Sampling tutorial (2019)

Exercise 1: Building blocks of sampling algorithms ....

There exist (as mostly is the case) many solutions; the students are likely to come up with many versions of if-then statements, or versions of rejection sampling.
The crucial common denominator is that the *frequency* of drawn samples must be proportional to the wanted probability.
Possible solutions are:

- 2: Draw α ~Uniform(0,1). If $\alpha < 0.5$, $x = -1$. Else: $x = +1$.
- 3: the acceptance probability shall be p, and because it is a probability, we have $p \in [0, 1]$. So the algorithm could be: Draw x. Draw α ~Uniform(0,1). If $\alpha < p$, accept x.
- 4: The xs of (3) are now the ± 1 from (2).
- 5 and 6: it results in a usual random walk, where the step-width is always integer. For $p = 0.5$ it is symmetric, for other ps the walker drifts into one direction.
- 7: yes, it is Markovian, since it only depends on the last former value in the chain.

Exercise 2: MCMC of Supernova data; measuring Ω_m, w_0, w_a ....

You will get scatter plots with points, since writing algorithms to set confidence contours at the right levels takes a bit of time. The results here depict what it would look like with confidence contours. Your scatter plots will have the same shapes and the density will increase towards the center of the distributions.

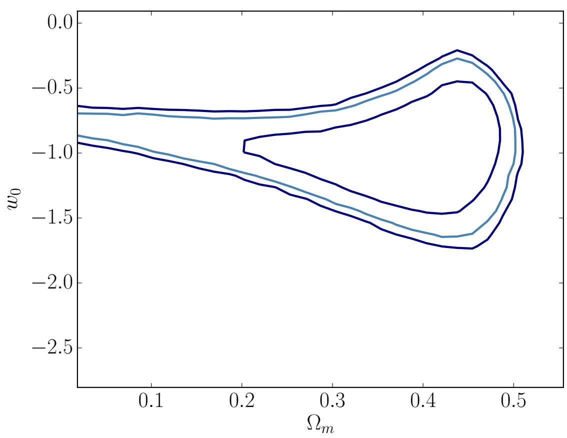

Fig. 1: Columns 0 and 1. Column 0 is Ω_m, column 1 is w_0 and column 2 is w_a. One can tell by looking at the axes and comparing to the values those parameters can theoretically take.

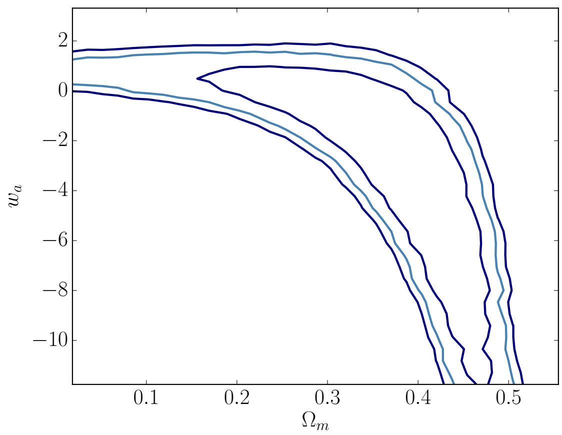

Fig. 2: Columns 0 and 2.

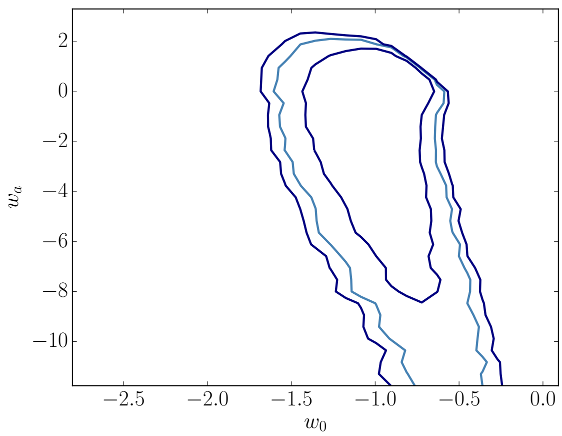

Fig. 3: Columns 1 and 2.

- 3: For three parameters, there are 3 1d marginals, and 3 2d marginals. For n parameters, there are n 1d marginals and $n(n-1)/2$ distinct 2d marginals.
- 5: Finding the maximum of the posterior. The chain given is one that stores auxilliary values at each parameter point. Not all chains do that, many chains store the parameters only. If one has a chain where only parameters are stored, then the maximum of the posterior lies where the density of samples is highest. The chain thus needs to be binned. The given chain does however store the log-likelihood. Since the priors are flat, the log-likelihood is proportional to the log-posterior. It is thus sufficient to find the row with the highest protocolled log-likelihood value.

Exercise 3: Uncleaned MC chain from a cosmic microwave background analysis ....

The weights arise from the accept/reject step: the weight of one point is counted up each time the sampler tried to leave this point, but its attempt was rejected since it landed in a region where the likelihood was too low. Without the weights, the sampler would not satisfy detailed balance. Additionally, the sampler would get stuck, once it found the peak.

The steep drop indicates the burn-in of the chain: the negative log-likelihood decreases. For a Gaussian likelihood (which is the case here), this simply indicates a decrease in χ^2. Once the sampler has reached

a region of acceptable χ^2, it begins to oscillate around the best-fitting solution. This then generates the plateau in the chain.
Removing the burn-in means 'throwing away all those early points which are not yet drawn from the oscillation around the best-fitting solution'.
When plotting two parameters of the chain against each other, one gets a cloudy distribution with holes. One can guess how the sampler walkes around, since the mixing in parameter space is bad. This is an indication of the chain not having converged yet. The jumping distance is still too small, because the sampler first needs to learn the approximate shape of the posterior. A typical user fault when running public MCMC codes will produce such cloudy chains, and it is hence important to be able to already tell from the plotted chains that a jumping distance is too low, and the code needs to be restarted with improved settings.

Exercise 4: Monte Carlo and the curse of dimensionality ....
The solution to this exercise is in the script/the slides.

Exercise 5: Gibbs sampling the straight line with dual errors (optional) ....
The resulting plots for this exercise should look like these.

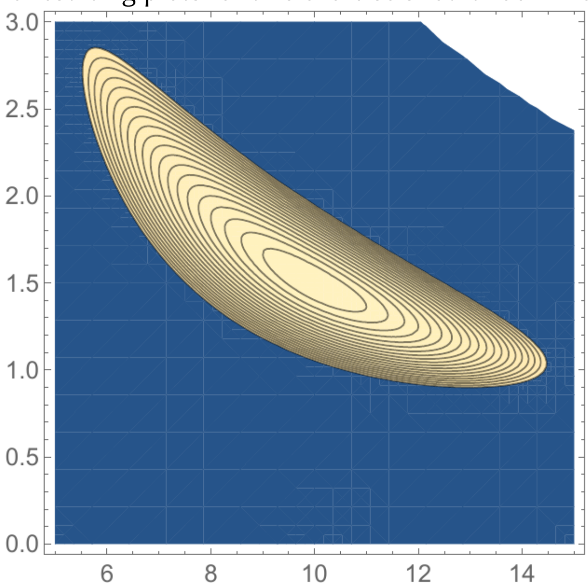

Fig. 4: Joint posterior of x and m, with indicative credibility contours.

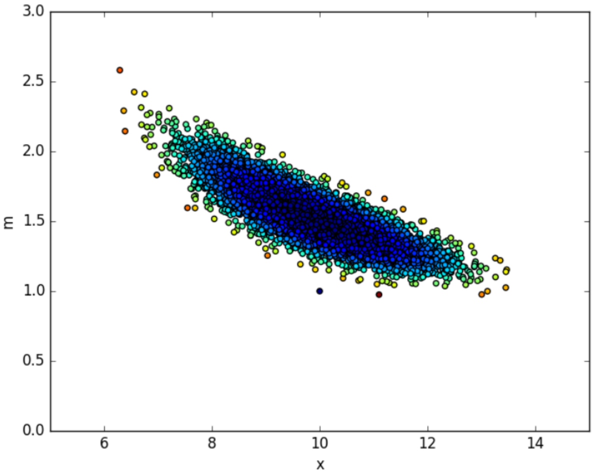

Fig. 5: Scatter plot of the joint x, m-posterior. The colour of the points indicates the posterior probability, with red being low probability and blue high probability.