Recycable Forecasting: DALI & the Fisher matrix

Elena Sellentin

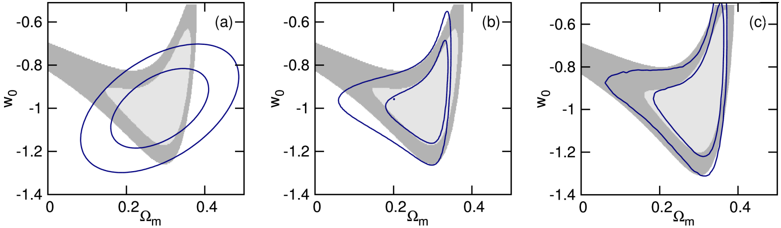
Université de Genève
Département de Physique Théorique

19. April 2017

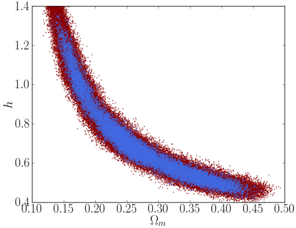
Forecasting

- Given a parametric model, I'd like to know to which extent a future data set can constrain its parameters.
- I'd also like to know how I have to optimize my survey such that I learn most.
- In the end, I'd like to recycle everything I calculated for the final pipeline of the true data analysis.
  - Metropolis-Hastings needed a proposal distribution.
  - HMC needed a cheap but faithful potential.

→ DALI & Fisher matrix



(Sellentin et al. 2014)



(Sellentin & Schäfer 2016)

Peak

Given a likelihood $L(\boldsymbol{p})$, its peak lies at

$$\nabla_{\boldsymbol{p}} L(\boldsymbol{p}) = 0. \tag{1}$$

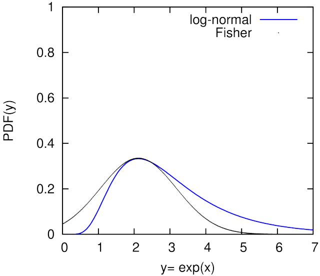
The second derivatives are then a first hunch of the width of the likelihood.

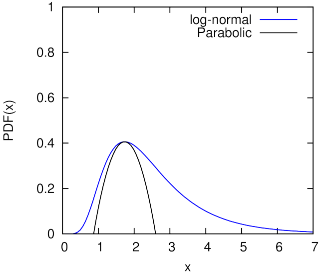
Forecasting: Assume you know the peak position.

The Fisher matrix

Locally, any peak can be approximated by a parabola. $\Rightarrow$ negative likelihoods.

$\Rightarrow$ parabolic expansion of log-likelihood.





Assume Gaussianly distributed data

$$\mathcal{G}(\boldsymbol{x}, \boldsymbol{\mu}(\boldsymbol{p}), \mathsf{C}) = \frac{1}{\sqrt{|2\pi \mathsf{C}|}} \exp\left(-\frac{1}{2}[\boldsymbol{x} - \boldsymbol{\mu}(\boldsymbol{p})]^T \mathsf{C}^{-1} [\boldsymbol{x} - \boldsymbol{\mu}(\boldsymbol{p})]\right), \qquad (2)$$

then the log-likelihood is

$$\mathcal{L} = \frac{1}{2} \mathrm{Tr}\left(\ln(\mathsf{C}) + (\boldsymbol{x} - \boldsymbol{\mu}) \boldsymbol{C}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)$$

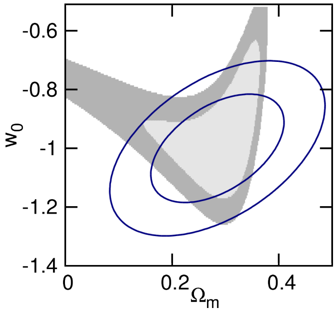
The Fisher-matrix is then

$$\begin{aligned} F_{\alpha\beta} &= \langle \mathcal{L}_{,\alpha\beta} \rangle |_{\hat{\boldsymbol{p}}} \\ &= \frac{1}{2} \mathrm{Tr}\left(\boldsymbol{C_0}^{-1} C_{,\alpha} \boldsymbol{C_0}^{-1} C_{,\beta}\right) + \boldsymbol{\mu}_{,\alpha} \boldsymbol{C_0}^{-1} \boldsymbol{\mu}_{,\beta} \end{aligned} \qquad (3)$$

Fisher approximation

$$L(\boldsymbol{X}|\boldsymbol{p}) \approx N \cdot \exp(-\frac{1}{2} F_{\alpha\beta} \Delta p_\alpha \Delta p_\beta) \tag{4}$$

$\beta \mathsf{F}^{-1}$ with $\beta \approx 0.01$ is a useful guess for MH's proposal distribution.



DALI

Q: Do I really always have to store all of my MCMC-samples in order to represent my posterior? Why can't I have an analytical expression for my non-Gaussian posterior?
A: You can. It is called DALI.

$$P \propto \exp\left[-\frac{1}{2}F_{\alpha\beta}\Delta p_\alpha \Delta p_\beta - \frac{1}{3!}S_{\alpha\beta\gamma}\Delta p_\alpha \Delta p_\beta \Delta p_\gamma - \frac{1}{4!}Q_{\alpha\beta\gamma\delta}\Delta p_\alpha \Delta p_\beta \Delta p_\gamma \Delta p_\delta \right]$$

DALI

Expand the *mean* or the *covariance matrix*:

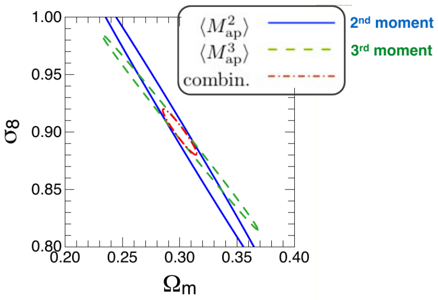
$$\boldsymbol{\mu}(\boldsymbol{p}) = \hat{\boldsymbol{\mu}} + \boldsymbol{\mu}_{,\alpha}\Delta p_\alpha + \frac{1}{2}\boldsymbol{\mu}_{,\alpha\beta}\,\Delta p_\alpha \Delta p_\beta + \frac{1}{3!}\boldsymbol{\mu}_{,\alpha\beta\delta}\Delta p_\alpha \Delta p_\beta \Delta p_\delta ...$$

$$\mathsf{C}(\boldsymbol{p}) = \hat{\boldsymbol{C}} + \mathsf{C}_{,\alpha}\,\Delta p_\alpha + \frac{1}{2}\mathsf{C}_{,\alpha\beta}\,\Delta p_\alpha \Delta p_\beta + \frac{1}{3!}\mathsf{C}_{,\alpha\beta\gamma}\,\Delta p_\alpha \Delta p_\beta \Delta p_\gamma ...$$

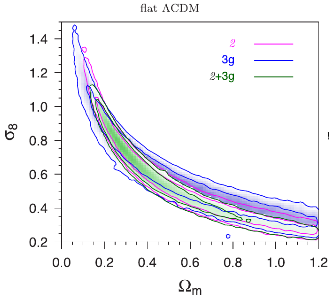
$$P \approx N \exp\left[-\frac{1}{2}(\boldsymbol{x} - \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_{,\alpha}\Delta p_\alpha - \frac{1}{2}\boldsymbol{\mu}_{,\alpha\beta}\Delta p_\alpha \Delta p_\beta - ...)C^{-1}(\boldsymbol{x} - \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_{,\alpha}\Delta p_\alpha - \frac{1}{2}\boldsymbol{\mu}_{,\alpha\beta}\Delta p_\alpha \Delta p_\beta - ...)\right]$$

Numerical issues

- Need rather precise derivatives for a good Fisher matrix and DALI
- Inaccuracies lead to reorientations of the Ellipses, and noses in DALI.

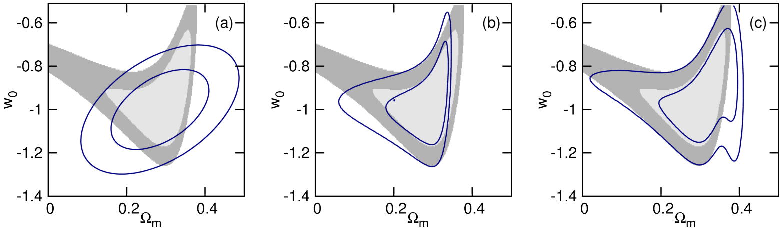


Kilbinger et al. 2005



Fu et al. 2014

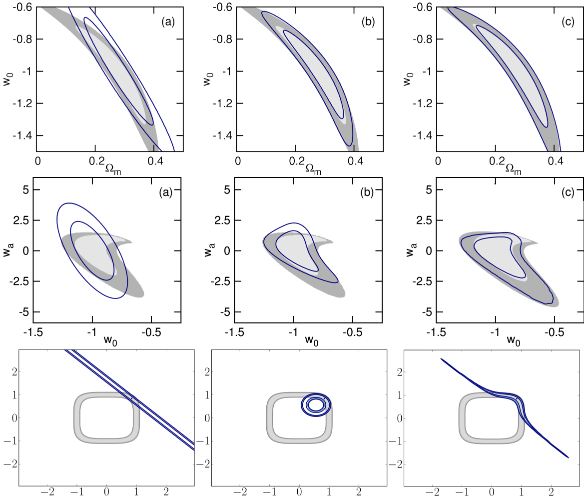
Noses



For Fisher matrices and DALI:

- Be careful with numerical instabilities in matrix inversions.
- Plot your derivatives together with the derived functions. Do they make sense?

Example applications

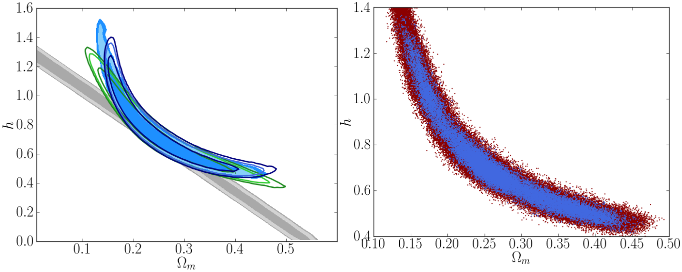


various Sellentin et al. 201x

Accelerating Hamiltonian Monte Carlo

- Use $\mathcal{L}$ as potential 'energy', add randomized kinetic energy

- Acceptance rate: 0.02 $\rightarrow 0.3 - 0.5$
- time/sample: $t_{leapfrog} \ll t_{MH}$ $\rightarrow$ speed up factor depends on leap frog steps
- Typical speed-ups: 100-300.
- Works also with real data!



Exercises

If your MH-MCMC runs, come and ask for the DALI-tensors to make it run in a circle as shown here: