Statistical Analysis in Cosmology

Francisco-Shu Kitaura

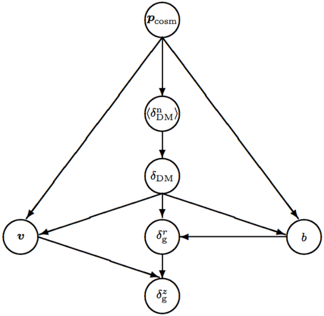
March 8, 2011

Introduction

- the cosmological large-scale structure encodes a wealth of information about the evolution and origin of the Universe
- the data are plagged by many observational effects (sky mask, radial selection, bias, shot noise ...)
- statistical treatment is necessary
- compare observations with theory
- study structure formation

Classes of uncertainty

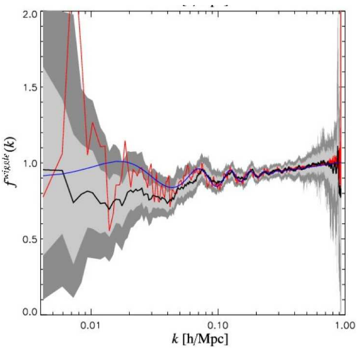
- intrinsic stochastic character: cosmic variance ▸ 1
- physical uncertainties: galaxy bias ▸ 2, nonlinear structure formation ▸ 3
- observational uncertainties: redshift distortions (linear, non-linear) ▸ 4, radial selection function (magnitud limited sample), sky mask (scanning strategy) ▸ 5
- mathematical/representation uncertainties: aliasing effects (mass assignment scheme, pixel window) ▸ 6



scheme: FK & Ensslin 2008

Cosmic variance

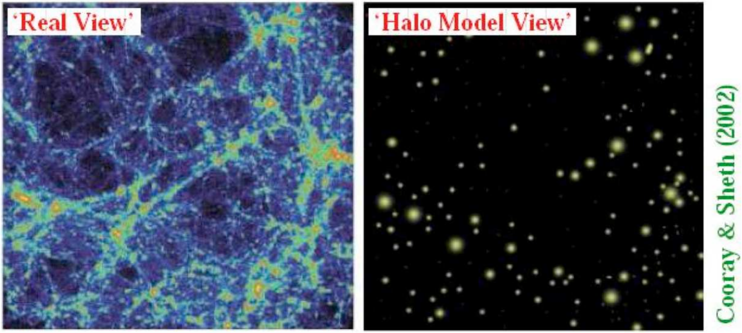
- intrinsic variance: particular realization of the Universe
- observational volume variance
- mathematical/representation (grid): number of k-vectors per mode.



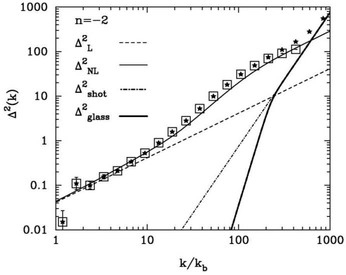
plot: Jasche, FK, Wandelt & Ensslin 2010

back

Bias

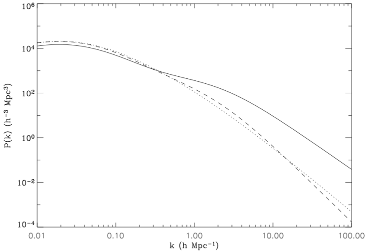


Nonlinear clustering



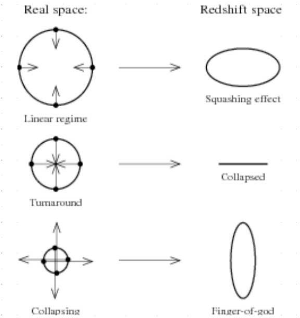
Smith et al 2010; Peacock and Dodds 94

▸ back

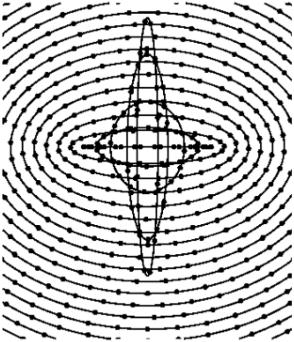


plot: Erdogdu et al 2004

Redshift distortions

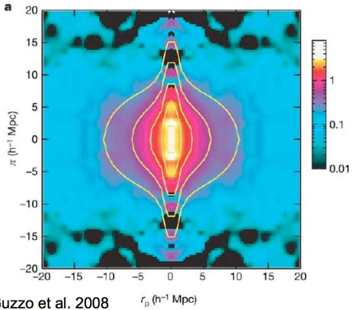


take care with this picture!
Nonlinear fogs are due to virialized random motions.



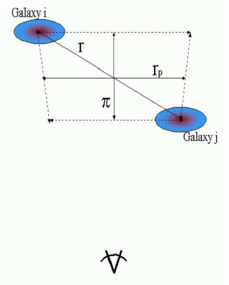
Kaiser 87; Ballinger et al 96; plots: Hamilton 97, 98

Redshift distortions



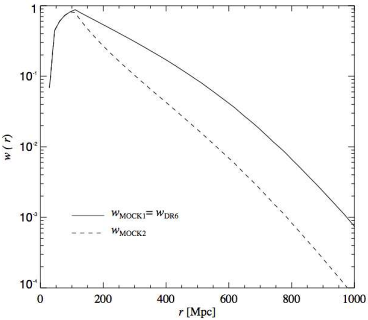
Guzzo et al. 2008

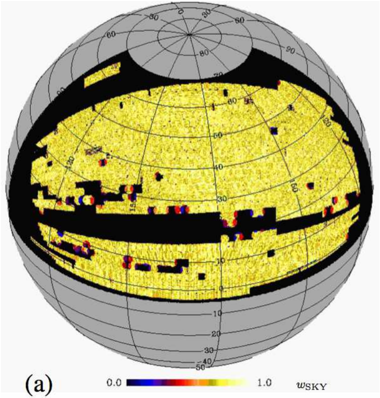
back



scheme: Stefanie Phleps

Completeness: radial selection function and sky mask

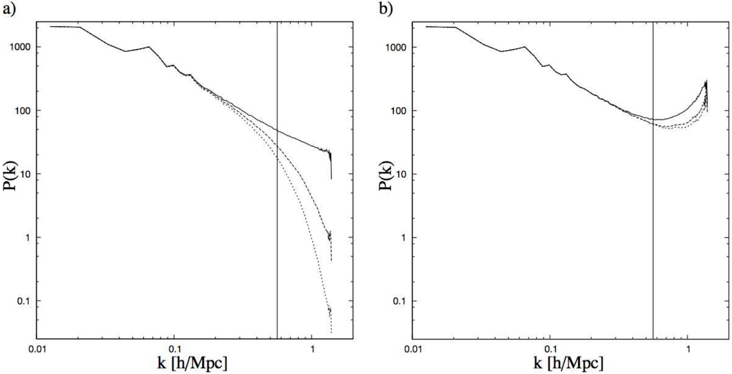




FK et al 2009

back

Aliasing: Mass assignment scheme



NGP, CIC, TSC

Hockney & Eastwood 81; plots: Jasche, FK & Ensllin 2010

Data model: signal degradation model

- Nonlinear data model: data $\mathbf{d}$ m-vector, signal $\mathbf{s}$ n-vector, $n >> m$

$$\mathbf{d} = R(\mathbf{s}) + \boldsymbol{\epsilon} \quad (1)$$

- Linear data model

$$\mathbf{d} = \mathbf{R}\mathbf{s} + \boldsymbol{\epsilon} \quad (2)$$

$$d_i = \sum_j R_{ij} s_j + n_i \quad (3)$$

- $\mathbf{R}$ response operator ($m \times n$ matrix) may include: mask, selection fnct, foregrounds, blurring fnct, PSF, pixel window ...
- $\boldsymbol{\epsilon}$ noise: random component, white noise, colored noise, attention: mask, selection fnct, pixel window etc. (see Jing 2005, FK et al 2009)

Estimate of the signal

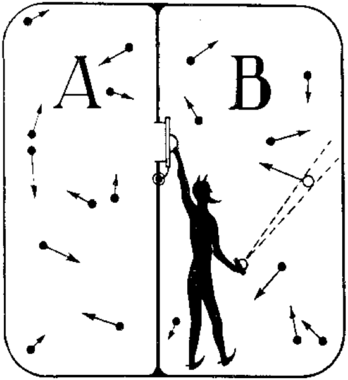
How do we get an estimate of $\mathbf{s}$: $\hat{\mathbf{s}}$?

Notions of information

- What is information?
- How do we quantify it?
- Etymology: (*Latin*) informare: *give form to the mind*
- Systems theory: *information is any type of pattern that influences the formation of other patterns*
- J. D. Bekenstein 2003: *the physical world is made of information itself*
- Relation between entropy and information: Maxwell's demon

Information and entropy: Maxwell's demon

- Container in thermal equilibrium divided into 2 parts A and B with a trapdoor
- Demon lets faster molecules pass from B to A
- Kinematic energy is reduced in B
- Violation of second law of thermodynamics?
- The information on the molecule velocities increases the overall entropy!
- Information and entropy are tightly related.



Information and entropy

- Entropy

$$S = -k_B \sum_i P_i \ln P_i \tag{4}$$

- density of states Ω

$$S = k_B \ln \Omega \tag{5}$$

Shannon's entropy

- Claude Elwood Shannon (April 30, 1916 February 24, 2001), father of information theory, electronic ingeneer worked for Bell Labs (Shannon 1948 *A mathematical theory of communication* ▸ link)
- Shannon's entropy

$$H(\mathbf{x}) = -\sum_i P(x_i) \log_b P(x_i) \tag{6}$$

units of entropy: b=2: bits; b=e: nats; b=10: dits

- conditional entropy

$$H(\mathbf{y}|\mathbf{x}) = \sum_{i,j} P(x_i, y_j) \log_b \frac{P(y_j)}{P(x_i, y_j)} \tag{7}$$

Shannon's entropy

- Kullback Leibler distance between two distributions (Jones test!)

$$D_{\mathrm{KL}}(P||Q) = \sum_i P_i \log_b \frac{P_i}{Q_j} \tag{8}$$

- Mutual information

$$I(x, y) = \sum_{i,j} P(x_i, x_j) \log_b \frac{P(x_i, x_j)}{P(x_i)P(y_j)} \tag{9}$$

Cramer-Rao inequality/lower bound

- Harald Cramer (September 25, 1893 October 5, 1985) Swedish mathematician, actuary, and statistician
- Calyampudi Radhakrishna Rao (born September 10, 1920) Indian statistician. Prof. em. at Penn State Univ. and Res. Prof. Univ. at Buffalo (still today! look at link)
- unbiased estimator

$$\langle \hat{s} \rangle = s \tag{10}$$

$$\langle \hat{s} - s \rangle \equiv \int \mathrm{d}d\, P(d|s)(\hat{s} - s) = 0 \tag{11}$$

- $\partial/\partial s \rightarrow$

$$\int \mathrm{d}d\,(\hat{s} - s)\frac{\partial P(d|s)}{\partial s} - \int \mathrm{d}d\, P(d|s) = 0 \tag{12}$$

Cramer-Rao inequality/lower bound

- Cauchy Schwarz inequality, inner product:

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle \tag{14}$$

$$\rightarrow \left[\int \mathrm{d}d \left(\frac{\partial \ln P(d|s)}{\partial s}\right)^2 P(d|s)\right]\left[\int \mathrm{d}d\,(\hat{s}-s)^2 P(d|s)\right] \geq 1 \tag{15}$$

- Mean Squared Error (MSE)

$$e^2(s) \equiv \langle(\hat{s}-s)^2\rangle = \int \mathrm{d}d\,(\hat{s}-s)^2 \tag{16}$$

- Fisher information

$$\mathcal{I}(s) \equiv \int \mathrm{d}d \left(\frac{\partial \ln P(d|s)}{\partial s}\right)^2 P(d|s) \equiv \langle\left(\frac{\partial \ln P(d|s)}{\partial s}\right)^2\rangle \tag{17}$$

Cramer-Rao inequality/lower bound

-
$$e^2\mathcal{I} \geq 1 \tag{18}$$
- If $e^2\mathcal{I} = 1 \rightarrow$ Minimum Variance Unbiased (MVU) estimator
- In general there is a statistical bias $\mathrm{B}(\hat{s}) \equiv \langle \hat{s} \rangle - s$

$$\mathrm{MSE}(\hat{s}) = \mathrm{VAR}(\hat{s}) + \mathrm{B}^2(\hat{s}) \tag{19}$$

$$\mathrm{VAR}(\hat{s}) \equiv \sigma^2(s) \equiv \langle (\hat{s} - \langle \hat{s} \rangle)^2 \rangle \tag{20}$$

Fisher information

- Score

$$\mathcal{S} \equiv \frac{\partial}{\partial s} \ln P(d|s) = \frac{1}{P(d|s)} \frac{\partial P(d|s)}{\partial s} \tag{21}$$

- Fisher information: variance of the score

$$\mathcal{I}(s) \equiv \langle \left(\frac{\partial \ln P(d|s)}{\partial s} \right)^2 \rangle \tag{22}$$

- regularity condition

$$\int \mathrm{d}d \, \frac{\partial^2 P(d|s)}{\partial s^2} = 0 \tag{23}$$

-

$$\mathcal{I}(s) = -\langle \frac{\partial^2 \ln P(d|s)}{\partial s^2} \rangle \tag{24}$$

Fisher information

- $$\mathcal{I}(s) = -\langle \frac{\partial^2 \ln P(d|s)}{\partial s^2} \rangle \tag{25}$$

- $$\begin{aligned} &= -\int \mathrm{d}d\, P(d|s) \left[\frac{\partial}{\partial s}\left(\frac{1}{P(d|s)}\frac{\partial P(d|s)}{\partial s}\right)\right] \\ &= -\int \mathrm{d}d\, P(d|s) \left[-\frac{1}{P(d|s)^2}\left(\frac{\partial P(d|s)}{\partial s}\right)^2 + \frac{1}{P(d|s)}\frac{\partial^2 P(d|s)}{\partial s^2}\right] \\ &= \int \mathrm{d}d\, P(d|s)\frac{1}{P(d|s)^2}\left(\frac{\partial P(d|s)}{\partial s}\right)^2 - \int \mathrm{d}d\, \frac{\partial^2 P(d|s)}{\partial s^2} \end{aligned} \tag{26}$$

Fisher information

- Generalization to Fisher matrix

$$\mathcal{I}(s)_{ij} = -\langle \frac{\partial^2}{\partial s_i \partial s_j} \ln P(d|s) \rangle \tag{27}$$

- Information may be seen to be a measure of the *sharpness*/curvature of the support curve ($\ln P(d|s)$) near the Maximum Likelihood estimate of **s**.
- What is the Maximum Likelihood?

Estimator of the signal

- expectation/mean/ensemble average of $\mathbf{s}$:

$$\mathrm{E}[\mathbf{s}] = \langle \mathbf{s} \rangle \tag{28}$$

- take care specify ensemble average!

$$\mathrm{E}[\mathbf{s}] = \int \mathrm{d}\mathbf{s}\,\mathrm{d}\mathbf{p}\,P(\mathbf{s},\mathbf{p}|\mathbf{d})\,\mathbf{s} = \langle \mathbf{s} \rangle_{(\mathbf{s},\mathbf{p}|\mathbf{d})} \tag{29}$$

Probability theory axioms

- Sum rule: *OR* (Venn diagram)

$$P(\mathbf{a}_1+\mathbf{a}_2|\mathbf{c}) = P(\mathbf{a}_1|\mathbf{c}) + P(\mathbf{a}_2|\mathbf{c}) - P(\mathbf{a}_1,\mathbf{a}_2|\mathbf{c}) \qquad (30)$$

- Product rule: *AND*

$$P(\mathbf{a},\mathbf{b}|\mathbf{c}) = P(\mathbf{a}|\mathbf{b},\mathbf{c})P(\mathbf{b}|\mathbf{c}) \qquad (31)$$

- Invariance under permutation of arguments

$$P(\mathbf{s},\mathbf{d}|\mathbf{p}) = P(\mathbf{s}|\mathbf{d},\mathbf{p})P(\mathbf{d}|\mathbf{p}) \qquad (32)$$

$$P(\mathbf{d},\mathbf{s}|\mathbf{p}) = P(\mathbf{d}|\mathbf{s},\mathbf{p})P(\mathbf{s}|\mathbf{p}), \qquad (33)$$

Probability distribution functions are always conditioned on some prior information!

Bayes theorem: the posterior/inference

- Thomas Bayes (c. 1702 17 April 1761) was an English mathematician and Presbyterian minister

$$\begin{aligned} P(\mathbf{s}|\mathbf{d},\mathbf{p}) &= \frac{P(\mathbf{s}|\mathbf{p})P(\mathbf{d}|\mathbf{s},\mathbf{p})}{P(\mathbf{d}|\mathbf{p})} \\ \text{posterior} &= \text{prior} \times \text{likelihood/evidence} \end{aligned} \tag{34}$$

- Likelihood: $\mathcal{L}(\mathbf{s}|\mathbf{d},\mathbf{p}) = \mathcal{P}(\mathbf{d}|\mathbf{s},\mathbf{p})$
- update of prior with posterior $\rightarrow$ learning algorithm
- Bayesian notion of information: information is encoded in conditional probability distribution functions

Evidence

- normalization of the posterior
- margnalization over the signal

$$P(\mathbf{d}|\mathbf{p}) = \int \mathrm{d}\mathbf{s}\, P(\mathbf{d},\mathbf{s}|\mathbf{p}) = \int \mathrm{d}\mathbf{s}\, P(\mathbf{s}|\mathbf{p})P(\mathbf{d}|\mathbf{s},\mathbf{p}) \qquad (35)$$

Bayesian inference steps

- Definition of the prior: knowledge of the underlying signal
- Definition of the likelihood: nature of the observed data
- Linking the prior to the likelihood: link signal to the data
- Bayes theorem: the posterior
- Maximization of the posterior: Maximum a posteriori: MAP
- Sampling the posterior: MCMC (read Neal 93), importance sampling, population Monte Carlo (M. Kilbinger)

Informative priors

- Gaussian prior (Wiener 1949, Rybicki & Press 1992, Zaroubi et al 1995) $\rightarrow$ Thikonov regularization
- Lognormal prior/nonlinear transformation (Tarantola & Valette 1982, FK et al 2010)
- Expanded Gaussian prior (Juszkiewiz et al 95, Bernardeau & Kofman 95, Colombi 94, FK 2010)

Gaussian prior

- Gaussian likelihood

$$P(\mathbf{d}|\mathbf{s},\mathbf{p}) \propto \exp\left(-\frac{1}{2}\mathbf{s}^{\dagger}\mathbf{S}^{-1}\mathbf{s}\right) \tag{36}$$

$\mathbf{S} = \langle \mathbf{s}\mathbf{s}^{\dagger}\rangle_{(\mathbf{s}|\mathbf{p})}$

- Wiener-filter $\hat{\mathbf{s}} = \mathbf{F}\mathbf{d}$ with
  $\mathbf{F} = (\mathbf{S}^{-1} + \mathbf{R}^{\dagger}\mathbf{N}^{-1}\mathbf{R})^{-1}\mathbf{R}^{\dagger}\mathbf{N}^{-1} = \mathbf{S}\mathbf{R}^{\dagger}(\mathbf{R}\mathbf{S}\mathbf{R}^{\dagger} + \mathbf{N})^{-1}$
  $\mathbf{N} \equiv \langle \boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\dagger}\rangle_{(\boldsymbol{\epsilon}|\mathbf{p})}$
- take care with the definition of the noise matrix

Non-informative priors

- Flat prior
- Jeffrey's prior (prior for the power-spectrum: FK & Ensslin 08)
- Entropic prior (Jaynes 63, Narayan & Nityananda 86, Skilling 89, FK & Ensslin 08)
- (Edwin Thompson Jaynes (Waterloo, Iowa, July 5, 1922 St. Louis, Missouri, April 30, 1998) Professor of Physics at Washington University in St. Louis) ▸ link

Flat prior

- improper prior: integral diverges to infinity
- maximization leads to maximum likelihood ML
- Gaussian likelihood

$$P(\mathbf{d}|\mathbf{s},\mathbf{p}) \propto \exp\left(-\frac{1}{2}\chi^2(\mathbf{s})\right) \tag{37}$$

$\chi^2(\mathbf{s}) \equiv \boldsymbol{\epsilon}^\dagger \mathbf{N}^{-1}\boldsymbol{\epsilon}$

- COBE-filter: $\mathbf{F} = (\mathbf{R}^\dagger\mathbf{N}^{-1}\mathbf{R})^{-1}\mathbf{R}^\dagger\mathbf{N}^{-1}$

Flat prior

- Poissonian likelihood

$$P(\mathbf{N}|\boldsymbol{\lambda}, \mathbf{p}) = \Pi_i \exp(-\lambda_i) \frac{\lambda_i^{N_i}}{N_i} \tag{38}$$

- Richardson-Lucy deconvolution algorithm (Richardson 1972, Lucy 1974, Shepp & Vardi 1982)

Lognormal prior

- continuity equation:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + \frac{1}{a}\rho\nabla_{\mathbf{r}} \cdot \mathbf{v} = 0, \tag{39}$$

- solution:

$$\rho = \langle\rho\rangle e^{s}, \;\; s = -\int \mathrm{d}t \frac{1}{a}\nabla_{\mathbf{r}} \cdot \mathbf{v}, \tag{40}$$

- if s is Gaussian at all times $\rightarrow$ ρ is lognormal distributed (Coles & Jones 91)
- when structures start to virialize the peculiar velocity field changes
- relax the Gaussian assumption Colombi 94: skewed lognormal model with the 1D Edgeworth exansion. (based on the skewed Gaussian Edgeworth expansion: Juszkiewicz, Bouchet & Colombi 93)

Multidimensional

- Let me introduce here the multidimensional case:

$$\Phi_i \equiv \ln \rho_i - \langle \ln \rho \rangle = s_i - \mu_i, \;\; \nu_i \equiv \sum_j S_{ij}^{-1/2} \Phi_i \, , \tag{41}$$

$s_i = \ln(\rho_i / \langle \rho \rangle) = \ln(1 + \delta_{\mathrm{M}i})$

Multidimensional

Multidimensional Edgeworth expansion

$$P(\mathbf{\Phi}) = \frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\mathbf{\Phi}} P(\boldsymbol{\nu}) = (\det(\mathbf{S}))^{-1/2} G(\boldsymbol{\nu}) \tag{42}$$

$$\times \Bigg[1 + \frac{1}{3!} \sum_{ijk} \langle \nu_i \nu_j \nu_k \rangle_{\mathrm{c}} h_{ijk}(\boldsymbol{\nu}) + \frac{1}{4!} \sum_{ijkl} \langle \nu_i \nu_j \nu_k \nu_l \rangle_{\mathrm{c}} h_{ijkl}(\boldsymbol{\nu})$$

$$+ \frac{1}{6!} \sum_{ijklmn} \left[\frac{1}{3!3!2} \sum_{j_1 \ldots j_6 \in [1,\ldots,6]} \tilde{\epsilon}_{j_1 \ldots j_6} \langle \nu_{i_{j_1}} \nu_{i_{j_2}} \nu_{i_{j_3}} \rangle_{\mathrm{c}} \langle \nu_{i_{j_4}} \nu_{i_{j_5}} \nu_{i_{j_6}} \rangle_{\mathrm{c}} \right]_{10} h_{ijklmn}(\boldsymbol{\nu}) + \ldots \Bigg] ,$$

Lognormal model

- 0th order: lognormal model ($G(\boldsymbol{\nu}) \rightarrow P(\boldsymbol{\delta}_{\rm M}|\mathbf{S})$)
  $\langle \ln \rho_i \rangle = \ln \langle \rho \rangle + \mu_i$

$$P(\boldsymbol{\delta}_{\rm M}|\mathbf{S}) = \frac{1}{\sqrt{(2\pi)^{N_{\rm cells}} \det(\mathbf{S})}} \prod_k \frac{1}{1+\delta_{{\rm M}k}} \times \exp\left(-\frac{1}{2} \sum_{ij} \left(\ln(1+\delta_{{\rm M}i}) - \mu_i\right) S^{-1}_{ij} \left(\ln(1+\delta_{{\rm M}j}) - \mu_j\right) \right) , \quad (43)$$

  multidimensional implementation for matter field reconstructions:
  FK, Jasche & Metcalf 2009; applied to SDSS DR7: Jasche, FK, Li & Ensslin 2010
- when $\delta_{\rm M} \ll 1 \rightarrow$ Gauss distribution

Maximum a posteriori

- Let us define the *energy* $E(\mathbf{s})$

$$E(\mathbf{s}) \equiv -\ln\left(P\left(\mathbf{s}|\mathbf{d},\mathbf{S}\right)\right), \tag{44}$$

- MAP

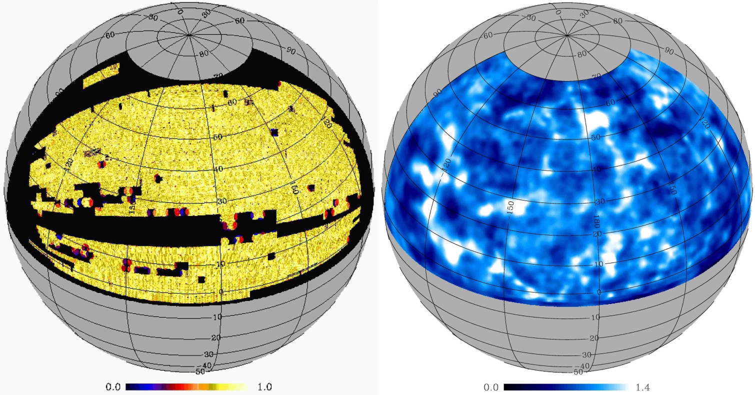
$$\frac{\partial E(\mathbf{s})}{\partial s_l} = 0, \tag{45}$$

- Krylov conjugate gradient schemes (FK & Ensslin 2008; Jasche, FK, Wandelt, Ensslin 2009; FK, Jasche, & Metcalf 2009)

$$s_i^{j+1} = s_i^j - \sum_k T_{ik}\frac{\partial E(\mathbf{s})}{\partial s_k}, \tag{46}$$

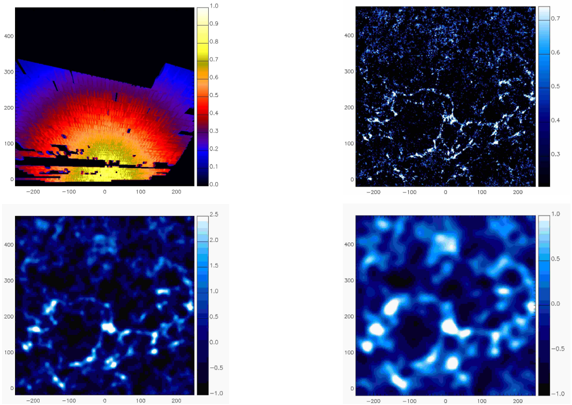
Results: Wiener filter reconstruction of the SDSS DR6

Wiener-filter with the ARGO code: FK, Jasche, Li, Ensslin, Metcalf, Wandelt, Lemson & White 2009



Results: detection of a super-void in the SDSS DR6

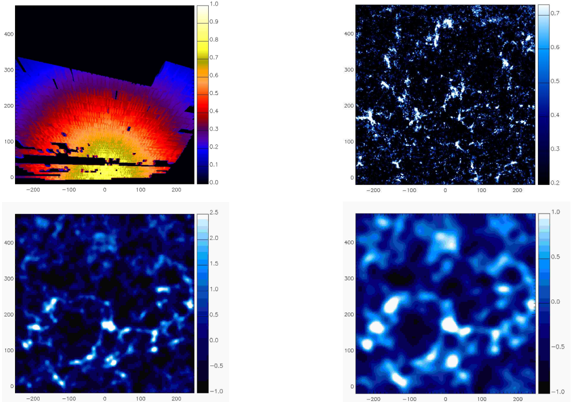
(about 250.000 galaxies from the main sample)



FK, Jasche, Li, Ensslin, Metcalf, Wandelt, Lemson & White 2009

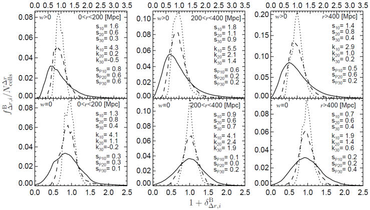
Results: detection of a super-void in the SDSS DR6

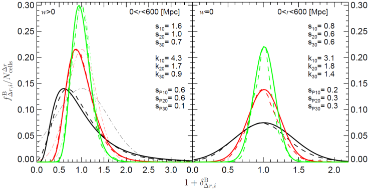
(about 250.000 galaxies from the main sample)$\rightarrow$ cluster prediction



FK, Jasche, Li, Ensslin, Metcalf, Wandelt, Lemson & White 2009

Results: matter statistics in the SDSS DR6

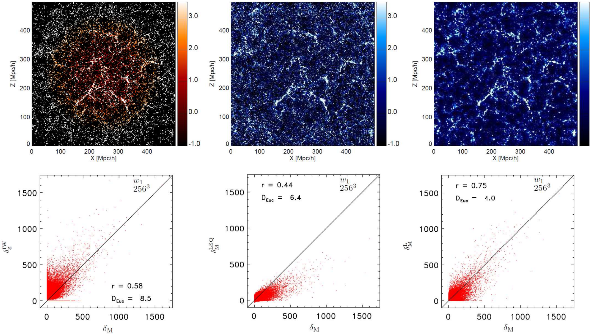




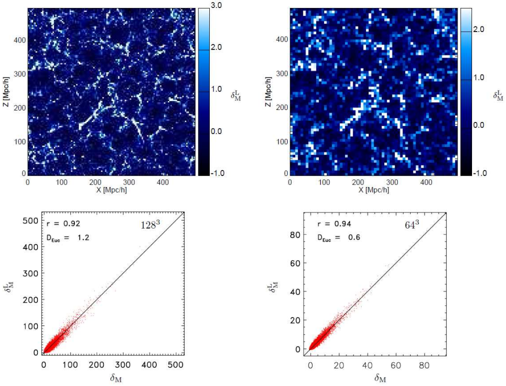
FK, Jasche, Li, Ensslin, Metcalf, Wandelt, Lemson & White 2009

Results: lognormal filter against Wiener filter and inverse weighting

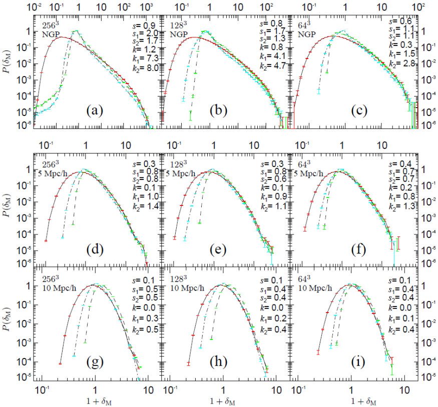
tests with the Millenium Run including selection function effects (about 350.000 mock galaxies)



Results: lognormal filter against Wiener filter and inverse weighting



Results: matter statistics in the lognormal reconstruction



FK, Jasche & Metcalf 2009 (upgrade of the ARGO code)

Markov Chains

- Andrey Markov (June 14, 1856 N.S. July 20, 1922) was a Russian mathematician
- Famous papers:
  - A.A. Markov. "Extension of the limit theorems of probability theory to a sum of variables connected in a chain". reprinted in Appendix B of: R. Howard. Dynamic Probabilistic Systems, volume 1: Markov Chains. John Wiley and Sons, 1971. (original in Russian 1906)
  - applied to language and vowels: ▸ link
- Gibbs-sampling, Metropolis-Hastings, Hybrid MCMC, Hamiltonian MCMC, etc

Gaussian distributions: Gibbs sampling

- Geman & Geman 84, Wandelt 04, Eriksen et al 07, FK & Ensslin 08, Jasche, FK, Wandelt & Ensslin 10, (read Neal 93)
- Power spectrum and map sampling

$$\mathbf{s}^{j+1} \sim P(\mathbf{s}|\mathbf{S}^j, \mathbf{d}) \quad (47)$$

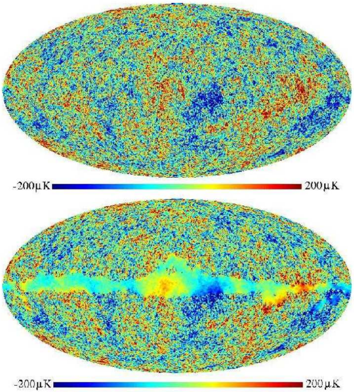
$$\mathbf{S}^{j+1} \sim P(\mathbf{S}|\mathbf{s}^{j+1}) \quad (48)$$

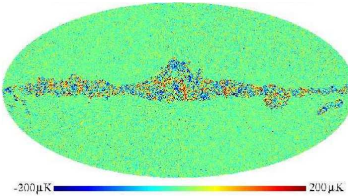
- Wiener filter

$$\mathbf{s}^j = \hat{\mathbf{s}}^j + \mathbf{y}^j \quad (49)$$

$\mathbf{y}^j = ((\mathbf{S}^j)^{-1} + \mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R})^{-1}((\mathbf{S}^j)^{-1/2}\mathbf{x}_1 + \mathbf{R}^\dagger \mathbf{N}^{-1/2}\mathbf{x}_2)$ with Gaussian random variates $\mathbf{x}_1$ and $\mathbf{x}_2$

Gaussian distributions: Gibbs sampling





Eriksen et al 2007

Sampling the posterior

- Hamiltonian sampling (Taylor et al 2010, Jasche & FK 2010)

$$H(\mathbf{s},\mathbf{p}) = K(\mathbf{p}) + E(\mathbf{s}), \tag{50}$$

- kinetic term with a given mass as the variance for the momenta

$$K(\mathbf{p}) = \frac{1}{2}\mathbf{p}^\dagger \mathbf{M}^{-1}\mathbf{p}, \tag{51}$$

- Marginalization over the momenta

$$P(\mathbf{s},\mathbf{p}) = \frac{e^{-H}}{Z_H} = \frac{e^{-K}}{Z_K}\frac{e^{-E}}{Z_E} = P(\mathbf{p})P(\mathbf{s}), \tag{52}$$

- Please note, that the kinteic PDF is a Gaussian
- Marginalization occurs by drawing momenta from a Gaussian and

Sampling the posterior

- Hamiltonian evolution equations: $(\mathbf{s},\mathbf{p}) \rightarrow (\mathbf{s}',\mathbf{p}')$

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = -\frac{\partial H}{\partial \mathbf{s}} = -\frac{\partial E}{\partial \mathbf{s}}, \tag{53}$$

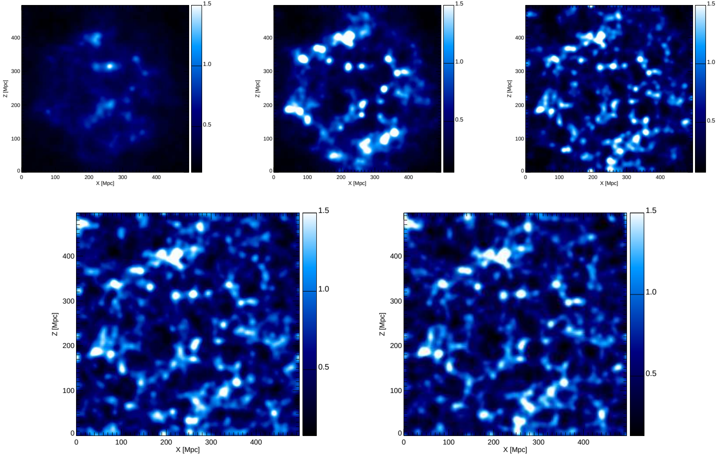
$$\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t} = \frac{\partial H}{\partial \mathbf{p}} = \mathbf{M}^{-1}\mathbf{p}, \tag{54}$$

- Metropolis-Hastings acceptance step

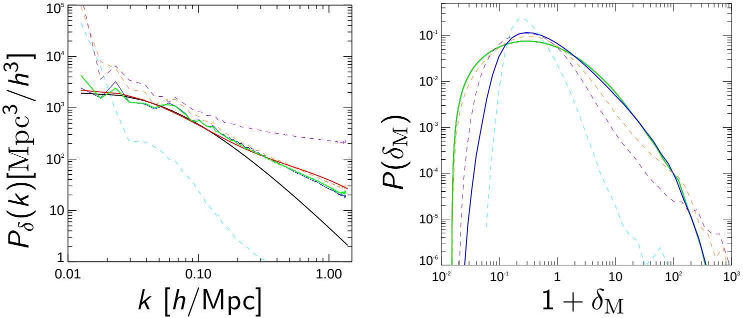
$$p_a = \min(1, e^{-\delta H}), \tag{55}$$

$\delta H = H(\mathbf{s}',\mathbf{p}') - H(\mathbf{s},\mathbf{p}) \rightarrow$ we do not care about the evidence!

Skewed matter statistics: FK tbs



Skewed matter statistics: FK tbs



Conclusions

- There is a need to compare observations with theory as precisely as possible.
- Observations are plagued by many uncertainties which require a statistical treatment.
- The Bayesian approach is flexible and clear.
- We have shown that we can deal with complex models in this framework.