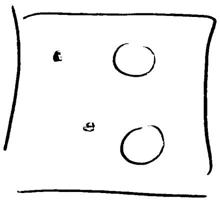
Bayesian Hierarchical Models (BHM)

3g	2.9	size
		material
		wearing
		impurities
3.5g	2	manufacturer
		noisy detector

BHM account for all sources of uncertainty and propagate the uncertainties from one step to the next

step ↓ submodel

Example



$L = \underline{\sigma_T} A T^4$

Collect random variables:

- A } population variability
- T
- L funct of R.V.
- $\hat{L}$ noisy detector
- σ_T uncertain

want $P(\sigma_T \mid \hat{L})$

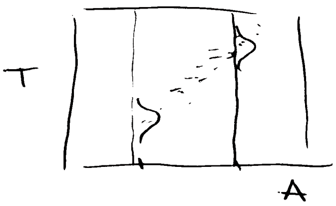
- detector: $P(\hat{L} \mid L) = G(\hat{L} \mid L)$
- physical law $L = \sigma_T A T^4$. δ_D

$$P(L \mid \sigma_T, A, T) = \delta_D(L - \sigma_T A T^4)$$

- $P(\sigma_T, A, T) = \underbrace{P(\sigma_T)}_{\pi(\sigma_T)} P(A, T)$

• $P(A, T)$ look at data!

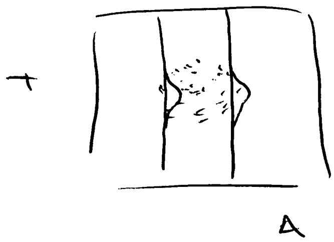
case A:



dependent

$P(A,T) \neq P(A)P(T)$

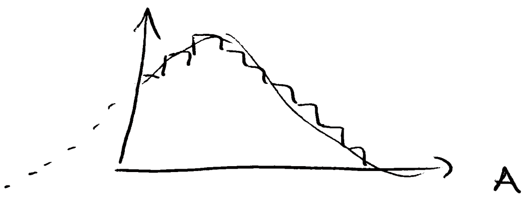
→ case B:



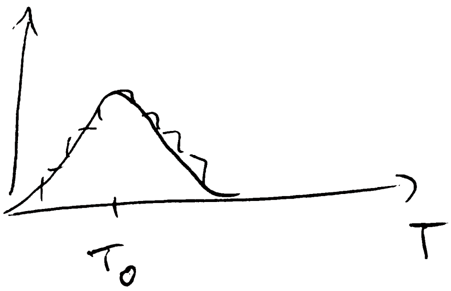
independent

$P(A,T) = P(A)P(T)$

• $P(A)$ $P(T)$ histogram data



$\Gamma(A \mid A_0, \nu)$



$G(T \mid T_0)$

$$P(\sigma_T \mid \hat{L}) = \int P(\sigma_T, A, T, L \mid \hat{L})\, dA\, dT\, dL$$

$$= \int \frac{P(\sigma_T, A, T, L, \hat{L})}{P(\hat{L})}\, dA\, dT\, dL$$

$$= \int \frac{P(\sigma_T, A, T \mid L, \hat{L})\, P(L, \hat{L})}{P(\hat{L})}\, dA\, dT\, dL$$

$$= \int \frac{P(\sigma_T, A, T \mid L, \cancel{\hat{L}})\, \overbrace{P(\hat{L} \mid L)}^{G}\, P(L)}{P(\hat{L})}\, dA\, dT\, dL$$

$$= \int \frac{P(\sigma_T, A, T, L)\, G(\hat{L} \mid L)}{P(\hat{L})}\, dA\, dT\, dL$$

$$= \int \frac{P(L \mid \sigma_T, A, T)\, P(A, \sigma_T, T)\, G(\hat{L} \mid L)}{P(\hat{L})}\, dA\, dT\, dL$$

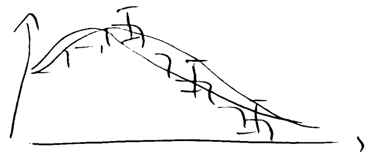
$$= \iiint \frac{\delta_D(L - \sigma_T A T^4)\, P(A, \sigma_T, T)\, G(\hat{L} \mid L)}{P(\hat{L})}\, dA\, dT\, dL$$

$$= \int_A \int_T \frac{P(\sigma_T, A, T)\, G(\hat{L} \mid \sigma_T A T^4)}{P(\hat{L})}\, dA\, dT$$

$$= \int_A \int_T \frac{P(\sigma_T)\, P(A)\, P(T)\, G(\hat{L} \mid \sigma_T A T^4)}{P(\hat{L})}\, dA\, dT$$

$$= \int_A \int_T \frac{\Gamma(\sigma_T)\, \Gamma(A \mid A_0, \nu)\, G(T \mid T_0)\, G(\hat{L} \mid \sigma_T A T^4)}{P(\hat{L})}\, d$$

$A \sim \pi(A \mid \nu, A_0)$

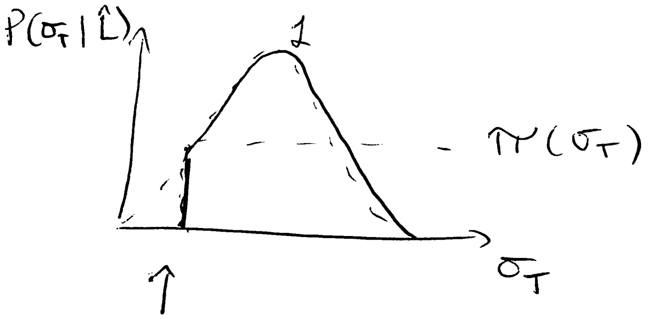


$\pi(\nu)$ $\pi(A_0)$

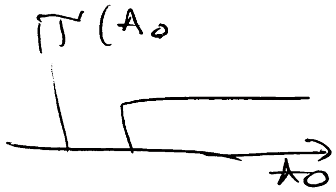
hyperparameters: part of the model or distribution that has some uncertainty.

$$\iint P(A)\,P(T)\,\pi(\sigma_T)\,G(\hat{L} \mid L)\,\underbrace{\pi(\nu)\,\pi(A_0)}\,dA\,dT\,d\nu\,dA_0$$

Priors

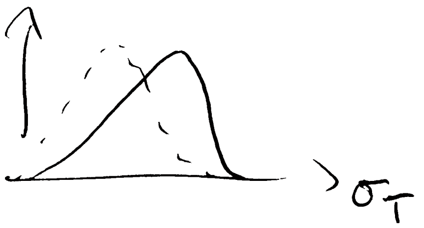


prior is visible!



prior $\to$ $\pi(A_0)$

$P(\sigma_T \mid \hat{L})$



$$P(\sigma_T \mid \hat{L}) \propto \iint \underbrace{\delta_D(L - \sigma_T A T^4)}_{\text{theory}} \underbrace{\pi(\sigma_T)\Gamma(A)G(T)}_{\text{priors}} \underbrace{G(\hat{L}|L)}_{\text{measurement}} \, dA\, dT$$

① $T \sim G(T|T_0)$

② $A \sim \Gamma(A)$

③ $\sigma_T \sim \pi(\sigma_T)$

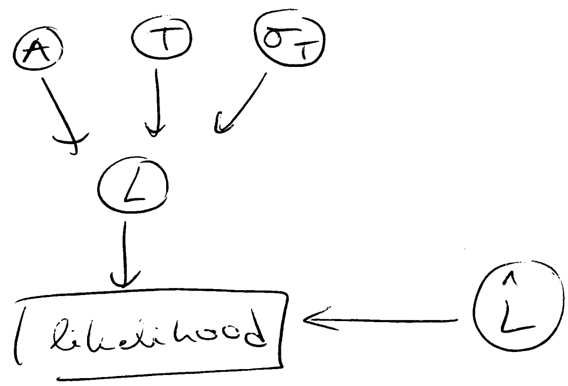
④ take σ_T, A, T samples, put them in δ_D

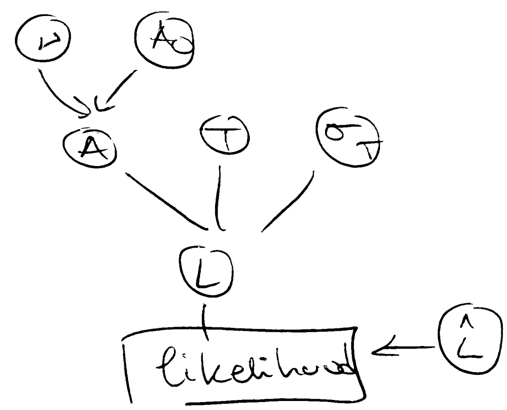
$L = \sigma_T A T^4$

⑤ take L sample and use to compute

$G(\hat{L}|L) \longrightarrow$ probability

⑥ accept - reject step for sample





n	A	T	σ_T
0	A_0	T_0	σ_{T_0}
1	A_1	T_1	σ_{T_1}
⋮	⋮	⋮	⋮

Example: WL

shear maps (noisy)

want. $P(\underset{\text{cosmic param.}}{\theta} \mid \hat{map})$

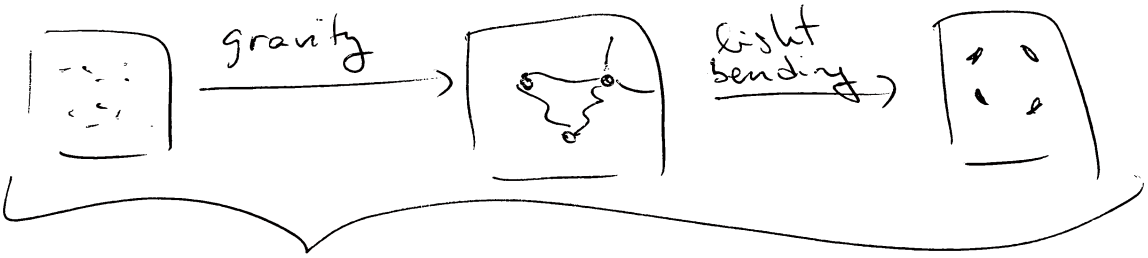
Collect random variables:

- θ (uncertain)
- map (initial fluct. δ)
- $\hat{map}$ (noisy)

$$P(\theta \mid \hat{map}) = \int P(\theta, map \mid \hat{map})\, d(map)$$

- $P(\theta, map \mid \hat{map}) \propto \underbrace{P(\hat{map} \mid map)}_{\substack{\text{detector}\\ \text{noise}\\ G}} P(map, \theta)$

- $P(map, \theta) = \underbrace{P(map \mid \theta)}_{\text{theory}} \underbrace{P(\theta)}_{\substack{\pi(\theta)\\ \text{prior}}}$

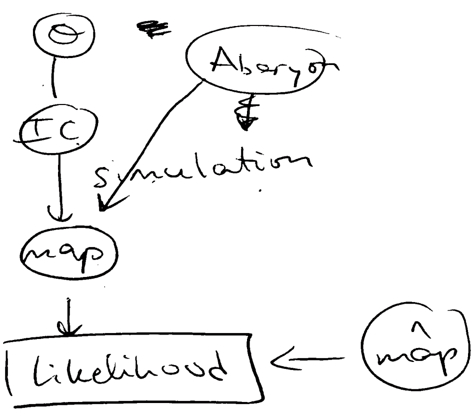


theoretical model
deterministic

$$\delta_D\left[IC - M(IC)\right]$$

$$P(\theta \mid \hat{map}) = \int \underbrace{P(\hat{map} \mid map)}_{\text{detector}} \underbrace{P(map \mid IC, \theta)}_{\text{simulator}} \underbrace{P(IC \mid \theta)}_{\text{theory}} \underbrace{P(\theta)}_{\text{prior}}\, dm$$

1. $\theta \sim \pi(\theta)$
2. $IC \sim G(IC \mid \theta)$
3. take θ, IC samples and put them in the S_D $\rightarrow$ gravity model (simulation) $\rightarrow$ map.
4. take map sample and compute $P(\hat{map} \mid map) \rightarrow$ prob.
5. accept-reject step.

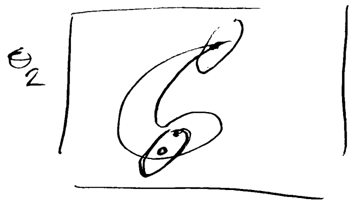


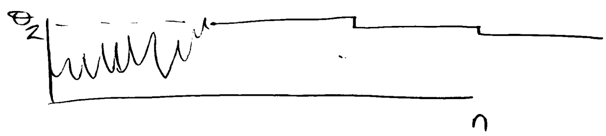
- rejection sampling. $\underline{n < 4}$

difficult to explore posterior randomly.

- MCMC $x_0 \curvearrowright x_1 \curvearrowright x_2$

  Metropolis - Hastings $\underline{n < 20}$





tunning of proposal distrib.

More efficient samplers: $\underline{10^6 \text{ param}}$

- Gibbs sampling : need all conditional distribu.
- HMC : need derivatives of posterior.

Gibbs sampling

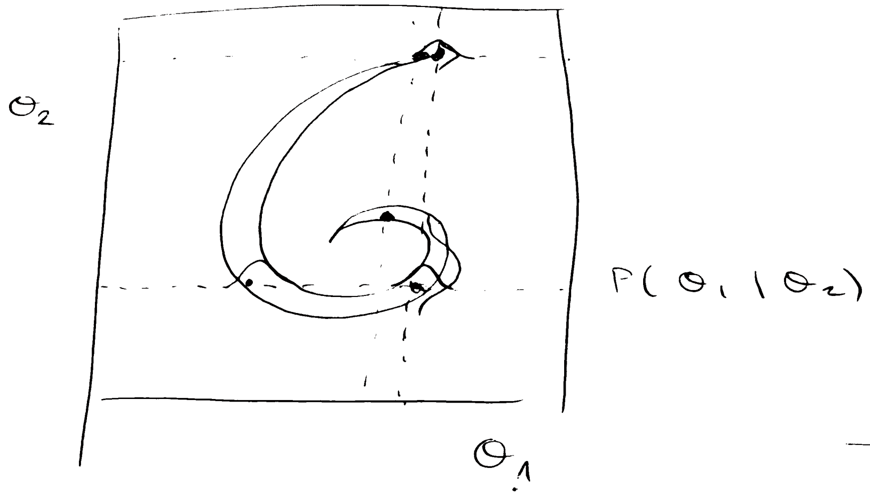
$$\vec{\theta} = (\underbrace{\theta_1, \theta_2, \theta_3, \theta_4}, \underbrace{\theta_5, \theta_6, \theta_7, \theta_8})$$

$$\vec{\theta} = \begin{pmatrix} \vec{\theta}_d = \vec{f} \\ \vec{\theta}_{8-d} = \vec{r} \end{pmatrix} \begin{matrix} \text{fixed} \\ \text{random} \end{matrix}$$

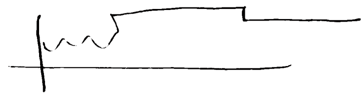
$$\vec{\theta}_d \sim P(\vec{\theta}_d \mid \vec{\theta}_{8-d})$$

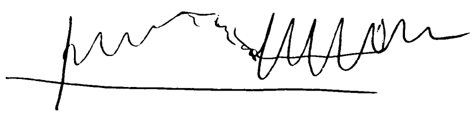
$$\vec{\theta}_{8-d} \sim P(\vec{\theta}_{8-d} \mid \vec{\theta}_d)$$

Gibbs:



MH





Hamiltonian Monte Carlo

use physics to solve statist. problem.

- build a potential from $P(x)$

$$\Psi = -\ln P(x)$$

- define kinetic energy

$$k = p_i^T M^{-1} p_i$$

- define Hamiltonian

$$H = k + \Psi$$

- evolve dynamical syst.

$$\dot{x} = \frac{\partial H}{\partial p} = M^{-1} p$$

$$\dot{p} = -\frac{\partial H}{\partial x} = -\frac{\partial \Psi}{\partial x}$$

- solve eq. of motion : Leapfrog algorithm

$$p_i\left(t+\frac{\varepsilon}{2}\right) = p_i(t) - \frac{\varepsilon}{2}\left(\frac{\partial p}{\partial x}\right)_{x(t)}$$

$$x_i(t+\varepsilon) = x_i(t) + \varepsilon\, p_i\left(t+\frac{\varepsilon}{2}\right)$$

$$p_i\left(t+\frac{\varepsilon}{2}\right) = p_i\left(t+\frac{\varepsilon}{2}\right) - \frac{\varepsilon}{2}\left(\frac{\partial p}{\partial x}\right)_{x(t+\varepsilon)}$$

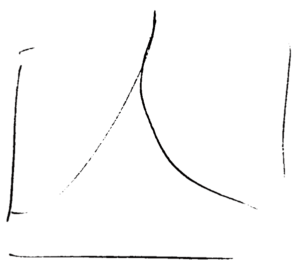
$$(x_i, p_i) \longrightarrow (x_{i+1}, p_{i+1})$$

$$2N \begin{cases} x \\ p \end{cases}$$

- H is conserved $\rightarrow$ acceptance is 100%

MH: $\min\left[1, \exp(-H^{i} + H^{i+\Delta})\right]$

it uses gradients to guide sampler to the interesting region.



HMC pseudo code:

for $i = 0$ to N_{MCMC}:

if $i = 0$ $\rightarrow$ choose starting point x_i

else $\rightarrow$ use current point

Draw random velocity $p_i \sim G(0, \mathbb{I})$

Leapfrog loop: use x_i and p_i as inial cond.

for $j = 0$ to N_L:

$(x_j, p_j) \rightarrow (x_{j+1}, p_{j+1})$

$(x_{NL}, p_{NL}) = (x_{j+1}, p_{j+1})$

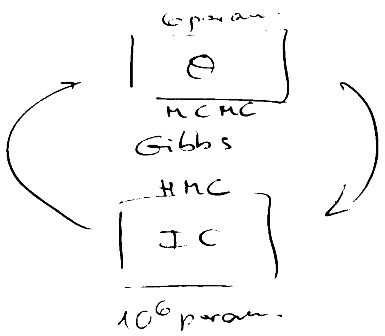
MH $\left\{ \begin{array}{l} \text{draw } \alpha \sim U(0,1) \text{ ; } R = \exp\left[-H_i + H_{NL}\right] \\ \text{if } \alpha > R \quad x_{NL} \text{ is rejected } \quad x_{i+1} = x_i \\ \text{if } \alpha < R \quad x_i = x_{NL} \text{ accept.} \end{array} \right.$

Parameters to tune:

$\left.\begin{array}{l} \varepsilon \text{ : timestep} \\ N_L \text{ : number of timesteps} \end{array}\right] \rightarrow$ acceptance rate.

M: mass matrix $\rightarrow$ efficientcy $\frac{1}{\mathrm{cor}(x)}$

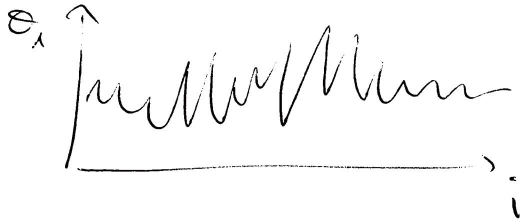
(updated on-the fly)

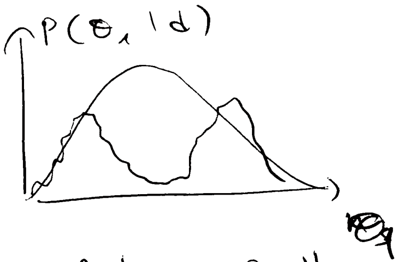


Tests and diagnostics for sampling

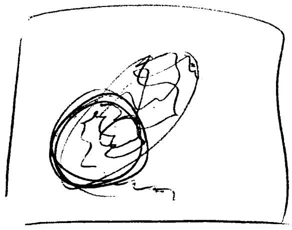
① numerical

② human (post processing of samples)

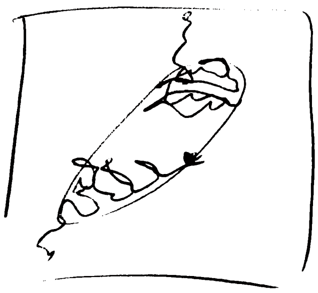




correlation length



resuming chains



only combine chains with different x_i when they are independently converged.