

# Back to normal

## Box-Cox transformations in cosmology

Benjamin Joachimi

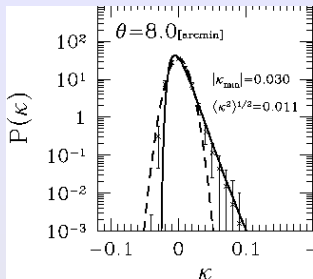
bj@roe.ac.uk



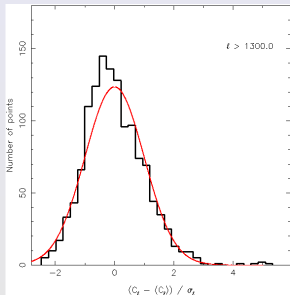
Institute for Astronomy  
University of Edinburgh

Astronomical Data Analysis VII, Cargese, Corsica  
May 17, 2012

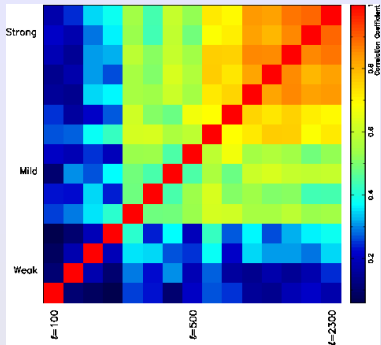
# Non-Gaussianity abounds in weak lensing data



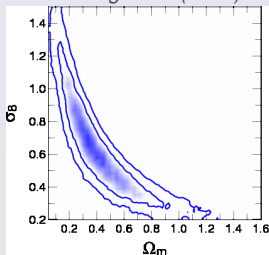
Taruya et al. (2002)



Kiessling et al. (2011)



Kiessling et al. (2011)



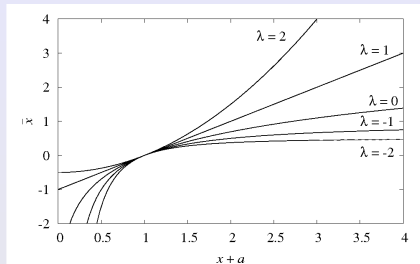
Schrabback et al. (2010)

- 1 Box-Cox transformations
- 2 Gaussianising the weak lensing convergence
- 3 Transformed power spectra & constraints on cosmology
- 4 Gaussianising posterior distributions
- 5 Normal parameters for arbitrary inference problems
- 6 Conclusions

*Together with:* Andy Taylor (IfA), Alina Kiessling (Caltech)

# The Box-Cox transformation

- proposed by *Box & Cox (1964)*
- in astronomy: only applied in CMB analysis by *Dineen & Coles (2005)*



*Goal:*

- transform data set  $\mathbf{x} \rightarrow \bar{\mathbf{x}}$  such that  $\bar{\mathbf{x}}$  is Gaussian distributed

For every element  $x_i$  in the data set

$$\bar{x}_i(\lambda, a) = \begin{cases} [(x_i + a)^\lambda - 1] / \lambda & \lambda \neq 0 \\ \ln(x_i + a) & \lambda = 0 \end{cases}$$

Box-Cox parameters:  $\lambda, a$

# Determining the Box-Cox parameters

Maximum likelihood (ML) approach:

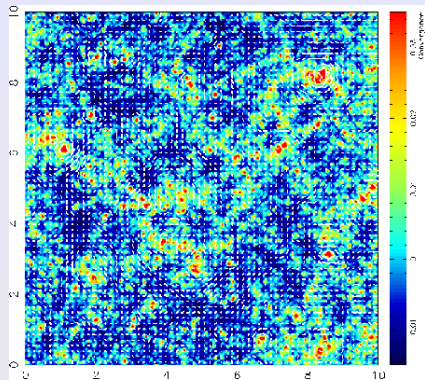
$$L(\mathbf{x}) = \bar{L}(\bar{\mathbf{x}}) \frac{\partial \bar{\mathbf{x}}}{\partial \mathbf{x}} = \bar{L}(\bar{\mathbf{x}}) \prod_{i=1}^n [x_i + a]^{\lambda-1}$$

- transformation should ensure that  $\bar{L}(\bar{\mathbf{x}})$  is a Gaussian
- Free parameters in  $L(\mathbf{x})$ :  $\lambda$ ,  $a$ , mean and covariance of Gaussian  $\bar{L}(\bar{\mathbf{x}})$
- compare this model for  $L(\mathbf{x})$  to the data:  
insert ML estimates of mean and covariance into expression for  $L(\mathbf{x})$   
→ Marginal log-likelihood

$$\mathcal{L}(\lambda, a) = -\frac{n}{2} \ln \det \text{Cov} \left[ \bar{\hat{\mathbf{x}}}(\lambda, a) \right]_{\text{ML}} + (\lambda - 1) \sum_{i=1}^n \ln(\hat{x}_i + a)$$

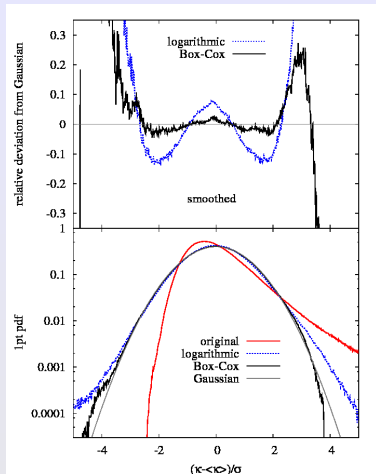
- maximise with respect to  $\lambda$ ,  $a$  → optimal Box-Cox parameters
- straightforward generalisation to higher-dimensional data  
(pair of  $\lambda$ ,  $a$  for each dimension)
- Bayesian approach yields very similar result (*Box & Cox 1964*)

# The convergence distribution



shear/convergence field

*Kiessling et al. (2011)*



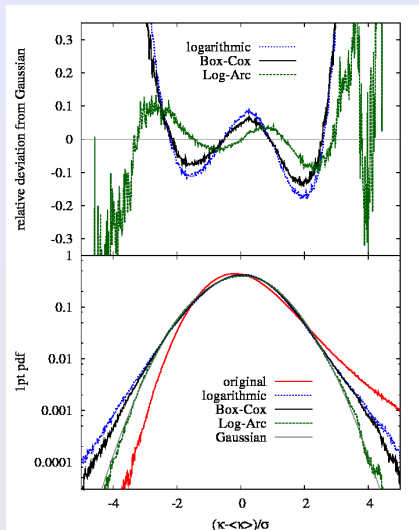
1point probability distribution for  $\kappa$

*Joachimi et al. (2011)*

→ study only the 1point distribution – simplification!

→ log-transform performs fairly well, but not optimal ( $\lambda = -2.2$ ;  $a = 0.08$ )

# Gaussianisation in presence of shape noise



1point probability distribution for  $\kappa$   
 now with realistic level of Gaussian  
 shape noise added

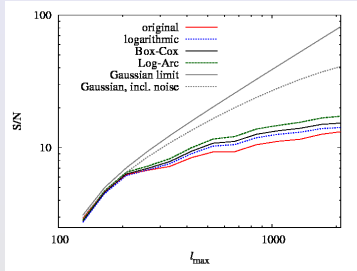
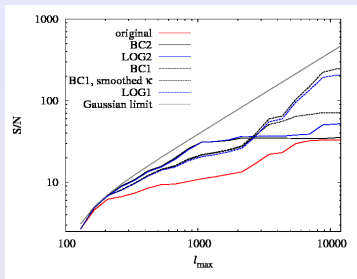
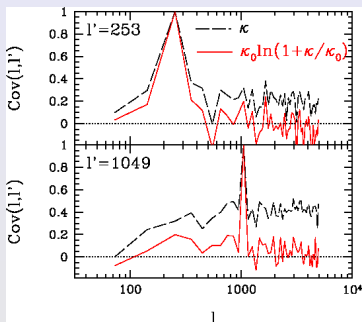
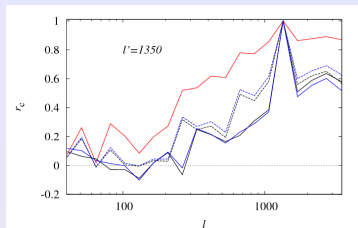
Log-Arc transformation:

$$\bar{x}(s, a) = \begin{cases} \arctan [s \ln(x + a)] / s & s \neq 0 \\ \ln(x + a) & s = 0 \end{cases}$$

→ Box-Cox formalism is versatile!

type	skew.	kurt.	$D_{KL} * 100$
original	0.83	3.11	3.8
log	-0.05	0.73	0.7
Box-Cox	-0.02	0.43	0.4
Log-Arc	-0.05	-0.02	0.2

# Covariances & signal-to-noise



Top:  $\sigma_8 = 0.81$ ; *Joachimi et al. (2011)*

Bottom:  $\sigma_8 = 0.76$ ; *Seo et al. (2011a)*

Top: S/N without shape noise

Bottom: S/N with shape noise



Taylor-expand:

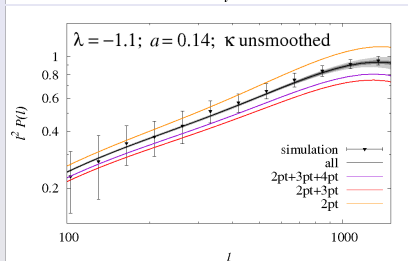
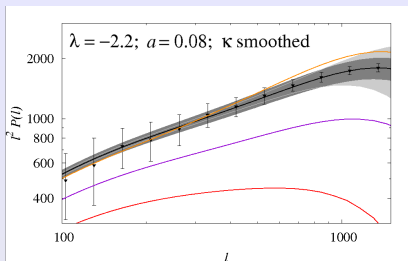
$$(\kappa + a)^\lambda = a^\lambda + \lambda a^{\lambda-1} \kappa + \frac{\lambda(\lambda-1)}{2} a^{\lambda-2} \kappa^2 + \dots$$

Express transformed power spectrum  $P_{\bar{\kappa}}$  in terms of  $\kappa$  statistics:

$$\begin{aligned} P_{\bar{\kappa}}(\ell) &= a^{2\lambda-2} \left\{ P_\kappa(\ell) + (\lambda-1) a^{-1} \int \frac{d^2 \ell_1}{(2\pi)^2} B_\kappa(\ell, \ell_1, |\ell - \ell_1|) \right. \\ &\quad \left. + (\lambda-1)(\lambda-2) a^{-2} P_\kappa(\ell) \int \frac{d^2 \ell_1}{(2\pi)^2} P_\kappa(\ell_1) + \dots \right\} \end{aligned}$$

- $\ell \rightarrow \infty$  in integrals  $\rightarrow$  need to smooth  $\kappa$  map (use Gaussian kernel)
- simulation discreteness noise and shape noise contribute also to low  $\ell$  in transformed power spectrum  $\rightarrow$  noise needs to be modelled

# Comparison to simulations



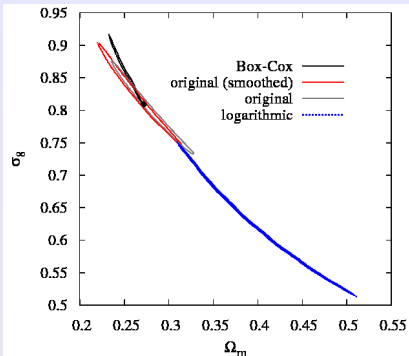
Transformed  $\kappa$  power spectra  
*light grey*: bispectrum uncertainty  
*dark grey*: higher-order uncertainty

- power spectra with non-linear corrections (*Smith et al. 2003*) – accurate to  $\ell \sim 1500$
- bispectra from perturbation theory with non-linear corr. (*Scoccimarro & Couchman 2001*)
- (initially) neglect trispectrum contributions to 4th-order term
- smoothing with kernel width 1.5 arcmin
- analytic computation of simulation and shape noise

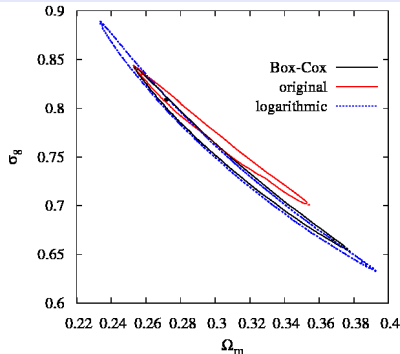
That's not enough!

→ include up to 6th order in  $\kappa$ , use lognormal model and rescale each  $\kappa$  order with simulation-based moment (factors of a few larger)

# Cosmological parameter constraints



noise-free



with shape noise

Disappointing performance because

- only integral over certain triangle, quadrangle, etc. configurations contribute to higher-order statistics
- only sum of different orders enters data vector
- shape noise makes distribution more Gaussian  $\rightarrow$  less to gain

Confirmed by recent work using numerical simulations (*Seo et al. 2011b*)

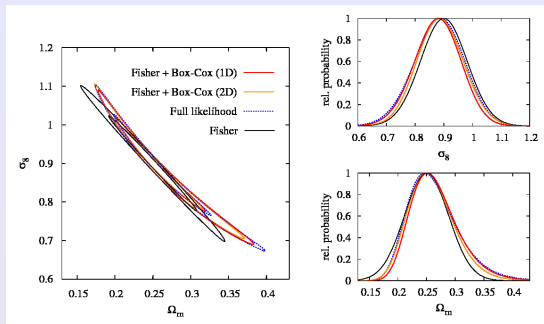
Fisher information matrix

$$F_{\mu\nu} = - \left\langle \frac{\partial^2 \ln L}{\partial p_\mu \partial p_\nu} \right\rangle = \sum_{\ell} \frac{\partial P_\kappa(\ell)}{\partial p_\mu} \text{Cov}^{-1} \{P_\kappa(\ell)\} \frac{\partial P_\kappa(\ell)}{\partial p_\nu}$$

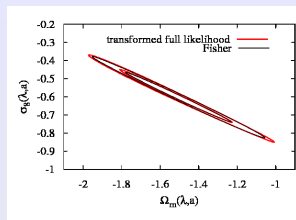
→ can only fully represent Gaussian posteriors ↔ elliptical contours

How to get more accurate posterior forecasts:

- 1 determine Box-Cox parameters for every posterior dimension e.g. via small MCMC sample
- 2 compute Fisher matrix  $\bar{F}$  and posterior maximum  $\bar{\mathbf{p}}_{\max}$  for transformed cosmological parameters (approximation in terms of original  $F$ )
- 3 transformed Gaussian posterior given by  $\bar{\mathcal{P}} \propto \exp \left\{ -\frac{1}{2} (\bar{\mathbf{p}} - \bar{\mathbf{p}}_{\max})^\tau \bar{F} (\bar{\mathbf{p}} - \bar{\mathbf{p}}_{\max}) \right\}$
- 4 inverse Box-Cox transformation yields original posterior  $\mathcal{P}$



2D and 1D posteriors for  $\Omega_m$  and  $\sigma_8$

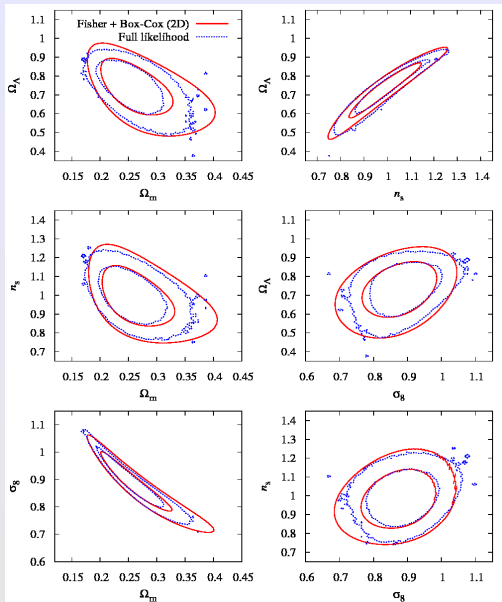


Transformed posterior and transformed Fisher matrix

*Joachimi & Taylor (2011)*

parameters changed	$D_{\text{KL}}(\Omega_m)$	$D_{\text{KL}}(\sigma_8)$	$D_{\text{KL}}(\Omega_m, \sigma_8)$
Fisher matrix	0.143	0.029	3.482
Box-Cox, fid. parameters	0.013	0.013	0.017
$\Omega_m : 0.25 \rightarrow 0.225$	0.008	0.008	0.025
$z_{\text{med}} : 0.9 \rightarrow 1.0; n_g : 20 \rightarrow 37.4'^{-2}$	0.005	0.005	0.019
$\ell_{\text{max}} : 10000 \rightarrow 8700$	0.015	0.015	0.019
$A_s : 100 \rightarrow 110 \text{ deg}^2$	0.013	0.012	0.016

# A higher-dimensional example



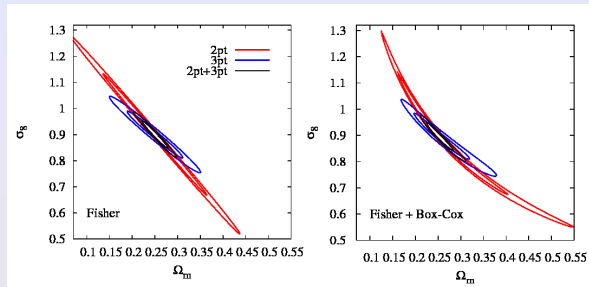
4D cosmic shear posterior  
 $\{\Omega_m, \sigma_8, n_s, \Omega_\Lambda\}$

Which set of contours is more trustworthy?

# Application: Bispectrum constraints



How much do 3pt statistics add to parameter constraints?



forecast of  $1\sigma$  and  $2\sigma$  constraints from cosmic shear power spectrum and bispectrum  
*Joachimi & Taylor (2011)*

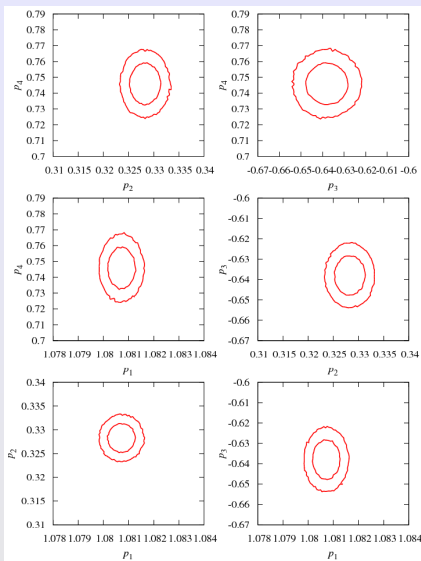
statistics	Fisher	B-C-F
2pt	$0.90^{+0.23}_{-0.23}$	$0.85^{+0.31}_{-0.27}$
3pt	$0.90^{+0.09}_{-0.09}$	$0.89^{+0.11}_{-0.12}$
2pt + 3pt	$0.90^{+0.05}_{-0.05}$	$0.90^{+0.06}_{-0.08}$

→ the stronger the constraints the better is the standard Fisher matrix

*But:* larger survey → more parameters with new degeneracies

Marginal  $2\sigma$  constraints on  $\sigma_8$

# Independent parameters for weak lensing



*Joachimi & Taylor, in prep.*

1 Box-Cox transformation of posterior  $\rightarrow$  multivariate Gaussian posterior

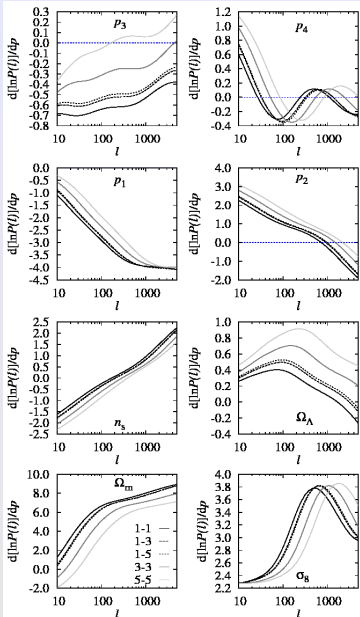
2 decorrelation via principal component analysis  $\rightarrow$  statistically independent Gaussian parameters



- speeds up MCMC sampling
- simplifies marginalisation
- highlights where parameter constraints come from



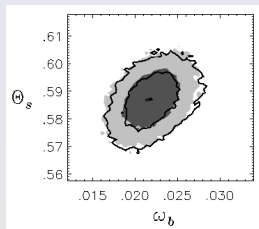
# Interpretation of new parameter set



Relative power spectrum derivatives for parameter set  $\{\Omega_m, \sigma_8, n_s, \Omega_\Lambda\}$  and corresponding normal parameters  $\{p_1, p_2, p_3, p_4\}$

*Joachimi & Taylor, in prep.*

Can one define physically motivated weak lensing normal parameters in analogy to CMB normal parameters (*Kosowsky et al. 2002, Sandvik et al. 2004*)?



Contours: MCMC  
Shaded area:  
Gaussian approx.  
*Chu et al. (2003)*

*Work in progress ...*

- Gaussianisation of the 1pt weak lensing convergence does NOT improve constraints on cosmology in cases with realistic noise
- Box-Cox transformations can be combined with Fisher matrices to produce accurate forecasts of posterior distributions
- Gaussianisation of posteriors may yield a recipe to define normal (statistically independent) parameters for arbitrary inference problems

→ Box-Cox transformations are a simple, effective, and versatile tool to Gaussianise uni- and multivariate distributions

## To-do list:

- Go beyond 1pt convergence Gaussianisation; treat convergence and shape noise distributions separately
- Investigate alternative ways to efficiently exploit cosmological information in higher-order statistics of large-scale structure
- Find physical weak lensing normal parameters