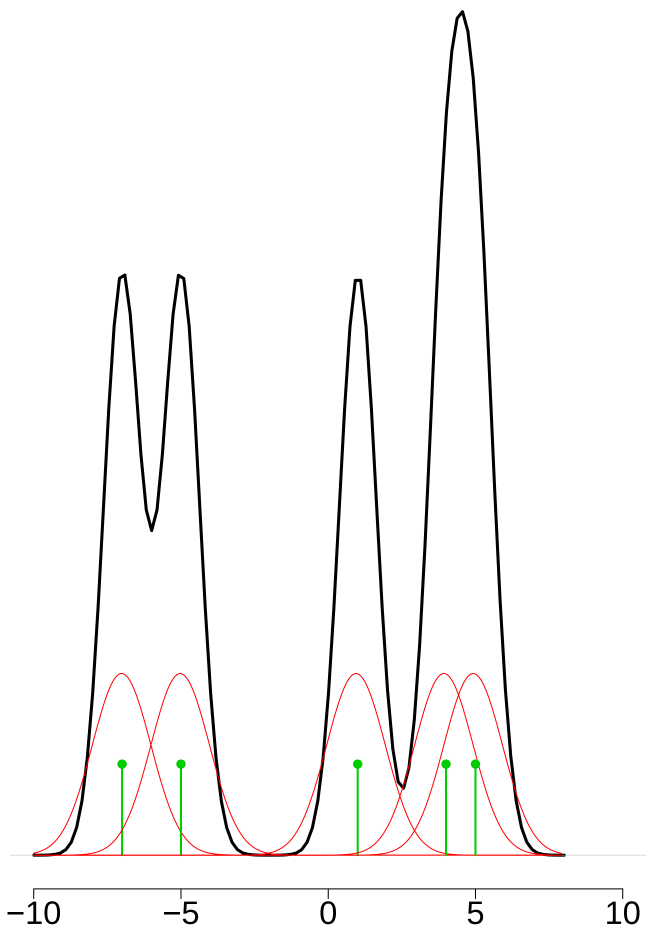
Parametric case: Fixed contribution of assumptions.

Nonparametric Density Estimation



Nonparametric case: Contribution of assumptions is controlled by λ_n. Optimally, $\lambda_n = o(n^{-1/(4+d)})$, where $d =$ dimension of data.

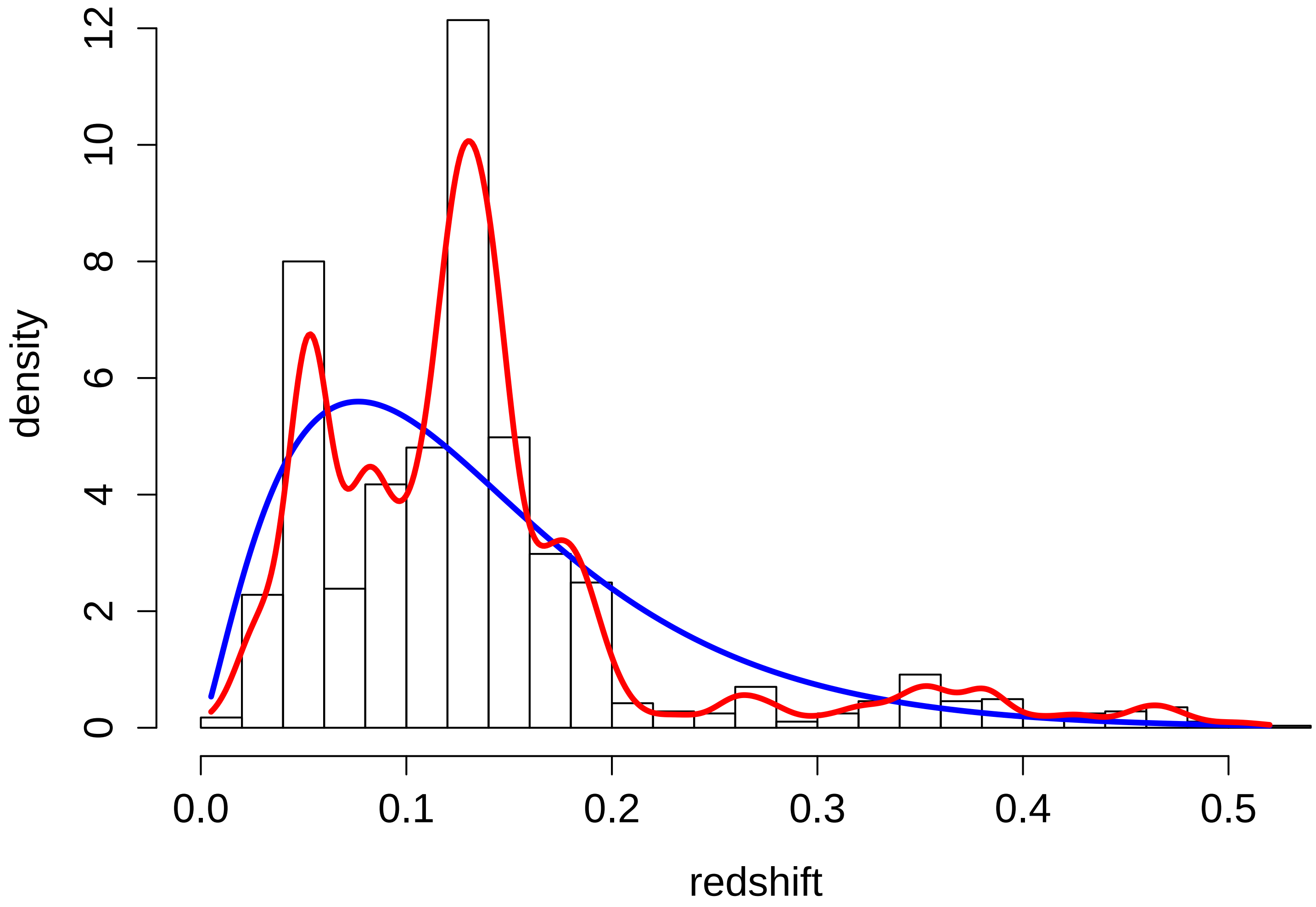
Nonparametric Density Estimation



Kernel density estimation puts a smooth mass at each data point.

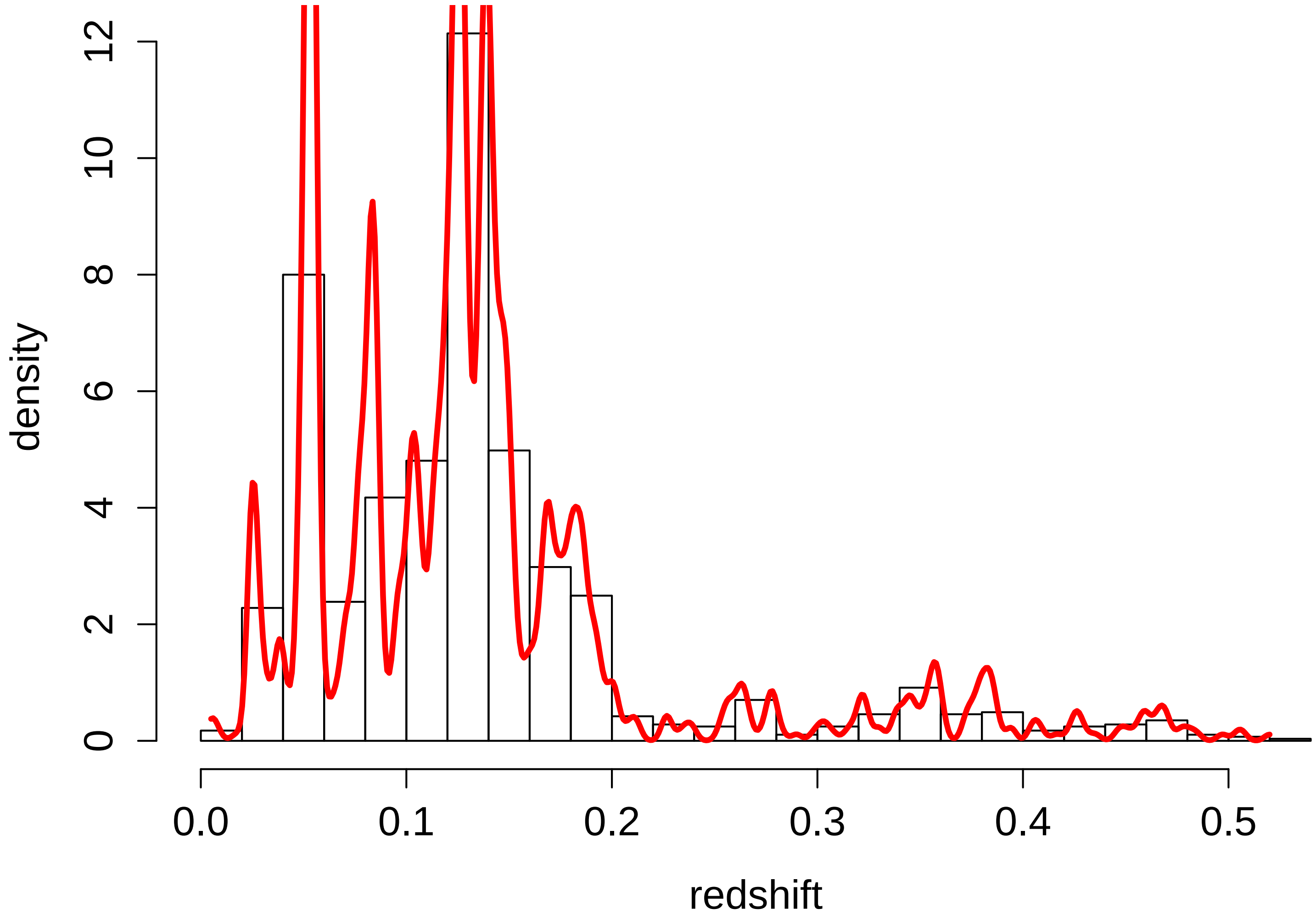
λ_n controls the width of the "bumps."

Nonparametric Density Estimation



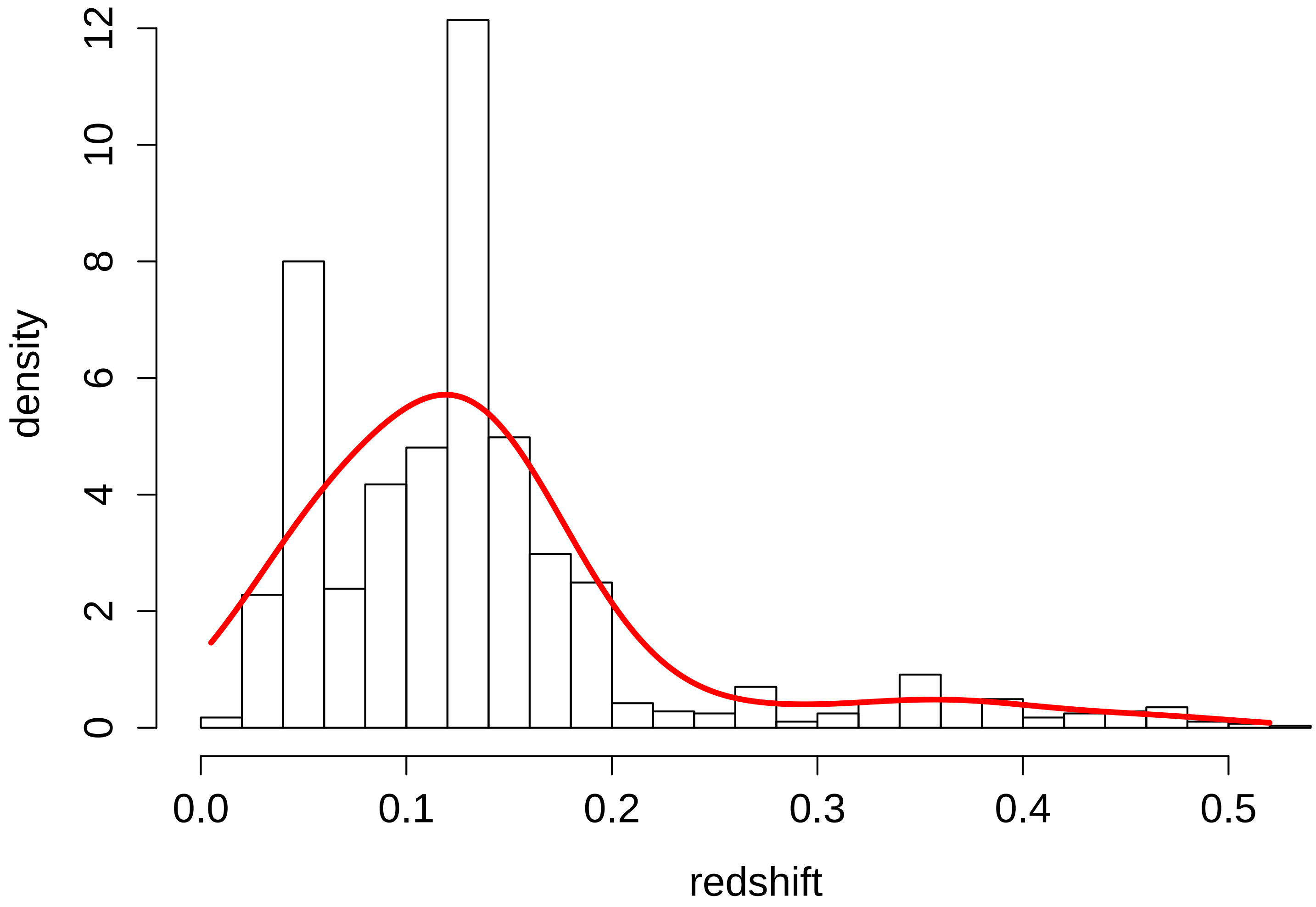
Parametric versus nonparametric estimate (kernel density estimate).

Nonparametric Density Estimation



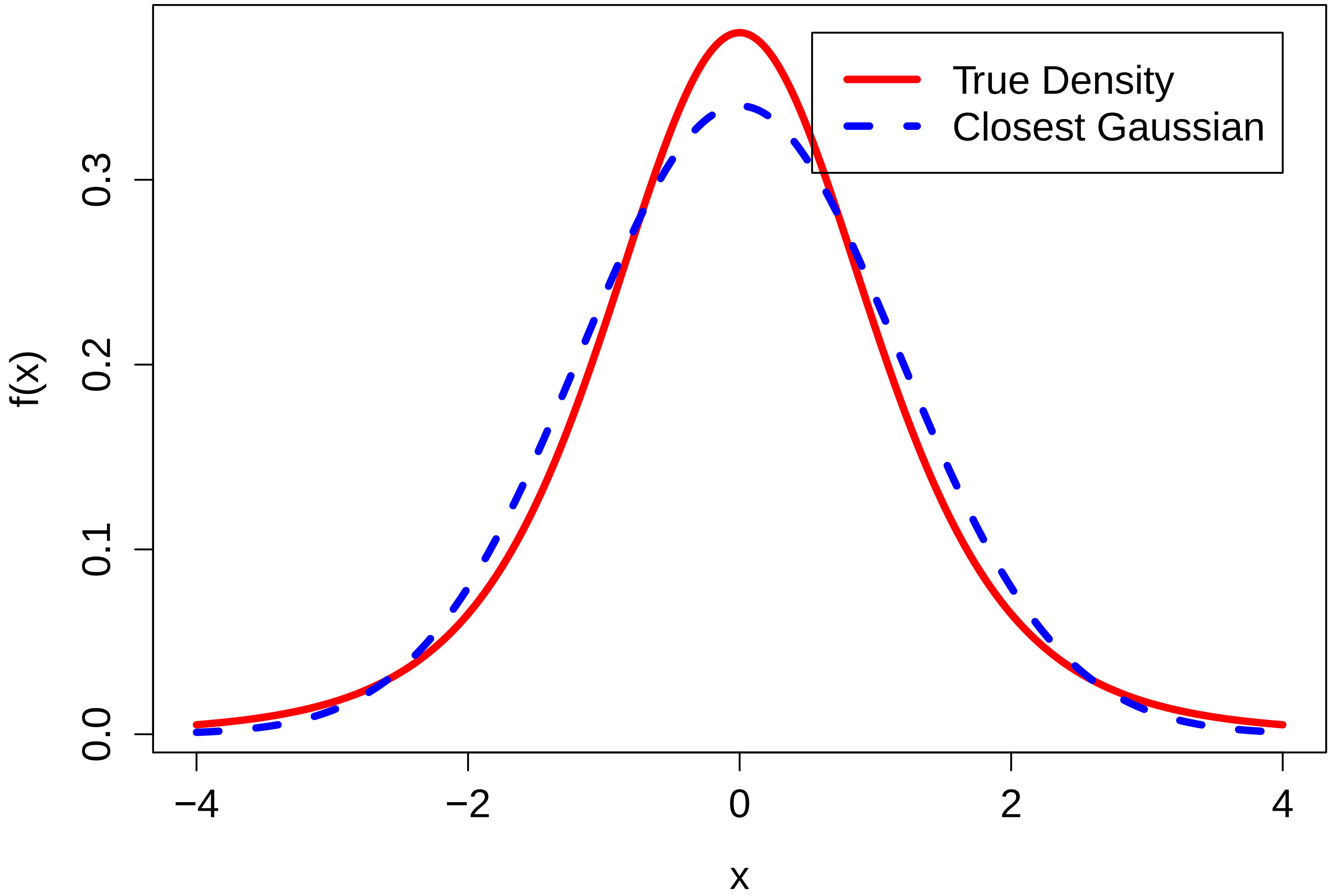
λ_n chosen too small, i.e. too much weight on data

Nonparametric Density Estimation



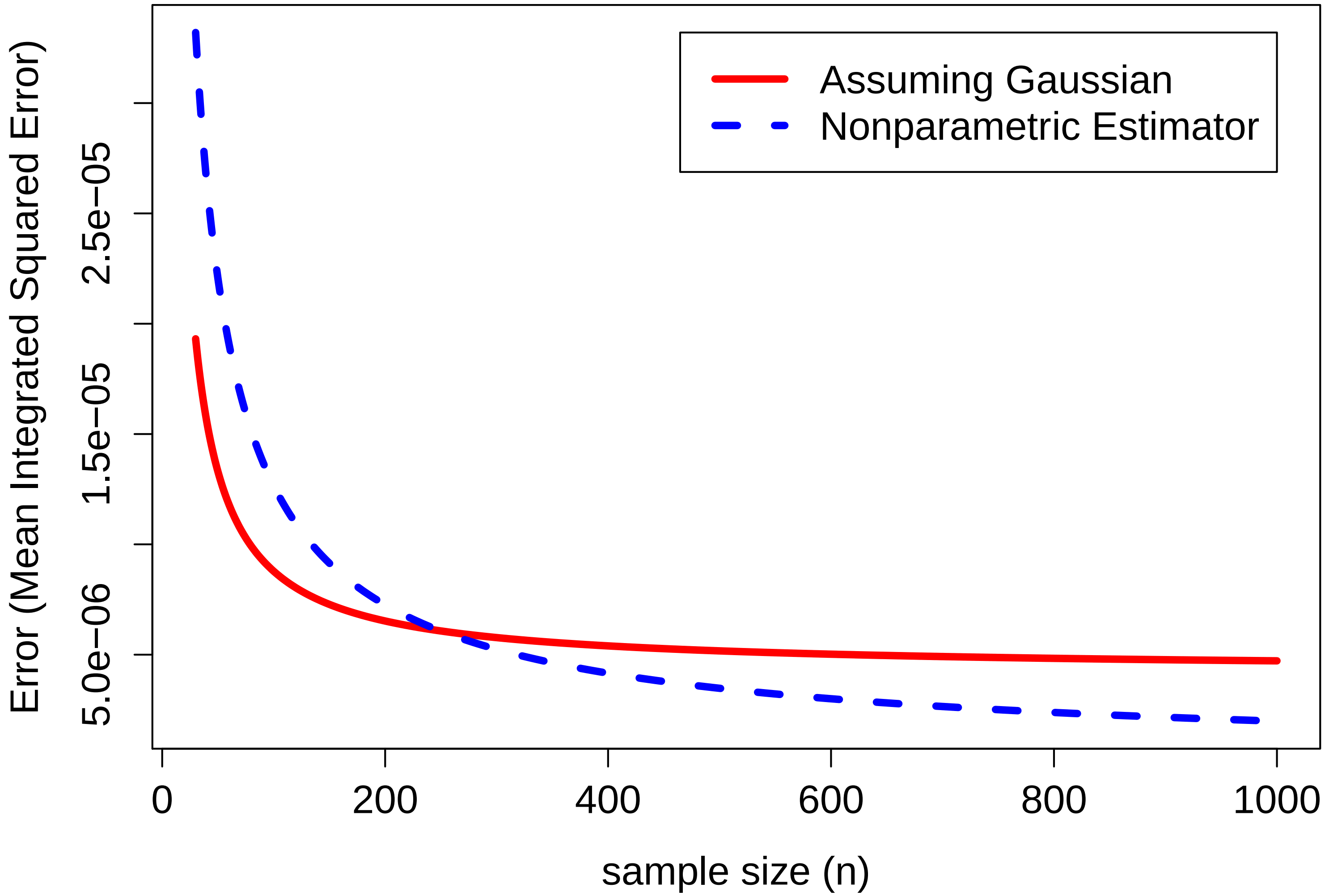
λ_n chosen too large, i.e. too much weight on assumptions

Nonparametric Density Estimation



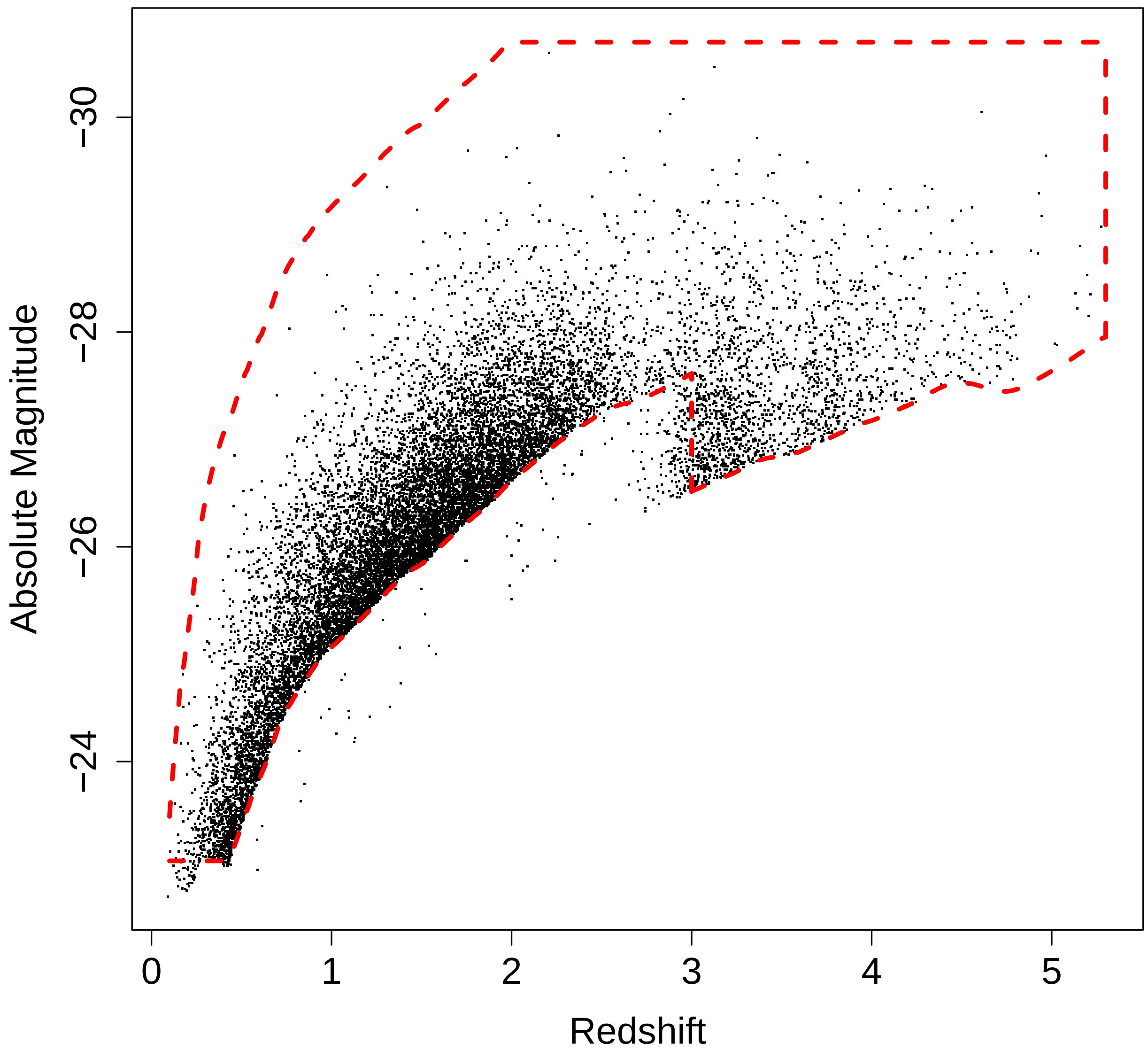
Truth is not quite a Gaussian distribution.

Nonparametric Density Estimation



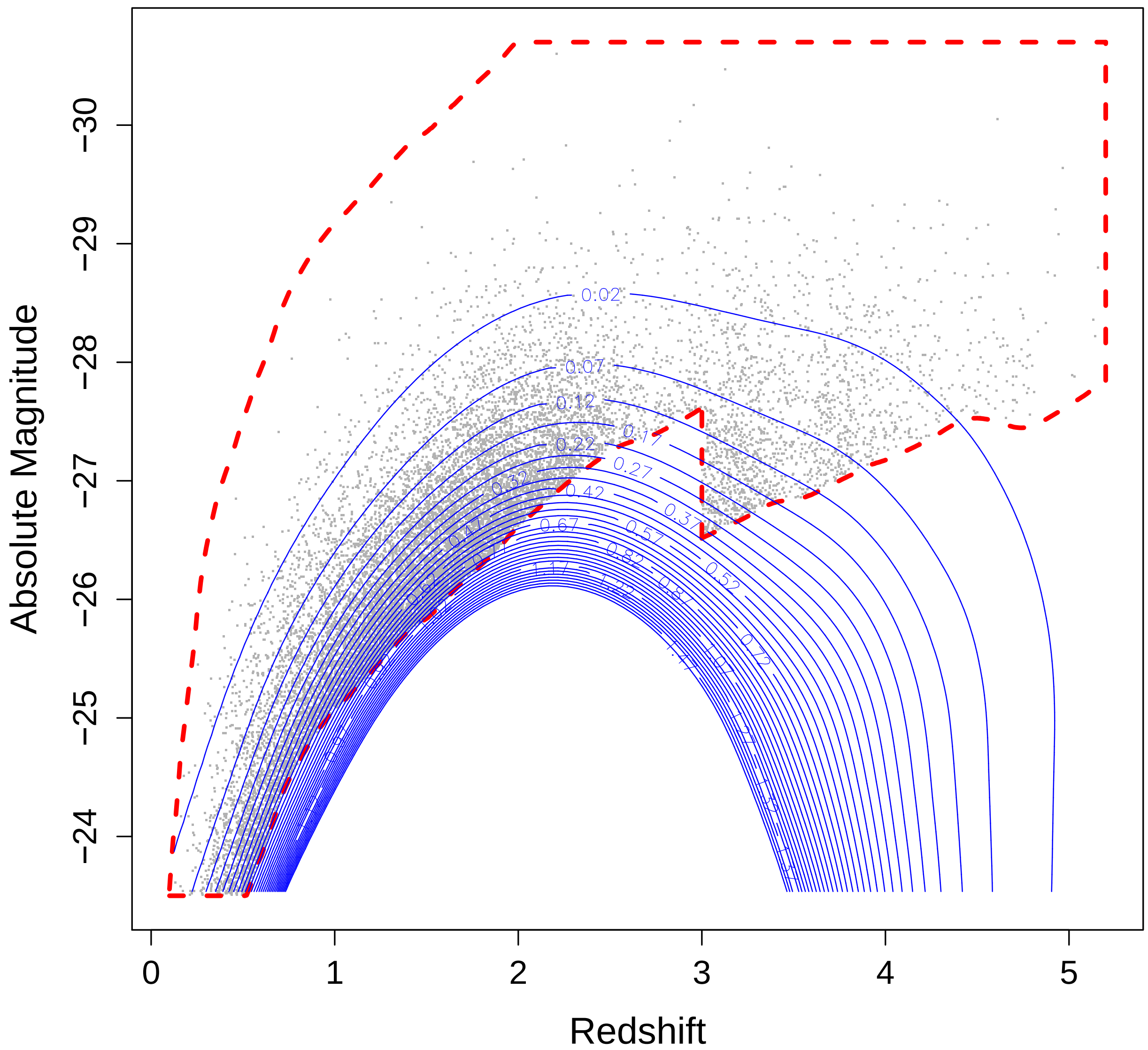
Even at moderate sample sizes, nonparametric estimator superior.

Bivariate Density Estimation



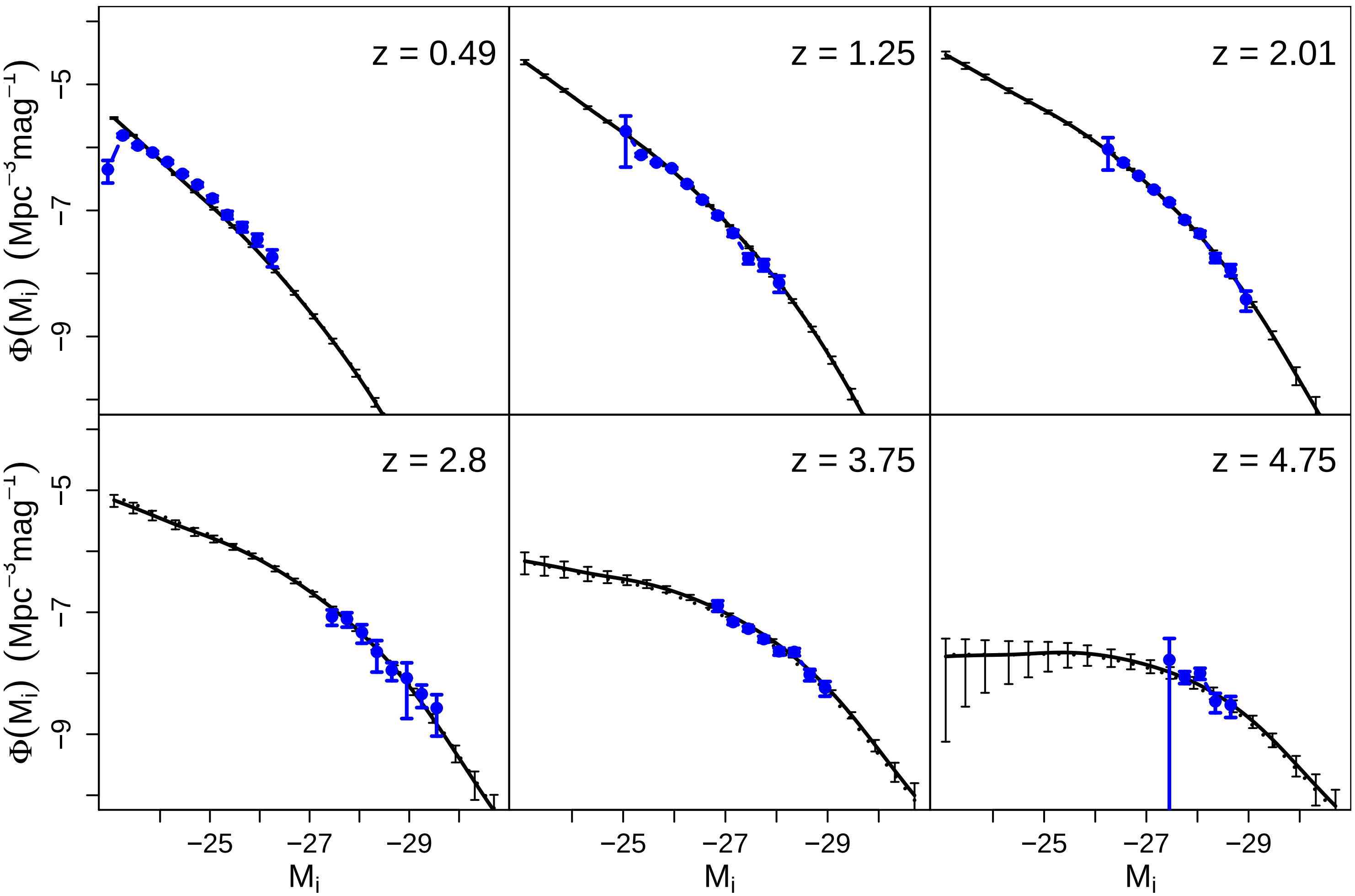
Sample of
15,057 SDSS quasars.
(Richards, et al. 2006)

Bivariate Density Estimation



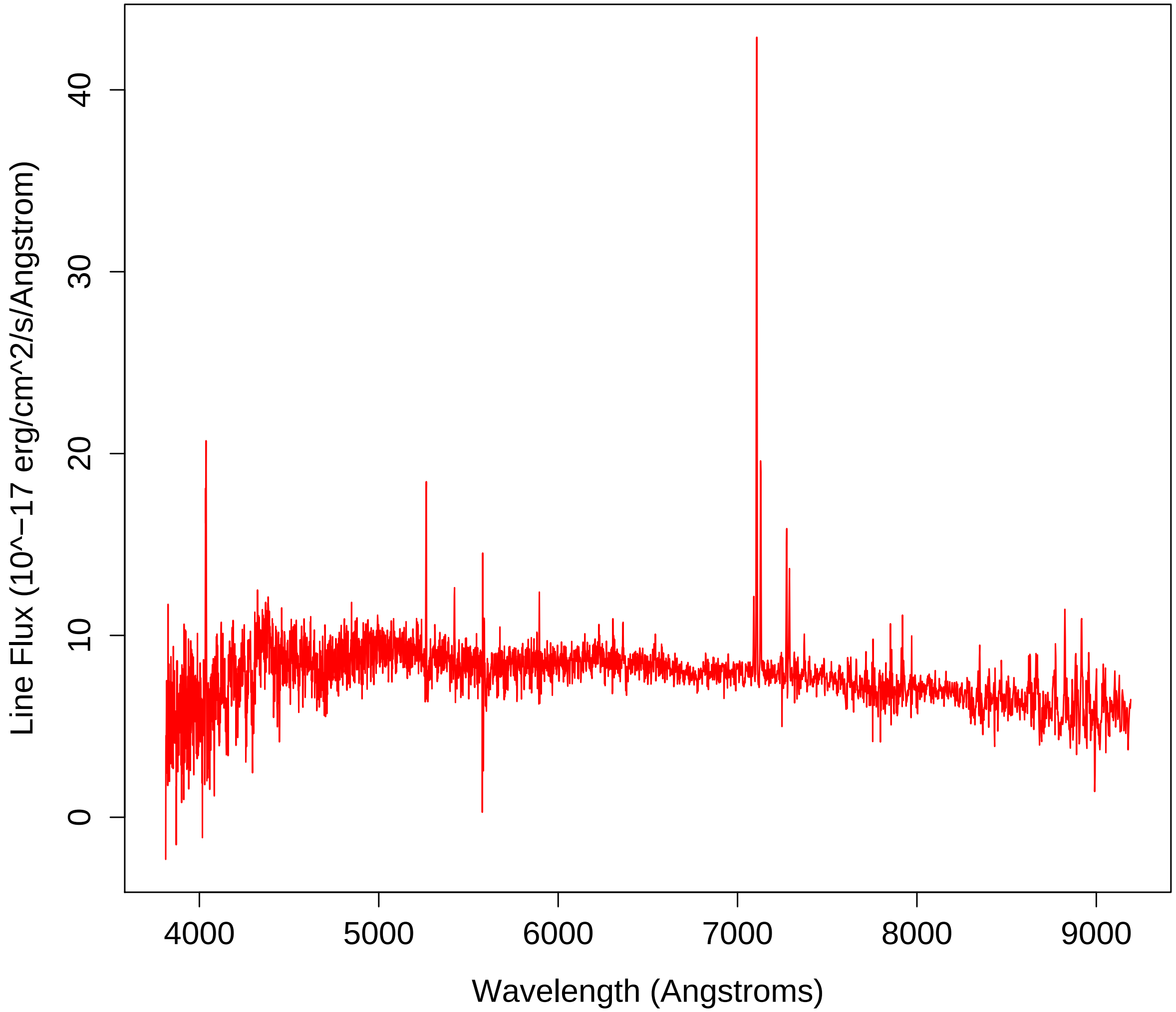
Bivariate luminosity function estimate (Schafer (2007))

Bivariate Density Estimation



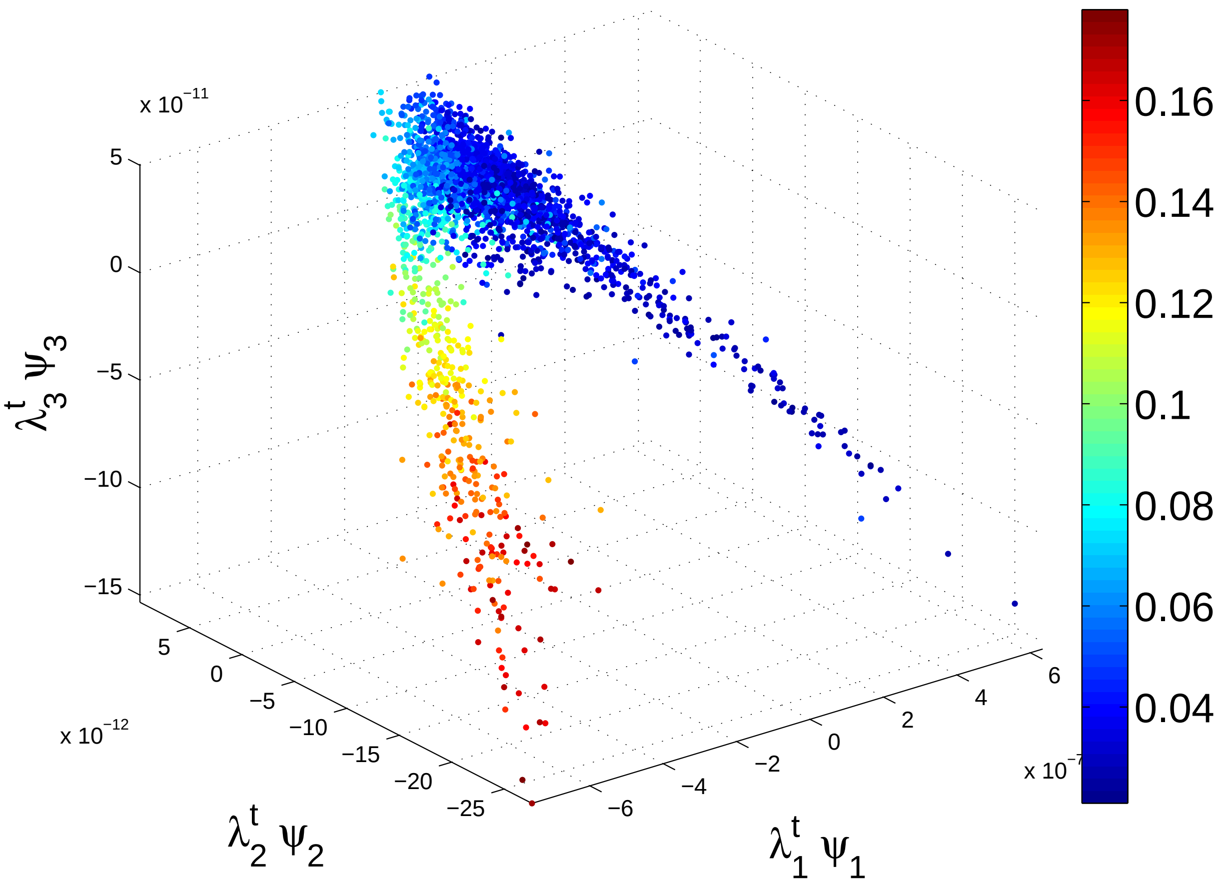
Cross-sections, compared with "standard" approach.

Working in Higher Dimensions



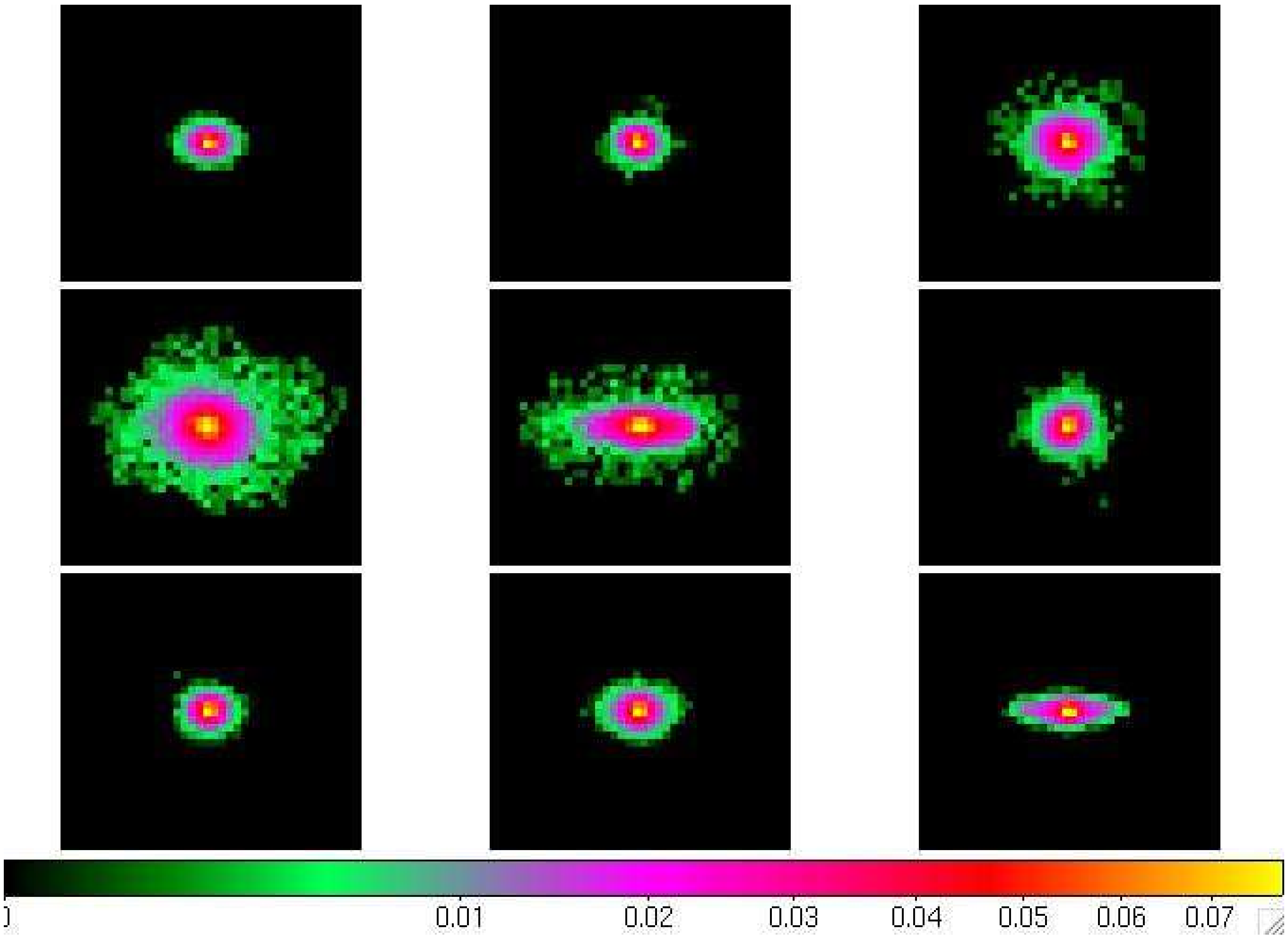
SDSS galaxy spectrum.

Working in Higher Dimensions



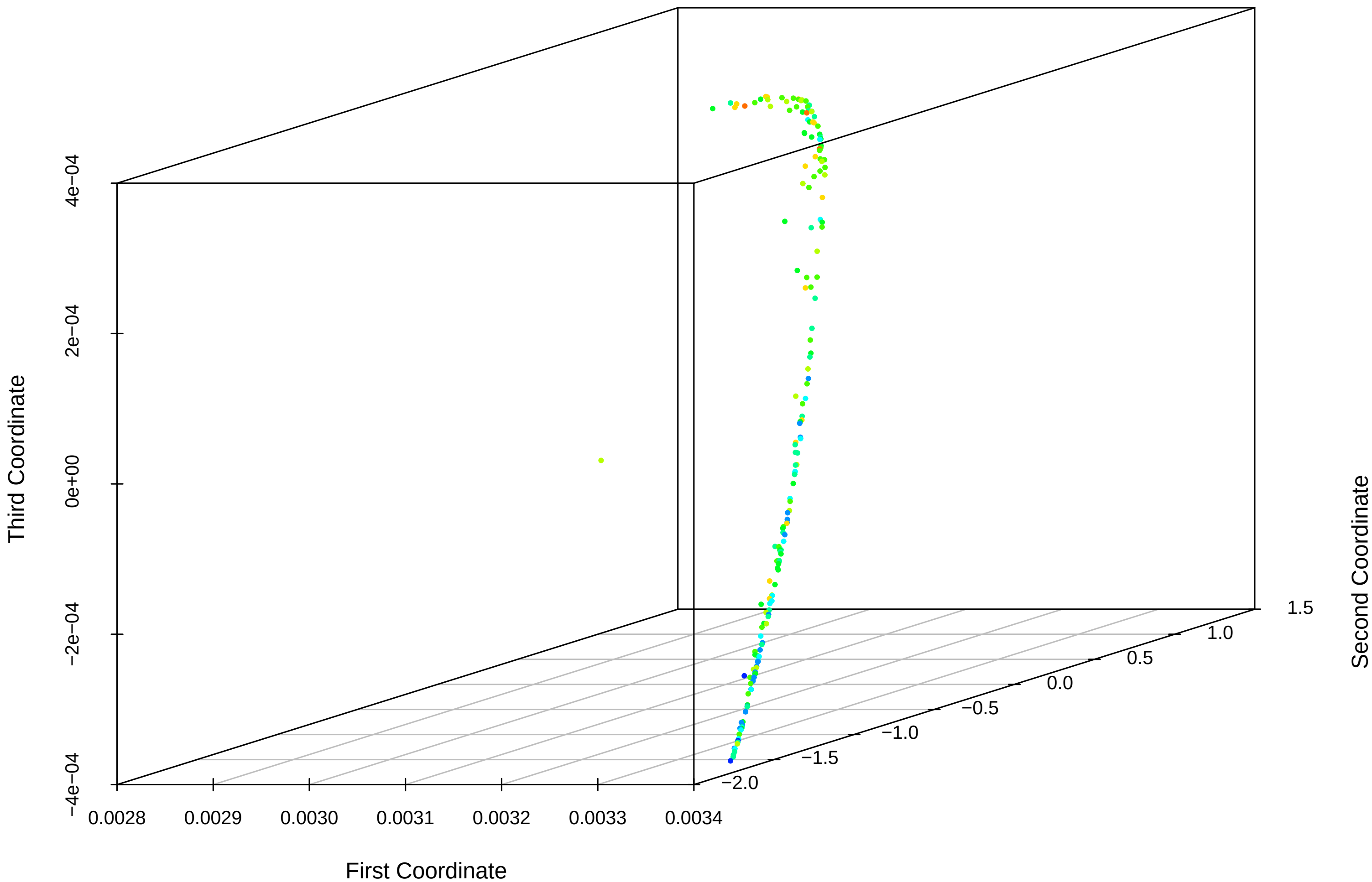
3,846 galaxy spectra, colored by redshift (Richards, Freeman, Lee, Schafer (2009a))

Working in Higher Dimensions



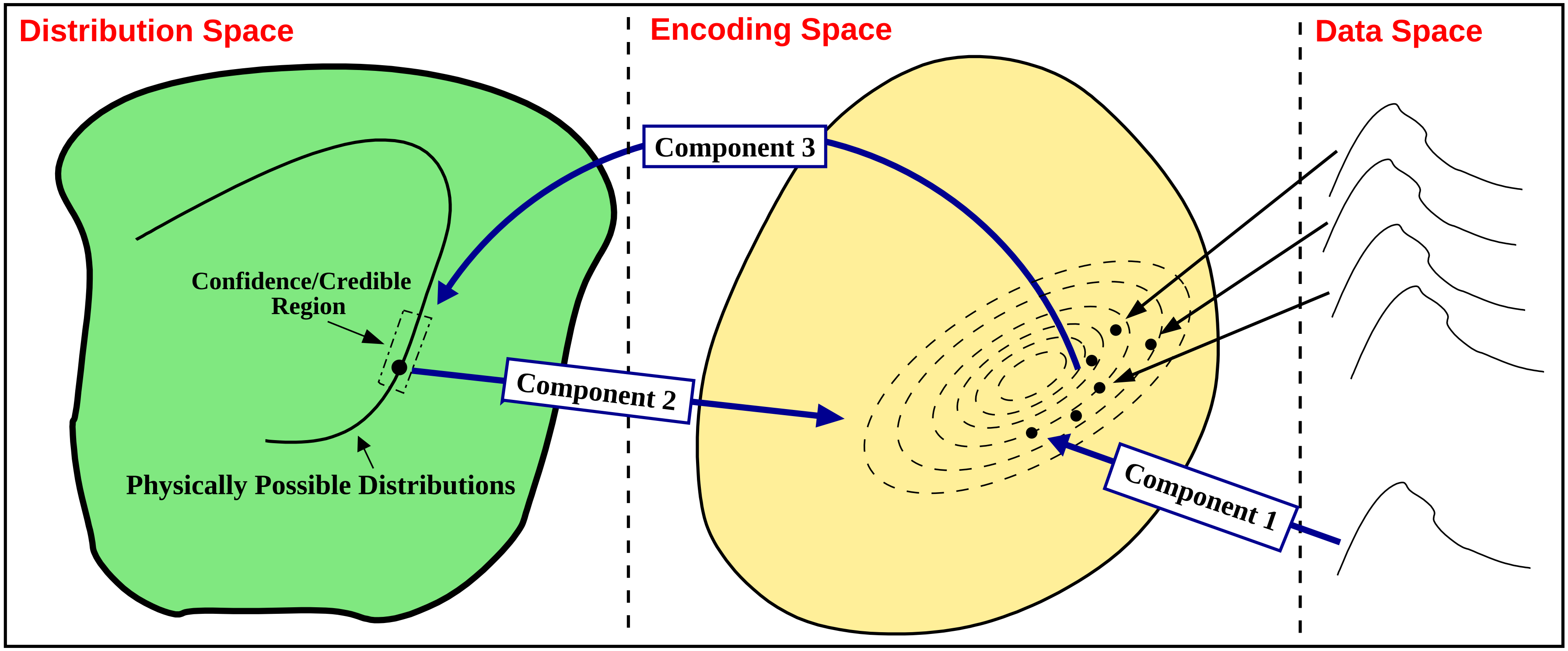
Examples of galaxy image data.

Working in Higher Dimensions



200 galaxies, colored by eccentricity.

The Big Picture



Once represented in low-dimensional space encoding space, nonparametric density estimation useful for comparing observations and theory

References

Buchman, Lee, and Schafer (2009). To appear in *Statistical Methodology.* `arXiv:0907.0199`

Richards, et al. (2006) *ApJ*. **131** 2766

Richards, Freeman, Lee, and Schafer (2009a). *ApJ*. **691** 32-42.

Schafer (2007). *ApJ* **661** 703-713.

Schafer and Stark (2009). *J. Amer. Stat. Assoc.*