Lecture 07 - More MCMC

How do we choose our transition/proposal distribution?

- We can choose anything that fulfills our 4 criteria: memoryless, irreducible, aperiodic, detailed balance
- A symmetric proposal distribution makes our life easier

⇒ Normal distribution is often a good choice.

⇒ σ then corresponds to our step size

small σ: takes long, samples precise

large σ: faster coverage, might miss details

In the long run we are fine either way though, as our MCMC is guaranteed to converge

Good rule of thumb: choose step size to have about 30% acceptance rate.

Metropolis-Hastings for high-dimensional spaces:

Two solutions:

1. Blockwise updating: Same as before, just now taking each step in the multidimensional space.
   => Can take very long to cover the whole space

2. Componentwise updating: Take steps only in one dimension at a time, accept or not, then repeat.
   => could end up doing the same dimension two times in a row if chosen randomly. In the long run this evens out.

Mixing and Convergence

Mixing: How fast do we cover the relevant space

Convergence: We are sampling from $p(x)$

Good mixing leads to faster convergence.

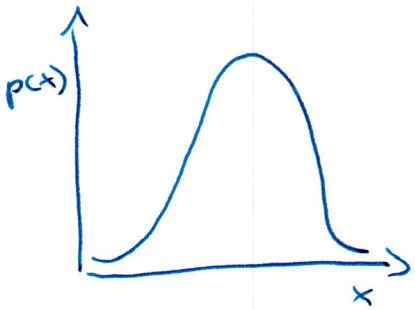
We know our sequence will converge as $n \to \infty$

We don't know how many iterations it will take.

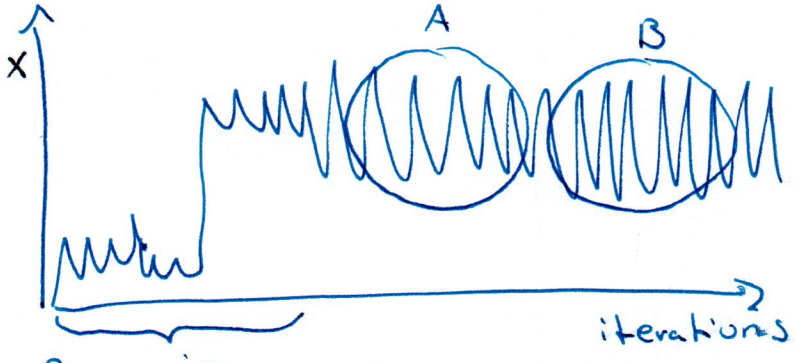
⇒ need to test for convergence.

There are formal and heuristic tests for convergence. Let's look at some heuristics to get more intuition.

Trace plot: Very useful for visual inspection



an easy target distribution



Burn in

After convergence subsamples A and B should be statistically the same (both ~p(x))

Inspecting the trace plot

↳ can show when we haven't converged yet.

↳ cannot show for sure we have converged.

possibly helpful visualizations:

- look at the whole plot
- divide into subsets (> 100 samples) and compare histograms
- start multiple chains from random start points and compare the trace plots.

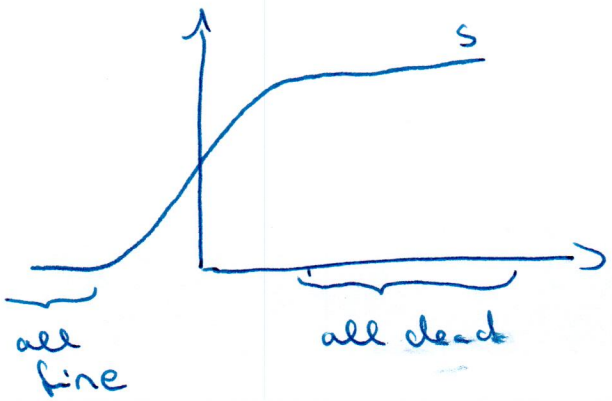
Example: Bioassay

We want to test some poison / pesticide.

⇒ Find the rate of death given the dose of the poison.

Likelihood: Once again Binomial ⇒ test n insects, get y deaths with unknown success / death rate θ.

How do we model θ?



s is the sigmoid:

$$\frac{\exp(x)}{1+\exp(x)} = \frac{1}{1+\exp(-x)}$$

We model θ as a sigmoid of the dose x_i, but we allow for a linear mapping of the dosage:

$$\theta = s(\alpha + \beta x_i)$$

α, β: parameter; x_i: dosage

θ now depends on x_i by means of α and β.

Prior: $p(\alpha, \beta) \propto$ uniform

Posterior: $p(\theta|D) = p(\alpha, \beta | x, n, k)$

$$\propto \prod_i s(\alpha+\beta x_i)^{k_i} \cdot (1 - s(\alpha+\beta x_i))^{n_i - k_i}$$

$\Rightarrow$ We have a two dimensional posterior and we can sample from it using MCMC!

Autocorrelation: How predictable is the future, given the past?

- MCMC gives us samples from $p(\cdot)$
- But, these samples are not iid!

Solution:

1. Reshuffle all samples
2. Thinning: Only take every n^{th} point.