Commentary on “Nonlinear Cosmostatistics” by Ben Wandelt

Christopher R. Genovese
Department of Statistics
Carnegie Mellon University

SCMA V 16 June 2011

The Problems

1. Detect primordial non-Gaussianity in the CMB

2. Infer the Hubble parameter $H(z)$ from the relation between void stretching and redshift.

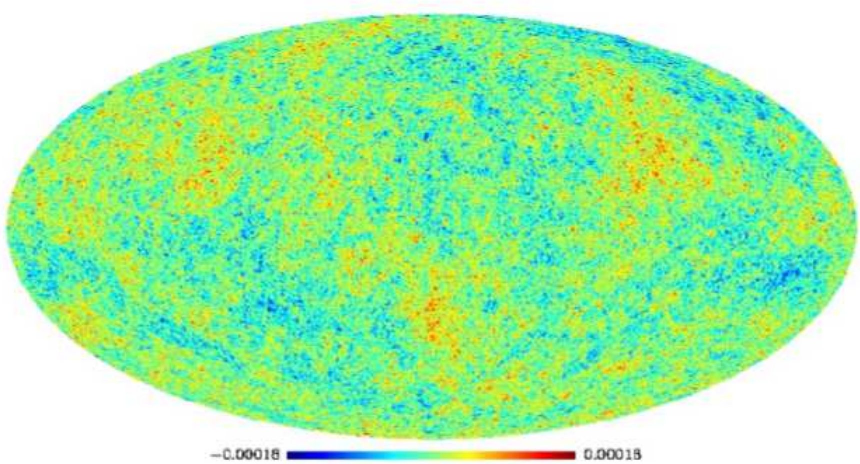
3. Improve photometric redshift correction

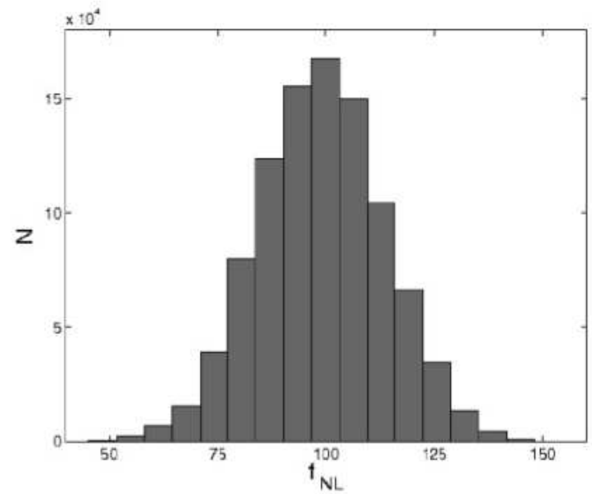
   (Alternately: Estimate galaxy positions and density field)

The Methods: Primordial Non-Gaussianity

- Theoretically motivated Bayesian hierarchical model
- Very large data size (order 10^7)
- MCMC with order 10^8 parameters

  Uses dynamical MCMC sampling technique to accelerate mixing.
- Output is posterior for f_{NL}





The Methods: Void Stretching

- Exploits isotropy and the well-known effects of cosmic expansion
- Unwraps one layer of integration (relative to, say, using D_L)
- Given the locations of particles that trace the matter distribution (e.g., galaxies):
  1. Identify voids in the matter distribution.
  2. "Average" the voids within distinct redshift shells.
  3. Fit an ellipsoid to each average.
  4. Map the stretching of the voids (relative to angular size on the sky) to an estimate of $H(z)$.
- The mapping from points $\rightarrow$ voids $\rightarrow$ shapes $\rightarrow$ stretching function $\rightarrow H(z)$ is a complicated *statistical* procedure.

  To assess the effectiveness of the method, we need to understand the statistical performance this procedure.

The Methods: Photo-z Correction

- Observe galaxy positions on the sky ($\boldsymbol{Y}$) and obtain (possibly corrected) photometric redshifts ($\boldsymbol{Z}_{\text{phot}}$)

- Parameters are the true redshifts $\boldsymbol{z}$ for each galaxy and the (non-parametric) mass density $\boldsymbol{m}$.

- Blockwise MCMC with order 10^7 parameters.

  Iteratively sample from two conditional distributions:

$$\boldsymbol{z}^{(k)} \leftarrow \boldsymbol{z} \mid \boldsymbol{m}^{(k-1)}, \boldsymbol{Y}, \boldsymbol{Z}_{\text{phot}}$$
$$\boldsymbol{m}^{(k)} \leftarrow \boldsymbol{m} \mid \boldsymbol{z}^{(k)}, \boldsymbol{Y}, \boldsymbol{Z}_{\text{phot}}$$

- Speed-up techniques for high-dimensions
  - Use dynamical MCMC for density field to speed up mixing
  - Exploit conditional independence for redshift to exploit parallelism

Key Ideas

- Simultaneous inference (e.g., redshift *and* density field)
- Physically meaningful (and yet often simplifying) priors
- Fast MCMC sampling techniques
- Block decomposition of parameter space

It is *dimensionality* rather than nonlinearity that is the fundamental challenge here.

Key Ideas

- Simultaneous inference (e.g., redshift *and* density field)
- Physically meaningful (and yet often simplifying) priors
- Fast MCMC sampling techniques
- Block decomposition of parameter space

It is *dimensionality* rather than nonlinearity that is the fundamental challenge here.

A Brief Look At Dynamical MCMC

- Random walk methods such as Metropolis-Hastings can move slowly, roughly a distance proportional to $\sqrt{m}$ in m steps.

  In high dimensions, this is far too slow to be practical.

- Embedding the problem in a dynamical system can produce a candidate that moves more easily to distant points.

  Simulate the dynamics for some (small) fixed time step.

  This candidate is deterministic but under certain conditions (reversibility and unit Jacobian) gives a valid – though not ergodic – chain.

- Dynamical (aka Hybrid or Hamiltonian) MCMC combines these two approaches . . .

A Brief Look At Dynamical MCMC (cont'd)

- Suppose $\pi(q)$ is (up to a proportionality constant) the posterior we want to sample from, in parameter vector $q = (q_1, \ldots, q_d)$.
- Augment the problem by introducing additional parameters $p = (p_1, \ldots, p_d)$, which we take to be standard Gaussian variables independent of each other and the q's. Later, we will throw away the p's.
- Writing $U(q) = -\ln \pi(q)$ and $T(p) = \frac{1}{2}\|p\|^2$, define the "Hamiltonian"

$$H(q,p) = U(q) + T(p) = -\ln \pi(q) + \frac{1}{2}\sum_{i=1}^{d} p_i^2.$$

This gives us an augmented posterior $\pi(q,p) \propto e^{-H(q,p)}$.

- The Hamiltonian induces "dynamics" via

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} = p_i$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} = \frac{\pi'(q)}{\pi(q)}.$$

The dynamics conserve H, preserve volume, and are reversible.

A Brief Look At Dynamical MCMC (cont'd)

The Hybrid Monte Carlo (HMC) algorithm (Duane et al. 1987, Neal 1996) samples from $\pi(q, p)$ by alternating the following two steps:

1. Gibbs step on the p's.

   Draw new standard Normal's IID for each p_i.

2. Metropolis step on q and p.

   Generate candidate state (q', p') by simulating the dynamics for a fixed time τ and negating the p_i's.

   (The negation makes the step reversible, ensuring that detailed balance holds.)

Because $\pi(q, p)$ factors, we can drop the p's from the sample to recover a sample from the original posterior $\pi(q)$.

A Brief Look At Dynamical MCMC (cont'd)

- Strengths
  - Can move more consistently in one direction

    After m steps, can move distance $\propto m$ rather than $\propto \sqrt{m}$ like Metropolis-Hastings.
  - Can change the probability density value more quickly

    Updating the p's changes the log density by order $\sqrt{d}$ rather than order 1 as for Metropolis-Hastings
  - Can move more easily to distant points

    Dynamics moves along $\pi(q, p)$ contours (up to discretization error)
- Potential Weaknesses
  - Performance can depend on discretization scheme
  - Can become trapped in isolated modes or near sharp gradients
  - $\pi(q)$ can change slowly with highly skewed distributions

Back to the methods. . .

Questions And Issues: Photo-z Correction

- Validation

  Good performance on simulations but would like to see results with real data where spectroscopic and photometric redshifts are available.

- Under what conditions are the input redshifts improved? Which method gives the best inputs?

- Could additional regularization constraints improve the density field estimator?

  Effectively producing a nonparametric estimate of the density field – typically need additional "smoothness" assumptions to get good performance. Is (approximate) isotropy enough?

- Sampling Performance

- Fragility and Generalizability (e.g., allowing redshift distribution to vary by galaxy type, more general priors)

- Possible downstream applications, e.g., active learning/experimental design

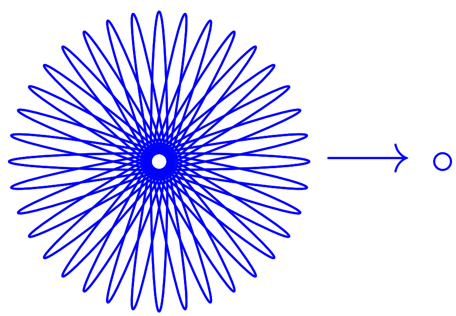
Questions And Issues: Voids

- This procedure comprises several interesting statistical problems that themselves merit study.

- Borrowing strength across redshift shells may improve estimation of $H(z)$.

- Choice of shells is a tuning parameter. Smoothing rather than binning?

- Estimating "average" void shape



- Sampling variation in the voids, leading to z-varying errors.

Bayesian Versus Frequentist Inference

- On the ground in Statistics, the Bayesian-Frequentist debate has mostly faded from view.
  - Nontrivial philosophical/conceptual differences certainly exist.
  - There are situations where each approach has an advantage.
  - Both approaches can be use successfully.

- But fundamentally: for hard problems there is no easy path.

  The main practical difference lies in *when* one needs to be creative.

- Addendum: Nonparametric and Bayesian methods complement each other.

Inference Versus Prediction

- Inference
  - Attempt to learn about the *true* data-generating mechanism.
  - Success determined average "distance" to the true mechanism.
  - Model and parameters are meaningful (even if not completely accurate)
  - Assessment of uncertainty is critical (and is sometimes an extra step)
  - Usually based on a stochastic, generative model for the data.
- Prediction
  - Attempt to produce an algorithm that can accurately predict new data
  - Success determined by prediction error alone
  - Models and parameters are a means to that end, not unique or meaningful.
  - Uncertainty captured by prediction error
  - Traditional statistical models are useful but not necessary.

Inference Versus Prediction (cont'd)

- This distinction is correlated with Statistics vs. Machine Learning, though not perfectly.
- Inference can be hard, especially in high dimensions or with complex data/structures.

  Ex: adaptive estimators versus confidence sets, high dimensional HPD's

  (In contrast, super-efficient convergence for classification under Tsybakov margin/low-noise condition.)
- Astronomy/Astrophysics/Cosmology has many problems of both types.
- When designing procedures for analyzing complex data, we need to understand what criteria a problem requires and what criteria the methods were designed to optimize.
- One challenge lies in combining methods based on different criteria (e.g., classification results feeding into an inferential regression).

Take-Home Points

- Impressive work both scientifically and methodologically
- Opens up new frontiers for high-dimensional Bayesian inference
- Statistical performance promising but needs further study
- Nonlinearity/Non-Gaussianity important but dimensionality is driving the difficulty of these problems
- Inference versus Prediction