Nonparametric Inference and the Dark Energy Equation of State

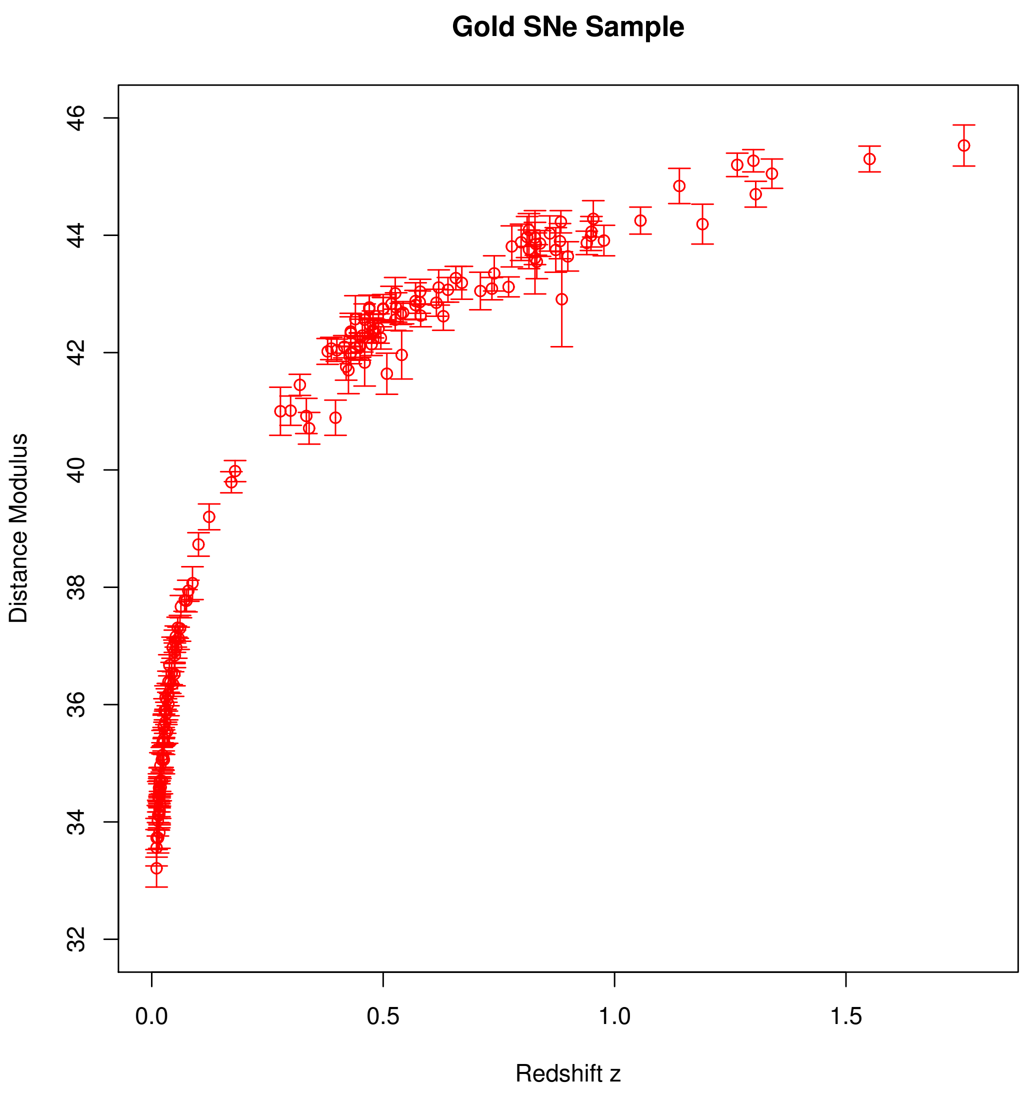
Christopher R. Genovese
Peter E. Freeman
Larry Wasserman

Department of Statistics
Carnegie Mellon University
http://www.stat.cmu.edu/~genovese/

Collaborators: Andy Connolly, Chris Miller, Bob Nichol

A Small-Sample Story for a Large-Sample World

- The flood of data in astronomy is only just beginning.
- Such data open up new questions and raise new challenges.
- Nonparametric methods are well suited to these new problems.
  - Cosmic Microwave Background (Genovese et al. 2004, Bryan et al. 2005)
  - Galaxy Evolution (e.g., Rojas et al. 2006)
  - Galaxy Spectra (work in progress)
  - Dark Energy (e.g., Daly and Djorgovski 2004, 2005; and below)
- So, in this terabyte age, I want to illustrate this potential with a data set of mere hundreds.



Gold SNe Sample
Distance Modulus
Redshift z
32
34
36
38
40
42
44
46
0.0
0.5
1.0
1.5

Road Map

1. Dark Energy
2. Nonparametric Inference
3. Derivative Estimation as an Inverse Problem
4. Inference for the Dark Energy Equation of State

Road Map

1. Dark Energy
2. Nonparametric Inference
3. Derivative Estimation as an Inverse Problem
4. Inference for the Dark Energy Equation of State

Preliminaries

- The Expanding Universe

  Scale factor $a(t)$ indicates relative expansion of the universe.
  ($a(t_0) = 1$ where t_0 is current age of universe.)

  Redshift z is an observable shift in the wavelength of light from a distant object that is induced by the expansion of the universe.

$$1 + z = \frac{\lambda_{\mathrm{obs}}}{\lambda_{\mathrm{emit}}} = \frac{a(t_{\mathrm{obs}})}{a(t_{\mathrm{emit}})}.$$

  Hubble parameter $H(t) = \frac{\dot{a}(t)}{a(t)}$. ($H_0 = H(t_0)$ is the Hubble "constant".)

- The Distance-Redshift Relation

  The relationship between objects' distances and redshifts contains fundamental information about the Universe's geometry.

  Hubble's Law, $z = H_0 d$, is reasonably accurate for small distances d.

Dark Energy

- Accelerating Expansion (Reiss et al. 2004, Perlmutter et al. 2004)

  Type Ia supernovae can serve as a "standard candle".

  Observations of many supernovae reveal that the expansion of the universe is *accelerating*.

  This conclusion is supported by other, independent, measurements, including the Cosmic Microwave Background (Spergel et al. 2003) and large-scale structure (Verde et al. 2002).

- Einstein's "mistake," Cosmological Constant, and Vacuum Energy
- This raises several puzzles. What's going on?
  - Mistaken assumptions, models, or data analysis
  - A failure of General Relativity
  - Anthropic Selection
  - Dark Energy

Dark Energy (cont'd)

- Dark Energy is a smoothly-distributed energy density that dominates the universe ($\sim$ 74% versus $\sim$ 4% for baryonic matter) and provides a negative pressure acting in opposition to gravity.
- What does the acceleration imply about dark energy?

  Let $\rho = \rho_{\rm matter} + \rho_{\rm radiation} + \rho_{\rm DE} + \cdots$ be the total energy density in the universe.

  Friedmann equation:

$$H^2(t) = \left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi G}{3}\rho - \frac{\kappa}{a^2(t)}$$

  or equivalently,

$$\dot{a}^2 = \frac{8\pi G}{3}a^2\rho - \kappa.$$

  Acceleration implies that $a^2\rho$ must increase.

  Neither matter ($\rho_{\rm matter} \propto a^{-3}$) nor radiation ($\rho_{\rm radiation} \propto a^{-4}$) can do this.
  A cosmological constant ($\rho_{\rm DE} \propto a^0$) could.

Dark Energy (cont'd)

- How do we quantify dark energy?

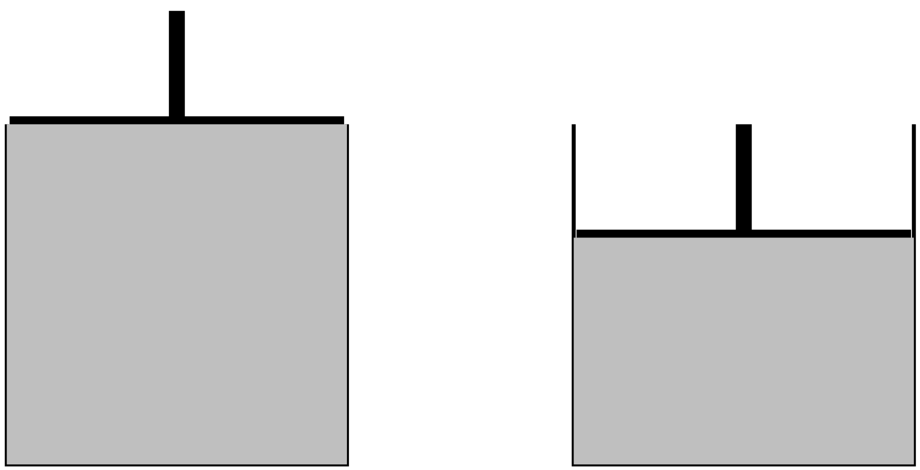
  We can attempt to make inferences about ρ directly.

  Alternatively, we can look at the equation of state (cf. ideal gas law).

  Let p_{DE} and ρ_{DE} be the pressure and energy density of dark energy, then the equation of state w relates these by

$$p_{\mathrm{DE}} = w\rho_{\mathrm{DE}}.$$

  For a cosmological constant, $w = -1$.



$$\text{Work} = -p_{\mathrm{DE}}\Delta V$$
$$\Delta\text{Energy} = \rho_{\mathrm{DE}}\Delta V$$
$$\implies p_{\mathrm{DE}} = -\rho_{\mathrm{DE}}$$

Dark Energy (cont'd)

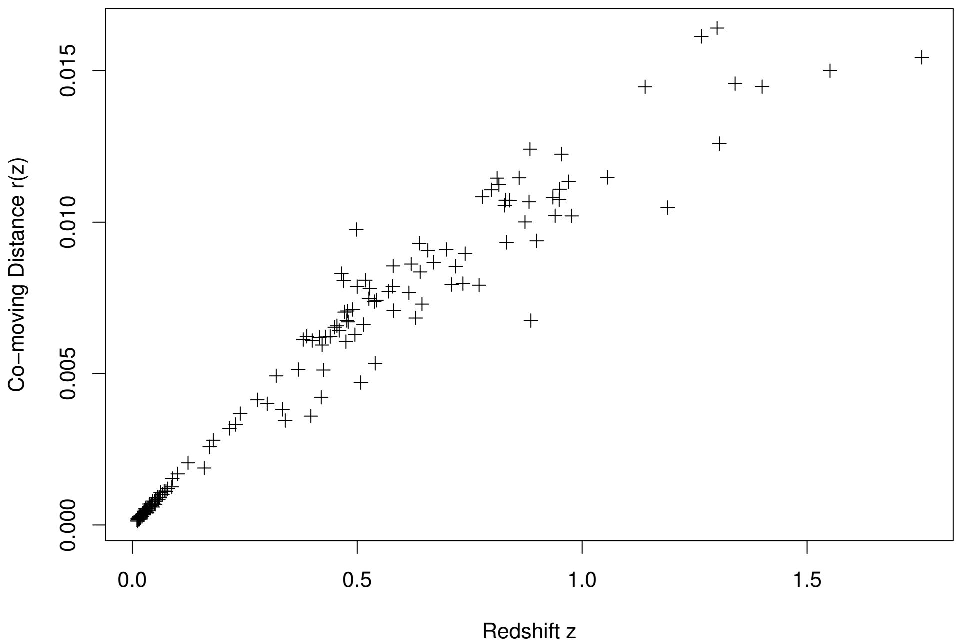
- The supernova data give us a way to infer the equation of state

  Roughly, we get

$$Y_i = r(z_i) + \sigma_i \epsilon_i, \qquad i = 1, \ldots, n,$$

where r is a measure of distance at each redshift z_i. Then,

$$w(z) = \frac{H_0^2 \Omega_M (1+z)^3 + \frac{2}{3} \frac{r''(z)}{(r'(z))^3}}{H_0^2 \Omega_M (1+z)^3 - \frac{1}{(r'(z))^2}} \equiv T(r', r'').$$



Road Map

1. Dark Energy
2. Nonparametric Inference
3. Derivative Estimation as an Inverse Problem
4. Inference for the Dark Energy Equation of State

Nonparametric Inference

Goal: make sharp inferences about unknown functions with a minimum of assumptions.

Constructing good estimators is important, but an accurate assessment of uncertainty is critical.

Why use nonparametric methods?

1. When we don't have a well-justified parametric (finite-dimensional) model for the object of interest.
2. When we have a well-justified parametric model but have enough data to go after even more detail.
3. When we can do as well (or better) more simply.
4. As a way of assessing sensitivity to model assumptions.

Aside: What's in a name?

The term "nonparametric" is unfortunate, although now firmly established.

- There is a parameter in these models – the unknown function.
- Loosely speaking, the contrast between nonparametric and parametric is an infinite-dimensinoal parameter versus a finite-dimensional parameter.
- We have only n data. But with a nonparametric analysis, the dimension of the fit grows with n; in a parametric analysis, it is fixed for all n.
- These methods go beyond the rank-based testing of classical nonparametric statistics.

The Nonparametric Regression Problem

Observe data (X_i, Y_i) for $i = 1, \ldots, n$ where

$$Y_i = f(X_i) + \epsilon_i,$$

where $\mathsf{E}(\epsilon_i) = 0$ and the X_is can be fixed (x_i) or random.

Leading cases: 1. $x_i = i/n$ and $\mathsf{Cov}(\boldsymbol{\epsilon}) \equiv \Sigma = \sigma^2 I$.

2. X_i IID g and $\mathsf{Cov}(\boldsymbol{\epsilon}) \equiv \Sigma = \sigma^2 I$.

Key Assumption: $f \in \mathcal{F}$ for some infinite dimensional space $\mathcal{F}$.

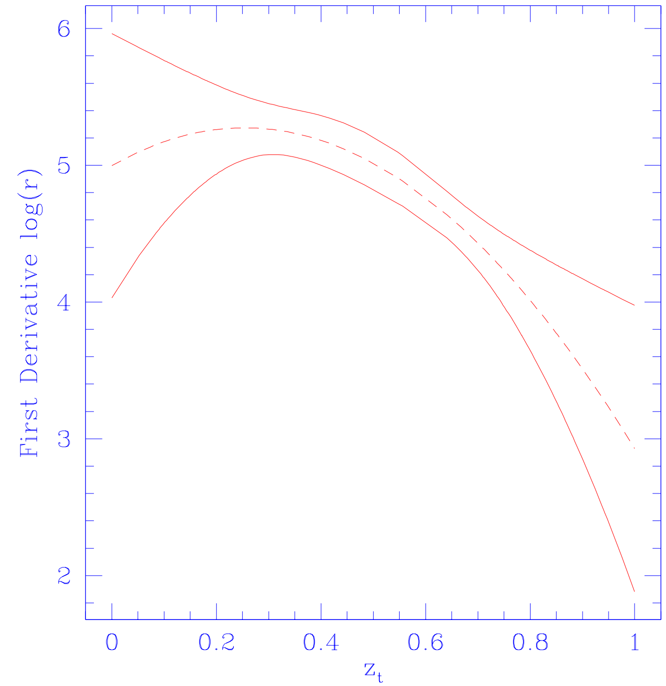
Examples

1. Sobolev: $\mathcal{F} \equiv \mathcal{W}_p(C) = \left\{ f \colon \int |f|^2 < \infty \text{ and } \int |f^{(p)}|^2 \le C^2 \right\}$
2. Lipschitz: $\mathcal{F} \equiv \mathcal{H}(A) = \{ f \colon |f(x) - f(y)| \le A|x - y|, \text{ for all } x, y \}$

Goal: Make inferences about f or about specific features of f.

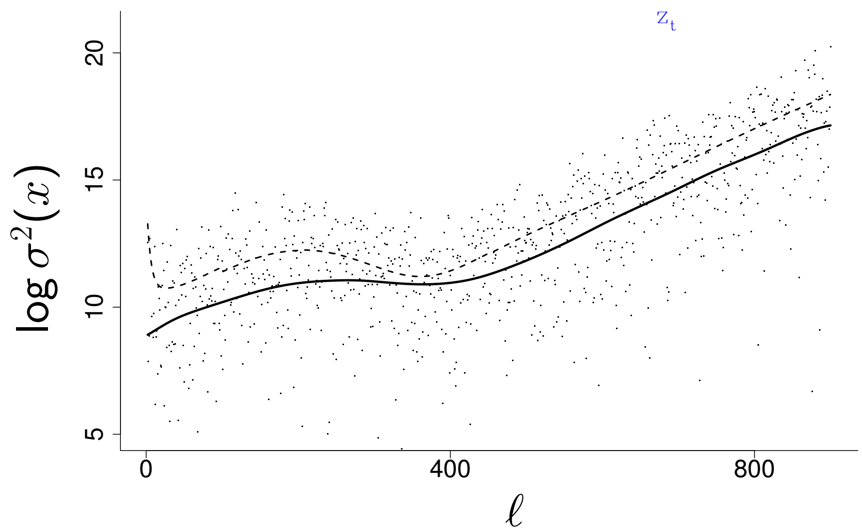
Variants of the Problem

- Inference for Derivatives of f
- Estimating Variance functions
- Regression in High dimensions
- Inferences about specific functionals of f



Related Problems:

- Density Estimation
- Spectral Density Estimation



Rate-Optimal Estimators

Choose a performance measure, or risk function, e.g.,
$R(\widehat{f}, f) = \mathsf{E} \int (\widehat{f} - f)^2$ or $R(\widehat{f}, f) = \mathsf{E}|\widehat{f}(x_0) - f(x_0)|^2)$

Want $\widehat{f}$ that minimizes worst-case risk over $\mathcal{F}$ (minimax).

But typically must settle for achieving the optimal minimax rate of convergence r_n:

$$\inf_{\widehat{f}_n} \sup_{f \in \mathcal{F}} R(\widehat{f}_n, f) \asymp r_n$$

In infinite-dimensional problems, $r_n \sqrt{n} \to \infty$.

For example, $r_n = n^{-\frac{2p}{2p+1}}$ on $\mathcal{W}_p$.

Rate-optimal estimators exist for a wide variety of spaces and risk functions.

Adaptive Estimators

It's unsatisfying to depend too strongly on intangible assumptions such as whether $f \in \mathcal{W}_p(C)$ or $f \in \mathcal{H}(A)$.

Instead, we want procedures to *adapt* to the unknown smoothness.

For example, $\widehat{f}_n$ is a *(rate) adaptive procedure* over the $\mathcal{W}_p$ spaces if when $f \in \mathcal{W}_p$

$$\widehat{f}_n \to f \text{ at rate } n^{-2p/2p+1}$$

without knowing p.

Rate adaptive estimators exist over a variety of function families and over a range of norms (or semi-norms).

Adaptive confidence sets?? Limited at best.

Inference Not So Easy

Using a rate-optimal smoothing parameter gives

$$\text{bias}^2 \approx \text{var}.$$

Loosely, if $\widetilde{f} = \mathsf{E}\widehat{f}$ and $s = \sqrt{\text{Var}\,\widehat{f}}$, then

$$\frac{\widehat{f} - f}{s} = \frac{\widehat{f} - \widetilde{f}}{s} + \frac{\widetilde{f} - f}{s} \approx \mathsf{N}(0,1) + \frac{\text{bias}}{\sqrt{\text{var}}}.$$

So, "$\widehat{f} \pm 2s$" undercovers.

Two common solutions in the literature:

– Bias Correction: Shift confidence set by estimated bias.

– Undersmoothing: Smooth so that var dominates bias^2.

Road Map

1. Dark Energy
2. Nonparametric Inference
3. Derivative Estimation as an Inverse Problem
4. Inference for the Dark Energy Equation of State

Derivative Estimation as an Inverse Problem

We can think of derivative estimation as an ill-posed inverse problem.

Suppose we have data

$$Y_i = r(z_i) + \sigma_i \epsilon_i$$

and want to make inferences about $f \equiv r'$. Then we can write (in vector form)

$$Y = Kf + \Sigma^{1/2}\epsilon$$

where the operator $K = (K_1, \ldots, K_n)$ maps functions to $\mathbb{R}^n$ and where $K_i = \int_0^{z_i}$.

Create an orthonormal basis $\phi_1, \ldots, \phi_n$ from the eigenfunctions of K^*K with associated eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$.

Here, K^* is the adjoint of K given by

$$K^* u = \sum_{i=1}^{n} u_i 1_{[0,z_i]}.$$

Derivative Estimation (cont'd)

Then,

$$f = \sum_{j=1}^{n} \beta_j \phi_j + f_\perp$$

$$= \sum_{j=1}^{n} \lambda_j^{-1/2} \langle u_j, Kf \rangle \phi_j + f_\perp,$$

where $u_j = K\phi_j / \|K\phi_j\|$. The $f_\perp$ component is not estimable.

Using an optimal shrinkage scheme,

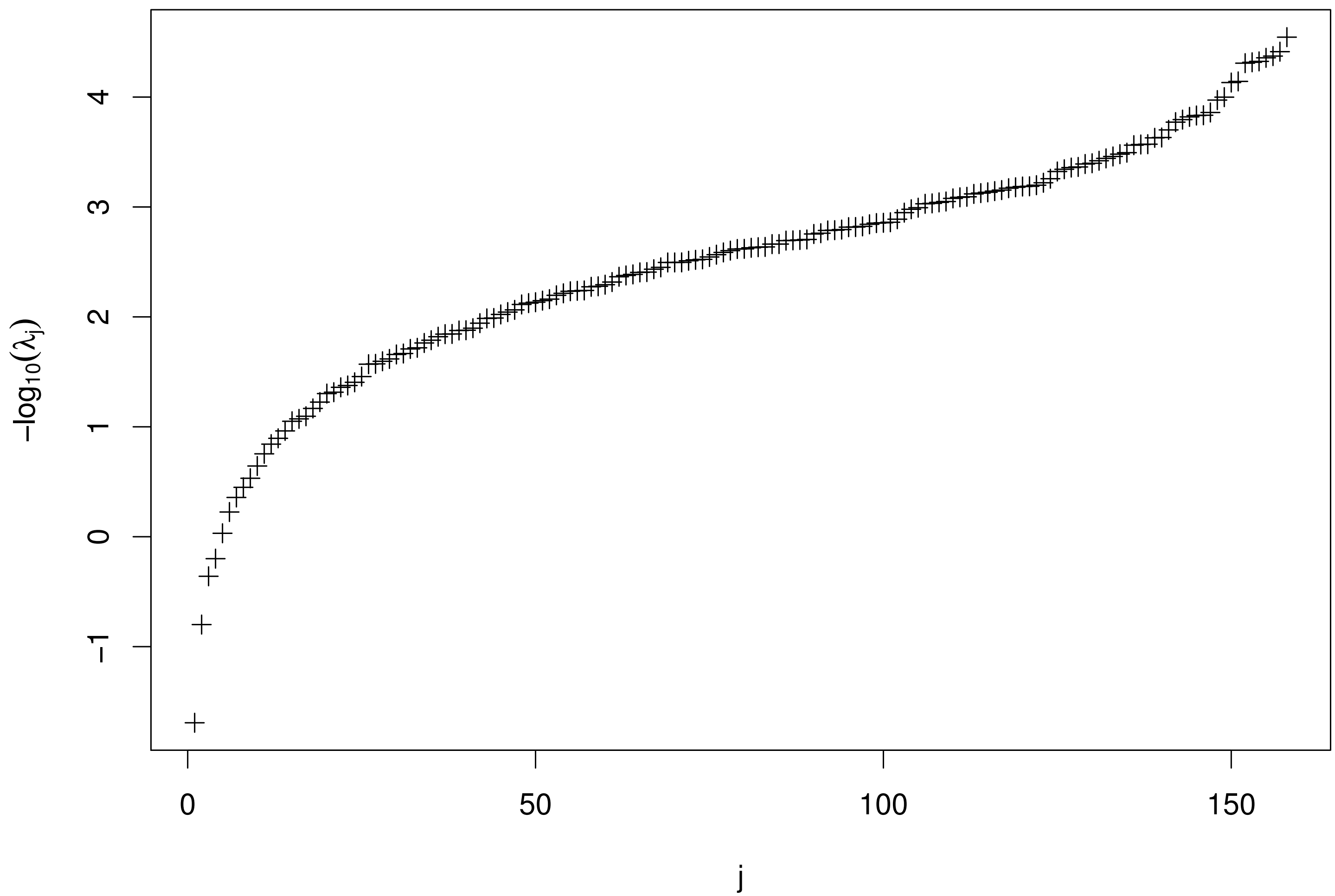
$$MSE \approx \sum_{j=1}^{n} \min(\beta_j^2, \lambda_j^{-1} \tau_j^2),$$

where $\tau_j^2 = \sum_k u_{jk}^2 \sigma_k^2$.

Large components at high order are bad news!

Derivative Estimation (cont'd)

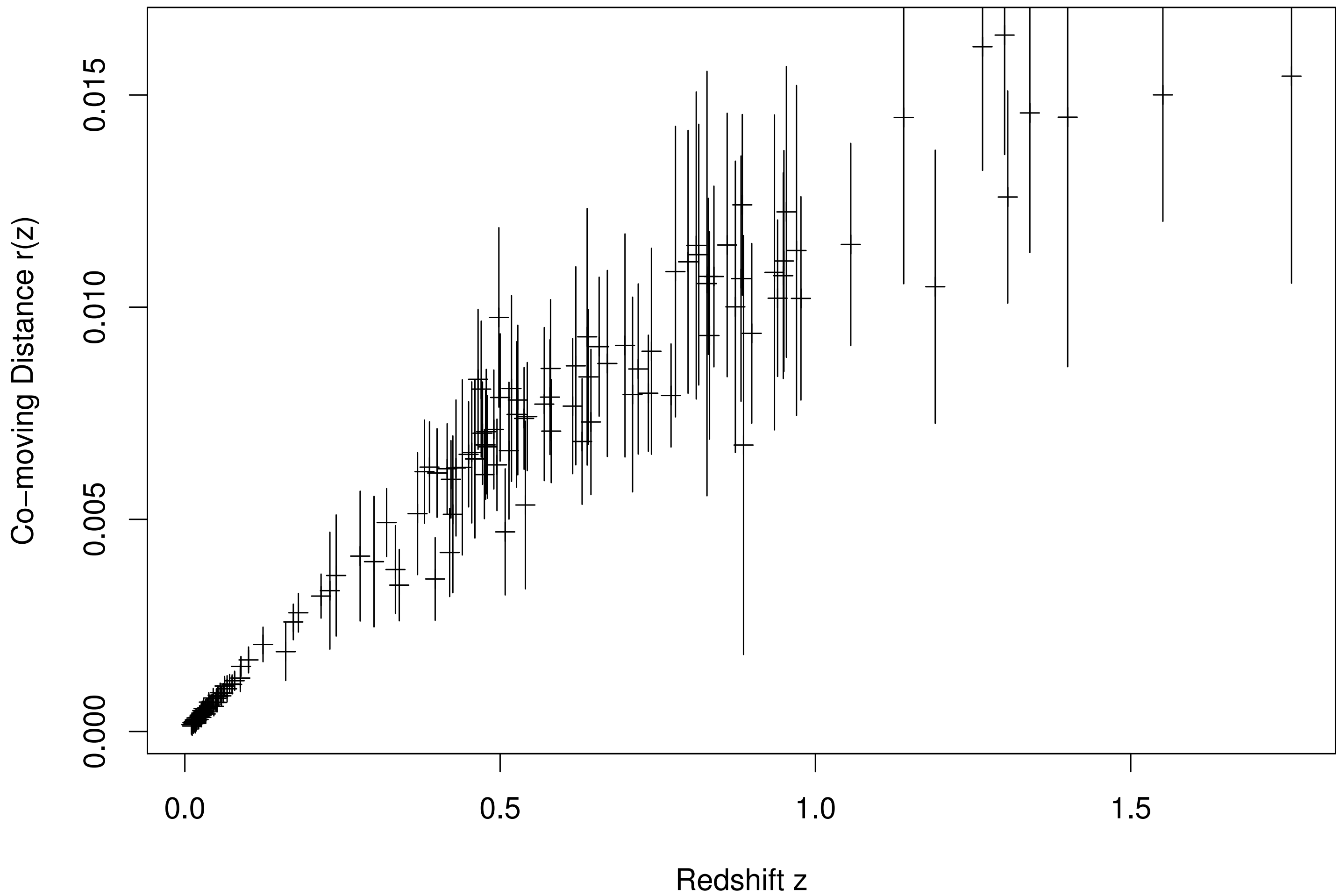
Associated Eigenvalues (as $-\log_{10}\lambda_j$) for the Supernova Data



Road Map

1. Dark Energy
2. Nonparametric Inference
3. Derivative Estimation as an Inverse Problem
4. Inference for the Dark Energy Equation of State

Inference for the Equation of State



Two approaches: Direct testing and nonparametric methods.

Inference (cont'd): Direct Testing

- Let $Q(z)$ be a fixed function, such as $Q(z) \equiv -1$.

  Then, the null hypothesis $H_0 : w = Q$ corresponds to the set of r' solving a differential equation:

  $$\frac{2}{3}(1+z)r''(z)+r'(z)(1+Q(z))-(r'(z))^3 Q(z)H_0^2\Omega_m(1+z)^3 = 0.$$

  This can be solved.

- The result is a family of solutions parameterized by $r'(0)$.

- Invert goodness of fit tests to generate a confidence set of those solutions in the family that are consistent with the data.

  Account for the uncertainty in $H_0^2\Omega_m$.

- Easily generalizes to any finite-dimensional family of $Q(z)$s.

Inference (cont'd): Nonparametric Methods

- Orthonormal basis expansion (singular functions or wavelet-vaguelete) or local polynomial regression
- Double reflection drastically reduces boundary bias in this problem.
- Minimize an unbiased estimate of risk $\widehat{R}$ to select tuning parameters
- Confidence bands via tube formula (Sun and Loader 1994).

  $$\widehat{f} \pm c\,\mathrm{se}(\widehat{f})$$

  where, c solves

  $$\alpha = 2(1 - \Phi(c)) + \kappa\phi(c),$$

  where Φ and ϕ are standard Gaussian cdf and density, respectively, and κ is a constant that depends on the procedure but not on f.
- Must account for bias. Commonly used methods of bias adjustment/estimation fail in simulations.

  Use a global estimate of bias to dilate the bands.

Inference (cont'd): Nonparametric Methods

Issue: What should the target of inference be?

- We can target the effective density $\rho_{\mathrm{DE}}/\rho_{\mathrm{crit}}$ or the equation of state w.
- The effective density can be estimated more precisely, but inference is harder: comparing growth rates of various components.
- The equation of state is harder to estimate (nonlinear functional of two derivatives), but inference is relatively straightforward ($w > -1$?).
- This is an empirical question that depends on accurate assessment of uncertainties: that is, good confidence sets.

Results

- Direct Testing: marginal rejection of $H_0 : w = -1$ with $p \approx 0.006$.

  But a 12% increase in standard errors eliminates the effect.

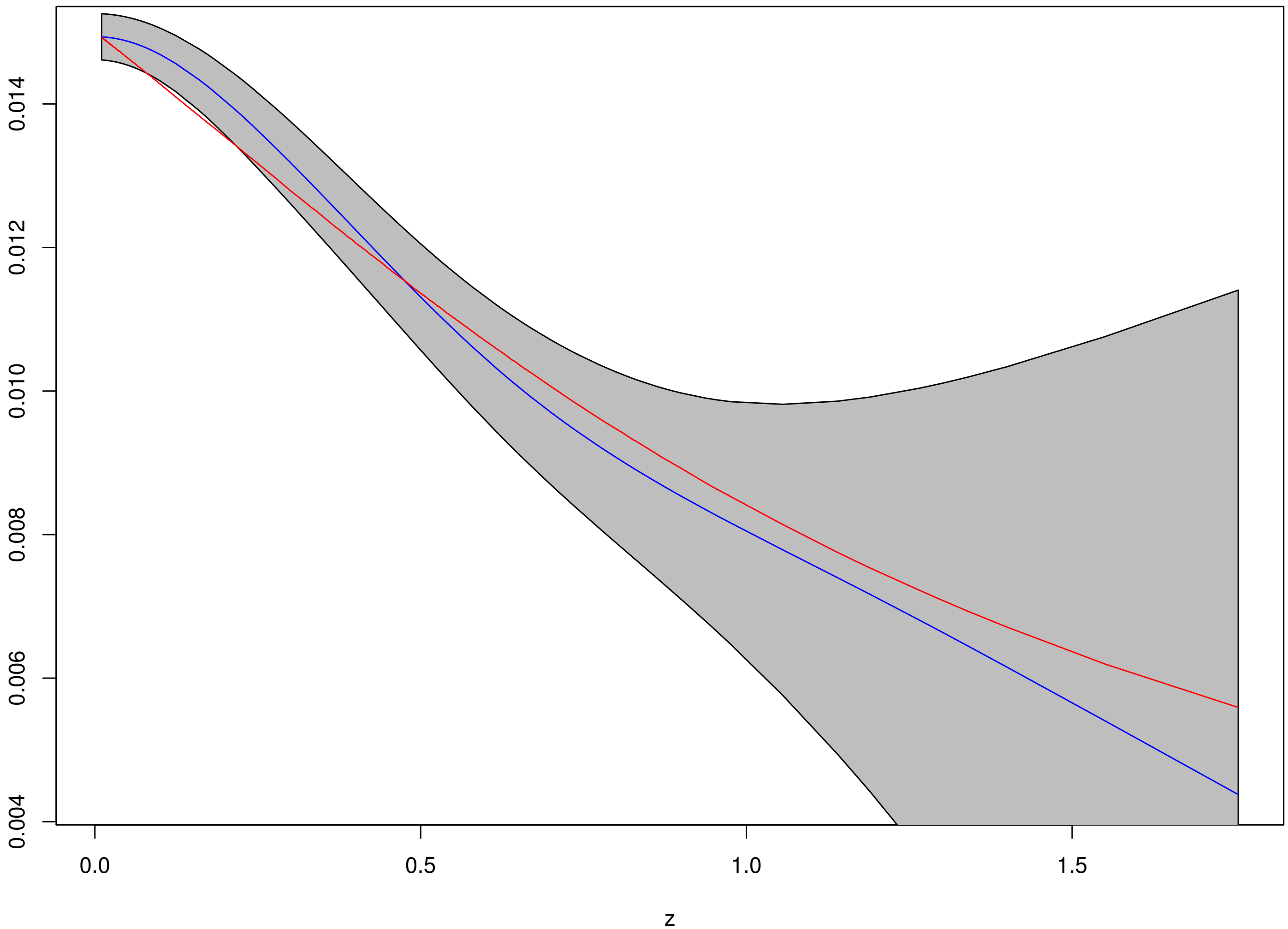
- Nonparametric methods: All the methods generally agree and give results consistent with what we would expect.

  Precision of the estimates is low at high redshift.

  Best fitting r' for $w = -1$ just outside of the confidence bands over small range.

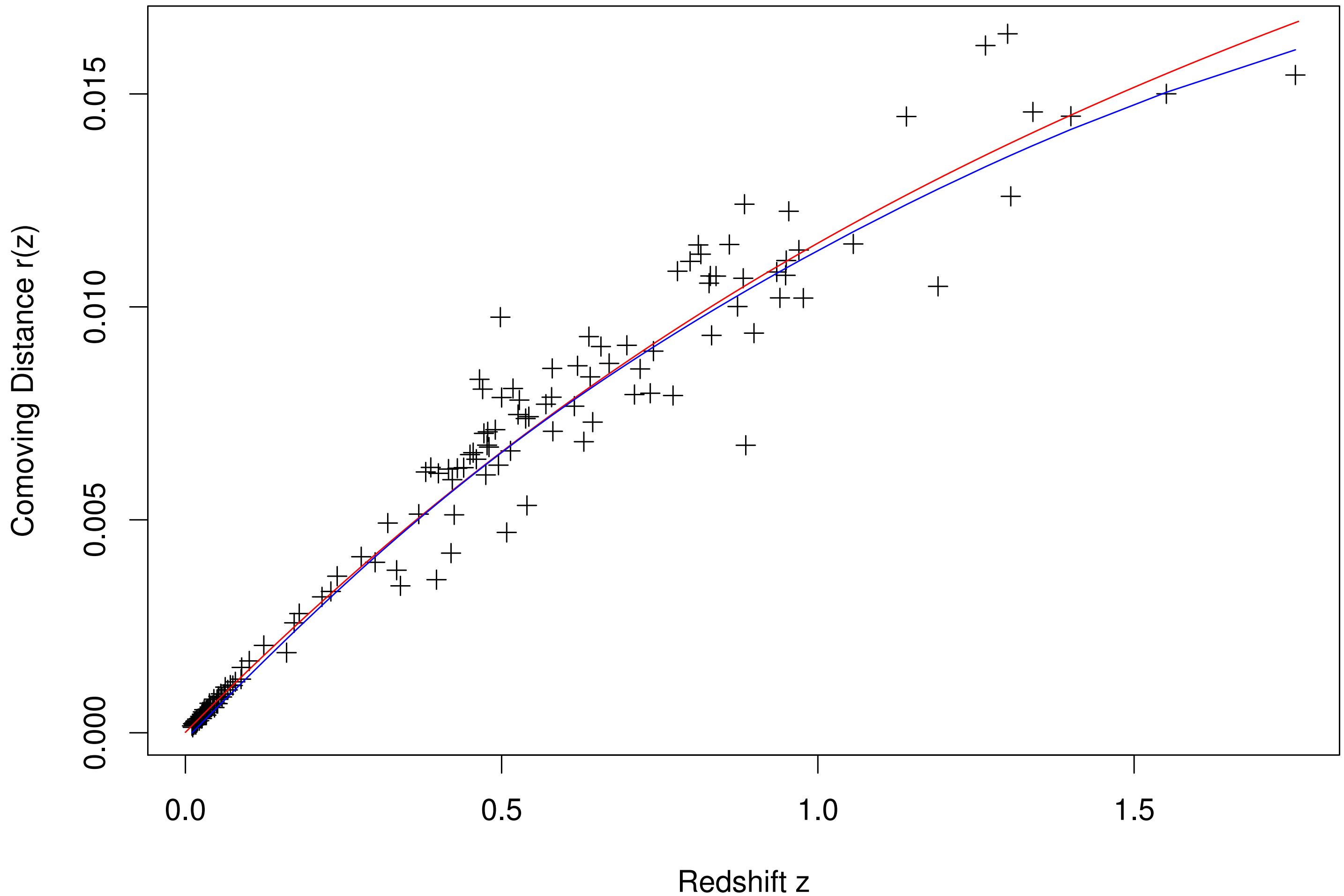
Results (cont'd)

$\widehat{r}'$ with confidence bands ($\widehat{r}'$ within band; $\widehat{r}_0'$ exits band)



Results (cont'd)

$\hat{r}$ from nonparametric and best fitting $\hat{r}_0$ from $w = -1$ solution.



Road Map

1. Dark Energy
2. Nonparametric Inference
3. Derivative Estimation as an Inverse Problem
4. Inference for the Dark Energy Equation of State

Take-Home Points

- Nonparametric methods can contribute to fundamental problems in cosmology and astrophysics.

- With the large data sets coming through the pipeline, we can eschew simpler parameterizations and go after the basic physics directly.

- The critical statistical problems focus on constructing inferences for the unknown function (e.g., confidence sets) and for complicated functionals.

Results (cont'd)

$\widehat{\rho}_{\mathrm{DE}}/\rho_{\mathrm{crit}}$ from nonparametric and best fitting from $w = -1$ solution.