Convergence tests for MCMC

Elena Sellentin

Sterrewacht
Leiden University, NL

Imperial Data Analysis Workshop (2018)

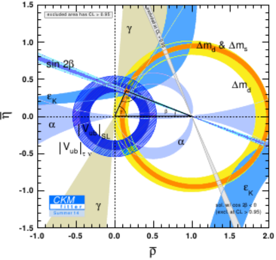
Bayesian Inference

$$\mathcal{P}(\boldsymbol{\theta}, \mathcal{M}|\boldsymbol{x}) = \frac{L(\boldsymbol{x}|\boldsymbol{\theta}, \mathcal{M})\pi(\boldsymbol{\theta})}{\epsilon(\boldsymbol{x}|\mathcal{M})} \tag{1}$$

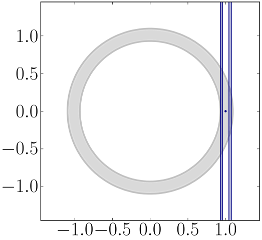
- $\mathcal{P}(\boldsymbol{\theta}, \mathcal{M}|\boldsymbol{x})$: the posterior.
- $L(\boldsymbol{x}|\boldsymbol{\theta}, \mathcal{M})$: the likelihood.
- $\pi(\boldsymbol{\theta})$: the priors.
- $\epsilon(\boldsymbol{x}|\mathcal{M})$: the evidence (‘marginal likelihood’).

→ The posterior can be difficult to obtain. Need sampling techniques (MCMC, Gibbs...)

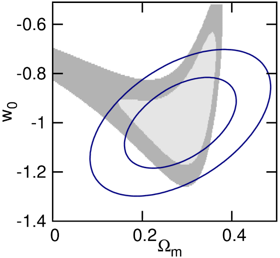
Examples



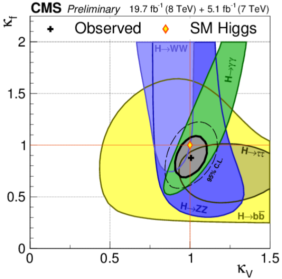
γ
Δm_d & Δm_s
sin 2β
Δm_d
ε_K
α
$|V_{ub}|_{SL}$
$|V_{ub}|_{\tau\nu}$
α
ε_K
γ
CKM fitter
$\bar{\eta}$
$\bar{\rho}$



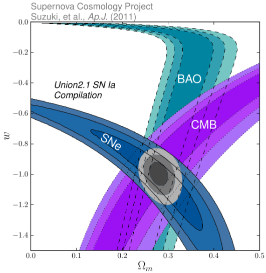
1.0
0.5
0.0
−0.5
−1.0
−1.0 −0.5 0.0 0.5 1.0



w_0
Ω_m
-0.6
-0.8
-1
-1.2
-1.4
0
0.2
0.4



CMS Preliminary 19.7 fb⁻¹ (8 TeV) + 5.1 fb⁻¹ (7 TeV)
Observed
SM Higgs
κ_f
κ_V
H→WW
H→γγ
H→ττ
H→ZZ
H→bb̄
95% C.L.



Supernova Cosmology Project
Suzuki, et al., Ap.J. (2011)
Union2.1 SN Ia Compilation
BAO
CMB
SNe
w
Ω_m

Detailed Balance

Equilibrium between the occupation of two states $\mathcal{P}_{i,j}$ is reached if

$$r_{i \to j}\mathcal{P}_i = r_{j \to i}\mathcal{P}_j, \tag{2}$$

and the transition probability from $\mathcal{P}_i$ to state $\mathcal{P}_j$ has rate $r_{i \to j}$.

$\Rightarrow$ MCMC uses the same principle as e.g. in photon emisson/absorption from electronic shells in atoms

→ Reaching equilibrium takes time!

Convergence

If the chain ran for long enough ('has converged'), then

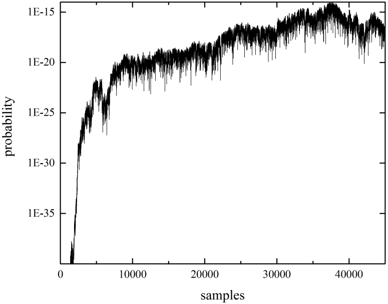
$$n(\boldsymbol{\theta}) \propto \mathcal{P}(\boldsymbol{\theta}).$$

It is $n(\boldsymbol{\theta})$ which matters, so the entire chain, and not single points.

Burn-in period

Reaching equilibrium needs interactions.

- Thermodynamics: put cold object into warmer environment
- Radiation physics: put a phosphorescent object into the dark
- MCMC: burn-in period (searching for the peak)

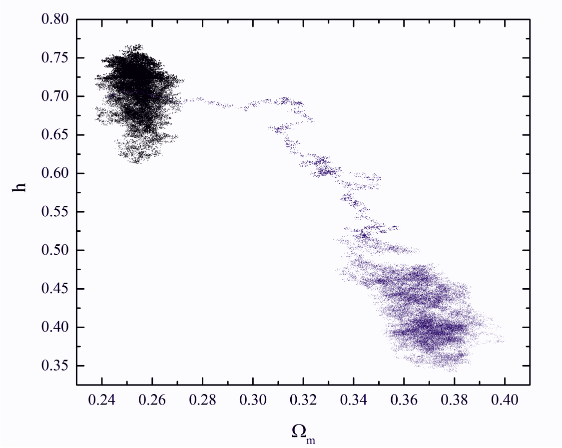


Plot: K. Wolz

Burn-in

Cut away the unrepresentative burn-in period:

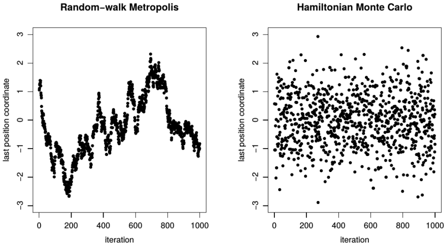
- Rat tails in likelihood plots
- log-likelihood steadily increases



Plot: K. Wolz

Monitor convergence

- measure the correlation length of chains
- thin highly correlated chains

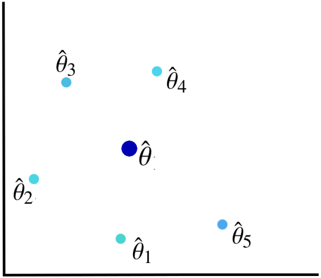


Neil 2012, arXiv: 1206.1901

Convergence Diagnostics

Gelman-Rubin Test: Intra-Chain variance vs. Inter-chain variance.

- Run M different chains with different starting points, let $m \in [1, M]$.
- mth chain: $\theta_1^m, \theta_2^m, \theta_3^m, ...\theta_{N_m}^m$.
- Discard the burnins.
- Calculate for each parameter θ, the posterior mean $\hat{\theta}_m = \frac{1}{N_m} \sum_i^{N_m} \theta_i^m$,
- ...and the intra-chain variance $\sigma_m^2 = \frac{1}{N_m - 1} \sum_i^{N_m} (\theta_i^m - \hat{\theta}_m)^2$.
- Calculate $\hat{\theta}$, the mean of all chains $\hat{\theta} = \frac{1}{M} \sum_m^M \hat{\theta}_m$.



Gelman-Rubin cntd.

- Compute how the individual means vary around the joint mean
  $B = \frac{N}{M-1} \sum_{m=1}^{M} (\hat{\theta}_m - \hat{\theta})^2$
- Compute the averaged variances of the chains
  $W = \frac{1}{M} \sum_{m=1}^{M} \sigma_m^2$
- Define $\hat{V} = \frac{N-1}{N} W + \frac{M+1}{MN} B$; under convergence, this is an unbiased estimator of the true variance. But if the chains have converged, then W is *also* an unbiased estimate of the true variance. Hence...
- ...test whether $R = \sqrt{\hat{V}/W} \approx 1$.
  If it is not, convergence has not been reached.
- Various refinements exist, see Gelman & Rubin (1992), Brooks & Gelman (1997).

Summary of equations

- mth chain: $\theta_1^m, \theta_2^m, \theta_3^m, ...\theta_{N_m}^m$.
- For each parameter θ, the posterior mean $\hat{\theta}_m = \frac{1}{N_m}\sum_i^{N_m} \theta_i^m$,
- For each parameter compute the intra-chain variance $\sigma_m^2 = \frac{1}{N_m - 1}\sum_i^{N_m}(\theta_i^m - \hat{\theta}_m)^2$.
- Calculate $\hat{\theta}$, the mean of all chains $\hat{\theta} = \frac{1}{M}\sum_m^M \hat{\theta}_m$.
- Compute how indiv. means scatter around the joint mean $B = \frac{N}{M-1}\sum_{m=1}^M(\hat{\theta}_m - \hat{\theta})^2$
- Compute the averaged variances of the chains $W = \frac{1}{M}\sum_{m=1}^M \sigma_m^2$
- Compute $\hat{V} = \frac{N-1}{N} W + \frac{M+1}{MN} B$
- ...test whether $R = \sqrt{\hat{V}/W} \approx 1$.
  If it is not, convergence has not been reached.