DIP: Diagnostics for Insufficiencies of Posterior calculations - a CMB application

Sebastian Dorn

N. Oppermann, and T. A. Enßlin

2013-12-13

Max-Planck-Institut für Astrophysik

Outline

DIP TEST
– Validation Concept –

WHY is posterior validation needed?

WHY is posterior validation needed?

$\Rightarrow$ Increasing level of complexity in Bayesian posterior calcuations prone to errors:

- Mistakes in the numerical implementation/ insufficient numerical precision

WHY is posterior validation needed?

⇒ Increasing level of complexity in Bayesian posterior calcuations prone to errors:

- Mistakes in the numerical implementation/ insufficient numerical precision
- Analytic approximations (denoted by $\sim$) in the posterior derivation might influence the posterior:

$$P(s|d) = \frac{P(d|s)P(s)}{P(d)} \approx \tilde{P}(s|d)$$

DIP – procedure:

[1] for proof see S. Dorn *et al.*, Phys.Rev.E.88.053303

DIP – procedure:

1. Sample values of s_{gen} from the prior $P(s)$.

[1] for proof see S. Dorn *et al.*, Phys.Rev.E.88.053303

DIP – procedure:

1. Sample values of s_{gen} from the prior $P(s)$.
2. Generate data d for every s_{gen} according to $P(d|s_{\text{gen}})$.

[1] for proof see S. Dorn *et al.*, Phys.Rev.E.88.053303

DIP – procedure:

1. Sample values of s_{gen} from the prior $P(s)$.
2. Generate data d for every s_{gen} according to $P(d|s_{\text{gen}})$.
3. Calculate a posterior curve for given data by determining $\tilde{P}(s|d)$.

[1] for proof see S. Dorn *et al.*, Phys.Rev.E.88.053303

DIP – procedure:

1. Sample values of s_{gen} from the prior $P(s)$.
2. Generate data d for every s_{gen} according to $P(d|s_{\text{gen}})$.
3. Calculate a posterior curve for given data by determining $\tilde{P}(s|d)$.
4. Calculate the posterior probability for $s \leq s_{\text{gen}}$ according to

$$x := \int_{-\infty}^{s_{\text{gen}}} \mathrm{d}s \; \tilde{P}(s|d) \;\in [0, 1] \,. \tag{1}$$

[1] for proof see S. Dorn *et al.*, Phys.Rev.E.88.053303

DIP – procedure:

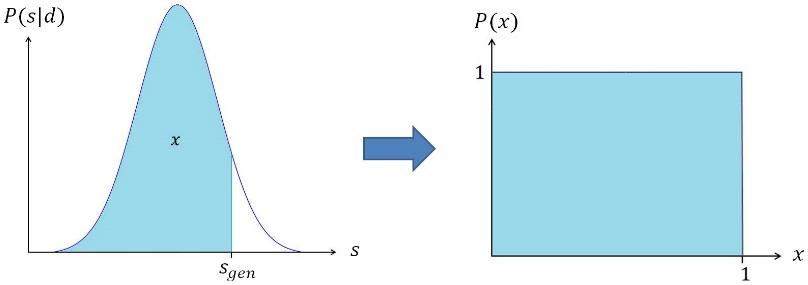
1. Sample values of s_{gen} from the prior $P(s)$.
2. Generate data d for every s_{gen} according to $P(d|s_{\mathrm{gen}})$.
3. Calculate a posterior curve for given data by determining $\tilde{P}(s|d)$.
4. Calculate the posterior probability for $s \leq s_{\mathrm{gen}}$ according to

$$x := \int_{-\infty}^{s_{\mathrm{gen}}} \mathrm{d}s\ \tilde{P}(s|d) \ \in [0,1]\,. \tag{1}$$

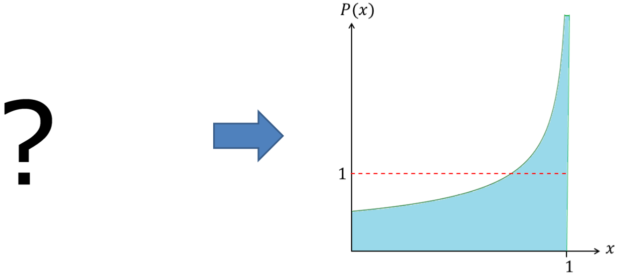
5. If the calculation of the posterior was correct, the distribution for x, $P(x)$, should be uniform[1] between 0 and 1.

[1] for proof see S. Dorn *et al.*, Phys.Rev.E.88.053303

DIP test: correct posterior (1D)



DIP test: incorrect posterior (1D)



$\Rightarrow$ What informations are encoded in the dip of the histogram?

DIP TEST

– Error diagnostics –

Henceforth we consider:

- $s \in \mathbb{R}$
- Gaussian posteriors (similar effects for other pdf's),

$$P(s|d) = \mathcal{G}(s_d, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{s_d^2}{2\sigma^2}\right),$$

  with $s_d = s - \bar{s}_d$ and $\bar{s}_d$ the data dependent maximum of the posterior and
- a wrongly determined value

$$x^\epsilon = \int_{-\infty}^{s_{\mathrm{gen}}} \mathrm{d}s \; P^\epsilon(s|d),$$

  where $P^\epsilon(s|d)$ is the distorted Gaussian.

Wrong variance

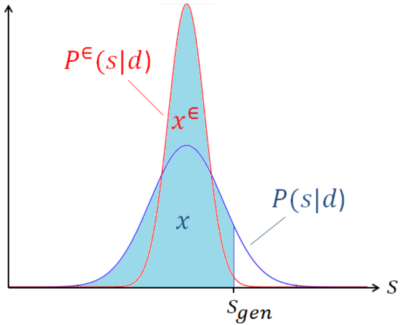
$$P^{\epsilon}(s|d) = \frac{1}{\sqrt{2\pi}\sigma(1+\epsilon)} \exp\left(-\frac{s_d^2}{2\sigma^2(1+\epsilon)^2}\right)$$

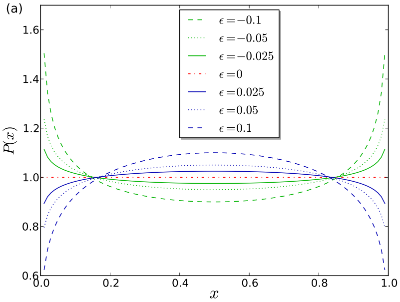
with $\epsilon > -1$. For $P(x)$ we obtain[2]

$$P(x) = (1+\epsilon)\exp\left(-\left[\mathrm{erf}^{-1}(2x-1)\right]^2\left[(1+\epsilon)^2 - 1\right]\right)$$

[2] see S. Dorn *et al.*, Phys.Rev.E.88.053303

Wrong variance





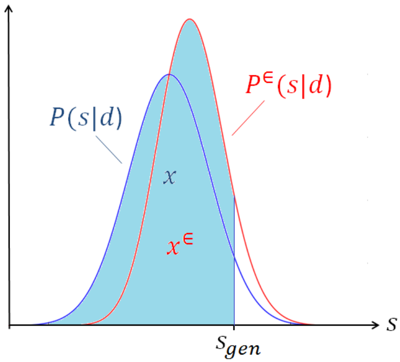
Wrong skewness

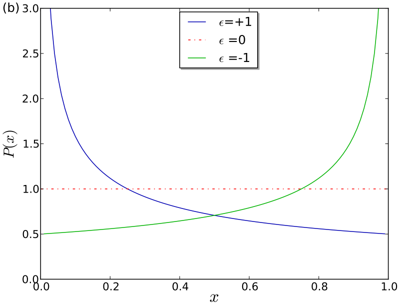
$$P^{\epsilon}(s|d) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{s_d^2}{2\sigma^2}\right)\left(1 + \mathrm{erf}\left(\frac{\epsilon s_d}{\sqrt{2}\sigma}\right)\right).$$

For $P(x)$ we obtain

$$P(x) = \begin{cases} 1/\left(2\sqrt{x}\right) & \text{if } \epsilon = 1 \\ 1/\left(2\sqrt{1-x}\right) & \text{if } \epsilon = -1 \end{cases}.$$

Wrong skewness





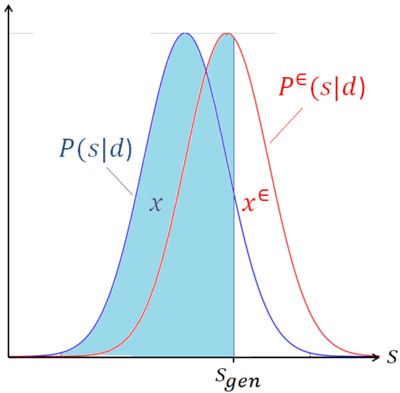
Wrong maximum position

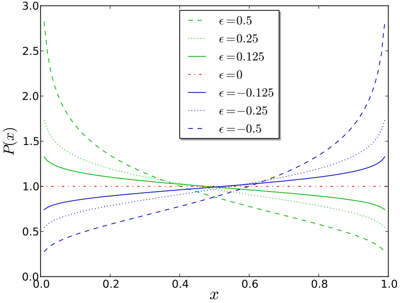
$$P^{\epsilon}(s|d) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(s_d - \epsilon)^2}{2\sigma^2}\right)$$

For $P(x)$ we obtain

$$P(x) = \exp\left(-\frac{1}{2}\left(\frac{\epsilon}{\sigma}\right)^2 - \sqrt{2}\left(\frac{\epsilon}{\sigma}\right)\ \mathrm{erf}^{-1}\left(2x - 1\right)\right)$$

Wrong maximum position





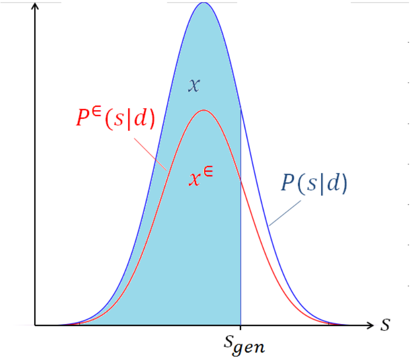
Wrong normalization

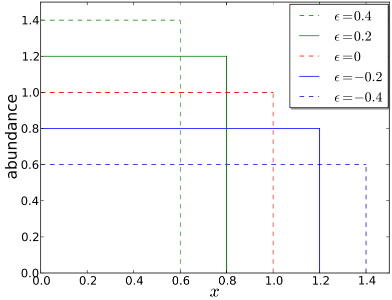
$$P^{\epsilon}(s|d) = \frac{1}{\sqrt{2\pi}\sigma(1+\epsilon)} \exp\left(-\frac{s_d^2}{2\sigma^2}\right)$$

For $P(x)$ we obtain

$$P(x) = 1 + \epsilon \qquad \text{for } x \in [0, 1-\epsilon].$$

Wrong normalization





What have we gained?

$\Rightarrow$ Connection between graphical effects & error-types of posteriors distortions

Graphical effect	Error-type
Flat distribution	–
"∪-(∩-)shape"	variance under-(over-)estimated
$x = 0$ $(x = 1)$ enhanced, purely left curved	too neg. (pos.) skewed
$x = 0$ $(x = 1)$ enhanced, left & right curved	too large (low) max. postition
x-interval smaller (greater) than one	too large (low) normalization

$\Rightarrow$ Quantitative errors on posterior pdf become estimateable $\rightarrow$ fitting formulae $P(x)$.

DIP TEST
– Example in 2D –

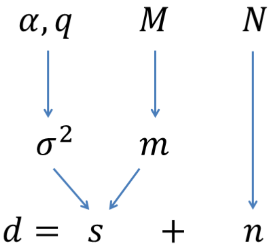
Remark: DIP test in higher dimensionen

- Considering $P(t|d),\ t \in \mathbb{R}^m,\ m \in \mathbb{N}$
- Histogram generation requires 1D posterior
  $\Rightarrow$ map $P(t|d)$ onto 1D, by

$$P(s|d) = \int \mathcal{D}t\ P(s|t,d)P(t,d).$$

- Infinitely many ways to perform this mapping
  $\Rightarrow$ a suite of tests are needed to probe $P(t|d)$.

Hierarchical model in 2D: Data model



- white noise: $n \hookleftarrow \mathcal{G}(n|N)$; mean: $m \hookleftarrow \mathcal{G}(m|M)$
- variance:

$$\mathcal{I}\left(\sigma^2, \alpha, q\right) := \frac{q^\alpha}{\Gamma(\alpha)} \sigma^{2-\alpha-1} \exp\left(-\frac{q}{\sigma^2}\right)$$

Hierarchical model in 2D: Inference

1. The posterior calculation for m, σ^2 yields

$$P\left(m, \sigma^2 | d\right) = \frac{\mathcal{G}(m, M)\ \mathcal{I}\left(\sigma^2, \alpha, q\right)\ \mathcal{G}\left(d - m, \sigma^2 + N\right)}{\int_0^\infty \mathrm{d}\sigma^2\ \mathcal{I}\left(\sigma^2, \alpha, q\right)\ \mathcal{G}\left(d, \sigma^2 + M + N\right)}.$$

Hierarchical model in 2D: Inference

1. The posterior calculation for m, σ^2 yields

$$P\left(m, \sigma^2 | d\right) = \frac{\mathcal{G}(m, M)\ \mathcal{I}\left(\sigma^2, \alpha, q\right)\ \mathcal{G}\left(d - m, \sigma^2 + N\right)}{\int_0^\infty \mathrm{d}\sigma^2\ \mathcal{I}\left(\sigma^2, \alpha, q\right)\ \mathcal{G}\left(d, \sigma^2 + M + N\right)}.$$

2. Artificial inclusion of an error by setting $\alpha \to \alpha(1 + \epsilon)$

Hierarchical model in 2D: Inference

1. The posterior calculation for m, σ^2 yields

$$P\left(m, \sigma^2 | d\right) = \frac{\mathcal{G}(m, M)\ \mathcal{I}\left(\sigma^2, \alpha, q\right)\ \mathcal{G}\left(d - m, \sigma^2 + N\right)}{\int_0^\infty \mathrm{d}\sigma^2\ \mathcal{I}\left(\sigma^2, \alpha, q\right)\ \mathcal{G}\left(d, \sigma^2 + M + N\right)}.$$

2. Artificial inclusion of an error by setting $\alpha \to \alpha(1 + \epsilon)$
3. Mappings onto 1D:

$$P\left(\sigma^2 | d\right) = \int \mathcal{D}m\ P(m, \sigma^2 | d)$$

$$P\left(m | d\right) = \int_0^\infty \mathrm{d}\sigma^2\ P(m, \sigma^2 | d)$$

Hierarchical model in 2D: Inference

1. The posterior calculation for m, σ^2 yields

$$P\left(m, \sigma^2|d\right) = \frac{\mathcal{G}(m, M)\ \mathcal{I}\left(\sigma^2, \alpha, q\right)\ \mathcal{G}\left(d - m, \sigma^2 + N\right)}{\int_0^\infty \mathrm{d}\sigma^2\ \mathcal{I}\left(\sigma^2, \alpha, q\right)\ \mathcal{G}\left(d, \sigma^2 + M + N\right)}.$$

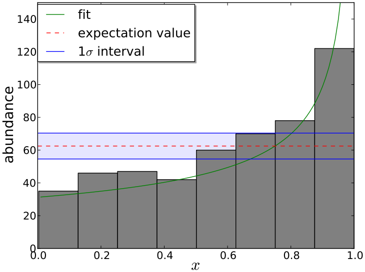
2. Artificial inclusion of an error by setting $\alpha \to \alpha(1+\epsilon)$
3. Mappings onto 1D:

$$P\left(\sigma^2|d\right) = \int \mathcal{D}m\ P(m, \sigma^2|d)$$
$$P\left(m|d\right) = \int_0^\infty \mathrm{d}\sigma^2\ P(m, \sigma^2|d)$$

4. Perform DIP test for the mapped posteriors.

Hierarchical model in 2D: Results

$\alpha = 2,\ q = 1,\ M = 1,\ N = 0.1$ and $\epsilon = 0.3$



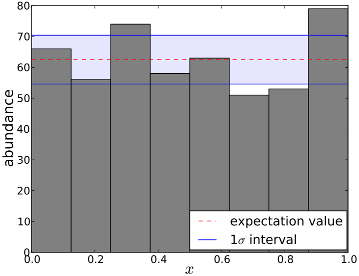
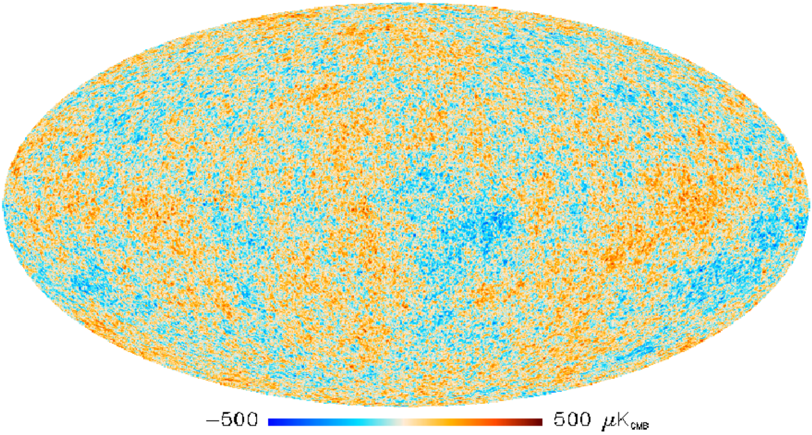


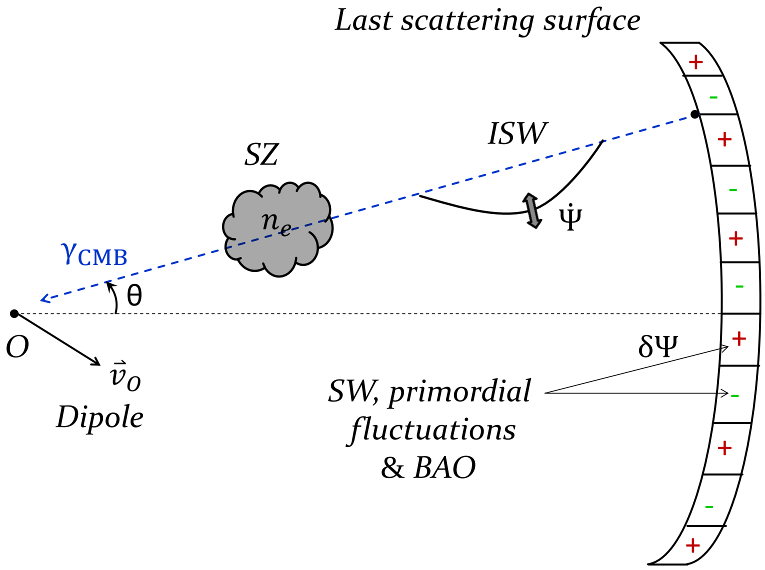
Figure: Left (right) histogram shows the unnormalized distribution of 500 x-values within eight bins as calculated from the m- (σ^2-) marginalized posterior. Fit: Skewness fitting formula with $\epsilon = 1$.

CMB APPLICATION
– Introduction to CMB non-Gaussianities –

Planck (SMICA) CMB map



source: Planck 2013 results. I.


Last scattering surface
ISW
SZ
n_e
$\dot{\Psi}$
γ_{CMB}
θ
O
$\vec{v}_O$
Dipole
δΨ
SW, primordial fluctuations & BAO
+
-

CMB TEMPERATURE ANISOTROPIES

– Statistics of the temperature anisotropies –

Characterization:

- Primordial graviational potential $\phi \to \Delta T/T$ is well described by a Gaussian distribution, i.e.

$$\phi \hookleftarrow \mathcal{G}(\phi, \Phi) := \frac{1}{\sqrt{|2\pi\Phi|}} \exp\left(-\frac{1}{2}\phi^\dagger \Phi^{-1} \phi\right).$$

Φ: 2-point function (covariance operator)

Characterization:

- Primordial graviational potential $\phi \to \Delta T/T$ is well described by a Gaussian distribution, i.e.

$$\phi \hookleftarrow \mathcal{G}(\phi, \Phi) := \frac{1}{\sqrt{|2\pi\Phi|}} \exp\left(-\frac{1}{2}\phi^{\dagger}\Phi^{-1}\phi\right).$$

  Φ: 2-point function (covariance operator)

- If ϕ purely Gaussian, statistics are determined by the two-point function.

Characterization:

- Primordial graviational potential $\phi \to \Delta T/T$ is well described by a Gaussian distribution, i.e.

$$\phi \hookleftarrow \mathcal{G}(\phi, \Phi) := \frac{1}{\sqrt{|2\pi\Phi|}} \exp\left(-\frac{1}{2}\phi^\dagger \Phi^{-1} \phi\right).$$

Φ: 2-point function (covariance operator)

- If ϕ purely Gaussian, statistics are determined by the two-point function.

However, there are deviations from Gaussianity!

Origins of non-Gaussianity:

1. primordial sources
   - Gaussian quantum fluctuations $\delta\phi$

     $\downarrow$

     non-linear inflation dynamics & non-linear GR

     $\downarrow$

     non-linear gravitational potential (curvature perturbation) $\varphi(\delta\phi,\ \delta\phi^2, \dots) \rightarrow \Delta T/T$

Origins of non-Gaussianity:

1. primordial sources
   - Gaussian quantum fluctuations $\delta\phi$
     $\downarrow$
     non-linear inflation dynamics & non-linear GR
     $\downarrow$
     non-linear gravitational potential (curvature perturbation)
     $\varphi(\delta\phi,\ \delta\phi^2, \dots) \rightarrow \Delta T/T$
2. non-primordial sources:
   - instrumental effects
   - residual foregrounds and point sources
   - 2nd order gravity effects
   - secondary CMB anisotropies, e.g. ISW, SZ, grav. lensing

Characterization of non-Gaussianity:

$\rightarrow$ need higher moments to describe deviations from Gaussianity

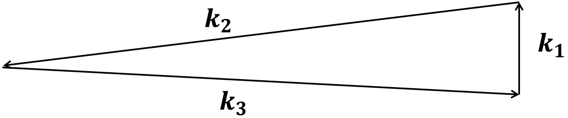
Characterization of non-Gaussianity:

→ need higher moments to describe deviations from Gaussianity
→ lowest order: three-point function / bispectrum:

$$\langle\varphi(k_1)\varphi(k_2)\varphi(k_3)\rangle_{(\varphi|C_l)}$$
$$= (2\pi)^3 \underbrace{\delta^{(3)}(k_1+k_2+k_3)}_{\text{triangle configuration}} \times \underbrace{f_{\mathrm{NL}}}_{\text{strength}} \times \underbrace{F_\varphi(|k_1|,|k_2|,|k_3|)}_{\text{shape of triangle}}$$

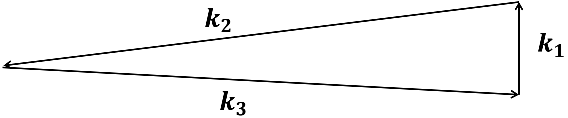
Local non-Gaussianity:

- The squeezed (local) shape of the bispectrum is defined as: $k_1 \ll k_2 \approx k_3$



Local non-Gaussianity:

- The squeezed (local) shape of the bispectrum is defined as: $k_1 \ll k_2 \approx k_3$



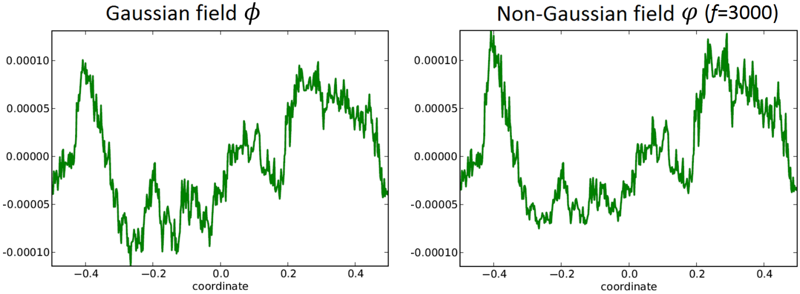
- For this local shape:

$$\frac{\Delta T}{T} \leftarrow \varphi = \phi + f_{\rm NL}\left(\phi^2 - \left\langle \phi^2 \right\rangle_{(\phi|\Phi)}\right)$$

$f_{\rm NL}$: non-Gaussianity parameter

CMB APPLICATION
– Posterior derivation –

$$d = \frac{\Delta T_{\mathrm{obs}}}{T_{\mathrm{CMB}}} = R\varphi + n \overset{\text{local type}}{=} R\left(\phi + f_{\mathrm{NL}}\left(\phi^2 - \langle\phi^2\rangle_{(\phi|\Phi)}\right)\right) + n$$



How can we reconstruct f_{NL} from given data d?

How likely is the resulting f_{NL}?

How can we reconstruct $f_{\rm NL}$ from given data d?

How likely is the resulting $f_{\rm NL}$?

$\rightarrow$ requested quantity: posterior $P(f_{\rm NL}|d)$

$\rightarrow$ used framework: information field theory

Information field theory:

likelihood prior

$$P(f|d) = \frac{P(d|f)P(f)}{P(d)} =: \frac{1}{Z} e^{-H(d,f)}$$

evidence

$$\rightarrow H(d,f) = -\ln[P(d|f)P(f)]$$

Z: partition function

- f_{NL}-posterior:

$$P(f|d) \propto P(d|f)P(f)$$
$$\propto \int \mathcal{D}\phi \; P(d,\phi|f) = \int \mathcal{D}\phi \; \exp(- \underbrace{H(d,\phi|f)}_{\text{contains terms } \propto \, \phi^4})$$

$(f = f_{\mathrm{NL}})$

- $f_{\rm NL}$-posterior:

$$P(f|d) \propto P(d|f)P(f)$$
$$\propto \int \mathcal{D}\phi \; P(d,\phi|f) = \int \mathcal{D}\phi \; \exp(-\underbrace{H(d,\phi|f)}_{\text{contains terms } \propto \, \phi^4})$$

$(f = f_{\rm NL})$

→ Impossible to perform path integration analytically!

<u>Solution:</u>

Taylor expansion of H in ϕ because:

$\phi \propto \mathcal{O}(10^{-5})$ and $P(\phi \approx 1) \approx 0$!

- around $m = \arg\min(H(d, \phi|f))$
- up to 2nd order in ϕ

$$P(f|d) \approx \int \mathcal{D}\phi \ \exp\left(-\left(H(d,m|f) + \frac{1}{2}(\phi - m)^\dagger D_{d,f}^{-1}(\phi - m)\right)\right)$$

$D_{d,f}$: inverse Hessian of $H(d,\phi|f)|_{\phi=m}$

$$P(f|d) \approx \int \mathcal{D}\phi \;\exp\left(-\left(H(d,m|f) + \frac{1}{2}(\phi - m)^\dagger D_{d,f}^{-1}(\phi - m)\right)\right)$$

$D_{d,f}$: inverse Hessian of $H(d,\phi|f)|_{\phi=m}$

$$\downarrow$$

Final f_{NL}-posterior:

$$P(f|d) \propto |2\pi D_{d,f}|^{\frac{1}{2}} \exp(-H(d,m|f))$$

$\rightarrow$ No numerically expensive sampling techniques necessary!

CMB APPLICATION
– Posterior validation –

Example: *Sachs-Wolfe*-limit

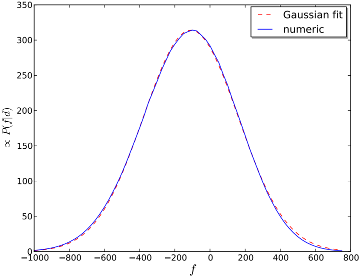
From now on we consider

$$N_{xy} = \sigma_n^2 \ \delta_{xy}$$
$$R(x,y) = -1/3 \ \delta(x-y)$$

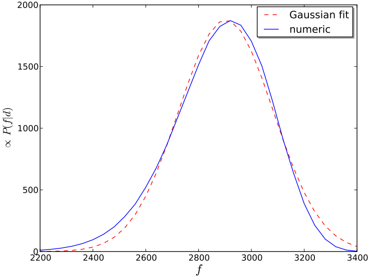
and study 1D-, 2D- flat sky and all sky toy-cases.

$$\Phi_{(l,m)(l',m')} = (C_l)_{\text{CMB}} \, \delta_{ll'} \delta_{mm'}.$$

Shape of the posterior (1D toy case)

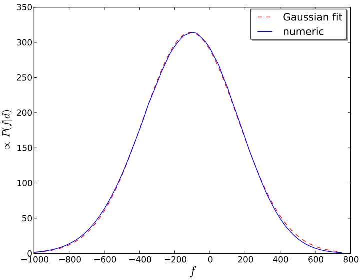


$f_{\mathrm{NL}}^{\mathrm{true}} = 3$

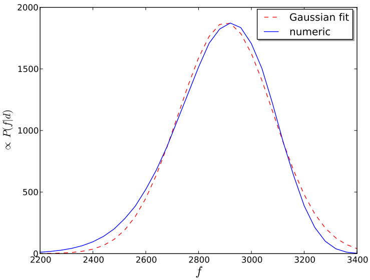


$f_{\mathrm{NL}}^{\mathrm{true}} = 3000$

Shape of the posterior (1D toy case)



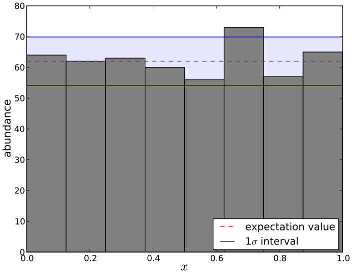
$f_{\mathrm{NL}}^{\mathrm{true}} = 3$



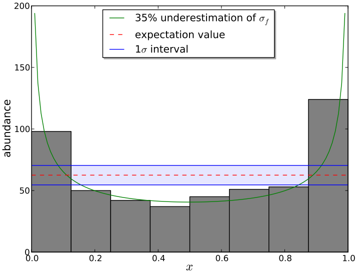
$f_{\mathrm{NL}}^{\mathrm{true}} = 3000$

How accurate is our posterior?

Sachs-Wolfe limit – 500 data realizations



2D flat sky



spherical harmonics

CONCLUDING REMARKS

Summary of the DIP test:

- DIP is a powerful posterior validation method
- Error diagnosis is possible
- Fitting formulae to estimate influence on posterior distribution
- Inspection by eye

Summary of the f_{NL} parameter:

- We derived a PDF for the f_{NL} parameter
- Precision of the posterior was validated (DIP–test) in the large-scale limit (1D, 2D)
- Gaussian shape of the PDF for small values of f_{NL}
- Monte Carlo sampling isn't necessary

Thank you for paying attention!

- DIP-test:
  S. Dorn *et al.*, Phys.Rev.E.88.053303 (2013)
- f_{NL}-posterior:
  S. Dorn *et al.*, Phys.Rev.D.88.103516 (2013)

Proof. We show here analytically that $P(x) = 1$ if $\tilde{P}(s|d) = P(s|d)$:

$$
\begin{aligned}
P(x) &= \int_{-\infty}^{\infty} ds \int \mathcal{D}d \; P(x,d,s) \\
&= \int_{-\infty}^{\infty} ds \int \mathcal{D}d \; P(x|d,s)P(d,s) \\
&= \int_{-\infty}^{\infty} ds \int \mathcal{D}d \; P(d,s)\; \delta\left(x - \int_{-\infty}^{s} ds' \; P(s'|d)\right) \\
&= \int_{-\infty}^{\infty} ds \int \mathcal{D}d \; P(d)P(s|d)\; \delta(x - x_d(s)),
\end{aligned}
$$

$$
\begin{aligned}
P(x) &= \partial_x \int_0^x dx' P(x') \\
&= \partial_x \int \mathcal{D}d \; P(d) \int_{-\infty}^{\infty} ds \; P(s|d) \underbrace{\int_0^x dx' \; \delta(x' - x_d(s))}_{\Theta(x - x_d(s))} \\
&= \partial_x \int \mathcal{D}d \; P(d) \int_{-\infty}^{s_d(x)} ds \; P(s|d) \\
&= \partial_x \int \mathcal{D}d \; P(d) \underbrace{x_d(s_d(x))}_{=x} = \partial_x x \int \mathcal{D}d \; P(d) \\
&= \partial_x x = 1
\end{aligned}
$$

$$P(x) = \int_{-\infty}^{\infty} \mathrm{d}s \int \mathcal{D}d\; P(d)\; \mathcal{G}\left(s - \bar{s}_d, \sigma^2\right) \times \delta\left(x - \int_{-\infty}^{s} \mathrm{d}\tilde{s}\; \mathcal{G}^{\epsilon}\left(\tilde{s} - \bar{s}_d, \sigma^2\right)\right)$$