

6 Markov Chain Monte Carlo

❖ basics about Markov chains

- consider a system with n (discrete) states
- the system evolves in discrete time steps
- the system has no memory; subsequent steps are independent
- at time t_n the system is with probabilities $\pi_j(t_n)$ in state j
- transition probabilities $P_{kj} = p(j \rightarrow k)$ determine one time step
- conservation of probability requires

$$\sum_k P_{kj} = 1$$

- evolution of probabilities per step

$$\pi_k(t_{n+1}) = \sum_j P_{kj} \pi_j(t_n)$$

Chapman-Kolmogorov equation



source: wikipedia

A.A. Markov

❖ example: weather forecast

states 1 = rainy 2 = sunny 3 = cloudy with matrix P of transition probabilities:

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} = \begin{pmatrix} 0.50 & 0.50 & 0.25 \\ 0.25 & 0.00 & 0.25 \\ 0.25 & 0.50 & 0.50 \end{pmatrix} \quad \text{with } p_{ij} = p_{\text{tomorrow, today}}$$

The columns of P are normalised. Check the evolution of probabilities, starting “sunny”:

$$\begin{aligned} \pi^T(t_0) &= (0, 1, 0) \\ \pi(t_1) &= P\pi(t_0) & \pi^T(t_1) &= (1, 0, 1)/2 \\ \pi(t_2) &= P\pi(t_1) & \pi^T(t_2) &= (3, 2, 3)/8 \\ \pi(t_3) &= P\pi(t_2) & \pi^T(t_3) &= (13, 6, 13)/32 \\ \pi(t_4) &= P\pi(t_3) & \pi^T(t_4) &= (51, 26, 51)/128 \\ \text{asymptotic} & & \pi^T(t_\infty) &= (0.4, 0.2, 0.4) \end{aligned}$$

❖ properties of a Markov chain

- initial conditions do not matter
- the transition probabilities define a stable **equilibrium distribution** π_j^*
- equilibrium means **detailed balance**: forward and backward steps with equal rates

$$p(j \rightarrow k)\pi_j^* = p(k \rightarrow j)\pi_k^* \quad \text{or} \quad P_{kj}\pi_j^* = P_{jk}\pi_k^*$$

- it follows that the elements of π^* don't change in a time step:

$$(P\pi^*)_j = \sum_k P_{jk}\pi_k^* = \sum_k P_{kj}\pi_j^* = \pi_j^* \sum_k P_{kj} = \pi_j^*$$

- π^* is the eigenvector of P with the largest eigenvalue
- the probabilities are constant, and subsequent steps sample the states with the probabilities of the equilibrium distribution π^*

sampling arbitrary densities by Markov Chain Monte Carlo (MCMC) →

MCMC: The Metropolis-Hastings algorithm to sample a density ρ

construct transition probabilities such that ρ is the equilibrium distribution of a Markov chain

- x, y : points in configuration space
- $P(y|x)$: probability for the transition $x \rightarrow y$
- ρ : PDF of the equilibrium distribution
- core of the algorithm: decomposition of $P(y|x)$
 - ▶ an arbitrary random step $q(y|x)$ from $x \rightarrow y$, and
 - ▶ the probability $\alpha(x, y)$ to accept this step

$$P(y|x) = q(y|x) \cdot \alpha(x, y) = q(y|x) \cdot \min \left[1, \frac{q(x|y)\rho(y)}{q(y|x)\rho(x)} \right]$$

- ▶ re-use the old point if the step is not accepted

show that $P(y|x)$ satisfies detailed balance for $\rho \rightarrow$

→ check the condition $P(y|x) \rho(x) = P(x|y) \rho(y)$ for all possible cases, i.e:

$$q(y|x)\rho(x) \min \left[1, \frac{q(x|y)\rho(y)}{q(y|x)\rho(x)} \right] = q(x|y)\rho(y) \min \left[1, \frac{q(y|x)\rho(x)}{q(x|y)\rho(y)} \right]$$

(1) $q(y|x)\rho(x) = q(x|y)\rho(y)$: OK - trivially satisfied

(2) $q(y|x)\rho(x) < q(x|y)\rho(y)$: OK

$$q(y|x)\rho(x) [1] = q(x|y)\rho(y) \left[\frac{q(y|x)\rho(x)}{q(x|y)\rho(y)} \right]$$

(3) $q(y|x)\rho(x) > q(x|y)\rho(y)$: OK

$$q(y|x)\rho(x) \left[\frac{q(x|y)\rho(y)}{q(y|x)\rho(x)} \right] = q(x|y)\rho(y) [1]$$

discussion →

- the Metropolis-Hastings algorithm is independent of the dimension of $\rho(x)$
- normalization of $\rho(x)$ is not needed
- arbitrary jump functions $q(y|x) \neq q(x|y)$ are allowed (directional steps)
- special case $q(y|x) = q(x|y)$: Metropolis algorithm

$$\alpha(x, y) = \min \left[1, \frac{\rho(y)}{\rho(x)} \right]$$

- proper PDF sampling requires sufficiently long running time (“burn in phase”)
 - ▶ for a given ρ “sufficiently long” depends on the jump-function
 - small steps $\rightarrow$ slow diffusion
 - large steps $\rightarrow$ danger to be trapped at sharp maxima
- MCMC has correlations between subsequent points which are absent in standard MC

applications $\rightarrow$

MCMC sampling of 1-dim PDFs

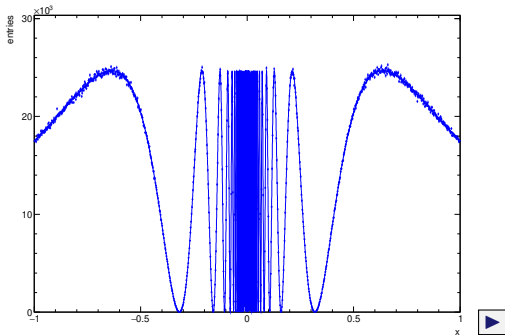
■ different PDFs in $-a < x < a$

- ▶ gaussian: $\exp(-x^2/2)$
- ▶ exponential: $\exp(-x), x > 0$
- ▶ singular density: $1/\sqrt{x}, x > 0$
- ▶ oscillating density: $\sin^2(1/x)$

■ uniform jump function $q(y|x)$

$$y = [x + \text{unif}(-s, s)] \bmod a$$

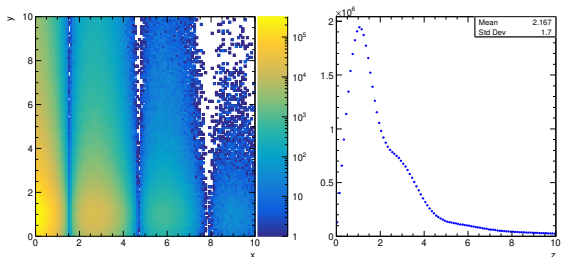
- ▶ jump range $0 < s < a$
- ▶ identify $-a$ and a



problems and exercises

Determine the distribution of $z = \sqrt{x^2 + y^2}$ when $x > 0$ is distributed according to $f(x) = \cos^2(x) e^{-x}$ and $y > 0$ according to $g(y) = 1/(1 + (y - 1)^2)$, i.e. use MCMC to calculate

$$h(z) = \int_0^\infty dx \int_0^\infty dy \frac{\cos^2(x) e^{-x}}{1 + (y - 1)^2} \delta(z - \sqrt{x^2 + y^2})$$



❖ MCMC sampling of bayesian posterior or the likelihood function

For independent measurements y_i , $i = 1, \dots, n$, with likelihood functions $L(a|y_i)$ and prior $p(a)$ for the parameter vector a , the bayesian posterior $p(a|y)$ is:

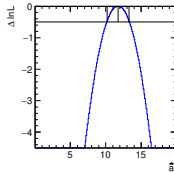
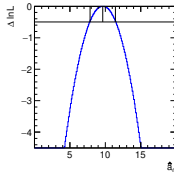
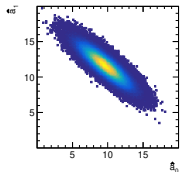
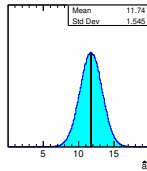
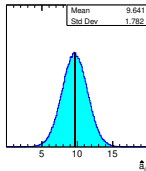
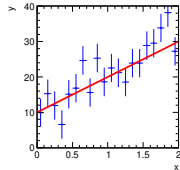
$$p(a|y) \propto p(a) \prod_{i=1}^n L(a|y_i)$$

- ▶ for a uniform prior the same functions are sampled
- ▶ for a linear fit the Jeffreys prior is uniform
- use of sampled information
 - ▶ bayesian: determine the distribution of a
 - ▶ frequentist: keep track of the maximum of the Likelihood as a function of a

❖ example: MCMC straight-line fit with uniform prior of $\langle y \rangle = a_0 + a_1 x$

■ 20 equidistant points in the range $x \in [0, 2]$ with $a_0 = a_1 = 10$

■ parameter errors $\sigma(\hat{a}_0) \approx 1.79$, $\sigma(\hat{a}_1) \approx 1.55$ from gaussian fluctuations with $\text{rms} = 4$



Summary on Markov Chain Monte Carlo

- elegant method to sample arbitrary PDFs
 - ▶ straightforward implementation
 - ▶ normalisation not required
- standard implementation: Metropolis(-Hastings) algorithm
 - ▶ performance depends on the choice of the jump-function
 - ▶ results are invariant when the running time is sufficiently long
 - ▶ a priori not obvious how to find out what qualifies for “sufficiently long”
- frequently used in Bayesian statistics
 - ▶ parameter estimate by marginalization of posterior distributions
 - ▶ transformation of variables
- sampling of the likelihood by MCMC also applicable for frequentist fitting