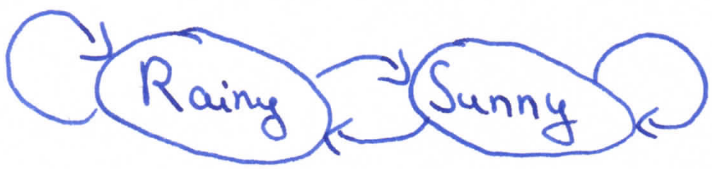
Lecture 07 - MCMC for Real

First a Bayesian warm up with the infectious disease example from Lecture 06 notes.

Now back to our Markov Chains.

Remember the Rainy - Sunny example:



This Markov Chain is irreducible and aperiodic
$\Rightarrow$ it has a unique stationary distribution.

For simplicity let's first assume all transition probabilitis are the same:

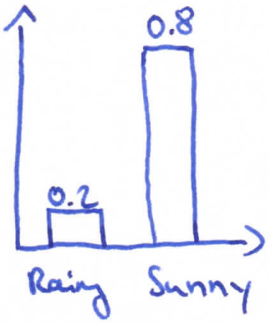
$$T = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

The stationary distribution is then $[0.5, 0.5] = s$

So a sampler would generate a chain with an equal proportion of Sunny and Rainy days in the long run.

Now comes the magic: We can modify this samples to generate any 2-state distribution we want!

Let's say we want:



we want
$\Rightarrow$ $S = [0.2, 0.8]$

So we need to guide our samples to spend more time in the sunny state.

Here is how we do it:

- current state: $x^{(i)}$
- ask sampler where it would like to so next $\Rightarrow$ proposal for new state $x^{(i+1)} = x^*$
- compute acceptance probability

$$a_{i*} = \frac{S[x^*]}{S[x^{(i)}]}$$

- Draw $u \sim \text{uniform}[0,1]$
- If $u < a_{i*} \Rightarrow x^{(i+1)} = x^*$

  Otherwise $\Rightarrow x^{(i+1)} = x^{(i)}$

Note, even if we reject we do get a sample!

Show example!

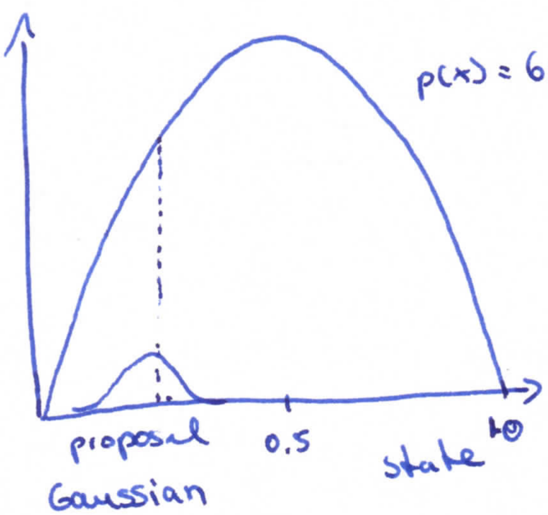
What if our initial transition matrix is not symmetric? $\Rightarrow$ We need to account for it when computing the acceptance probability

$$a_{i*} = \frac{S[x^*] \cdot T[x^* | x^{(i)}]}{S[x^{(i)}] \cdot T[x^{(i)} | x^*]}$$

Note that there is nothing here that limits us to apply this scheme to only 2 states. In fact we can apply this scheme to sample from continuous distributions!

Show example!

Intuition: What happens for Gaussians with different σ?



So let's see how and why this magic works:

We define: s^* : The stationary distribution we want

T^* : The transition matrix we want

T : The transition matrix we have

A : The acceptance probability

We assume that T is irreducible and aperiodic

For our target Markov Chain we want detailed balance:

$$s^*(x^{(i)}) \cdot T^*(x^{(i-1)} | x^{(i)}) = s^*(x^{(i-1)}) \cdot T^*(x^{(i)} | x^{(i-1)})$$

Now we rewrite T^* in terms of T and A:

$$T^*(x^{(i)} | x^{(i-1)}) = T(x^{(i)} | x^{(i-1)}) \cdot A(x^{(i)} | x^{(i-1)})$$

We haven't defined A yet, but we can see that it needs to make up for the difference between T^* and T, while leading to our desired stationary distibution.

So what we have now is:

$$s^*(x^{(i)}) \cdot T(x^{(i-1)}|x^{(i)}) \cdot A(x^{(i-1)}|x^{(i)})$$

$$= s^*(x^{(i-1)}) \cdot T(x^{(i)}|x^{(i-1)}) \cdot A(x^{(i)}|x^{(i-1)})$$

To make things a bit simpler I'm going to assume $T(x^{(i-1)}|x^{(i)}) = T(x^{(i)}|x^{(i-1)})$. This means that T is symmetric. This is often the case as we design T, but if not you can just put it back into the following lines.

With this simplification we have

$$s^*(x^{(i)}) \cdot A(x^{(i-1)}|x^{(i)}) = s^*(x^{(i-1)}) \cdot A(x^{(i)}|x^{(i-1)}).$$

Now remember that A is what transforms our T to T^* and intuitively we want to visit states where s^* is high more often, so T^* should propose states more often where s^* is high.

A possible first guess could be $A(x^{(i)}|x^{(i-1)}) = \dfrac{s^*(x^{(i)})}{s^*(x^{(i-1)})}$

this would let us accept more if $s^*(x^{(i)}) > s^*(x^{(i-1)})$. However, if we substitute this in the equation above we get

$$s^*(\cancel{x^{(i)}}) \cdot \frac{s^*(x^{(i-1)})}{s^*(\cancel{x^{(i)}})} = s^*(\cancel{x^{(i-1)}}) \cdot \frac{s^*(x^{(i)})}{s^*(\cancel{x^{(i-1)}})}$$

$$\Leftrightarrow s^*(x^{(i-1)}) = s^*(x^{(i)})$$

Now that would be a very boring s^*!

We don't want all these terms to cancel, but we do want to preserve our intuition of accepting more where s^* is high. It dawns on us that accepting with prob >1 doesn't do us any additional good, so let's try

$$A(x^{(i)} | x^{(i-1)}) = \min\left(1, \frac{s^*(x^{(i)})}{s^*(x^{(i-1)})}\right)$$

Now we have three different cases to look out for:

① $\frac{s^*(x^{(i)})}{s^*(x^{(i-1)})} = 1 \Leftrightarrow s^*(x^{(i)}) = s^*(x^{(i-1)})$

$\Leftrightarrow A(x^{(i)} | x^{(i-1)}) = A(x^{(i-1)} | x^{(i)}) = 1$

we end up with the same result as above when we put this in and every thing is consistent.

② $\frac{s^*(x^{(i)})}{s^*(x^{(i-1)})} > 1 \Leftrightarrow$ $s^*(x^{(i)}) > s^*(x^{(i-1)})$

$A(x^{(i)} | x^{(i-1)}) = 1$

$A(x^{(i-1)} | x^{(i)}) = \frac{s^*(x^{(i-1)})}{s^*(x^{(i)})} < 1$

Let's plug this in:

$$s^*(x^{(i)}) \cdot A(x^{(i-1)} | x^{(i)}) = s^*(x^{(i-1)}) \cdot A(x^{(i)} | x^{(i-1)})$$

$$\Leftrightarrow s^*\cancel{(x^{(i)})} \cdot \frac{s^*(x^{(i-1)})}{s^*\cancel{(x^{(i)})}} = s^*(x^{(i-1)}) \cdot 1$$

$$\Leftrightarrow \quad s^*(x^{(i-1)}) = s^*(x^{(i-1)})$$

This works! And case 3 works analog to case 2.

So by choosing our acceptance probability well, we have created a new Markov Chain that fulfills detailed balance ☺!

Some things to keep in mind:

- It's easier to choose T to be symmetric.
- Otherwise A becomes: $A(x^{(i)}|x^{(i-1)}) = \min\left(\frac{S^*(x^{(i)})}{S(x^{(i-1)})} \cdot \frac{T(x^{(i)}|x^{(i-1)})}{T(x^{(i-1)}|x^{(i)})}\right)$
- For A we only need the ratio $\frac{S^*(x^{(i)})}{S(x^{(i-1)})}$

  ⇒ S^* does not need to be normalized

  ⇒ we can sample from our posterior without having to compute the evidence!

- often T is chosen to be uniform or Gaussian in shape. The width of T is important to tune to give good mixing and convergence.