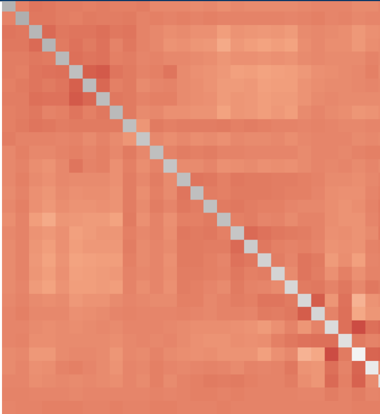
Application of Bayesian Graphs to SNIa Data Analysis and Compression

based on arXiv:1603.08519 & 1604.04631



Cong Ma

Purple Mountain Obs. & LUTH, Obs. Paris–Meudon

COSMO21 — 25/05/2016, Chania

Cosmological inference with SN data

- Background
- Bayesian inference and graphical models

Bayesian data compression

- Method of compression
- Validating & using compression

Discussion

Most of this talk is based on Ma, Corasaniti, Bassett (`arXiv:1603.08519`)

What SNIa data aren't, and what they are

They aren't...

- Standard candles
- Static table of z vs. $\mu(z)$
- "Error bar" on each $\mu(z)$

They are:

- "Standardizable" candles
- Reduced data from light-curves, incl. m (mag.), X (stretch), C (color)...
- Correlated covariance among all of them

SNIa standardization

SN dependencies:

- Tripp (1998) relation: $\mu = m + \alpha X + \beta C - M$
- Host-galaxy stellar mass: M_{stellar}

(Joint Light-Curve Analysis [JLA], Betoule+ `arXiv:1401.4064`)

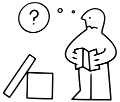
"Assembly" procedure:

$$\mu_{\mathsf{d}} = m + \alpha X + \beta C - (M + \Delta_M),$$

Δ_M always set to zero if $M_{\mathsf{stellar}} < 10^{10} \mathrm{M}_{\odot}$.

Standardization as affine transformation of data:

$$\vec{\mu}_{\mathsf{d}} = \mathbf{J}(\alpha, \beta)\vec{v} - \vec{M}_{\mathsf{d}}(\Delta_M),$$
$$\Rightarrow \mathbf{C}_{\mathsf{d}} = \mathbf{J}\mathbf{C}_{\vec{v}}\mathbf{J}^{\mathsf{T}}.$$



©Ikea

When cosmological model meets data

Cosmological model fitting specs:

$\vec{\boldsymbol{\mu}}_{\mathrm{t}}(\vec{\boldsymbol{\theta}})$: model prediction parameterized by $\vec{\boldsymbol{\theta}}$ (i.e. Ω_{M}, etc)

$\vec{\boldsymbol{\mu}}_{\mathrm{d}}(\vec{\boldsymbol{\varphi}})$, $\mathbf{C}_{\mathrm{d}}(\vec{\boldsymbol{\varphi}})$: standardized data, $\vec{\boldsymbol{\varphi}} = (\alpha, \beta, \Delta_M)$

$\vec{\boldsymbol{\Theta}}$: joint variable $\vec{\boldsymbol{\Theta}} = (\vec{\boldsymbol{\theta}}, \vec{\boldsymbol{\varphi}})$ (i.e. cos. + standardization)

Question:

We assume a Gaussian uncertainty model.

Is it fine to simply use the χ^2 expression

$$\chi^2(\vec{\boldsymbol{\Theta}}) = (\vec{\boldsymbol{\mu}}_{\mathrm{t}} - \vec{\boldsymbol{\mu}}_{\mathrm{d}})^{\mathsf{T}} \mathbf{C}_{\mathrm{d}}^{-1} (\vec{\boldsymbol{\mu}}_{\mathrm{t}} - \vec{\boldsymbol{\mu}}_{\mathrm{d}}) \text{ ?}$$

When cosmological model meets data

Cosmological model fitting specs:

$\vec{\mu}_{\text{t}}(\vec{\theta})$: model prediction parameterized by $\vec{\theta}$ (i.e. Ω_{M}, etc)

$\vec{\mu}_{\text{d}}(\vec{\varphi})$, $\mathbf{C}_{\text{d}}(\vec{\varphi})$: standardized data, $\vec{\varphi} = (\alpha, \beta, \Delta_M)$

$\vec{\Theta}$: joint variable $\vec{\Theta} = (\vec{\theta}, \vec{\varphi})$ (i.e. cos. + standardization)

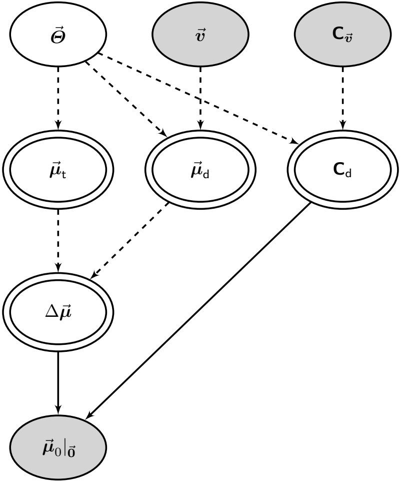
Question:

We assume a Gaussian uncertainty model.
Is it fine to simply use the χ^2 expression

$$\chi^2(\vec{\Theta}) = (\vec{\mu}_{\text{t}} - \vec{\mu}_{\text{d}})^{\mathsf{T}} \mathbf{C}_{\text{d}}^{-1} (\vec{\mu}_{\text{t}} - \vec{\mu}_{\text{d}}) \text{ ?}$$

Answer: No.

Graphical model: evidence & inference



3 kinds of variables/nodes:

F: "free" ones, to be inferred ($\vec{\Theta}$)

E: "evident" ones, with evidence (data or belief, gray here)

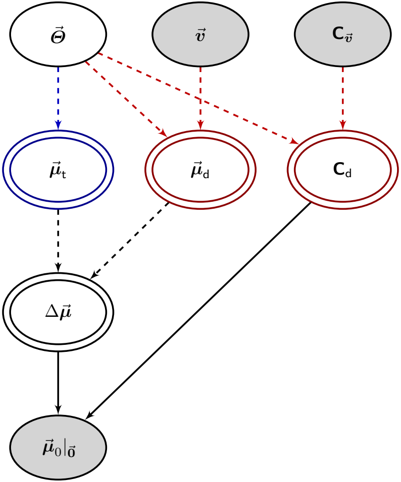
H: our "intermediate" ones, could be merely computing device

Meaning of "inference"

$$P(F \mid E) = \int \frac{P(F, E, H)}{P(E)} \, dH$$

Further, very accessible reading:
D'Agostini (`arXiv:physics/0511182`)

Graphical model: joint probability



Joint probability (chain rule)

$$P(\text{all}) = \prod_i P\big(X[i]\,\big|\,\text{parents of } X[i]\big)$$

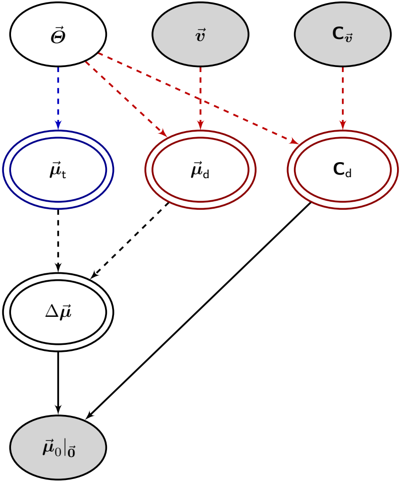
$$\begin{aligned} P(\text{all}) = {} & P(\vec{\mu}_0 \,|\, \Delta\vec{\mu}, \mathbf{C}_\text{d}) \times P(\Delta\vec{\mu} \,|\, \vec{\mu}_\text{t}, \vec{\mu}_\text{d}) \\ & \times P(\vec{\mu}_\text{t} \,|\, \vec{\Theta}) \\ & \times P(\vec{\mu}_\text{d} \,|\, \vec{\Theta}, \vec{v}) P(\mathbf{C}_\text{d} \,|\, \vec{\Theta}, \mathbf{C}_{\vec{v}}) \\ & \times P(\vec{\Theta}) P(\vec{v}) P(\mathbf{C}_{\vec{v}}) \end{aligned}$$

($\Delta\vec{\mu} = \vec{\mu}_\text{t} - \vec{\mu}_\text{d}$, diff. of theory vs. data)

Also read:

Kjælruff & Madsen (2013, Springer, 2nd ed.)

Graphical model: arrows, conditional probability



The "network of arrows" express how variables depend on other ones, i.e. conditional dependence.

Example

- Red dashed arrows: standardization (deterministic)
  $P(\vec{\mu}_{\rm d}, \mathbf{C}_{\rm d} \,|\, \vec{\Theta}, \vec{v}, \mathbf{C}_{\vec{v}}) = \delta(\dots)$
- **Solid** arrows $\Delta\vec{\mu} \rightarrow \vec{\mu}_0 \leftarrow \mathbf{C}_{\rm d}$: error model (non-deterministic)
  $P(\vec{\mu}_0 \,|\, \Delta\vec{\mu}, \mathbf{C}_{\rm d}) =$ Gauss function . . .

Results for Bayesian inference

Expression for the posterior:

$$\ln P(\vec{\boldsymbol{\Theta}} \,|\, E) = -\ln Z - \frac{1}{2}\left[\ln\det \mathbf{C}_\mathsf{d}(\vec{\boldsymbol{\Theta}}) + \chi^2(\vec{\boldsymbol{\Theta}})\right] + \ln P(\vec{\boldsymbol{\Theta}}),$$

$$\chi^2(\vec{\boldsymbol{\Theta}}) = (\Delta\vec{\boldsymbol{\mu}})^\mathsf{T}\, \mathbf{C}_\mathsf{d}^{-1}\, (\Delta\vec{\boldsymbol{\mu}})\,.$$

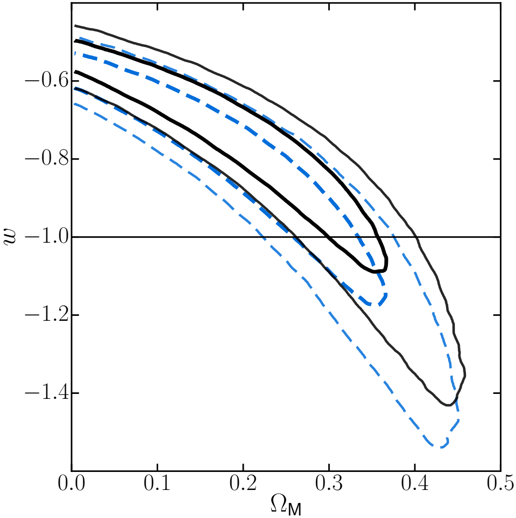
In other words. . .

(Log) posterior = −normalization + "likelihood" + prior
(for this fairly simple case of JLA)

A consequence of Gauss function

$$f(\vec{\boldsymbol{x}}) = \frac{1}{\sqrt{(2\pi)^n \det \mathbf{C}}} \exp\left[(\vec{\boldsymbol{x}} - \vec{\boldsymbol{x}}_0)^\mathsf{T}\, \mathbf{C}^{-1}\, (\vec{\boldsymbol{x}} - \vec{\boldsymbol{x}}_0)\right]$$

Example: wCDM parameter constraints

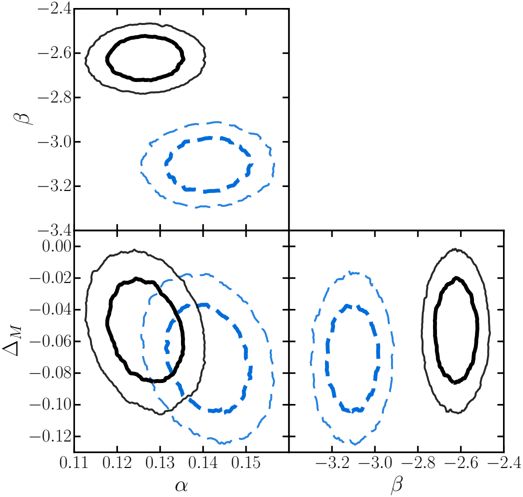


Flat wCDM, posterior of $w - \Omega_\mathrm{M}$, marginalized over everything else

Legends

- ——— Bayesian
- – – – "χ^2"

Example: (mostly) generic $(\alpha, \beta, \Delta_M)$

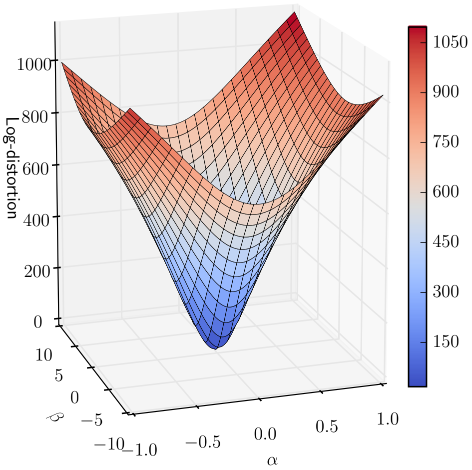


Another section through the posterior parameter space: SNIa standardization parameters.

Lowered significance of two-population by host-galaxy stellar mass (non-zero Δ_M).

- Sullivan+ (`arXiv:1003.5119`): 3.7 σ
- JLA (w/ systematics): 3 σ
- Bayesian: 2.4 σ

"χ^2" = prior distortion



$$
\begin{aligned}
\ln P(\vec{\Theta} \mid E) \sim &- \frac{1}{2}\chi^2(\vec{\Theta}) \\
&- \frac{1}{2} \ln \det \mathbf{C}_{\mathsf{d}}(\vec{\Theta}) \\
&+ \ln P(\vec{\Theta}),
\end{aligned}
$$

Using just χ^2 is equivalent to applying this distortion to the prior on (α, β).

No surprise the "χ^2" analysis "pulls" (α, β) away from zero.

Motivations for compression

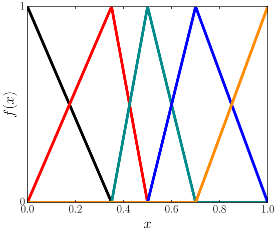
Before compression

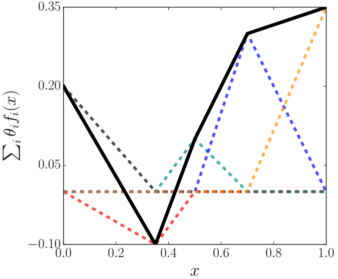
- (α, β)-dependent covariance
- $n \sim 100$ — 1000
- $\mathcal{O}(n^3)$ evaluation of log-prob.

After compression

- Constant covariance
- $n \sim$ several tens
- $\mathcal{O}(n^2)$ complexity; $\mathcal{O}(n^3)$ factorization only once

Compression needs fixing, too





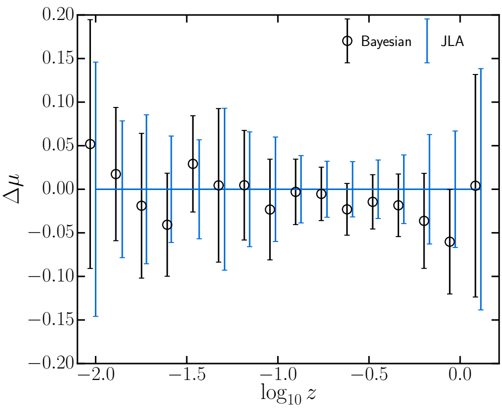
Compression method

- Discrete linear compression in the log-z space
- Lossy, but by design
- Another inference problem, so same cautions apply

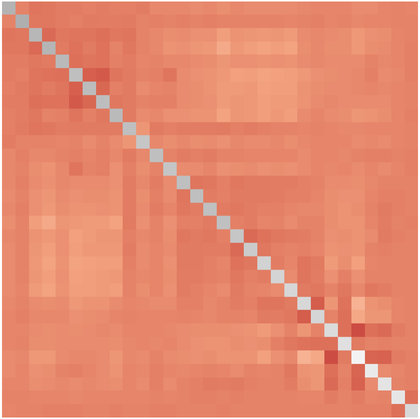
Compression output

- Posterior of combination coefficients $\rightarrow$ compressed distance moduli
- Posterior covariance $\rightarrow$ compressed covariance

Comparison of compression schemes

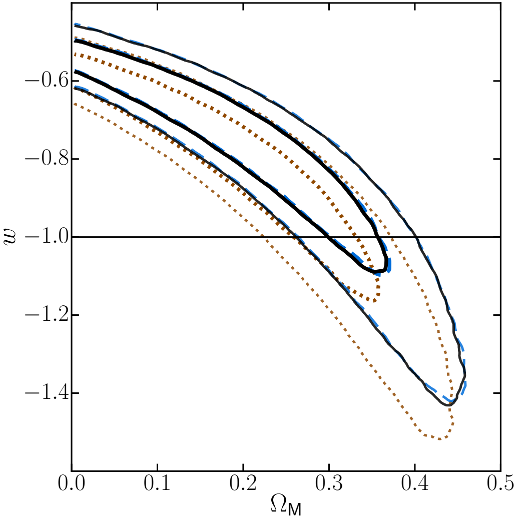


Bayesian vs. JLA compressions



Diff. in covariance structure

Bayesian compression validated



Performance of compression

- wCDM &co. as test models
- Result: minimal difference (by KL, visual, see text)

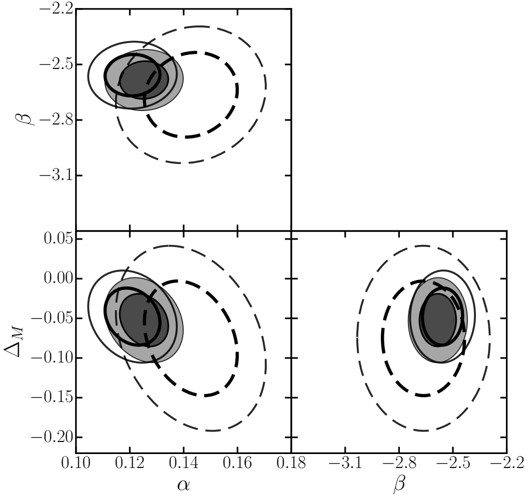
Legends

- ——— Bayes., full data
- – – – Bayes., compressed
- ········· Betoule+ (χ^2), compressed

Warning

Don't use JLA Tables F.1/F.2!

Application: simple cross-validation



Set-up of cross-validation:

- Lower z-cut: [0.01 – 0.114]
- Higher z-cut: [0.082 – 1.3]
- N. of points 166, 599

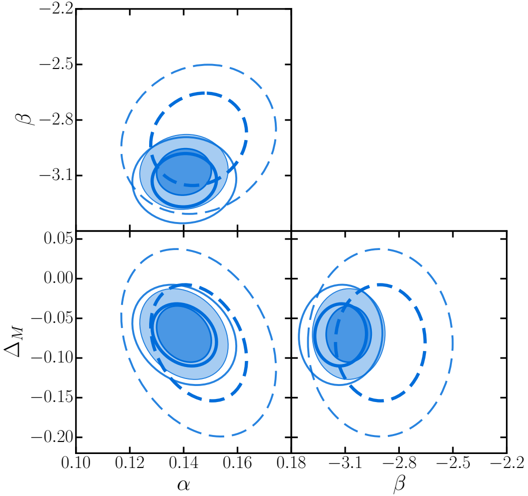
Legends

- – – – Lower z-cut
- ——— Higher z-cut

Method used:

- Baysian compression

Application: simple cross-validation



Set-up of cross-validation:

- Lower z-cut: [0.01 – 0.114]
- Higher z-cut: [0.082 – 1.3]
- N. of points 166, 599

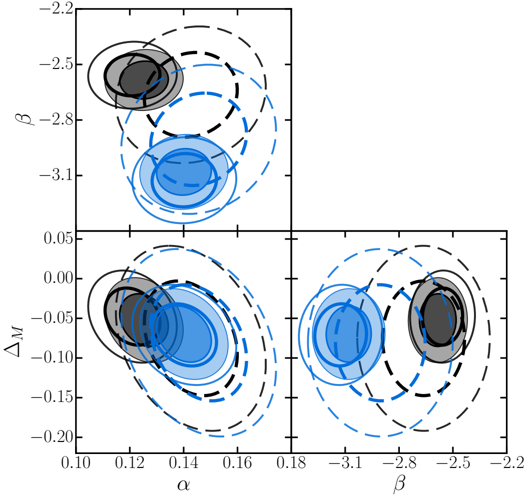
Legends

- – – – Lower z-cut
- ——— Higher z-cut

Method used:

- "χ^2" compression

Application: simple cross-validation



Set-up of cross-validation:

- Lower z-cut: [0.01 – 0.114]
- Higher z-cut: [0.082 – 1.3]
- N. of points 166, 599

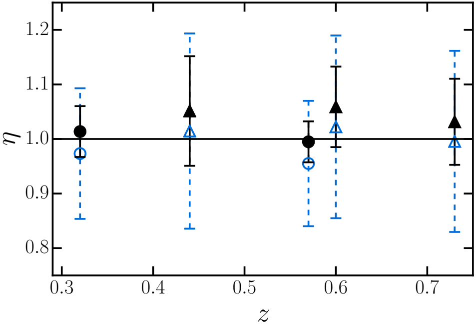
Legends

- – – – Lower z-cut
- ——— Higher z-cut

Method used:

- Baysian compression
- "χ^2" compression

Application: Etherington relation



(CM & PSC, `arXiv:1604.04631`)

Goal: testing

$$\eta = \frac{d_{\mathsf{L}}}{(1+z)^2 d_{\mathsf{A}}} = 1 \quad ?$$

Challenges:

- BAO gives d_{A} at certain effective z
- d_{L} not directly given by SNIa there

With log-z compression we don't need to cherry-pick SN data or use very narrow bands.

Summary

1. To use SNIa data correctly, we need more than χ^2.
2. Graphical models help us do this.
3. A correction leads to a shift in parameters and re-interpretation of SN standardization.
4. Compression simplifies data and computation, and provide other uses, but must be practised with the correct statistical method.

Code release

- https://gitlab.com/congma/libsncompress
- https://gitlab.com/congma/sn-bayesian-model-example

Discussion: is JLA self-consistent?

1. JLA (Betoule+ arXiv:1401.4064) → Mosher+ (.4065), for SALT2 model training & error modelling.
2. Mosher+ → Marriner+ (arXiv:1107.4631), for "intrinsic dispersion" calculation using the tool SALT2mu.
3. SALT2mu might seem to be discarding an equivalent of the $\ln \det \mathbf{C}$ term as well...?