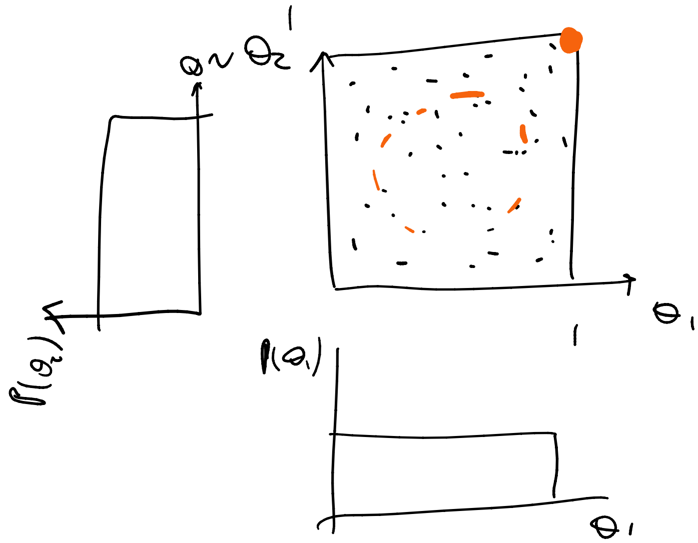
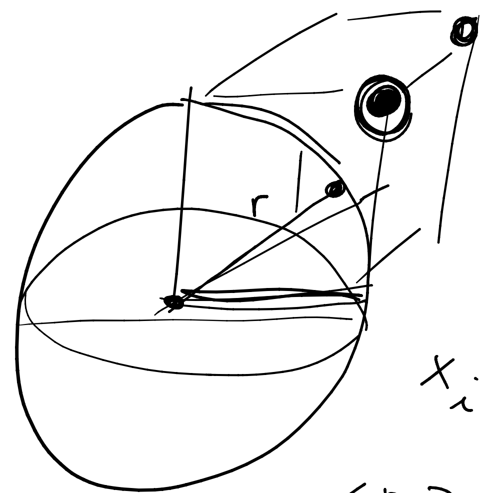
Lecture 3

Thursday, July 7, 2022 9:06 AM

Sensitivity analysis: change priors, see how your conclusions change! Recommended!





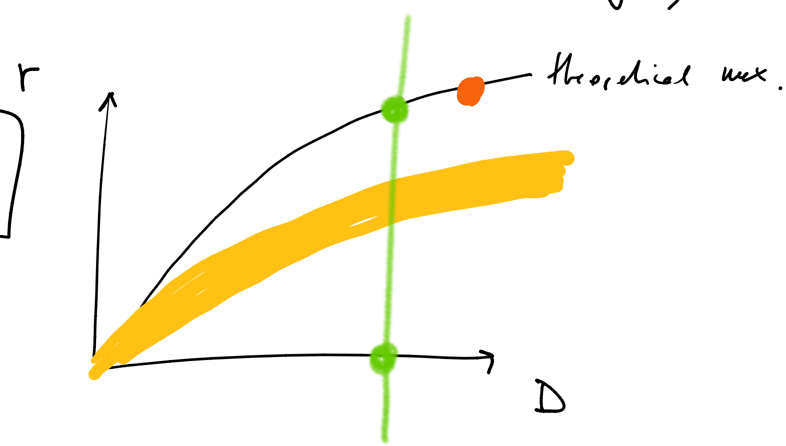
$$r = \sum_{i=1}^{D} x_i^2$$

$$x_i \sim U(0,1)$$

$$\langle r \rangle = \sqrt{D/3}$$

$\boxed{\text{Curse of dimensionality}}$

$(D \gtrsim 10)$



MCMC – <u>Markov Chain Monte Carlo</u>

$$p(\theta|d) \propto \mathcal{L}(\theta)\, p(\theta)$$

A Markov Chain is a sequence of random variables $\{X_0, \dots, X_n\}$ such that the prob. of element $t+1$ only depends on the value of X_t.

For $t \to \infty$, MC to converge to a stationary state with stationary distribution π

$$\pi(A) = \int_{\Omega} \pi(dy)\, \underbrace{p(y, A)}\, dy \qquad \forall A \subseteq \Omega$$

is the transition prob. for moving from y to $dy \subseteq A$

The MC being <u>irreducible</u> (any point in Ω can be reached from any point) and <u>aperiodic</u>

<u>Asymptotic convergence theorem</u>:

Let $p(x, A)$ be a transition prob from x to $A \subseteq \Omega$ for an irreducible, aperiodic MC w. stationary distribution π; Then for almost every point $x \subset \Omega$:

$$\lim_{n \to \infty} \sup_{A \subseteq \Omega} \left| p^n(x, \cdot) - \pi(\cdot) \right| = 0 \quad \forall A \text{ measurable}$$

Sufficient condition is for the chain to satisfy <u>detailed balance</u>:

"Jumping rule": $T(\theta^{(t)}; \theta^{(t+1)})$:

$$\boxed{\underbrace{p(\theta^{(t)} \mid d)}\, T(\theta^{(t)}; \theta^{(t+1)}) = p(\theta^{(t+1)} \mid d) \times T(\theta^{(t+1)}; \theta^{(t)})}$$

$$\underbrace{\qquad\qquad}_{\text{target distribution}} \qquad \times\, T\left(\theta^{(t+1)};\theta^{(t)}\right)$$

(this is a <u>reversible</u> chain).

<u>Metropolis-Hastings MCMC</u> :

(1) Initialize $t=0$; $\theta^{(0)} \sim p(\theta)$

distributed in this variable

(2) The next candidate point : $\theta^{(t+1)} \sim Q\left(\cdot \mid \theta^{(t)}\right)$

(3) Accept $\theta^{(t+1)}$ with probability

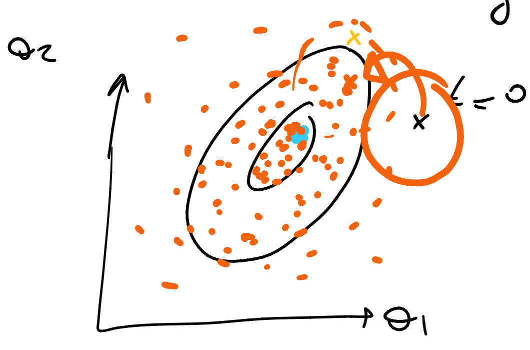
ratio

$$\min\left[1,\ \boxed{\frac{p\left(\theta^{(t+1)}\mid d\right) Q\left(\theta^{(t)}\mid\theta^{(t+1)}\right)}{p\left(\theta^{(t)}\mid d\right) Q\left(\theta^{(t+1)}\mid\theta^{(t)}\right)}}\ \right]$$

accept/reject

(4) Go back to (2).

e.g. $Q\left(\cdot\mid\theta^{(t)}\right) = N\left(\theta^{(t)},\Sigma\right)$



<u>MH-MCMC</u> :

$$T\left(\theta^{(t+1)};\theta^{(t)}\right) = \min\left[1,\ \text{Ratio}\right] Q\left(\theta^{(t+1)}\mid\theta^{(t)}\right)$$

$$= \frac{1}{p(\theta^{(t)}|d)} \cdot \min\Big[p(\theta^{(t)}|d)\, Q(\theta^{(t+1)}|\theta^{(t)}),\; p(\theta^{(t+1)}|d)\, Q(\theta^{(t)}|\theta^{(t+1)}) \Big]$$

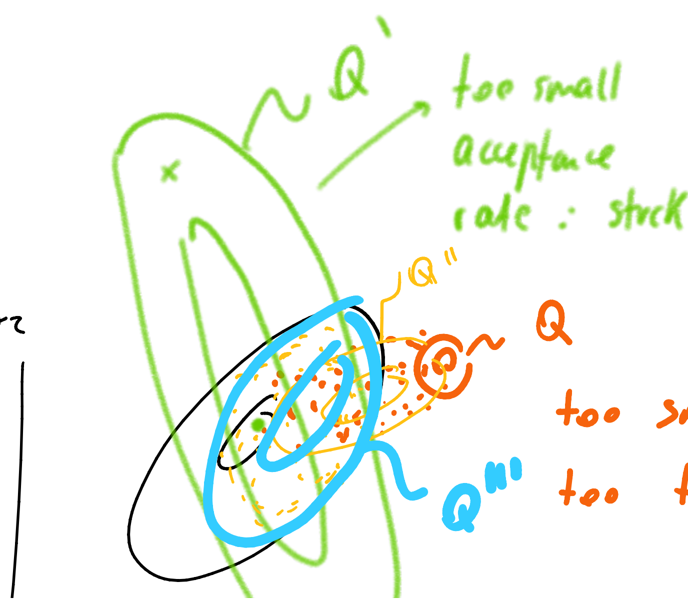
$$\Rightarrow \frac{T(\theta^{(t)};\theta^{(t+1)})}{T(\theta^{(t+1)};\theta^{(t)})} = \frac{p(\theta^{(t+1)}|d)}{p(\theta^{(t)}|d)}$$

$\Rightarrow$ detailed balance!

<u>Notice</u>: in the accept/reject step

$$\frac{p(\theta^{(t+1)}|d)}{p(\theta^{(t)}|d)} = \underbrace{\frac{\mathcal{L}(\theta^{(t+1)})\, p(\theta^{(t+1)})}{\mathcal{L}(\theta^{(t)})\, p(\theta^{(t)})}}_{\text{use logs in practice}}$$

<u>Practice</u>: θ_2

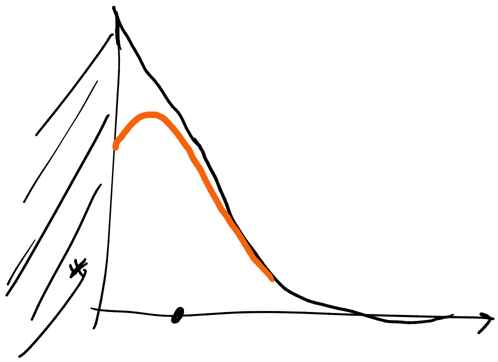


too small $\rightarrow$ diffusion

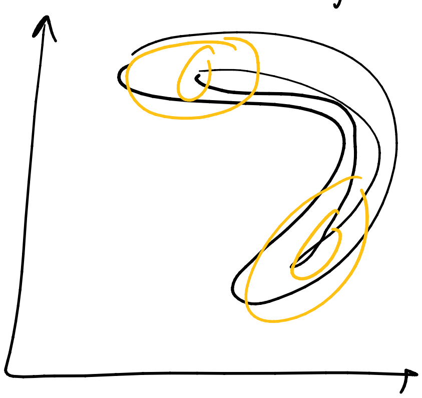
too large acceptance rate

$\downarrow$

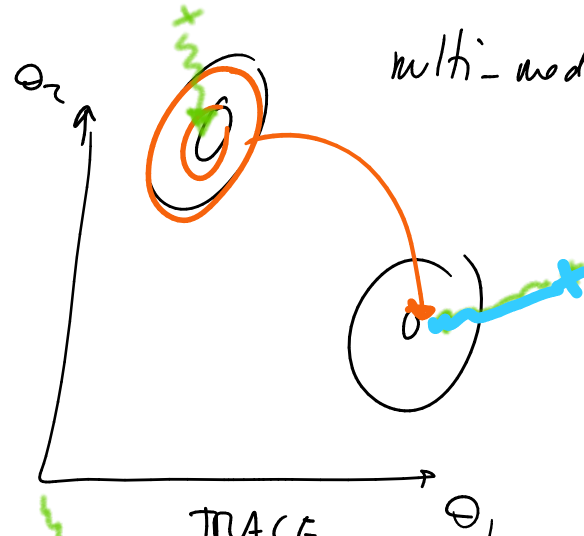
$$\frac{\#\ \text{accepted steps}}{\#\ \text{of candidate proposals}}$$

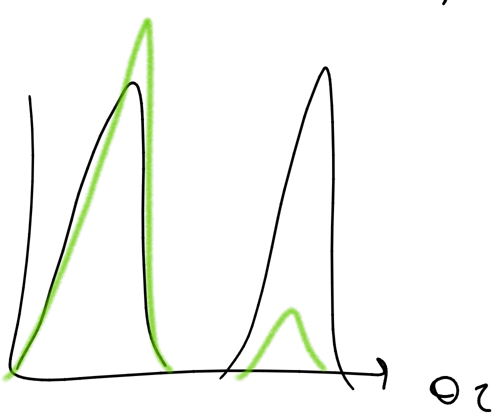


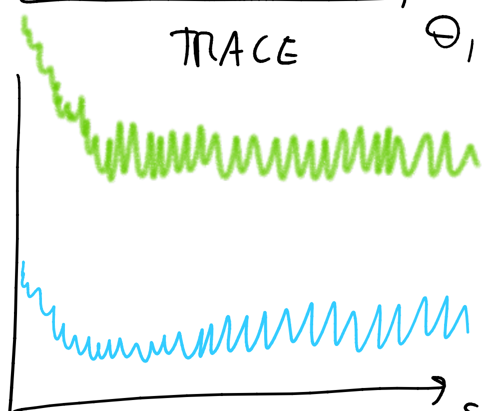
degeneracies



multi-modal

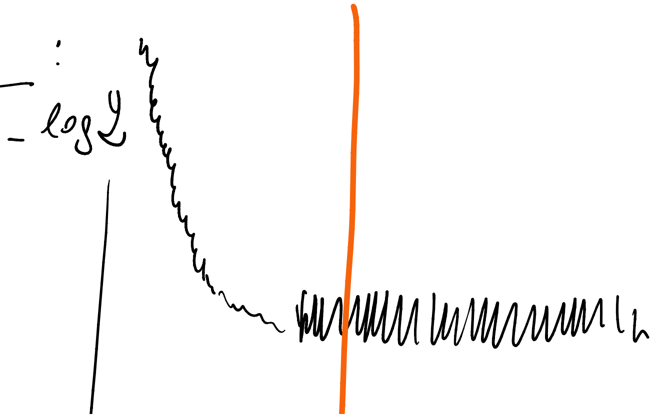


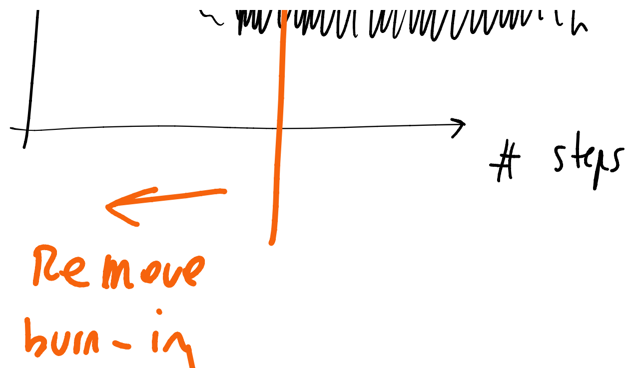


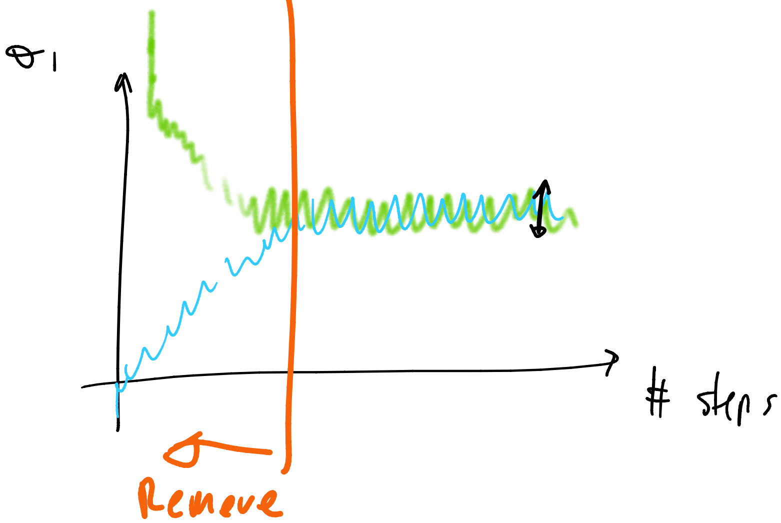


Watch out for:

BURN-IN:







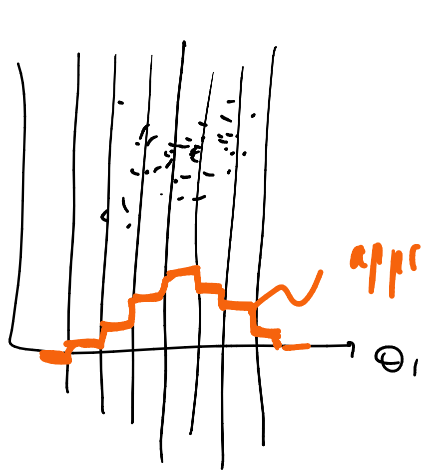
Convergence:

assess using R-number (Gelman & Rubin 1992)

$$\langle \theta \rangle = \int p(\theta|d)\,\theta\, d\theta \simeq \frac{1}{N}\sum_{i=1}^{N} \theta^{(i)}$$

$$\theta^{(i)} \sim p(\theta|d)$$

$$p(\theta_1|d) = \int d\theta_2\, p(\theta_1,\theta_2|d)$$



approx. to the 1-D marginal $p(\theta_1|d)$