The inadequacy of N-point functions to describe non-linear density fields

Julien Carron

ETH Zurich

September 6, 2012

M.C. Neyrinck, A. Szalay

ApJ, 738:86, 2011, 1105.4467

Phys. Rev. Lett. 108,071301, 2012, 1201.1000

ApJ, 738:86, 2012, with M.C. Neyrinck, 1201.1444

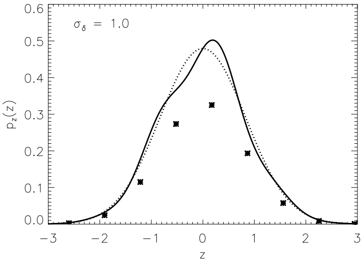
...

Context

- Description of fields through *N*-point moments and statistical inference from these moments :

$$\langle \rho(x_1) \cdots \rho(x_N) \rangle .$$

  How useful are they? How much information do they contain ?
- Non-linear matter density or convergence field $\approx$ lognormal.
- But the lognormal one-point pdf is known to be moment-indeterminate (Coles and Jones 91, Stieltjes 1894!).



Context

- For

$$\rho = (\rho(x_1), \cdots, \rho(x_d)),$$

define

$$p(\rho) = p^{LN}(\rho)\left[1 + \epsilon \sin\left(\pi\boldsymbol{\omega}\cdot\xi^{-1}_{\ln\rho}\left(\ln\rho - \bar{\ln}\rho\right)\right)\right]$$

$\boldsymbol{\omega}$ any vector of integer, and $|\epsilon| < 1$.

- All these measures have identical N-point moments

$$m_{\mathbf{n}} = \langle \rho(x_1)^{n_1} \cdots \rho(x_d)^{n_d} \rangle, \quad n_i = \cdots, -1, 0, 1, \cdots.$$

all all orders.

$$m_{\mathbf{n}} = \exp\left(\frac{1}{2}\mathbf{n}\cdot\xi_{\ln\rho}\mathbf{n} - \frac{1}{2}\sigma^2_{\ln\rho}\cdot\mathbf{n}\right)$$

Context

- *Tailed* distributions :

$$\left\langle e^{c|x|} \right\rangle = \infty \quad \forall c > 0,$$

- For indeterminate measures : the characteristic function is not a moment generating function.

$$\left\langle e^{itx} \right\rangle \neq \sum_n i^n t^n m_n, \quad t \neq 0.$$

No way to express the density in terms of the N-point moments.

Outline and aims :

1. Show and discuss how to understand the independent information content of the hierarchy of N-point moments.
2. Present some exact results for lognormal fields.
   Compare these results to (Mark's) N-body simulations results on the statistical power of the spectrum of (log-)density field, and discuss them in this light.

Fisher information.

$$F_{\alpha\beta} := \left\langle \frac{\partial \ln p}{\partial \alpha} \frac{\partial \ln p}{\partial \beta} \right\rangle$$

To each weight function (density) p and with parameters $\alpha, \beta, \cdots$ is assigned a positive matrix (the covariance matrix of the score functions), in a way such that:

- it is additive for independent variables,
- it can only be reduced with data transformation
- it vanishes if the parameters do not impact p.

$\rightarrow$ Meaningful *absolute* measure of information contained in p on the parameters .

Two matrix inequalities

- *Cramer - Rao bound* ($\sim$ statistical point of view):

$$\Sigma_{ij} \geq \sum_{\alpha\beta} \frac{\partial O_i}{\partial \alpha} \left[F^{-1}\right]_{\alpha\beta} \frac{\partial O_j}{\partial \beta}.$$

  Σ covariance matrix of the unbiased estimators $\hat{\mathbf{O}}$.

- *Information inequality* ($\sim$ information theoretic point of view):

$$F_{\alpha\beta} \geq \sum_{i,j} \frac{\partial O_i}{\partial \alpha} \left[\Sigma^{-1}\right]_{ij} \frac{\partial O_j}{\partial \beta}.$$

The independent information content of the moments.

Moment of order n :

$$m_n := \langle x^n \rangle , \quad \mu := m_1, \quad \sigma^2 := m_2 - m_1^2$$

Gaussian variables have a simple structure :

$$F_{\alpha\beta} = \underbrace{\frac{1}{\sigma^2}\left(\frac{\partial \mu}{\partial \alpha}\right)\left(\frac{\partial \mu}{\partial \beta}\right)}_{[F_1]_{\alpha\beta}} + \underbrace{\frac{1}{2}\left(\frac{\partial \ln \sigma^2}{\partial \alpha}\right)\left(\frac{\partial \ln \sigma^2}{\partial \beta}\right)}_{[F_2]_{\alpha\beta}} .$$

No independent information in higher order statistics, $F_n = 0 \quad n > 2$

Q :

What is the independent information content F_n of m_n for any $p(x)$?

The independent information content of the moments.

Approximate the score function with the orthogonal polynomials $P_n(x)$

$$s_n(\alpha) := \langle \partial_\alpha \ln p \, P_n(x) \rangle \,,$$

where

$$\langle P_n(x) P_m(x) \rangle = \delta_{mn}.$$

(E.g. use Gram-Schmidt on $1, x, x^2, \cdots$).

A :

The independent information content of m_n is $s_n(\alpha)s_n(\beta)$.

$$[F_n]_{\alpha\beta} := s_n(\alpha)s_n(\beta)$$

$$F_{\leq N} := \sum_{n=0}^{N} F_n.$$

Properties of the expansion.

- With
$$s_n(\alpha) = \langle \partial_\alpha \ln p \, P_n(x) \rangle = \sum_{k=0}^{n} C_{nk} \frac{\partial m_k}{\partial \alpha},$$
the series
$$\sum_{n=1}^{N} s_n(\alpha)\, P_n(x) =: s_{\leq N}(x,\alpha) \approx \partial_\alpha \ln p(x,\alpha),$$
gives the best approximation of the score function through polynomials, in the least squares sense.

- For any N we recover the RHS of the information inequality :
$$[F_{\leq N}]_{\alpha\beta} = \sum_{n=1}^{N} s_n(\alpha) s_n(\beta) = \sum_{i,j=1}^{N} \frac{\partial m_i}{\partial \alpha} \left[\Sigma^{-1}\right]_{ij} \frac{\partial m_j}{\partial \beta}$$
$$\Sigma_{ij} = m_{i+j} - m_i m_j, \quad \Sigma^{-1} = C^T C$$

Properties of the expansion.

The mean squared residual

$$\langle (\partial_\alpha \ln p - s_{\leq N}(x,\alpha)) \, (\partial_\beta \ln p - s_{\leq N}(x,\beta)) \rangle = F_{\alpha\beta} - [F_{\leq N}]_{\alpha\beta}$$

are the bits of information absent from the set of moments m_1 to m_N.

If the residual goes to zero as $N \to \infty$, then all of the information is contained within the hierarchy of moments.

Completeness vs incompleteness of the information

- Score function is a polynomial (fields of max. entropy with constrained moments) :

$$F_{\leq N} = F$$

- Moment-determinate measures :

$$\lim_{N \to \infty} F_{\leq N} = F$$

(Polynomials form a complete basis set)

- Moment-indeterminate measures :

$$\lim_{N \to \infty} F_{\leq N} \leq F$$

- Criteria for uniqueness tightly linked to decay rate [Freud, 1971]. E.g.

$$\langle e^{cx} \rangle < \infty \quad \text{for some } c > 0,$$

guarantees determinacy and thus convergence to F . Holds for any number of variables.

An example

- Information on p itself, $\alpha, \beta = p(x), p(y)$: polynomials and Christoffel-Darboux kernel as information :

$$s_n(\alpha) = P_n(x),$$

$$[F_{\leq N}]_{\alpha\beta} = \sum_{n=0}^{N} P_n(x)P_n(y),$$

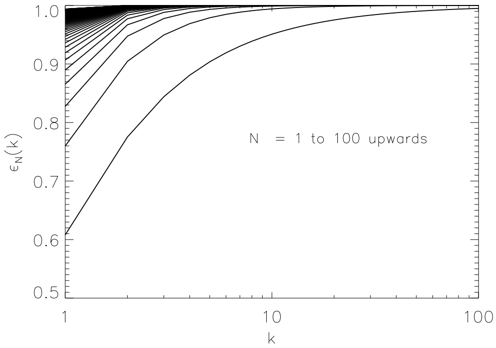
$$F_{\alpha\beta} = \frac{\delta^D(x-y)}{p(x)}.$$

- But $\lim_{N\to\infty} F_{\leq N}$ finite for indeterminate measures. Very consistent.

Another example

- Gamma (generalised χ^2) distribution

$$p(x,k) = \frac{1}{\Gamma(k)} x^{k-1} e^{-x}, \quad x > 0, k > 0$$



$$s_n^2(k) = \frac{1}{n} \frac{\Gamma(k)\Gamma(n)}{\Gamma(k+n)}, \quad \sum_{n=1}^{\infty} s_n^2(k) = \psi_1(k) = F_{kk} \quad \text{Trigamma function}$$

Lognormal family

- x lognormal $\leftrightarrow A := \ln x$ Gaussian. Two free parameters :

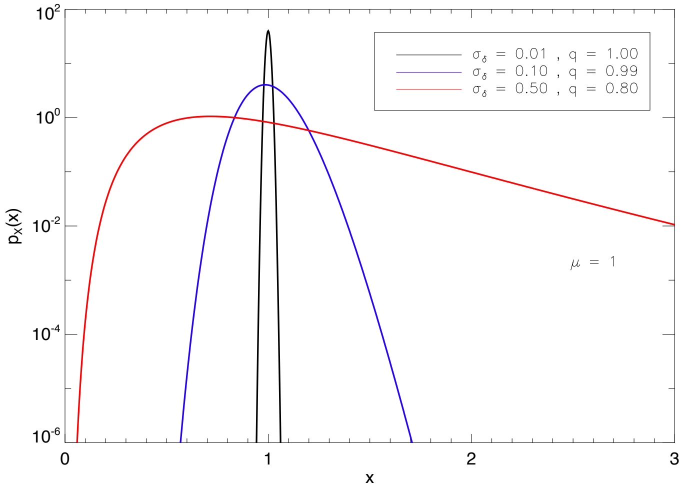
$$(\sigma_A,\ \mu_A)(\alpha) \leftrightarrow (\sigma,\ \mu)(\alpha)$$

  In the following $\mu = 1, A = \ln(1+\delta)$.

- Key quantity $q := \left(1+\sigma_\delta^2\right)^{-1} \in (0,1)$,

$$\begin{cases} q \to 1 & \text{Sharply peaked, } \approx \text{Gaussian} \\ q \to 0 & \text{Heavily tailed.} \\ q = \frac{1}{2} \to \sigma_\delta = 1. \end{cases}$$

What is to be expected ?



Log. score function + huge range = no good.

Exact result in terms of q-series :

$$s_n^2(\sigma_\delta^2) = q^2 \frac{q^n}{1-q^n} (q:q)_{n-1} \, \psi_n^2(q)$$

$$s_n^2(\ln \mu) = \frac{q^n}{1-q^n} (q:q)_{n-1} \, .$$

where

$$(t:q)_n := \prod_{k=0}^{n-1} \left(1 - tq^k\right) \qquad q - \text{Pochammer symbol},$$

$$\psi_n(q) := \sum_{k=1}^{n-1} \frac{q^k}{1-q^k} \qquad \text{Lambert Series}$$

Sketch of derivation :

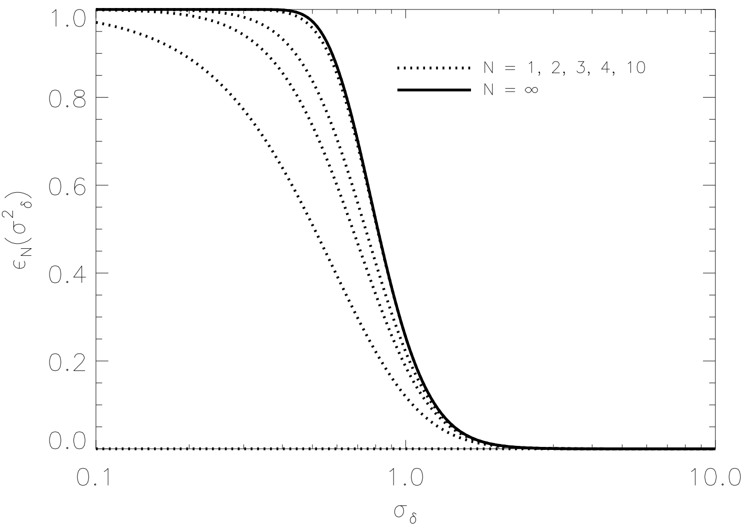
Using $m_{i+j} = m_i m_j q^{-ij}$ show

$$\left\langle P_n(q^i x)\right\rangle = \frac{1}{m_i}\left\langle P_n(x)x^i\right\rangle \to \langle P_n(tx)\rangle = \sum_{k=0}^{n} C_{nk} t^k m_k =\propto (t:q)_n$$

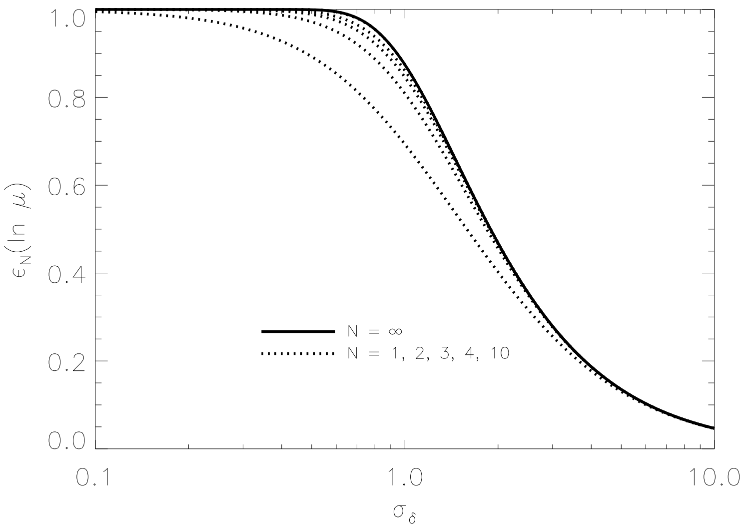
and combine with $m_n = \mu^n q^{-n(n-1)/2}$ and with

$$\begin{aligned} s_n &= \sum_{k=0}^{n} C_{nk}\frac{\partial m_k}{\partial \alpha} \\ &\propto \sum_{k=0}^{n} C_{nk}\, k\,(k-1)\, m_k \quad (\alpha = \sigma_\delta^2) \\ \text{or } &\propto \sum_{k=0}^{n} C_{nk}\, k\, m_k \quad (\alpha = \ln \mu) \end{aligned}$$

Cumulative efficiencies, $\alpha \sim \sigma_\delta^2$



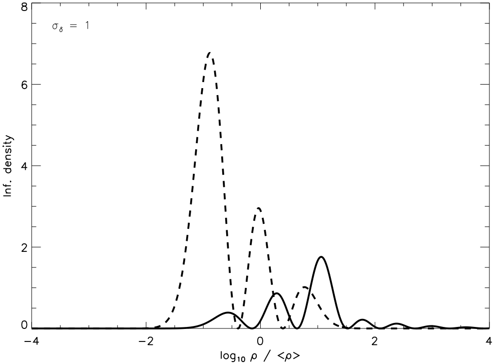
Cumulative efficiencies, $\alpha \sim \ln \mu$



Where is the information gone ?

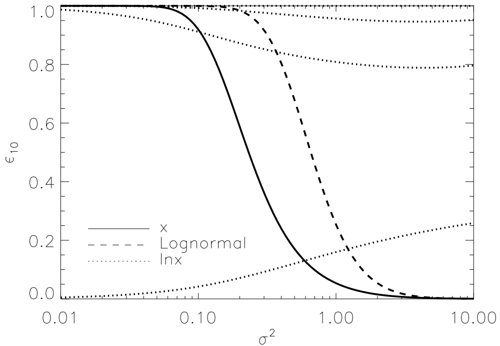
Information density :

$$f_{\alpha\beta} = p\frac{\partial \ln p}{\partial \alpha}\frac{\partial \ln p}{\partial \beta} = \frac{1}{p}\frac{\partial p}{\partial \alpha}\frac{\partial p}{\partial \beta}$$



Most of the information is the underdense regions, inaccessible to the moments dominated by the peaks. $\rightarrow$ combination of two effects.

Generic behavior, seen in simulations



$$x = 1 + \frac{\kappa}{\kappa_0}, \quad \kappa \quad \text{convergence field}$$

$$p(x, \sigma) = \frac{Z}{x} \exp\left[-\frac{1}{2\omega^2}\left(\ln x + \frac{\omega^2}{2}\right)^2 \left(1 + \frac{A}{x}\right)\right].$$

Pdf fit to simulations [Das & Ostriker 2006].

Several variables and fields

- Decomposition conceptually identical for any number d of variables : approximate $\partial_\alpha \ln p$ through orthogonal polynomials in d variables, and collect the terms of order N. There are $\binom{N+d-1}{N}$ independent polynomials of order N.

$$P_{\mathbf{n}}(x), \quad \mathbf{n} = (n_1, \cdots, n_d) \leftrightarrow n \text{ is the only change.}$$

- Hard !
- For independent variables, information adds up order by order (score functions adds up).

Multivariate lognormal

- For uncorrelated fiducial, but parameter creating correlations :

$$P_{\mathbf{n}}(x) = \prod_{i=1}^{d} P_{n_i}(x_i), \quad \langle P_{\mathbf{n}}(x) P_{\mathbf{m}}(x) \rangle = \delta_{\mathbf{nm}}$$

- Score function couples variables in pairs :

$$\ln p = -\frac{1}{2} \sum_{i,j} (A - \bar{A})_i \xi^{-1}_{A,ij} (A - \bar{A})_j + \mathrm{cst}$$

- $\rightarrow s_{\mathbf{n}}(\alpha) = 0$ if $\mathbf{n}$ has more than 2 nonzero indices ($since$ $\langle P_n(x) \rangle = 0$ for $n \neq 0$).

Multivariate lognormal

- Exact result for uncorrelated fiducial

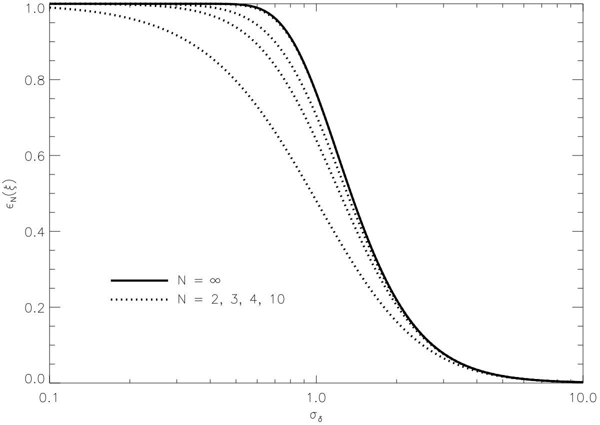
$$F_{\leq N} = \underbrace{F^{\sigma}\epsilon_N(\sigma^2)}_{\text{same as uncorrelated}} + \underbrace{F^{\xi}\epsilon_N(\xi)}_{\text{info. from correlations}}$$

with

$$\epsilon_N(\xi) := \sigma_A^4 \sum_{n=2}^{N} \sum_{i=1}^{n-1} s_i^2(\ln\mu) s_{n-i}^2(\ln\mu),$$

$$F^{\sigma}_{\alpha\beta} = d\frac{\partial\sigma_A^2}{\partial\alpha}\frac{\partial\sigma_A^2}{\partial\beta}\left(\frac{1}{4\sigma_A^2} + \frac{1}{2\sigma_A^4}\right)$$

$$F^{\xi}_{\alpha\beta} = \frac{1}{2\sigma_A^4}\sum_{i\neq j}\frac{\partial\xi_{A,ij}}{\partial\alpha}\frac{\partial\xi_{A,ij}}{\partial\beta}$$



$$\epsilon_N(\xi) \to \frac{\ln^2\left(1+\sigma_\delta^2\right)}{\sigma_\delta^4} = \left(\frac{\sigma_A}{\sigma_\delta}\right)^4 = \epsilon_2(\xi).$$

Comparing to N-body simulations

- The density field in N-body simulation is correlated.
  ? Is it possible at all to compare ?

Some hand-waving arguments :

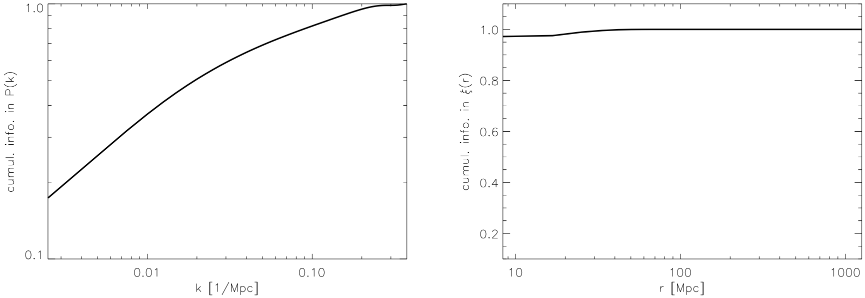
- For $\ln \sigma_8^2$ alike parameter, $\partial_\alpha \ln P \approx \mathrm{cst}$, the correlation structure is completely irrelevant :

$$F_{\alpha\beta}^{P_A} = \frac{V}{2} \int \frac{d^3k}{(2\pi)^3} \frac{\partial \ln P_A}{\partial \alpha} \frac{\partial \ln P_A}{\partial \beta},$$

- First term in ξ/σ^2 expansion of the covariance matrix between the N-point functions. Correct derivatives, approximate covariance matrices.
- Constancy of the improvement factor suggests a common mechanism. Uncorrelated model is a simple parameter independent model.

Numerical arguments

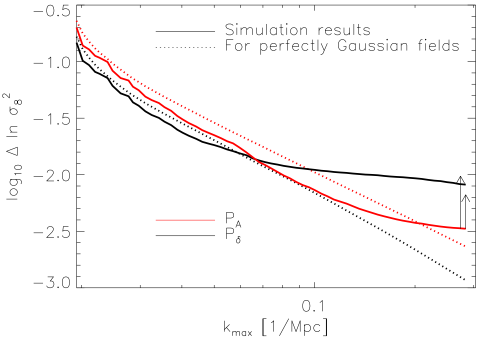
E.g. exact distribution of the information within the second order statistics of the 1D ΛCDM lognormal field, $\sigma_\delta^2 = 1$, on the amplitude of P_A :



95% of the information within the variance. No independent information from the correlations. Uncorrelated model gives $\epsilon_2 \simeq 0.12$ accurate to some 15%.

It works

Comparing improvement ratios of statistical power of P_A to that of P_δ.
$k_{\max} \sim 0.3/\mathrm{Mpc}$, $V = 2.2 Gpc^3$.



- Found in the simulations : $\ln \sigma_8^2$: 2.5, n_s: 2.4
- Uncorrelated model predictions : $2.0 - 2.9$ (range reflects some ambiguity in the value of the variance $\sigma_A^2 = \ln(1 + \sigma_\delta^2)$

Uncorrelated fiducial model

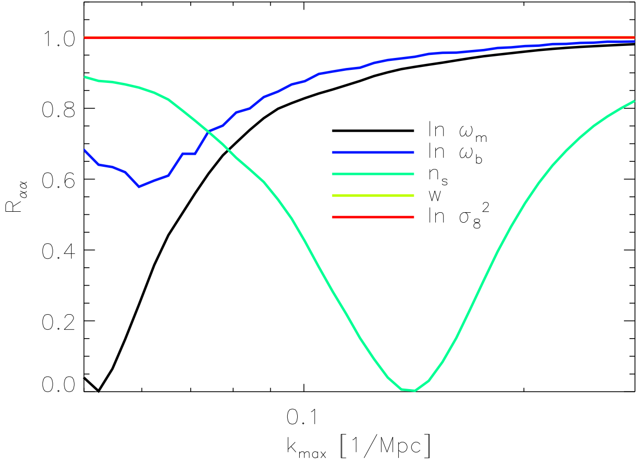
Introduce parameter dependence through the model with uncorrelated fiducial. In Fourier space :

$$\frac{[F_{\le N}]_{\alpha\beta}}{F^{P_A}_{\alpha\beta}} = \underbrace{\epsilon_N(\xi)\left[1 - R_{\alpha\beta}\right]}_{\text{from correlations}} + \underbrace{\epsilon_N(\sigma^2_\delta)\left(1 + \frac{1}{2}\sigma^2_A\right) R_{\alpha\beta}}_{\text{from the variance, uncorrelated model}}$$

where

$$R_{\alpha\beta} = \frac{\int dk\, k^2 \partial_\alpha \ln P_A \int dk\, k^2 \partial_\beta \ln P_A}{\left(\int dk\, k^2\right) \int dk\, k^2 \partial_\alpha \ln P_A \partial_\beta \ln P_A}$$

calibrates the part of the information coming from the correlations at nonzero lag to that from the variance.



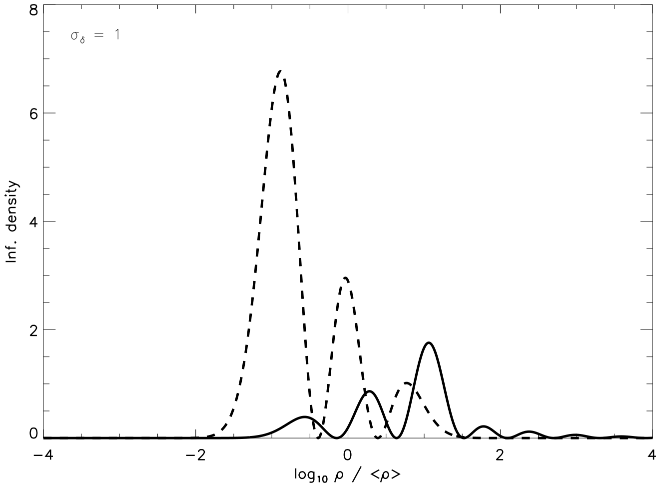
→ No real change to the generic factor of improvement, as found in the simulations.

Conclusions

- Non intuitive statistics at work in fields with high tails. Polynomials form a poor basis set of functions for such densities.
- There is a lot of information in voids that are uneasy to catch with moments. The very deeply non-linear lognormal field is a very non Gaussian field with useless N-point function hierarchy.
- Success of the lognormal model to reproduce N-body simulation results at the level of the spectrum, showing a more powerful statistical inference of standard cosmological parameters from the (noise-free) log-density field.
- While the matter power spectrum becomes a poor descriptor of the field, the one-point pdf is very important.

Freud, G. (1971).
Orthogonal Polynomials.
Pergamon Press Ltd., Headington Hill Hall, Oxford.

Information density



Comparison to SPT

Since

$$\sum_{n=1}^{N} s_n^2 = \sum_{i,j=1}^{N} \frac{\partial m_i}{\partial \alpha} \left[\Sigma^{-1}\right]_{ij} \frac{\partial m_j}{\partial \alpha}$$

$$\Sigma_{ij} := m_{i+j} - m_i m_j,$$

we can obtain s_n from $m_1 - m_{2n}$ (knowledge of the full shape of p_X not required).

$\rightarrow$ We can compare the lognormal predictions on the information content of the first few moments of the density fluctuation field δ with perturbation theory [Bernardeau 2004].

Comparison to SPT

SPT :

$$\langle \delta^n \rangle = m_n = m_n^{Gauss} + \sigma_\delta^{2(n-1)} S_n$$

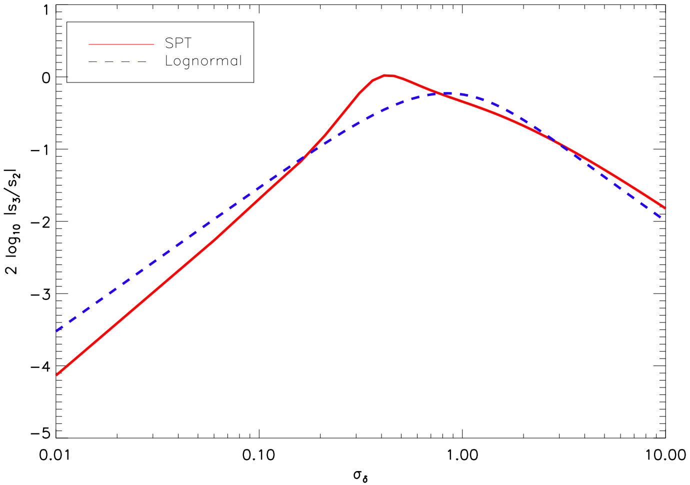
with variance

$$\sigma_\delta^2(R) = \frac{1}{2\pi^2} \int_0^\infty dk\, k^2 P(k, \alpha) |W(kR)|^2 .$$

S_n very weakly dependent on cosmology $\rightarrow$ model parameter independent comparison to lognormal predictions possible.

With S_n up to S_6 as given in [Bernardeau 2004], we evaluated s_3/s_2 (vanishes for a Gaussian distribution), for ΛCDM universe.

Comparison to SPT



Impact of noise