Introduction to Bayesian Statistics - 9

Edoardo Milotti

Università di Trieste and INFN-Sezione di Trieste

Our next important topic: Bayesian estimates often require complex numerical integrals. How do we confront this problem?

enter the Monte Carlo methods!

1. acceptance-rejection sampling
2. importance sampling
3. statistical bootstrap
4. Bayesian methods in a sampling-resampling perspective
5. Introduction to Markov chains and to Random Walks (RW)
6. Detailed balance and Boltzmann's H-theorem
7. The Gibbs sampler
8. More on Gibbs sampling
9. Simulated annealing and the Traveling Salesman Problem (TSP)
10. The Metropolis algorithm
11. Image restoration and Markov Random Fields (MRF)
12. The Metropolis-Hastings algorithm and Markov Chain Monte Carlo (MCMC)
13. The efficiency of MCMC methods
14. Affine-invariant MCMC algorithms (emcee)

11. Image restoration and Markov Random Fields (MRF) (ctd.)

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. PAMI-6, NO. 6, NOVEMBER 1984 721

Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images

STUART GEMAN AND DONALD GEMAN

Abstract—We make an analogy between images and statistical mechanics systems. Pixel gray levels and the presence and orientation of edges are viewed as states of atoms or molecules in a lattice-like physical system. The assignment of an energy function in the physical system determines its Gibbs distribution. Because of the Gibbs distribution, Markov random field (MRF) equivalence, this assignment also determines an MRF image model. The energy function is a more convenient and natural mechanism for embodying picture attributes than are the local characteristics of the MRF. For a range of degradation mechanisms, including blurring, nonlinear deformations, and multiplicative or additive noise, the posterior distribution is an MRF with a structure akin to the image model. By the analogy, the posterior distribution defines another (imaginary) physical system. Gradual temperature reduction in the physical system isolates low energy states ("annealing"), or what is the same thing, the most probable states under the Gibbs distribution. The analogous operation under the posterior distribution yields the maximum *a posteriori* (MAP) estimate of the image given the degraded observations. The result is a highly parallel "relaxation" algorithm for MAP estimation. We establish convergence properties of the algorithm and we experiment with some simple pictures, for which good restorations are obtained at low signal-to-noise ratios.

The Ising model as an example of Markov Random Field

The model describes a system of spins that point only in the +*z* or -*z* direction, so that their value can only be ±1.

The Hamiltonian includes only the interaction with the external magnetic field and the interaction between neighboring spins

$$H = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - B \sum_i \sigma_i$$

The corresponding lattice magnetization is

$$M = \langle \sigma_i \rangle$$

a quantity that ranges between -1 and +1

The Bragg-Williams approximation

This is a simple mean–field approximation

$$\langle \sigma_i \sigma_j \rangle \approx \langle \sigma_i \rangle \langle \sigma_j \rangle$$

correlations are ignored

so that the Hamiltonian can be restated in terms of an effective magnetic field

$$H \approx -J \sum_i \sigma_i \sum_{\langle j \rangle_i} \sigma_j - B \sum_i \sigma_i \approx (-Jz\langle \sigma_i \rangle - B) \sum_i \sigma_i = -B_{\text{eff}} \sum_i \sigma_i$$

Then, the partition function is

$$Z = \sum_{\text{configurations}} \exp\left(\frac{B_{\text{eff}} \sum_i \sigma_i}{kT} \right) = \prod_i \left(e^{B_{\text{eff}}/kT} + e^{-B_{\text{eff}}/kT} \right)$$

$$= \left[2\cosh\left(B_{\text{eff}}/kT \right) \right]^N = \left[2\cosh\left(\beta B_{\text{eff}} \right) \right]^N$$

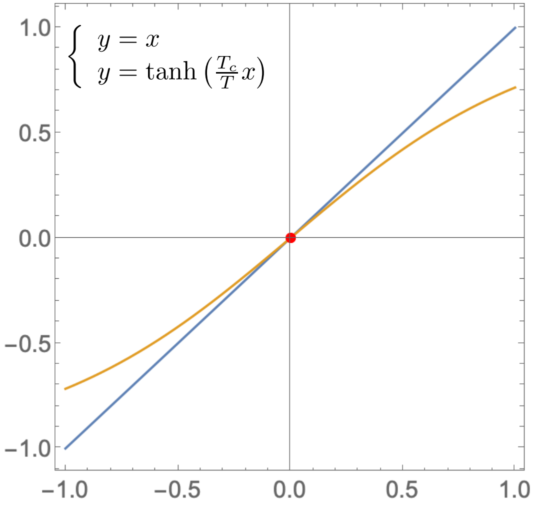
Therefore, the magnetization can be obtained as follows

$$
\begin{aligned}
M &= \frac{1}{NZ} \sum_{\text{configurations}} \sigma_i e^{\beta B_{\text{eff}} \sigma_i} = \frac{1}{NZ} kT \frac{\partial}{\partial B_{\text{eff}}} \sum_{\text{configurations}} e^{\beta B_{\text{eff}} \sigma_i} \\
&= kT \frac{\partial}{\partial B_{\text{eff}}} \ln Z = \tanh\left(\beta B_{\text{eff}}\right) \\
&= \tanh\left[\beta\left(B + JzM\right)\right]
\end{aligned}
$$

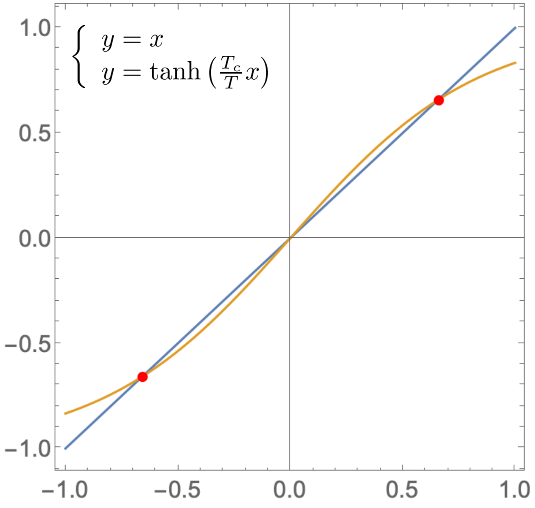
i.e., the magnetization is the solution of the nonlinear equation

with B field → $$M = \tanh\left[\beta\left(B + JzM\right)\right]$$

no B field → $$M = \tanh\left(\beta zJM\right)$$



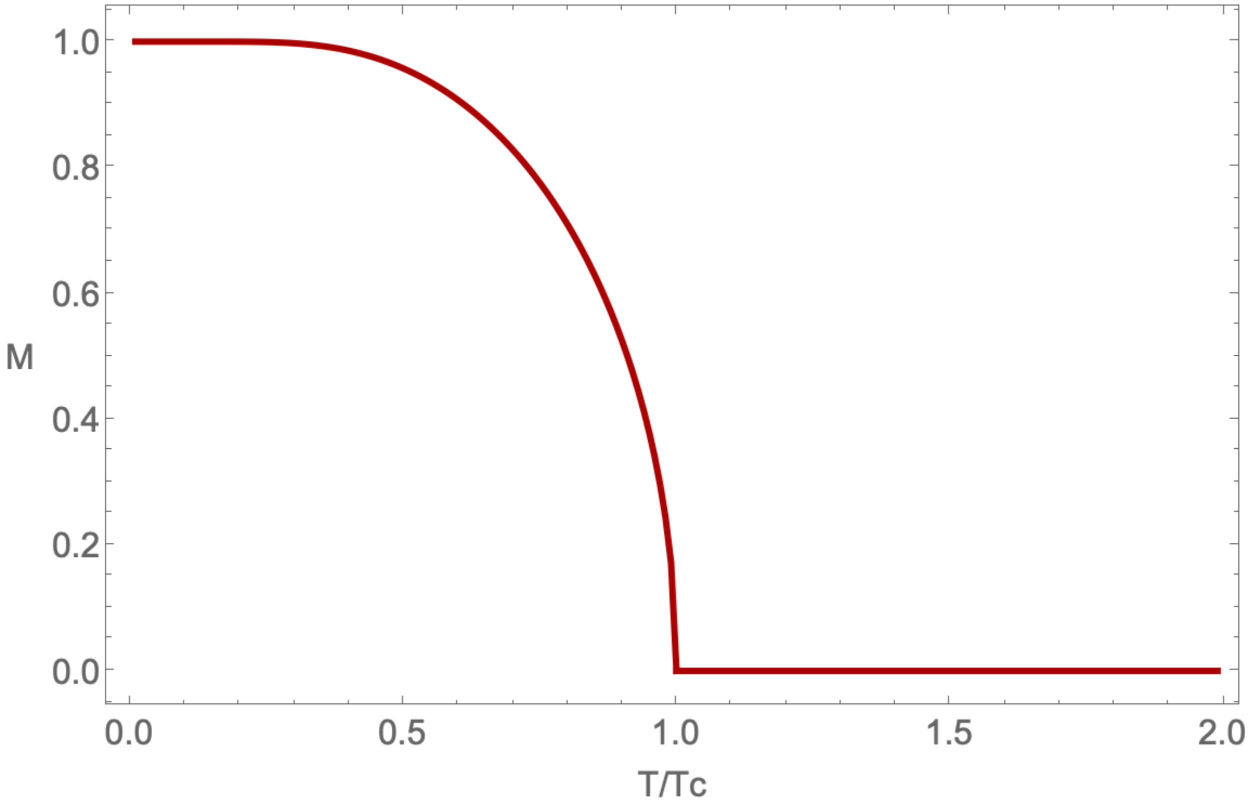
above critical temperature



below critical temperature

$$M = \tanh\left(\beta z J M\right)$$

Critical temperature $$T_c = \frac{zJ}{k}$$



STATISTICAL PHYSICS

90 years of the Ising model

Ernst Ising's analysis of the one-dimensional variant of his eponymous model (*Z. Phys* 31, 253–258; 1925) is an unusual paper in the history of early twentieth-century physics. Its central result — demonstrating that a linear chain of two-state spins cannot undergo a phase transition at finite temperature — is correct, if somewhat trivial compared with other physics breakthroughs published in the 1920s. But it is Ising's fateful extension of his conclusions to two and three dimensions that proved spectacularly wrong and, paradoxically, earned him an enduring association with the model that now bears his name.

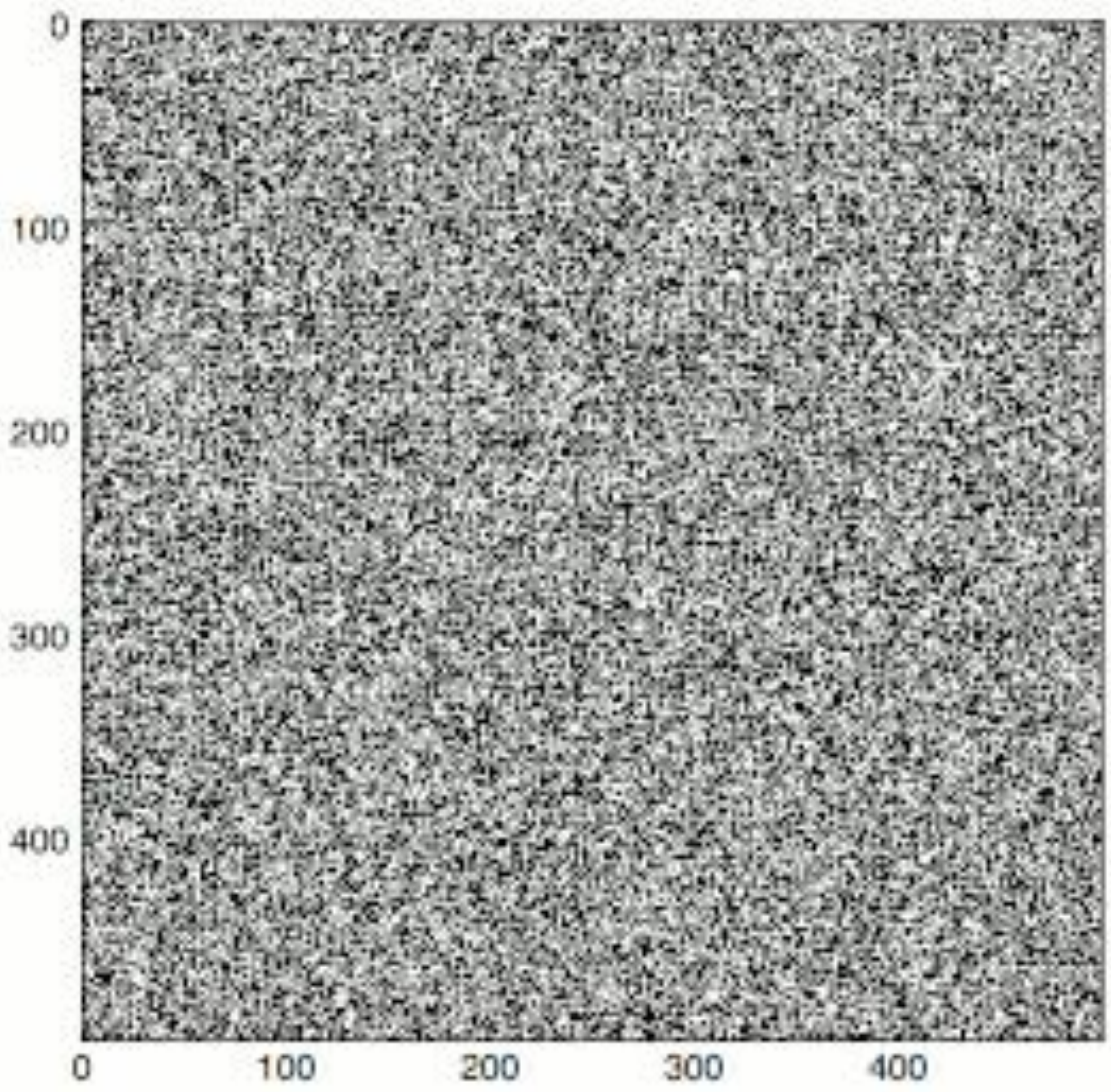
A possible reason for Ising's unexpected celebrity is that his erroneous conclusions betray a superficial understanding of what turned out to be some of the deepest and far-reaching problems to be addressed in twentieth-century physics. The Hamiltonian of the model is simple to write down — it describes a network of spins interacting with each other through a coupling that only applies if the spins are next to each other — but the physics it displays is rich and non-trivial: not only does it provide an intuitive device for illustrating the essential features of phase transitions and critical phenomena, it neatly encapsulates the main traits of the many-body problem that has come to dominate areas such as condensed-matter physics. The broader class of spin models it belongs to was used to uncover concepts such as universality, renormalization, symmetry-breaking and emergence. Ising can perhaps be forgiven for not predicting all of that.

Famously, the two-dimensional version for the model was solved analytically by Lars Onsager in the early 1940s (*Phys. Rev.* 65, 117; 1944), a result that is rightly considered a towering achievement among many significant contributions made over the years by the likes of Peierls, Bethe, Yang, Kadanoff (see page 995) Fisher and Wilson, just to name a handful. But the three-dimensional lattice has never been solved exactly, in spite of a multitude of attempts and false dawns — including a claim by John Maddox (who would later become the editor of *Nature*) made at a conference in Paris in 1952.

Although the 3D model is thought by some to be analytically intractable (and has also been claimed to belong to the NP-complete category of computational decision problems), progress has continued and recent numerical techniques based on conformal field theory have shed further light on the structure of the problem (*J. Stat. Phys.* 157, 869–914; 2014). Nevertheless, the real value of the Ising model and its many derivatives lies precisely in the complexity they encapsulate. These have found use in fields as disparate as condensed-matter physics, physical chemistry, neuroscience and, more broadly, the study of so-called complex systems.

Ising studied a deceptively simple model that, unknown to him at the time, captures the essential physics of an extremely wide category of problems. He may have been wrong in his 1925 paper, but he tripped over a veritable physics goldmine.

ANDREA TARONI



Quench of an Ising system on a two-dimensional square lattice (500 × 500) with inverse temperature $\beta = 10$, starting from a random configuration

(from https://en.wikipedia.org/wiki/Ising_model)

The Markov property of Markov Random Fields

The Ising model is an example of Markov Random Field (MRF). **What do we mean by "Markov property" in this case?**

This property corresponds to the locality of the spin interactions and the dependence of the probability of finding a certain spin with a given value only on the state of the neighboring spins.

$$P(X_s = x_s | X_r = x_r, r \in S, r \neq s) = P(X_s = x_s | X_r = x_r, r \in G_s)$$

graph node at *s*

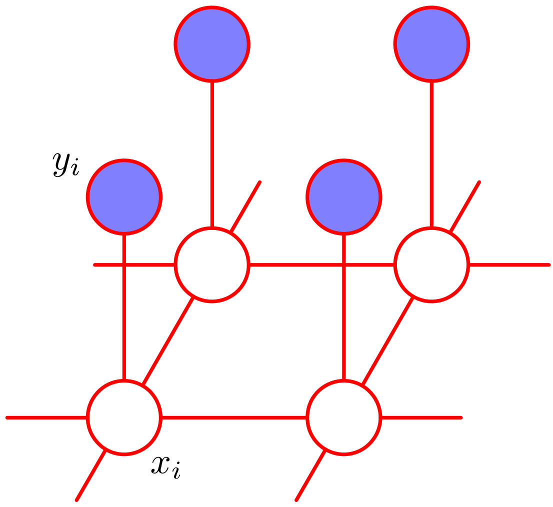
takes on this value

index *r* belongs to index set S

index *r* belongs to the set of neighbors of *s*

Choice of the *data* term

An undirected graphical model representing a Markov random field for image de-noising, in which x_i is a binary variable denoting the state of pixel i in the unknown noise-free image, and y_i denotes the corresponding value of pixel i in the observed noisy image.



The *data* term in the Hamiltonian must contain information on the background but also on the image data.

It would be natural to start with something like

$$H = -J \sum_{\langle i,j \rangle} x_i x_j - \sum_i y_i x_i$$

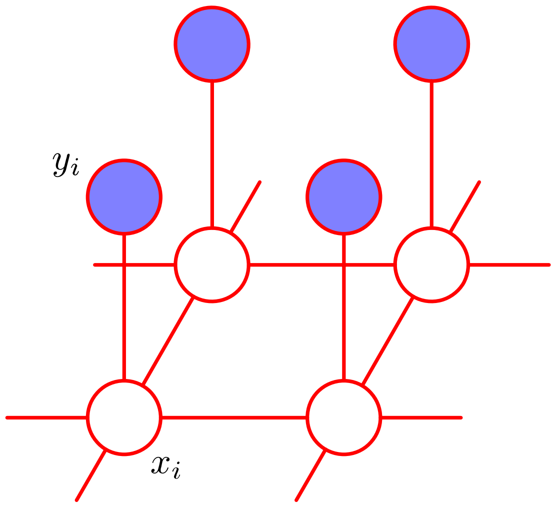
where the y represent the image data and act like a sort of local magnetic field. However, this does not take into account the natural fluctuations of the image data. Here we model these fluctuations with a quadratic term which leads to the following Hamiltonian

$$H = -J \sum_{\langle i,j \rangle} x_i x_j - \frac{1}{2\sigma^2} \sum_i (y_i - x_i)^2 \sim -J \sum_{\langle i,j \rangle} x_i x_j - \frac{1}{\sigma^2} \sum_i y_i x_i$$

Choice of the *data* term

The *data* term in the Hamiltonian must contain information on the background but also on the image data.

An undirected graphical model representing a Markov random field for image de-noising, in which x_i is a binary variable denoting the state of pixel i in the unknown noise-free image, and y_i denotes the corresponding value of pixel i in the observed noisy image.



It would be natural to start with something like

$$H = -J \sum_{\langle i,j \rangle} x_i x_j - \sum_i y_i x_i$$

This is the same as the Hamiltonian of the Random Field Ising Model (RFIM)

where the y represent the image data and act like a sort of local magnetic field. Note that this can be obtained from a form of chi square minimization

$$H = -J \sum_{\langle i,j \rangle} x_i x_j - \frac{1}{2\sigma^2} \sum_i (y_i - x_i)^2 \sim -J \sum_{\langle i,j \rangle} x_i x_j - \frac{1}{\sigma^2} \sum_i y_i x_i$$

In practice the Hamiltonian is taken as follows:

$$H = -\zeta \sum_i x_i - \beta \sum_{\langle i,j \rangle} x_i x_j - \eta \sum_i y_i x_i$$

where

- the first parameter (zeta) is just the global field, that corresponds to a prior assumption on the most likely spin direction
- the second parameter (beta) is the coupling constant between neighboring spins
- the third parameter (eta) is the local interaction between the measured pixel (spin) values and the MRF spins

The "physical" likelihood determined by Markov Random Fields

With the assumption that the image structure behaves like the magnetization islands in an Ising spin system, we find that the likelihood is given by the Maxwell-Boltzmann distribution with the given configuration energy

$$P(\mathbf{d}|\mathbf{x}) = \frac{1}{Z_n} \exp\left[-\beta H_n(x,d)\right]$$

noise Hamiltonian

Therefore, with a prior specified as

$$P(\mathbf{x}) = \frac{1}{Z_0} \exp\left[-\beta H_0(x)\right]$$

We find the posterior

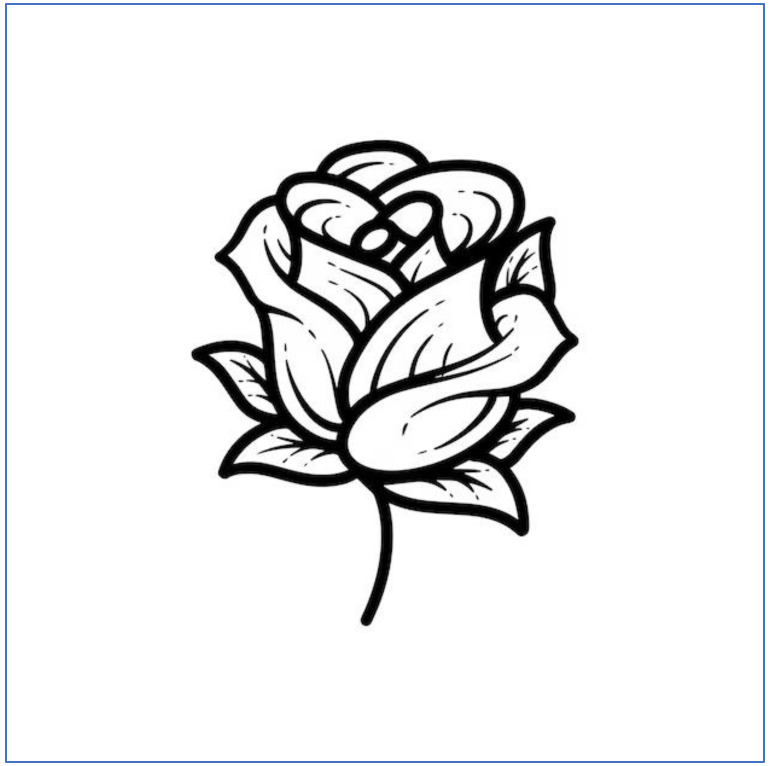
$$P(\mathbf{x}|\mathbf{d}) = \frac{1}{Z} \exp\left[-\beta H(x,d)\right] \qquad \text{with} \qquad H(x,d) = H_n(x,d) + H_0(x)$$

2. Bayesian paradigm. In real scenes, neighboring pixels typically have similar intensities, boundaries are usually smooth and often straight, textures, although sometimes random locally, define spatially homogeneous regions, and objects, such as grass, tree trunks, branches and leaves, have preferred relations and orientations. Our approach to picture processing is to articulate such regularities mathematically, and then to exploit them in a statistical framework to make inferences. The regularities are rarely deterministic; instead, they describe correlations and likelihoods. This leads us to the Bayesian formulation, in which prior expectations are formally represented by a probability distribution. Thus we design a distribution (a "prior") on relevant scene attributes to capture the tendencies and constraints that characterize the scenes of interest. Picture processing is then guided by this prior distribution, which, if properly conceived, enormously limits the plausible restorations and interpretations.

from Geman & Graffigne, Proc. Int. Congress of Mathematicians, Berkeley 1986

Simple example

proceed to https://github.com/edymil/Bayes-TS



BW image

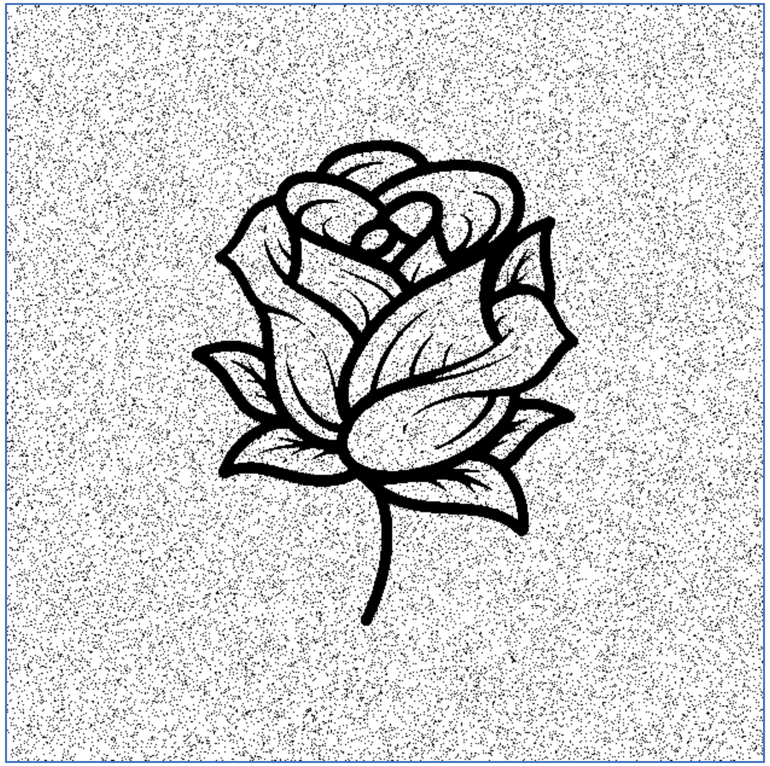
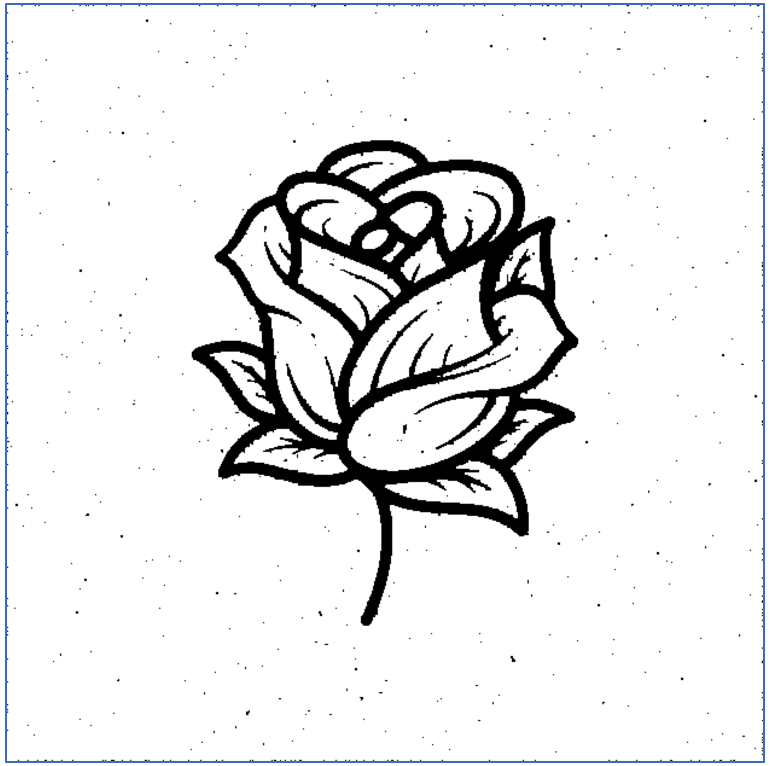


image + noise



processed image

Simple example (code of MRF function)

```
# here we define the sequential minimization of the zero-temperature Random Field Ising Model (RFIM)
# zeta: global b-field coupling
# beta: spin-spin coupling
# eta: coupling with input (noisy) image pixels (spins)
# repeats: number of repeats of the lattice minimization sweep
#
# here we assume a black and white image, where -1 means 0 and 1 means white
#
# any input image must be preprocessed before using this function

def MRF(bitmap,new_bitmap,zeta=0,beta=2.0,eta=1.5,repeats=1):

    # find the lattice size in x and y
    nx = bitmap.shape[0]
    ny = bitmap.shape[1]

    # the local interaction energy is defined as an inline function
    enn = lambda i,j: new_bitmap[i,j]*(new_bitmap[i,j-1]+new_bitmap[i,j+1]+new_bitmap[i-1,j]+new_bitmap[i+1,j]) # local interaction energy with nearest neighbors

    # loop over repeats
    for nrepeat in range(repeats):

        # loop over the whole lattice; notice that the boundary spins are left out (this defines the boundary condition)
        for i in range(1,nx-1):
            for j in range(1,ny-1):
                E0 = -zeta*new_bitmap[i,j] - beta*enn(i,j) - eta*bitmap[i,j]*new_bitmap[i,j] # total local energy before spin flip
                if E0 > 0: new_bitmap[i,j] *= -1 # the spin flip is accepted only if the new total local energy is lower after spin flip

    return new_bitmap
```

This method is not restricted to square lattices

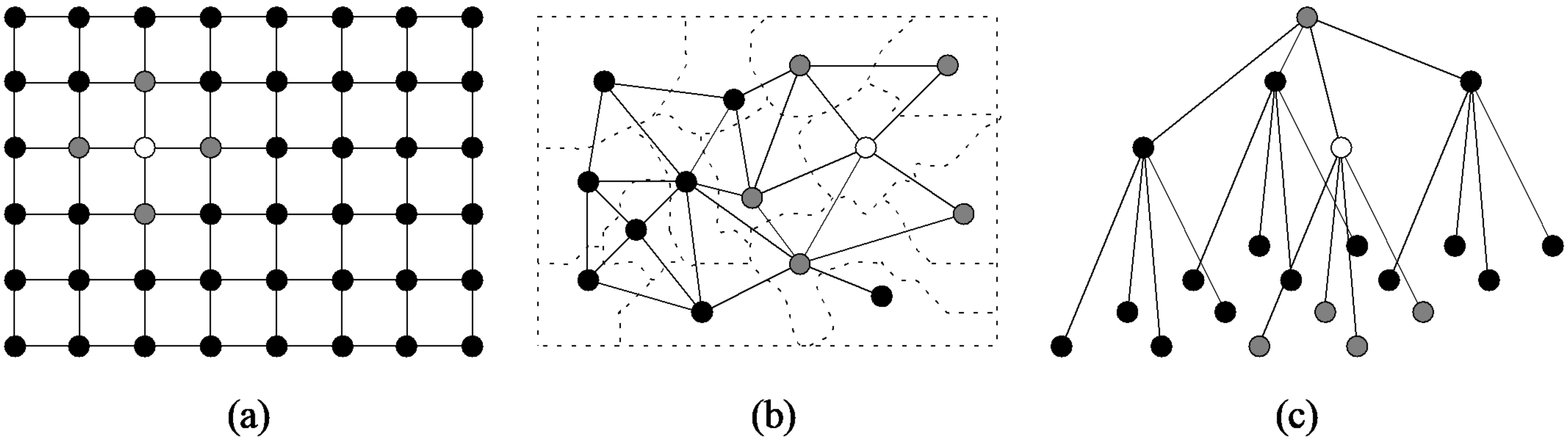


FIGURE 1. Three typical graphs supporting MRF-based models for image analysis: (a) rectangular lattice with first-order neighborhood system; (b) non-regular planar graph associated to an image partition; (c) quad-tree. For each graph, the grey nodes are the neighbors of the white one.

12. The Metropolis-Hastings algorithm and Markov Chain Monte Carlo

In our analysis of the Metropolis algorithm, we found that

$$T(C \to C') = \min\left[1, \exp\left(-\frac{(E'-E)}{kT}\right)\right]$$

Moreover, we found that the algorithm preserves detailed balance

$$T(C \to C')P(C) = T(C' \to C)P(C')$$

where P(C) is the stationary probability of configuration C. Indeed, at equilibrium we found that, if E' > E,

$$P(C)\exp\left(-\frac{(E'-E)}{kT}\right) = P(C')$$

$$\frac{P(C')}{P(C)} = \exp\left(-\frac{(E'-E)}{kT}\right)$$

Boltzmann's distribution

In summary

$$P(C)\exp\left(-\frac{(E'-E)}{kT}\right) = P(C')$$

$$\frac{P(C')}{P(C)} = \exp\left(-\frac{(E'-E)}{kT}\right)$$

$$T(C \to C') = \min\left[1, \frac{P(C')}{P(C)}\right]$$

$$T(C \to C') = \min\left[1, \exp\left(-\frac{(E'-E)}{kT}\right)\right]$$

This definition of the transition probability is the starting point for an important further step, the Metropolis-Hastings algorithm.

Notice that we only need the ratio P(C')/P(C)*:* ***the partition function (the normalization integral) has disappeared****. In this, the method is similar to the acceptance-rejection method.*

we define the transition probability – which includes a proposal function q –

$$P(\mathbf{x} \to \mathbf{y}) = q(\mathbf{x}, \mathbf{y})\alpha(\mathbf{x}, \mathbf{y})$$

and the target density

$$\boldsymbol{\pi}(\mathbf{x})$$

and we take the state

$$\mathbf{x} = \mathbf{x}_n$$

next we choose randomly another state $\mathbf{y}$ and we accept it with probability

$$\alpha(\mathbf{x}, \mathbf{y}) = \min\left\{1, \frac{\boldsymbol{\pi}(\mathbf{y})q(\mathbf{y}, \mathbf{x})}{\boldsymbol{\pi}(\mathbf{x})q(\mathbf{x}, \mathbf{y})}\right\}$$

Note that if the proposal function q is symmetrical, then the acceptance probability takes on the simpler form

$$\alpha(\mathbf{x}, \mathbf{y}) = \min\left\{1, \frac{\boldsymbol{\pi}(\mathbf{y})q(\mathbf{y}, \mathbf{x})}{\boldsymbol{\pi}(\mathbf{x})q(\mathbf{x}, \mathbf{y})}\right\} \longrightarrow \min\left\{1, \frac{\boldsymbol{\pi}(\mathbf{y})}{\boldsymbol{\pi}(\mathbf{x})}\right\}$$

and it depends on the target density only.

The M-H algorithm defines a Markov chain, and it is easy to show that **detailed balance holds**. The transition probability is

$$P(\mathbf{x} \to \mathbf{y}) = q(\mathbf{x}, \mathbf{y})\alpha(\mathbf{x}, \mathbf{y}) = q(\mathbf{x}, \mathbf{y}) \min \left\{ 1, \frac{\boldsymbol{\pi}(\mathbf{y})q(\mathbf{y}, \mathbf{x})}{\boldsymbol{\pi}(\mathbf{x})q(\mathbf{x}, \mathbf{y})} \right\}$$

- case $\dfrac{\boldsymbol{\pi}(\mathbf{y})q(\mathbf{y}, \mathbf{x})}{\boldsymbol{\pi}(\mathbf{x})q(\mathbf{x}, \mathbf{y})} \geq 1$

$$\Rightarrow \quad \alpha(\mathbf{x},\mathbf{y}) = 1; \quad \alpha(\mathbf{y},\mathbf{x}) = \frac{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})}{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})} \quad \Rightarrow \quad \begin{aligned} P(\mathbf{x} \to \mathbf{y}) &= q(\mathbf{x},\mathbf{y}) \\ P(\mathbf{y} \to \mathbf{x}) &= q(\mathbf{y},\mathbf{x}) \frac{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})}{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})} \end{aligned}$$

$$\Rightarrow \quad \begin{aligned} \pi(\mathbf{x})P(\mathbf{x} \to \mathbf{y}) &= \pi(\mathbf{x})q(\mathbf{x},\mathbf{y}) \\ \pi(\mathbf{y})P(\mathbf{y} \to \mathbf{x}) &= \pi(\mathbf{y})q(\mathbf{y},\mathbf{x}) \frac{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})}{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})} = \pi(\mathbf{x})q(\mathbf{x},\mathbf{y}) \end{aligned}$$

- case $\dfrac{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})} < 1$

$$\Rightarrow \alpha(\mathbf{x},\mathbf{y}) = \frac{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})}; \quad \alpha(\mathbf{y},\mathbf{x}) = 1 \Rightarrow \begin{array}{l} P(\mathbf{x}\to\mathbf{y}) = q(\mathbf{x},\mathbf{y})\dfrac{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})} \\ P(\mathbf{y}\to\mathbf{x}) = q(\mathbf{y},\mathbf{x}) \end{array}$$

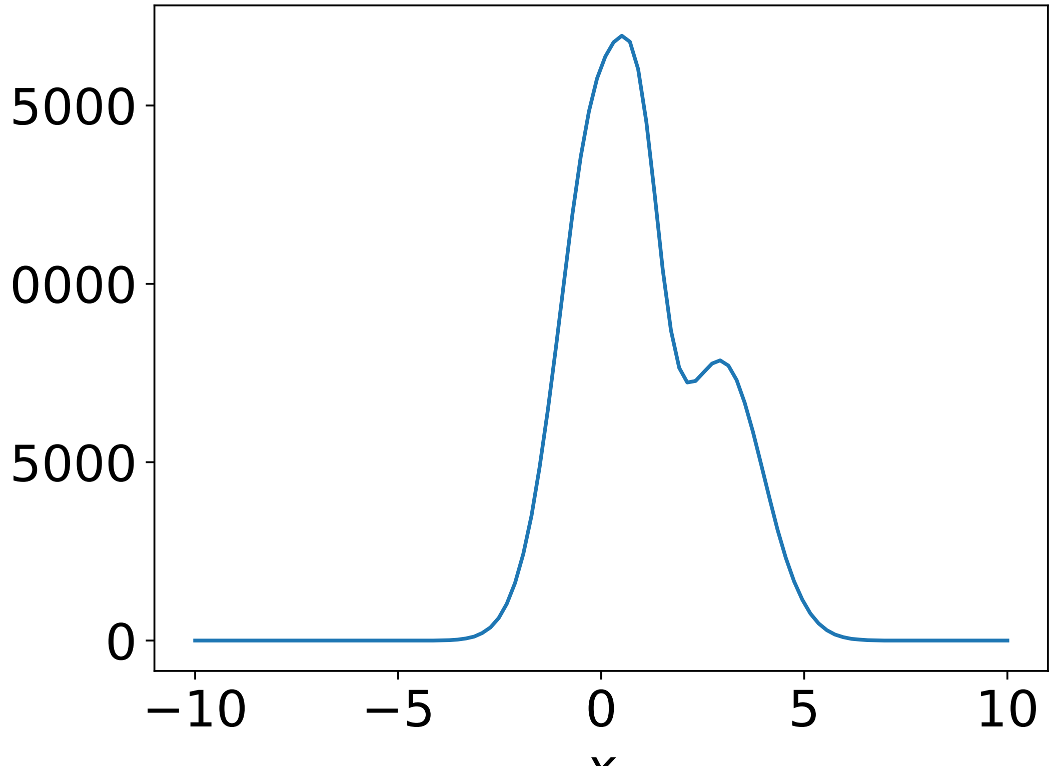
$$\Rightarrow \begin{array}{l} \pi(\mathbf{x})P(\mathbf{x}\to\mathbf{y}) = \pi(\mathbf{x})q(\mathbf{x},\mathbf{y})\dfrac{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})} = \pi(\mathbf{y})q(\mathbf{y},\mathbf{x}) \\ \pi(\mathbf{y})P(\mathbf{y}\to\mathbf{x}) = \pi(\mathbf{y})q(\mathbf{y},\mathbf{x}) \end{array}$$

Detailed balance holds in both cases and therefore $\pi(\mathbf{x})$ is stationary

The following figure shows a simulation with the MCMC algorithm and the distribution

$$p(x) = \frac{0.6}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) + \frac{0.3}{\sqrt{2\pi}} \exp\left(-\frac{(x-3)^2}{2}\right) + \frac{0.1}{\sqrt{0.5\pi}} \exp\left(-\frac{(x-1)^2}{0.5}\right)$$

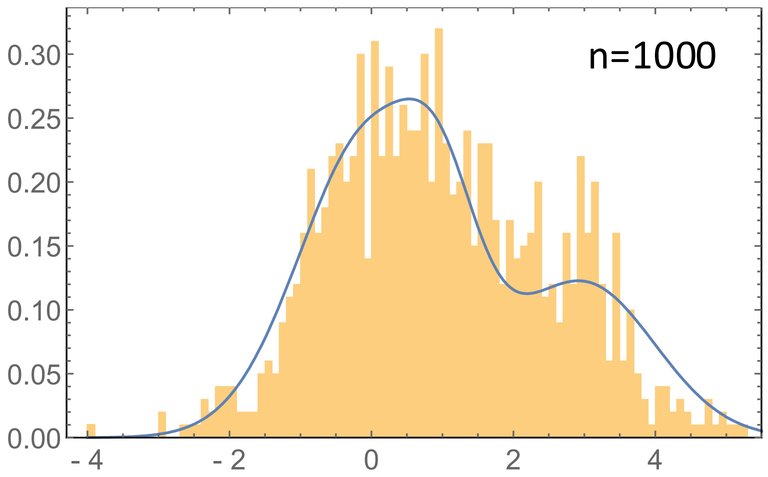
(a three-component mixture model)

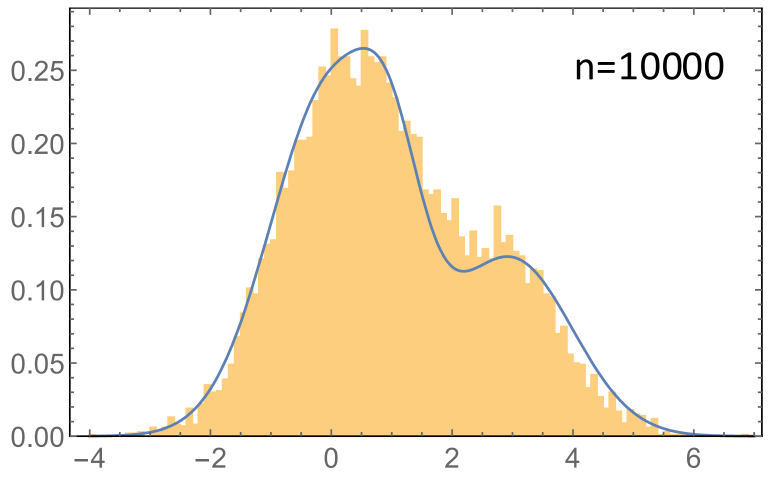


```
nrmax = 40 000;

xr = Table[0, {nrmax}];
xr[[1]] = -4;

nr = 1;
While[nr < nrmax,
 xtry = xr[[nr]] + RandomReal[NormalDistribution[0, 1]];
 If[ pdf[xtry] / pdf[xr[[nr]]] > RandomReal[] , nr ++; xr[[nr]] = xtry];
]
```

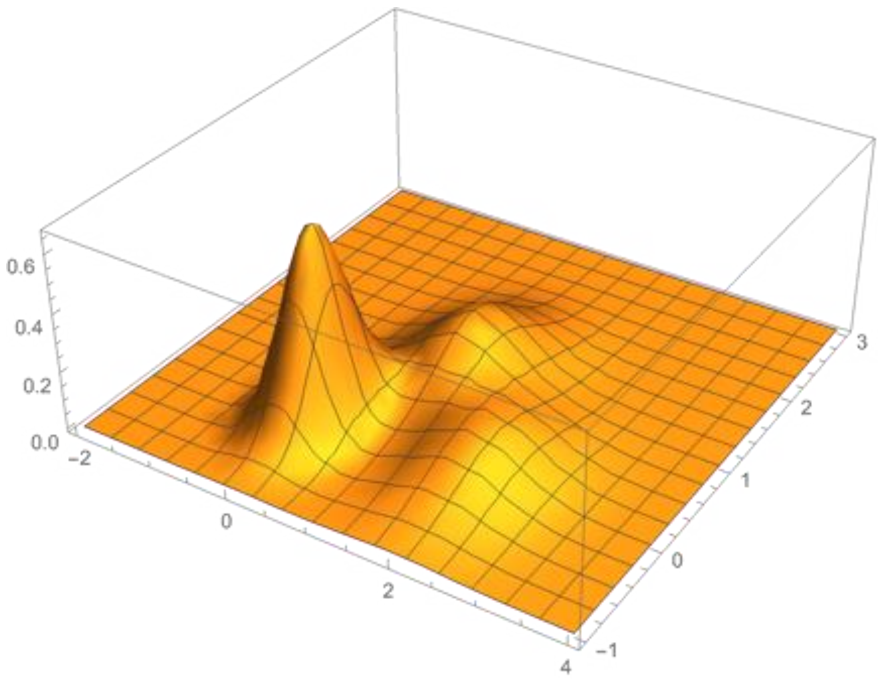


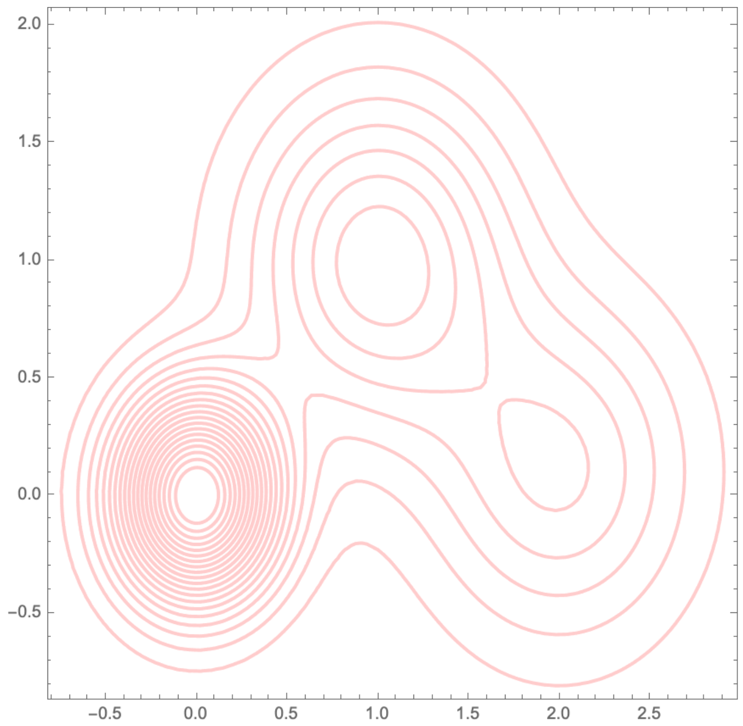


MCMC simulation of a 2D three-component mixture model

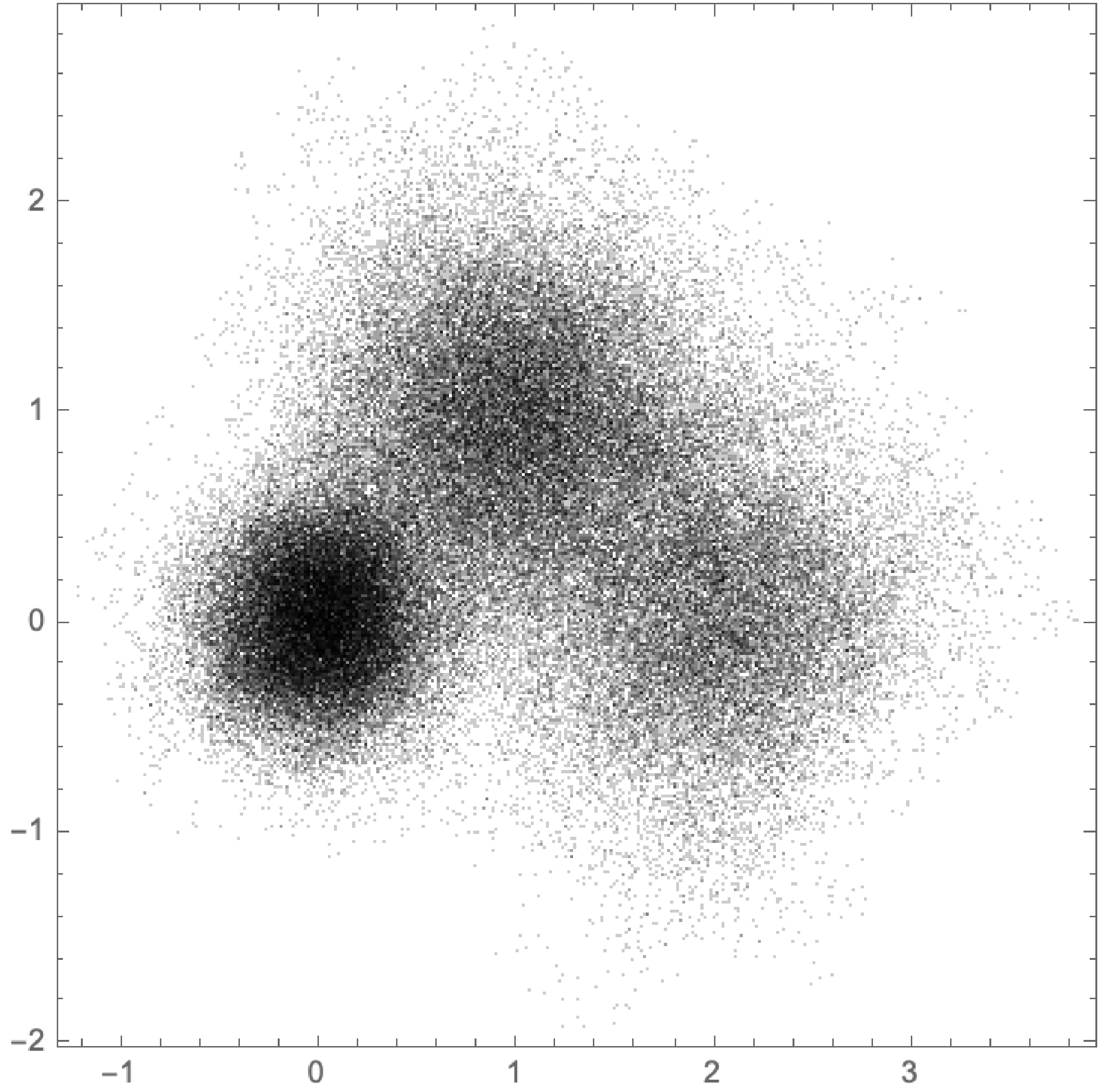
$$p(x,y) = \sum_{i=1}^{3} \frac{\alpha_i}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(x-\mu_{x,i})^2 + (y-\mu_{y,i})^2}{2\sigma_i^2}\right]$$

$$\begin{array}{llll}
\alpha_1 = 0.5; & \mu_{x,1} = 0; & \mu_{y,1} = 0; & \sigma_1 = 0.3; \\
\alpha_2 = 0.3; & \mu_{x,2} = 1; & \mu_{y,2} = 1.; & \sigma_2 = 0.5; \\
\alpha_3 = 0.2; & \mu_{x,3} = 2; & \mu_{y,3} = 0.1; & \sigma_3 = 0.5;
\end{array}$$

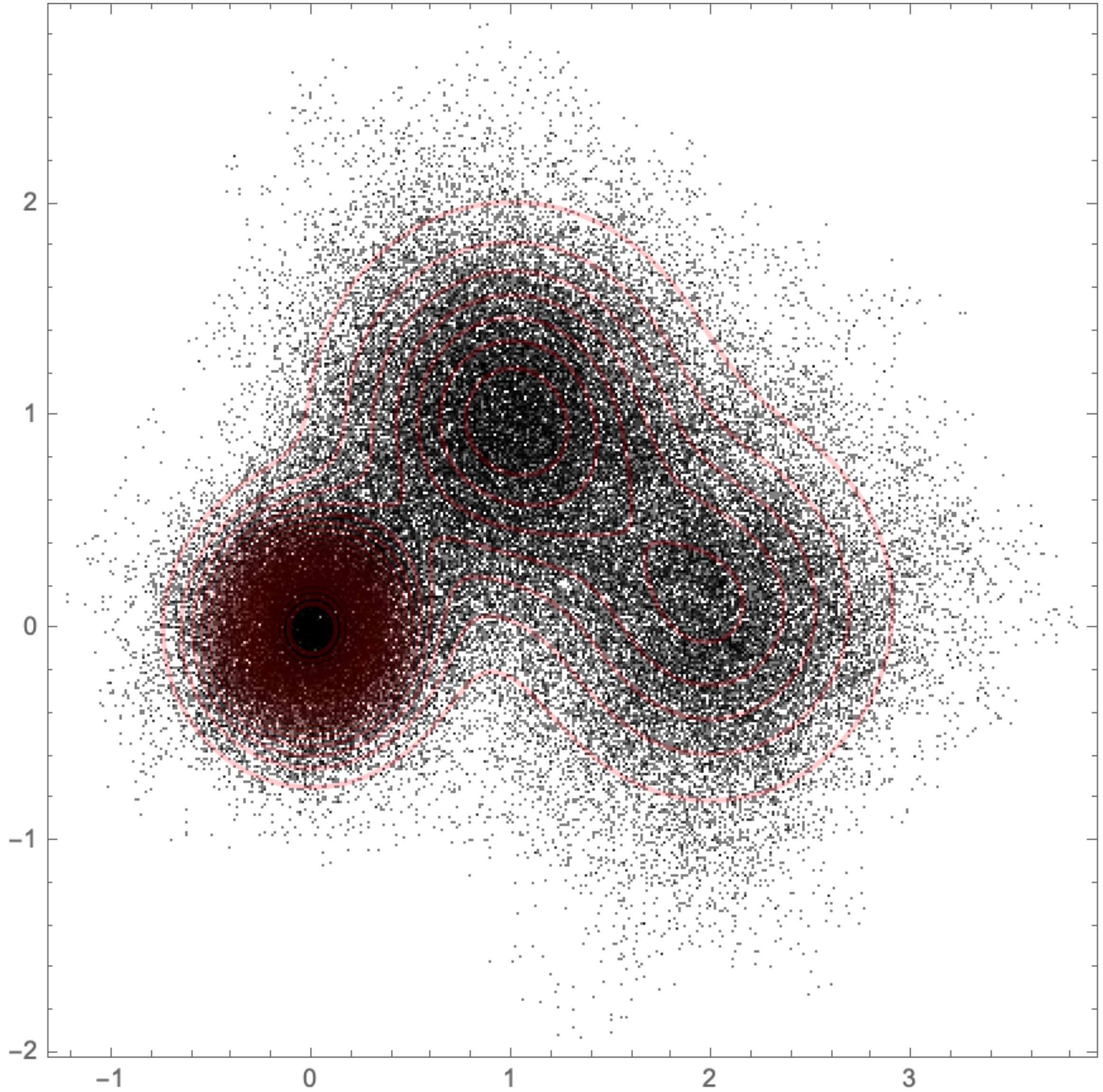


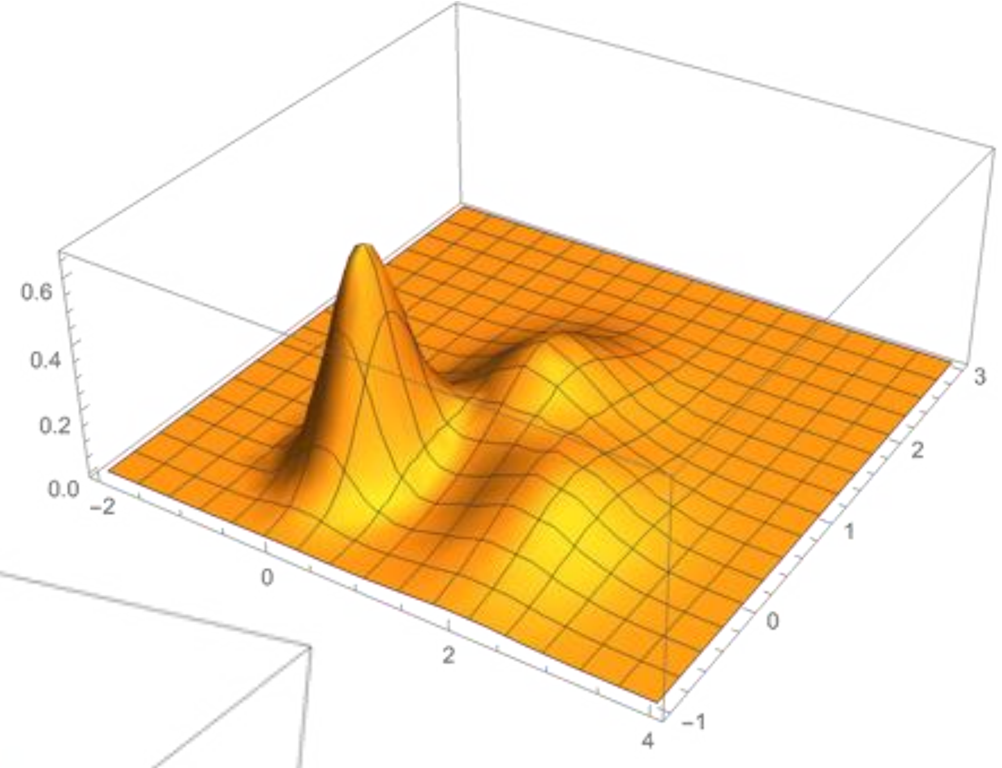


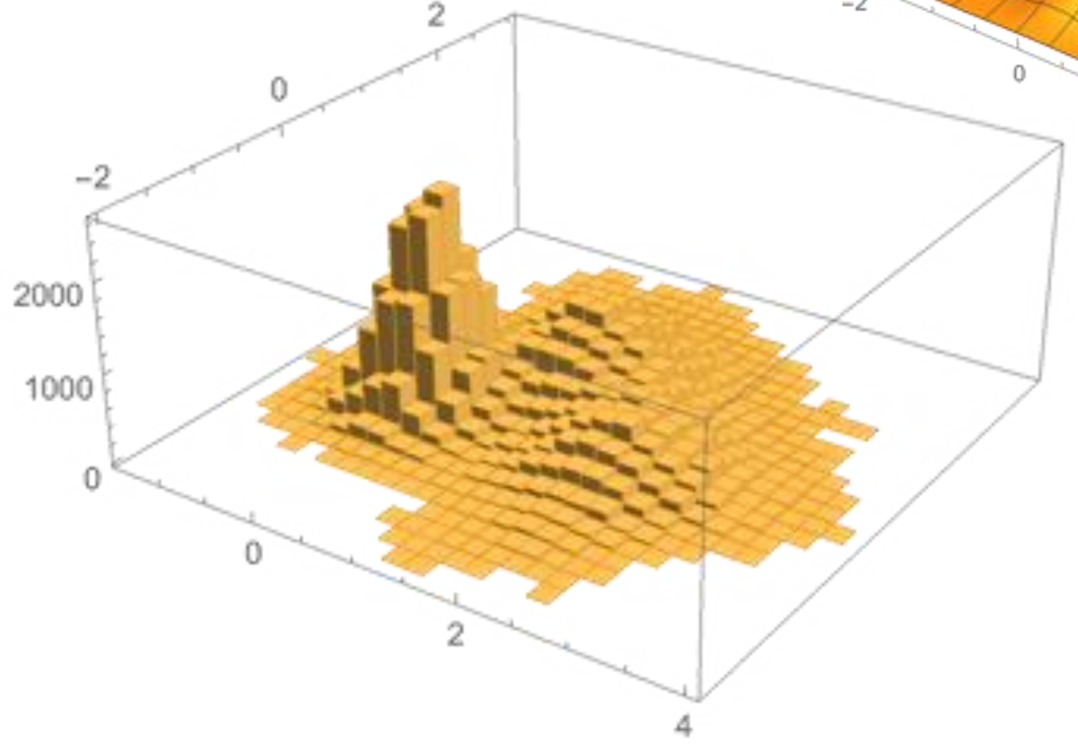
100000 steps



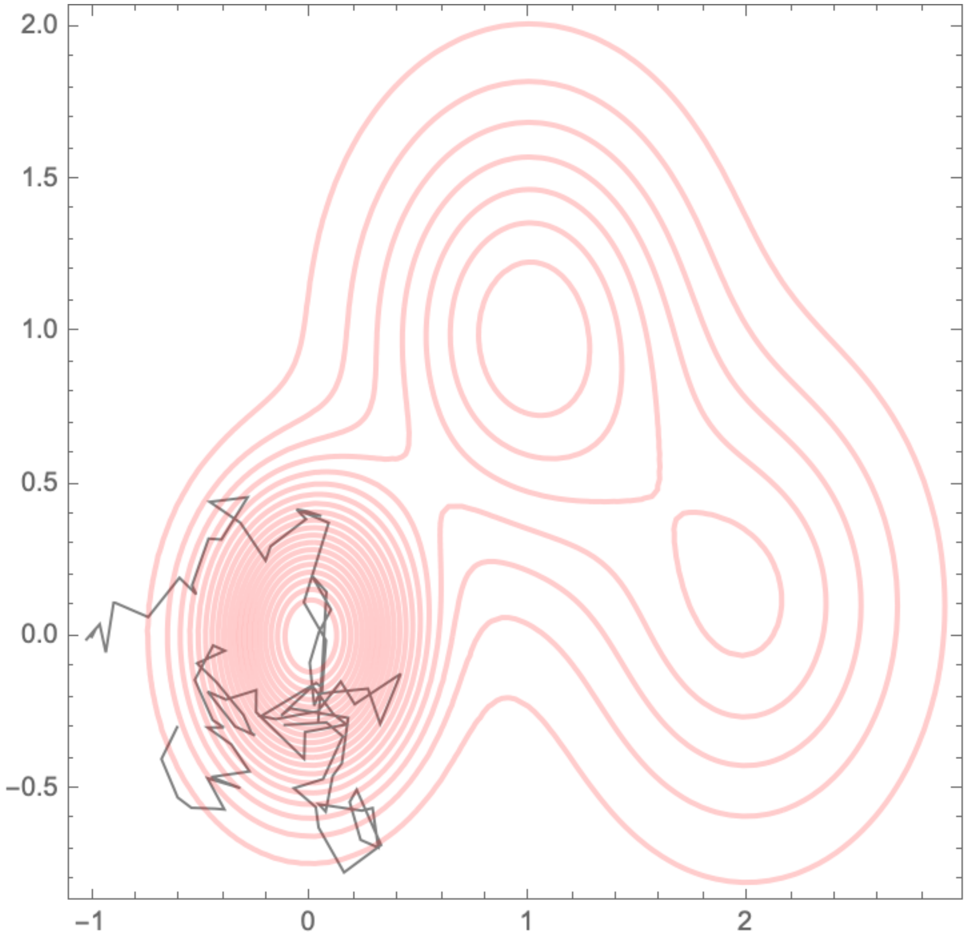
100000 steps



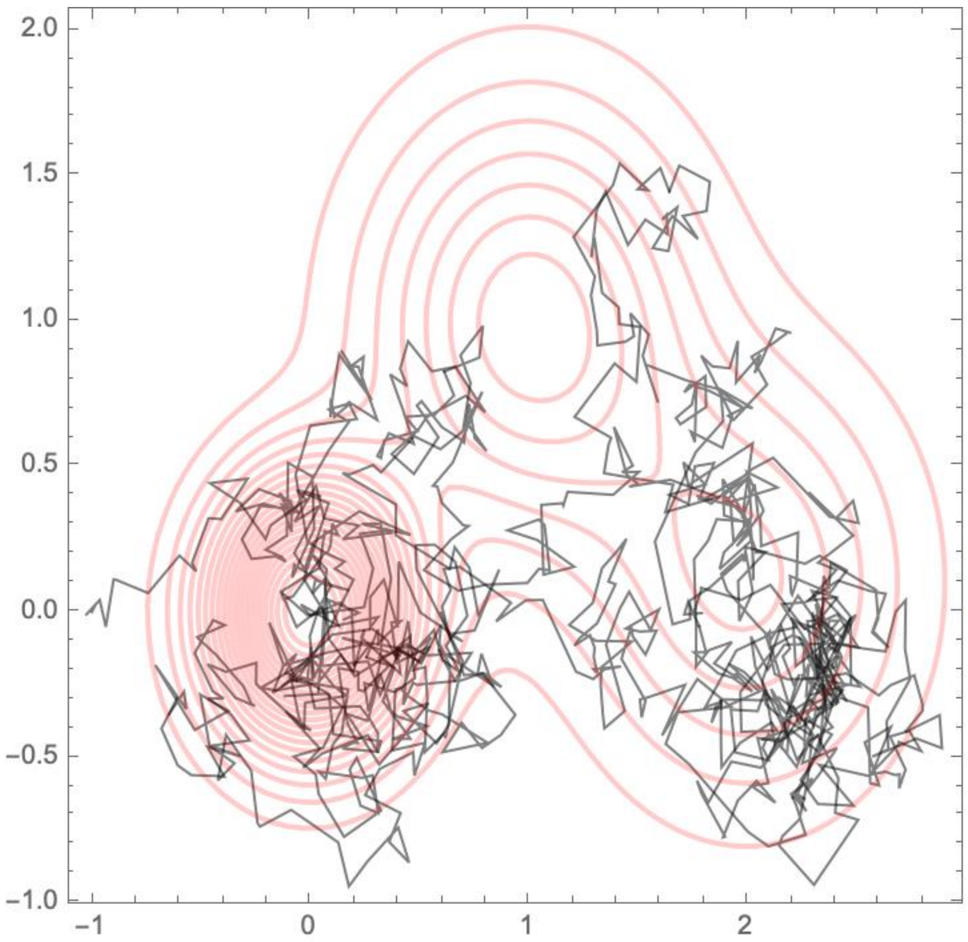

0.6
0.4
0.2
0.0
−2
0
2
4
3
2
1
0
−1


2
0
−2
2000
1000
0
0
2
4

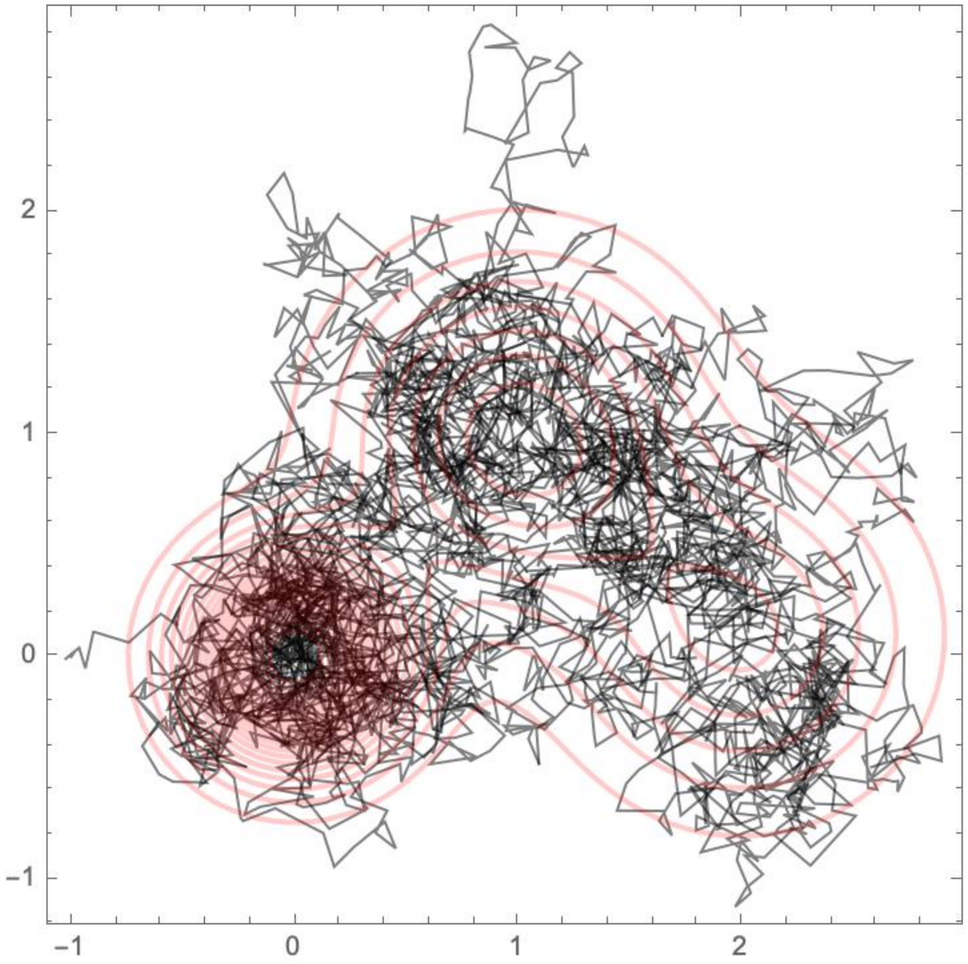
100 steps



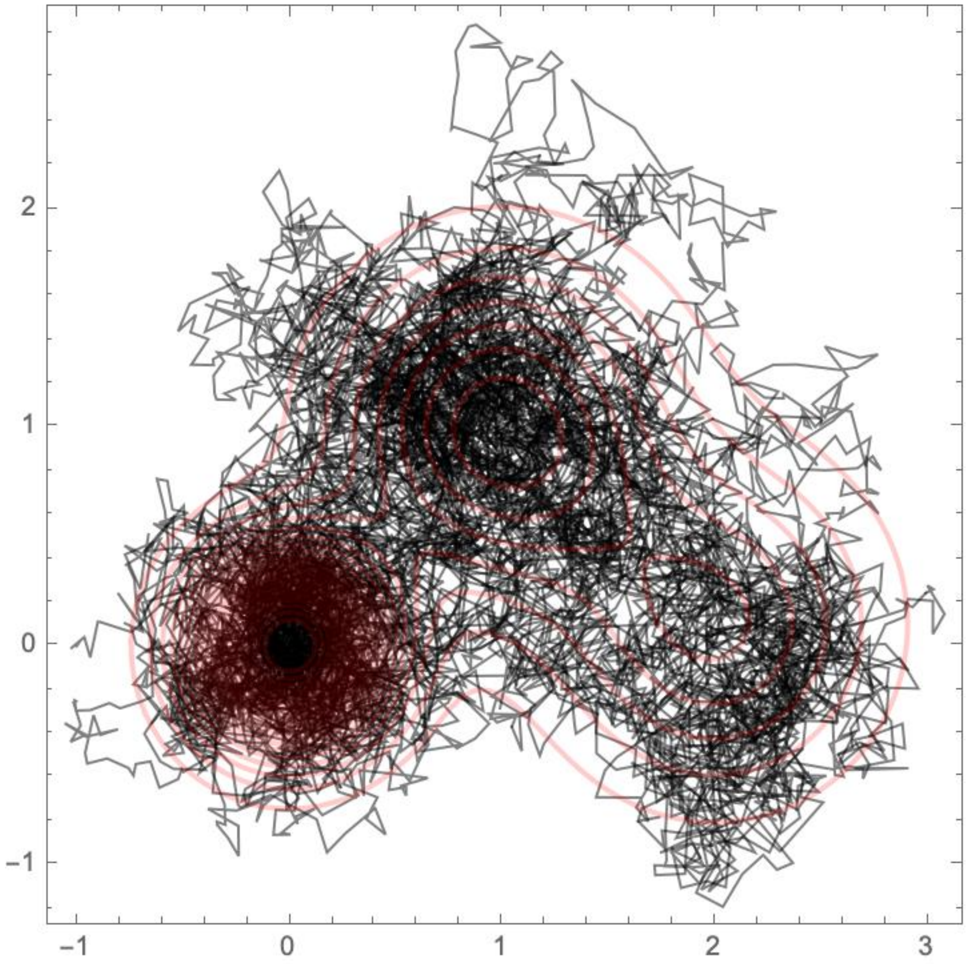
1000 steps



4000 steps



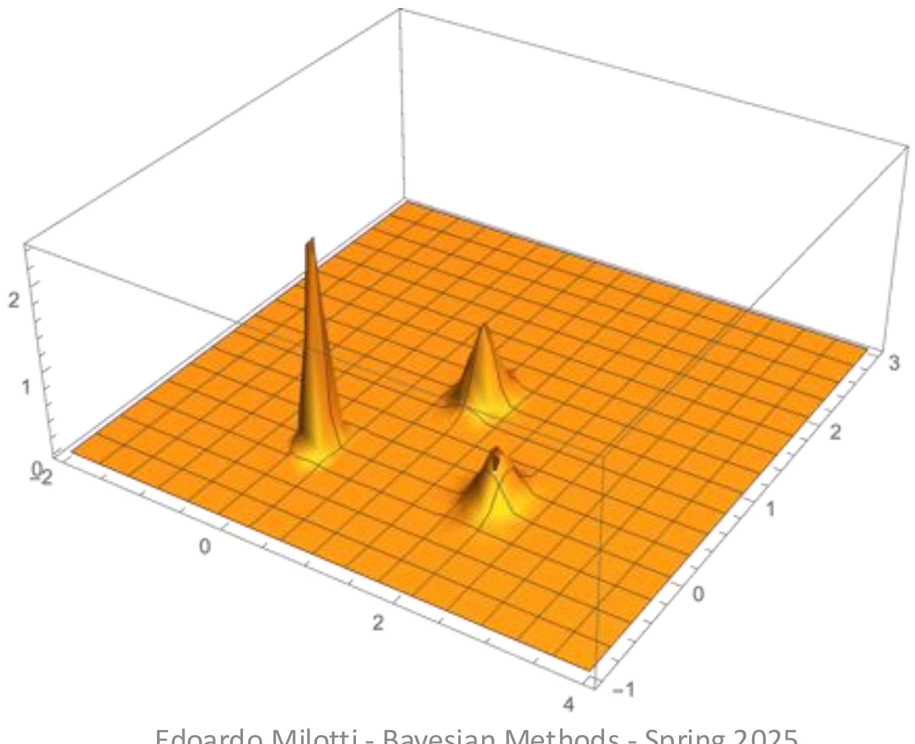
10000 steps



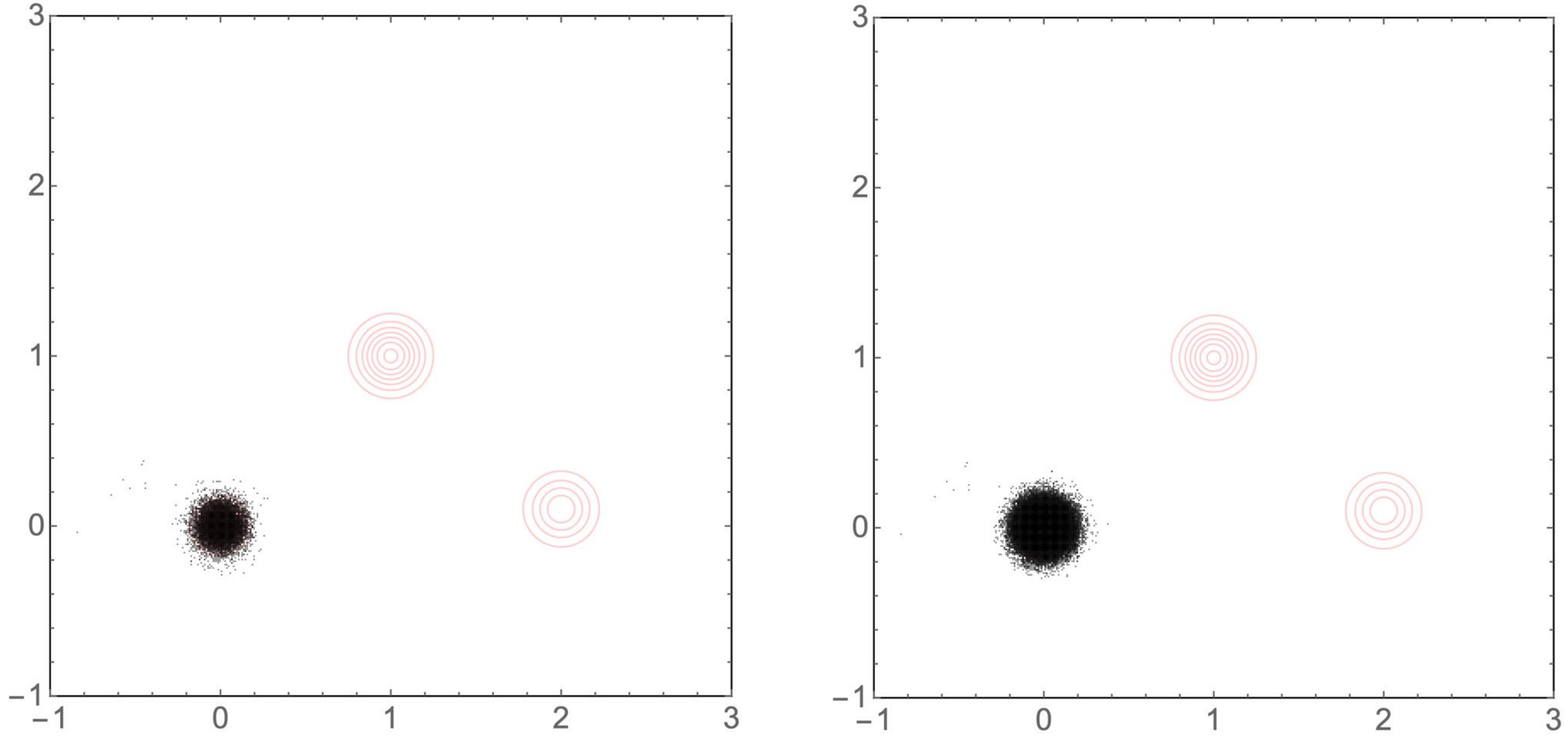
Notice that when the peaks are very narrow, the random walker may have problems visiting all of the peaks

$$p(x,y) = \sum_{i=1}^{3} \frac{\alpha_i}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(x-\mu_{x,i})^2 + (y-\mu_{y,i})^2}{2\sigma_i^2}\right]$$

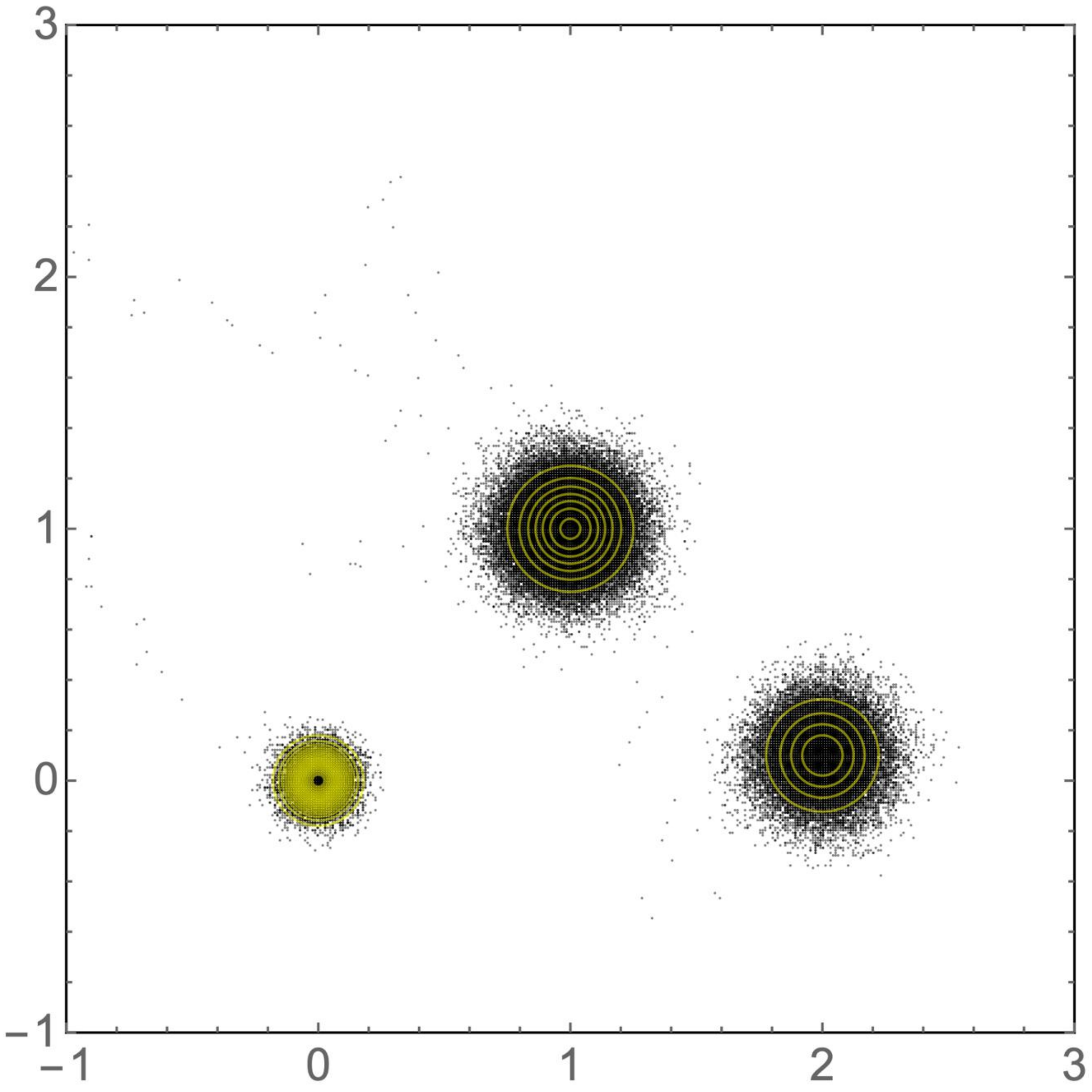
$$\begin{array}{llll}
\alpha_1 = 0.5; & \mu_{x,1} = 0; & \mu_{y,1} = 0; & \sigma_1 = 0.0725; \\
\alpha_2 = 0.3; & \mu_{x,2} = 1; & \mu_{y,2} = 1.; & \sigma_2 = 0.125; \\
\alpha_3 = 0.2; & \mu_{x,3} = 2; & \mu_{y,3} = 0.1; & \sigma_3 = 0.125;
\end{array}$$



With isolated, narrow peaks, increasing the number of steps may not suffice

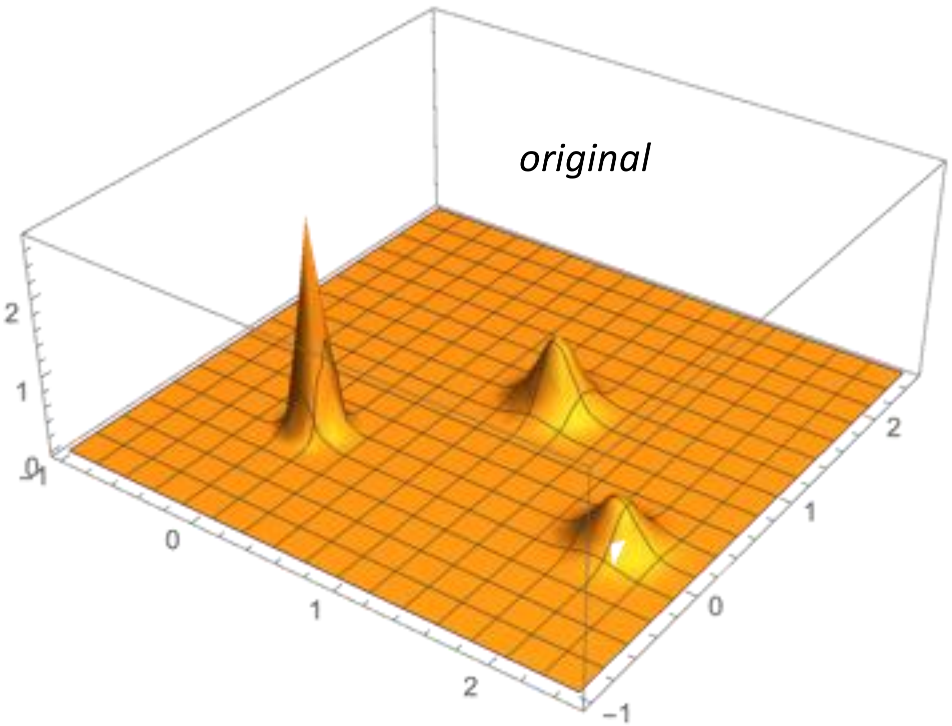


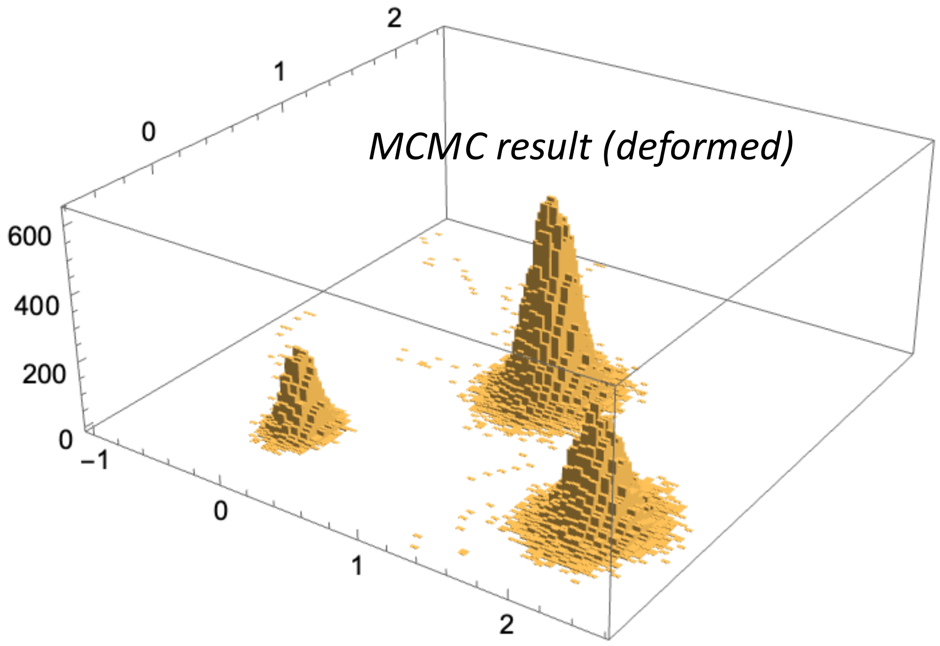
100000 steps, subdivided into 10 parallel chains with random starting points



The starting points of the chains are uniformly distributed in the plot region, however the "regions of influence" of each peak vary considerably.

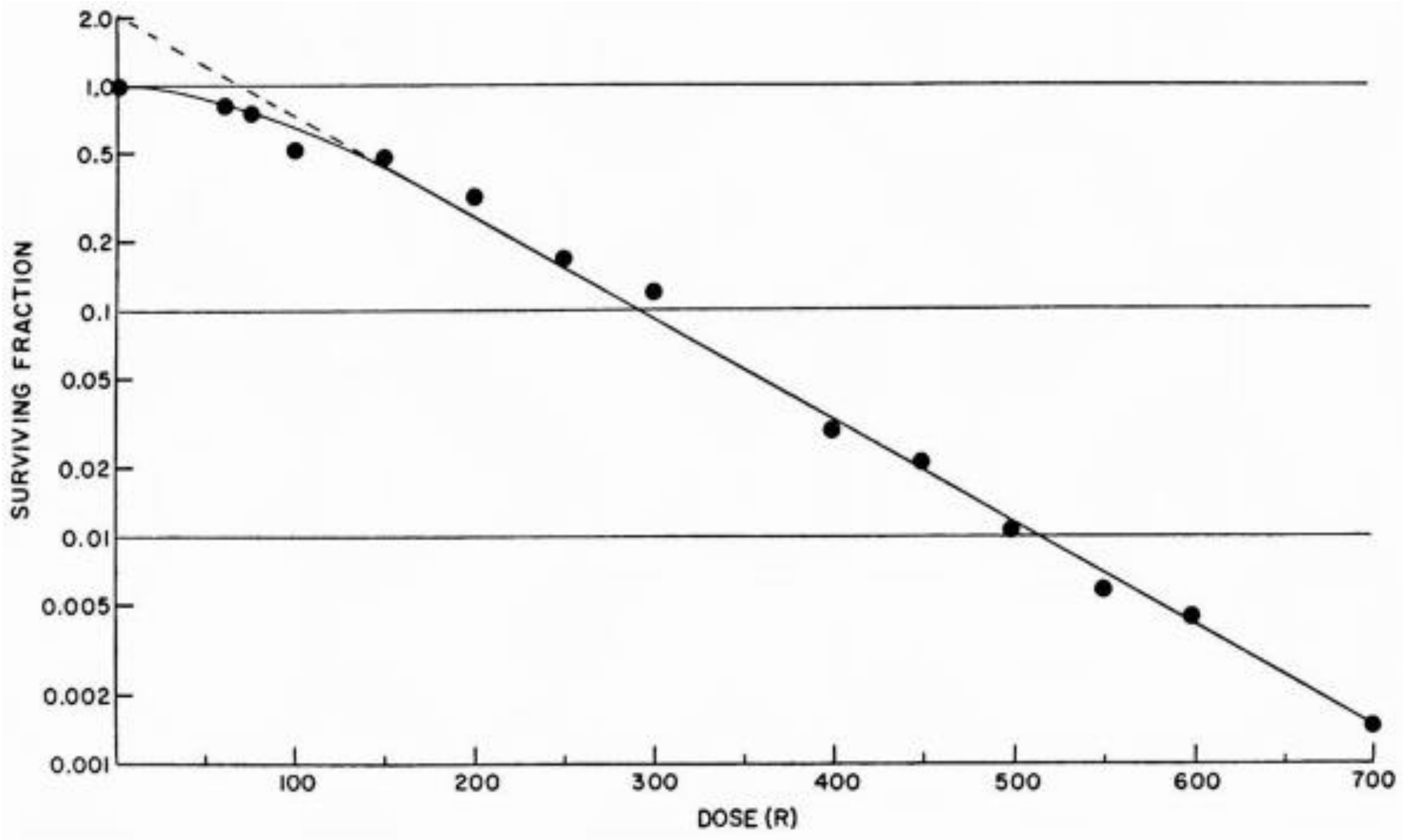
This leads to more chains being attracted into the lower peaks, with the result that the distribution is somewhat deformed (wrong alpha's in the mixture model)





Many techniques have been developed to avoid these pitfalls

Example of application of the MCMC technique in radiobiology



Survival curve for HeLa cells in culture exposed to x-rays. (From Puck TT, Markus PI: Action of x-rays on mammalian cells. *J Exp Med* 103:653-666, 1956)

Phenomenology: the linear-quadratic law

$$S(D) \approx e^{-\alpha D - \beta D^2}$$

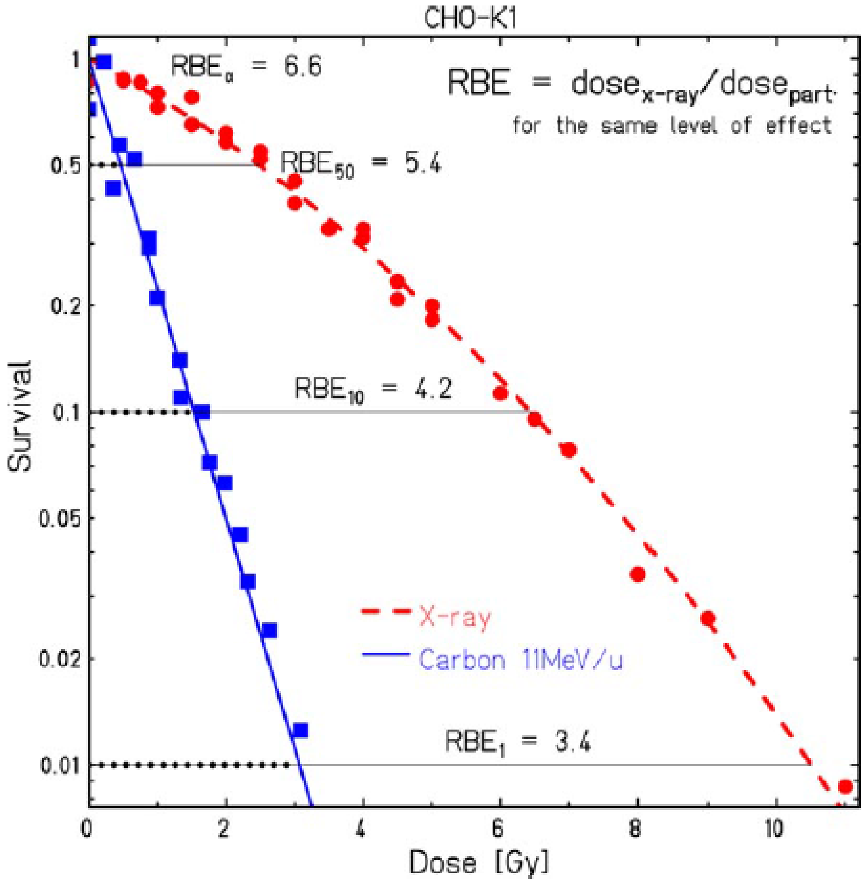


Fig. 1. Clonogenic survival curves illustrating the higher efficiency of the carbon ions compared with X-rays [10] (courtesy of the author, dr. Wilma K. Weyrather).

Example: Target theory

Simple Poisson model:

Probability of hitting *n* times a given target, when the average number of good hits is *a*:

$$P(n) = \frac{a^n}{n!} e^{-a}$$

Probability missing the target: $P(0) = e^{-a}$

Average number of hits: $a = D/D_0$

$$\Rightarrow \quad S(D) = P(0, D) = e^{-D/D_0}$$

Multitarget model, asymptotic behavior and threshold effect.

If there are multiple targets, say *n* targets, all of which must be hit to kill a cell, then the probability of missing at least one of them – i.e., the survival probability – is

$$S(D) = 1 - (1 - e^{-D/D_0})^n$$

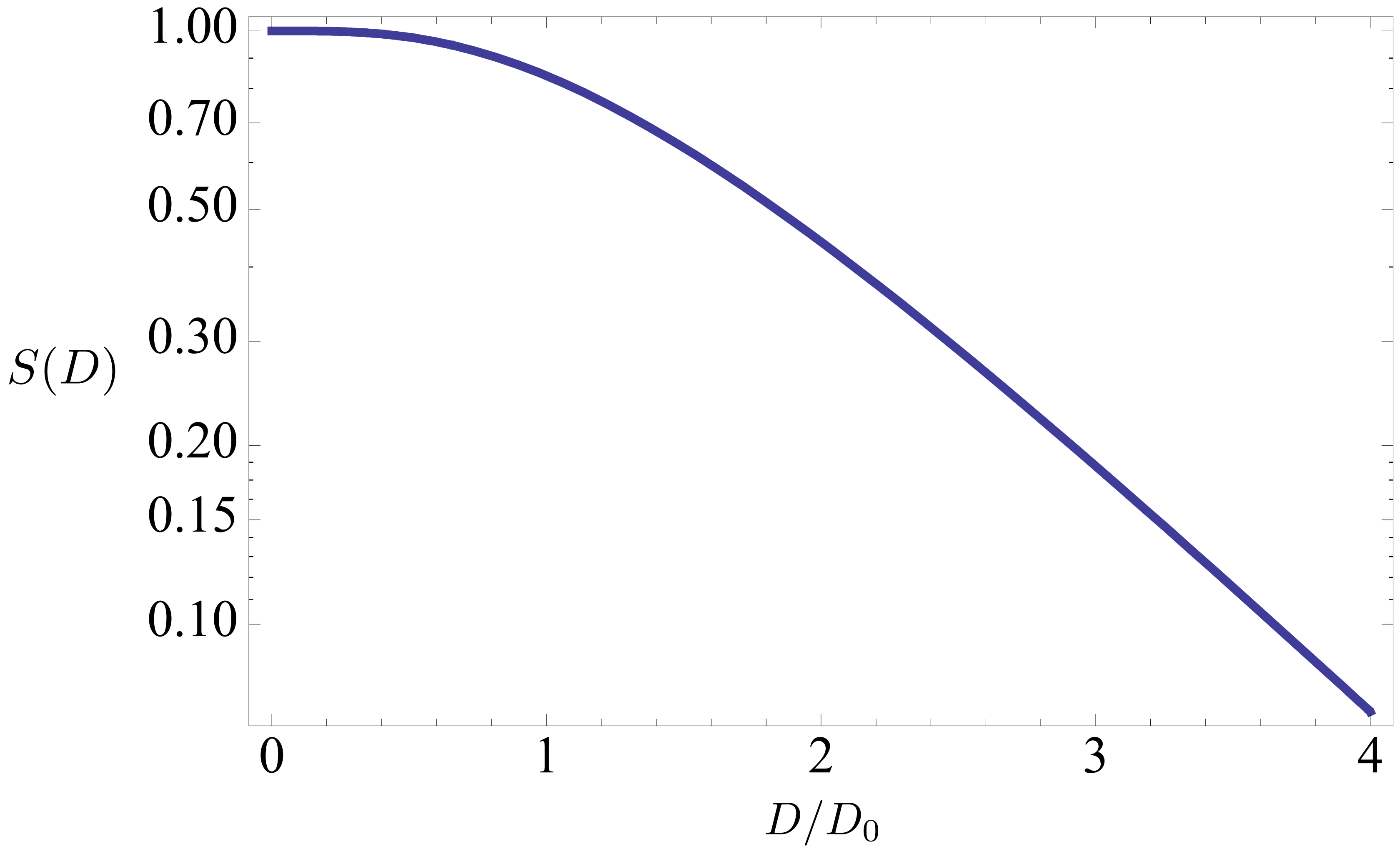
then, for large dose

$$S(D) \approx n e^{-D/D_0}$$

i.e.,

$$\ln S(D) \approx \ln n - D/D_0$$

which is a linear relation with intercept ln *n*, and slope $-1/D_0$.


1.00
0.70
0.50
0.30
0.20
0.15
0.10
$S(D)$
0
1
2
3
4
D/D_0

Notice that

$$\left[\frac{d}{dD} e^{-\alpha D - \beta D^2}\right]_{D=0} = (-\alpha - 2\beta D) e^{-\alpha D - \beta D^2}\Big|_{D=0} = -\alpha$$

and that

$$\frac{d}{dD}\left[1 - (1 - e^{-D/D_0})^n\right]_{D=0} = -n\,\frac{e^{-D/D_0}}{D_0}(1 - e^{-D/D_0})^{n-1}\Big|_{D=0} = 0$$

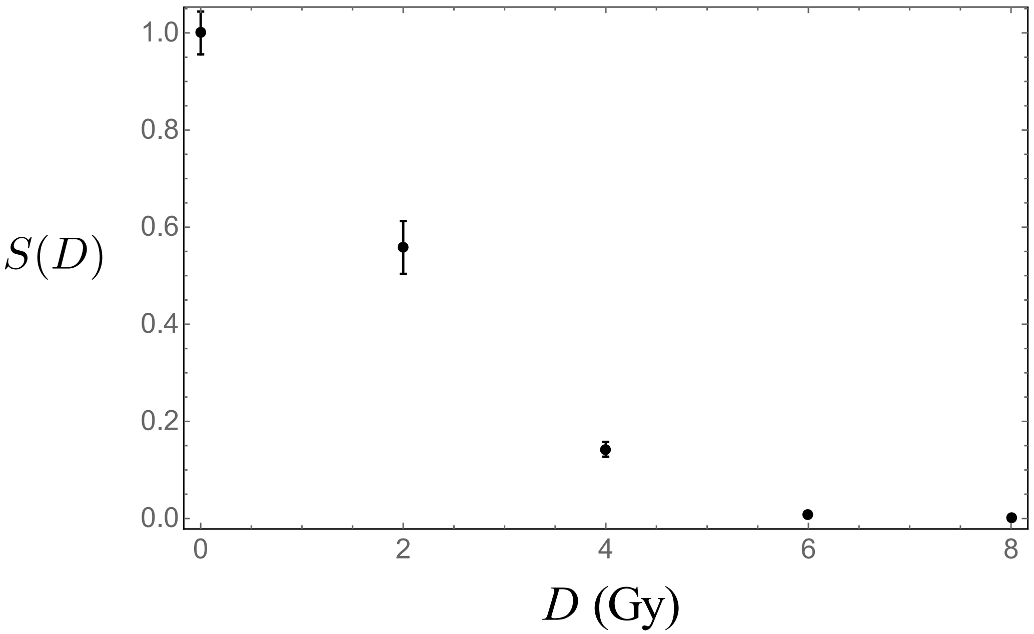
The derivatives differ in the origin, and the multitarget model fails to reproduce the observed linear-quadratic law.

The RCR (Repairable-Conditionally Repairable Damage) model

In this case the surviving fraction is

$$S = \exp(-aD) + bD\exp(-cD)$$

This is a 3-parameter expression, which is not easy to fit to data when the data set is small.



1a. Simple Gaussian likelihood for the LQ model

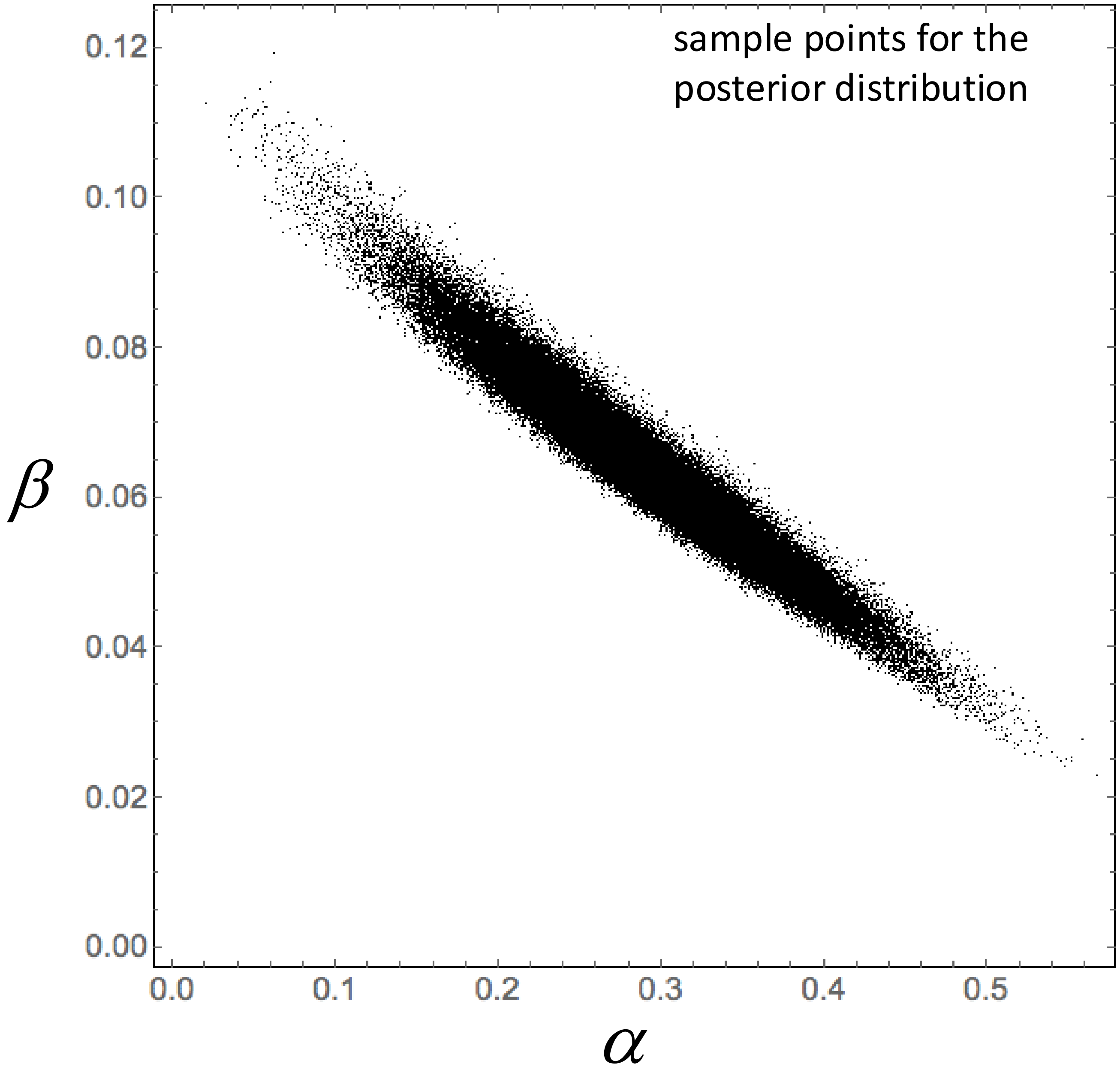
$$L(\{S_k\}|\alpha, \beta) = \prod_k \exp\left(-\frac{(S_k - S(D_k; \alpha, \beta)^2}{2\sigma_k^2}\right)$$

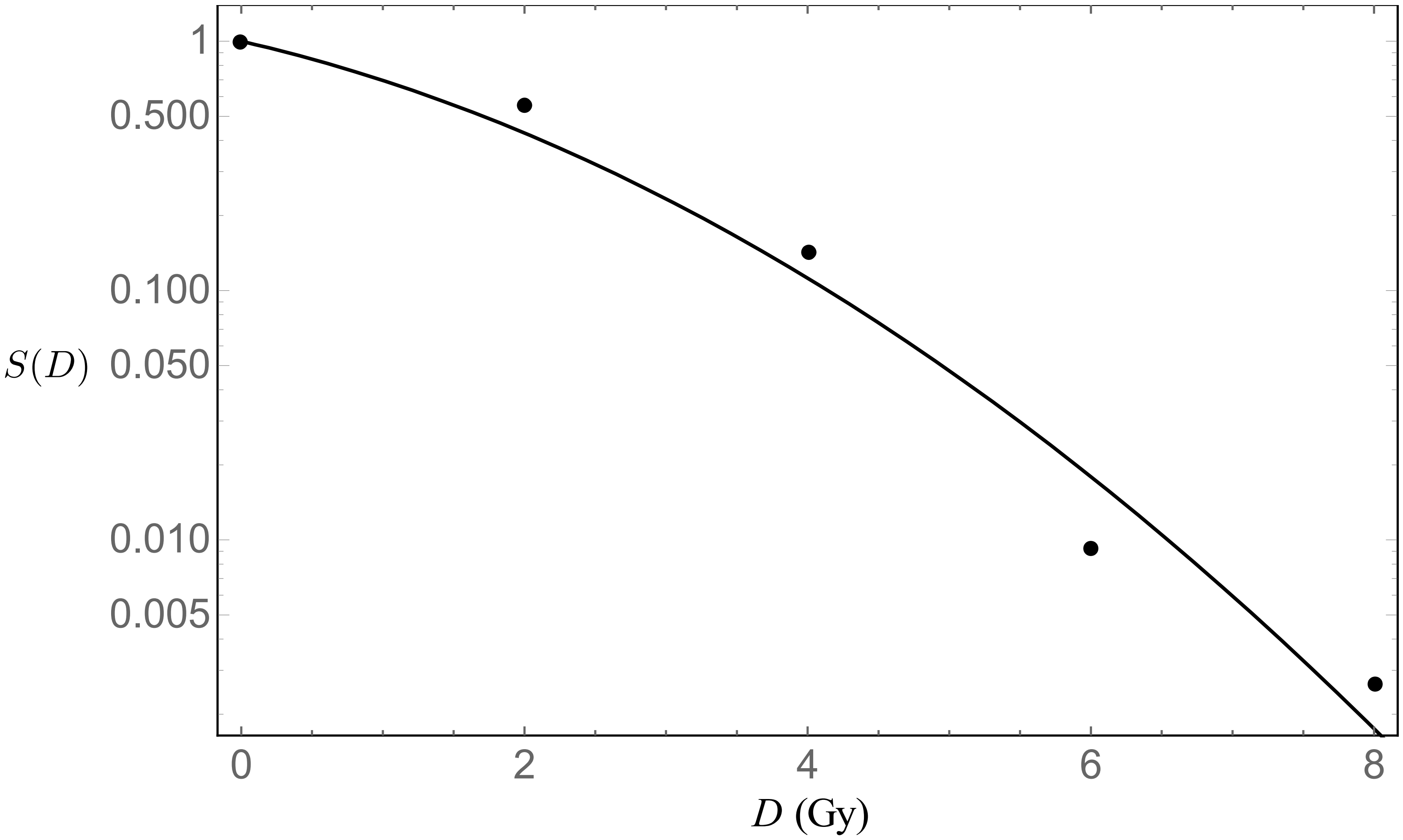
1b. Chose exponential priors for the parameters

1c. Complete posterior pdf

$$p(\alpha, \beta|\{S_k\}, I) = \left[\prod_k \exp\left(-\frac{(S_k - S(D_k; \alpha, \beta)^2}{2\sigma_k^2}\right)\right] \overbrace{\exp(-0.1\alpha) \exp(-0.1\beta)}^{\text{a rather arbitrary choice of priors}}$$

1d. Use MCMC to find the MAP estimate (and any moment of the pdf)


sample points for the posterior distribution
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0.0
0.1
0.2
0.3
0.4
0.5
β
α



1
0.500
0.100
0.050
0.010
0.005
$S(D)$
0
2
4
6
8
D (Gy)

2a. Simple Gaussian likelihood for the RCR model

$$L(\{S_k\}|a,b,c) = \prod_k \exp\left(-\frac{(S_k - S(D_k;a,b,c)^2}{2\sigma_k^2}\right)$$

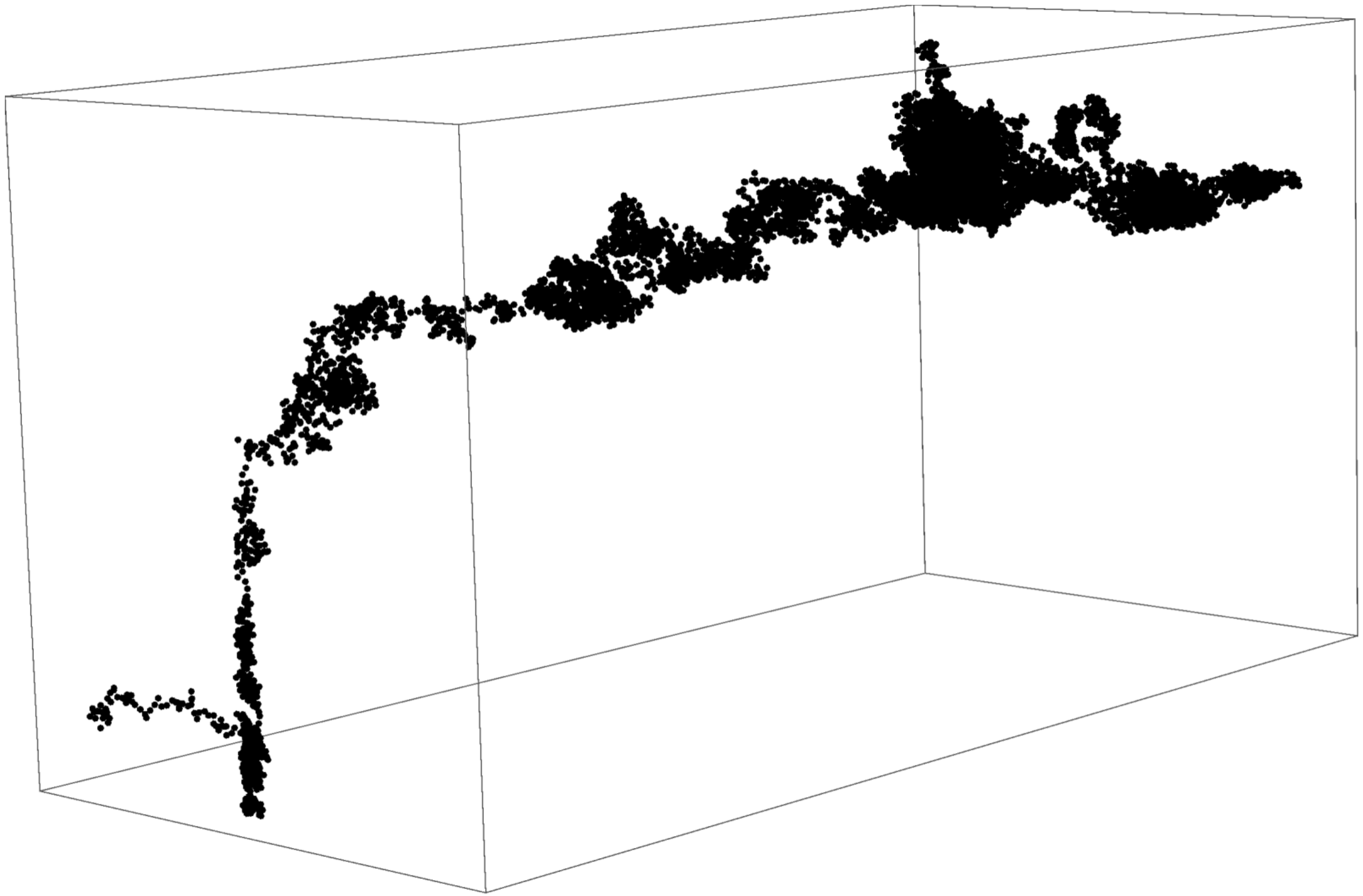
2b. Chose exponential priors for the parameters

2c. Complete posterior pdf

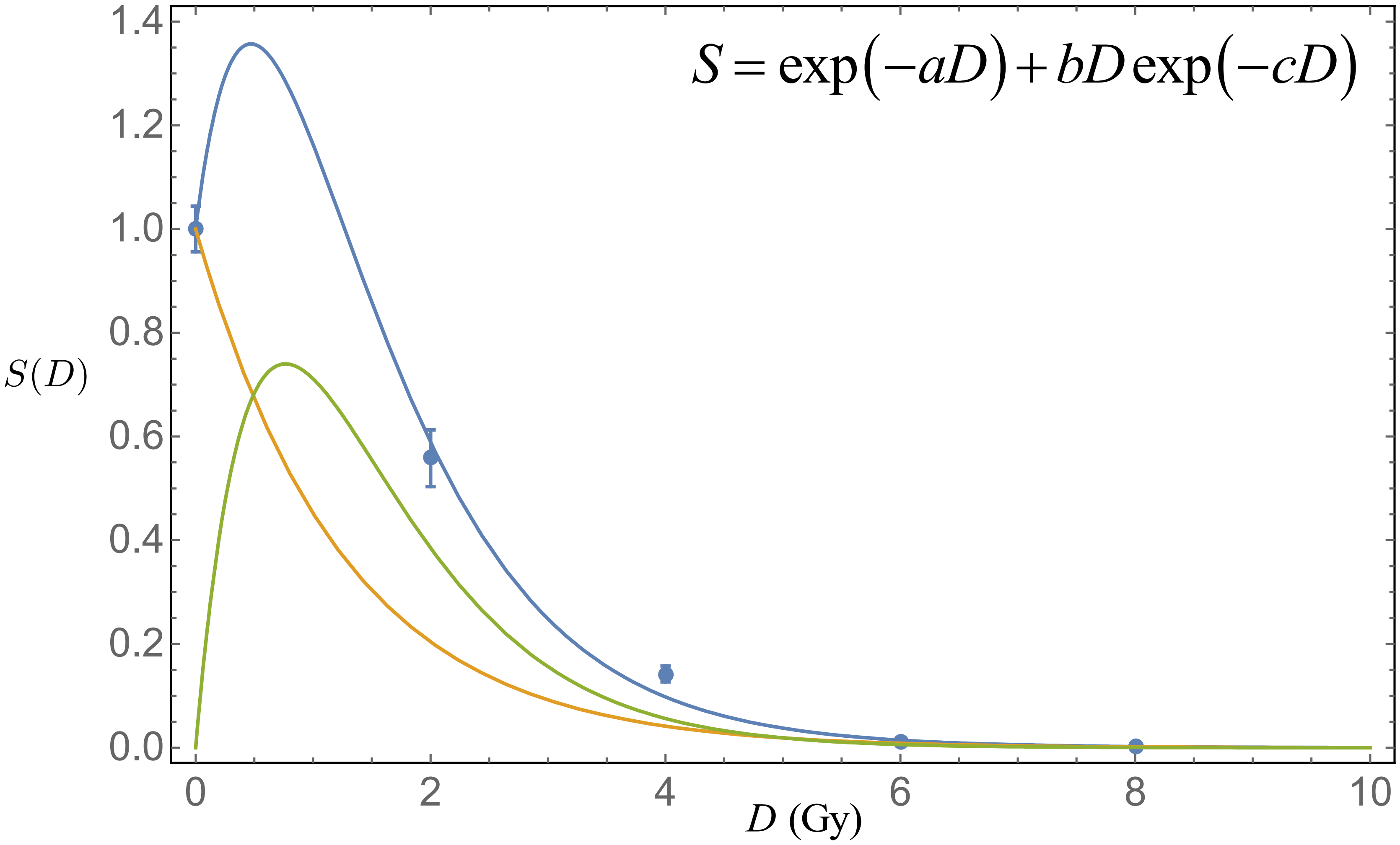
$$p(a,b,c|\{S_k\},I) = \left[\prod_k \exp\left(-\frac{(S_k - S(D_k;a,b,c)^2}{2\sigma_k^2}\right)\right] \overbrace{\exp(-0.2a)\exp(-0.2b)\exp(-0.2c)}^{\text{yet another arbitrary choice of priors}}$$

2d. Use MCMC to find the MAP estimate (and any moment of the pdf)

Path in (a,b,c) space

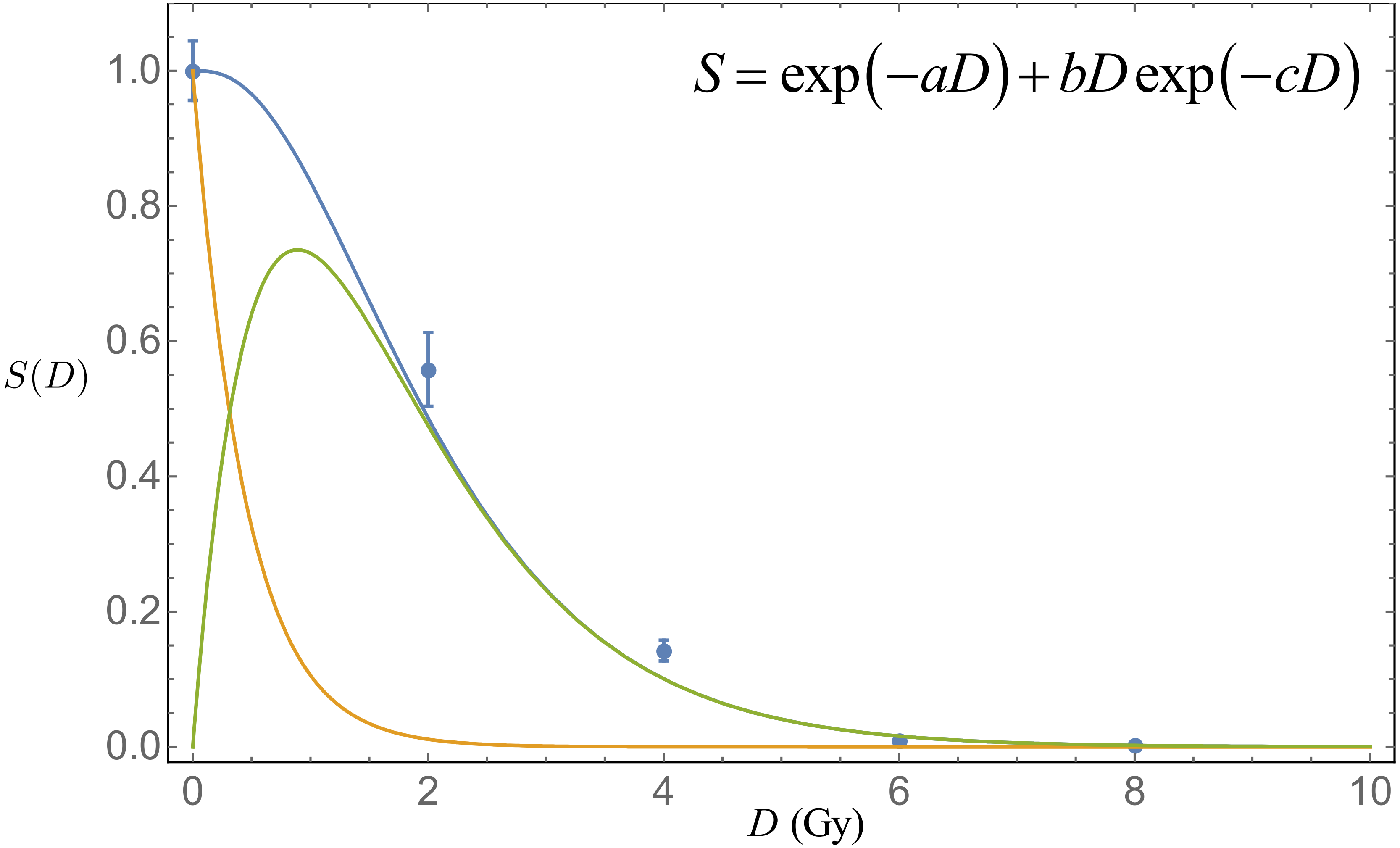


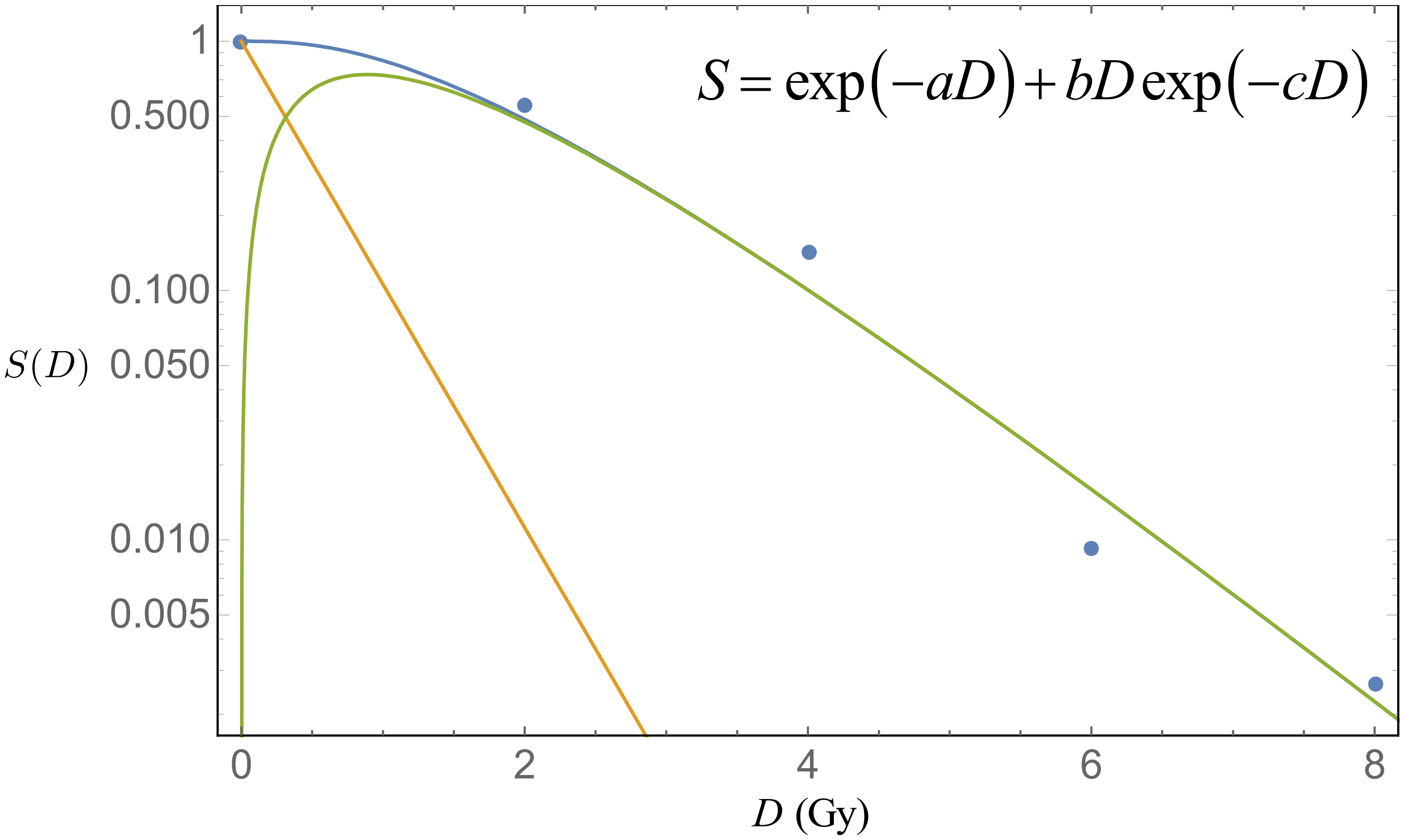
Fit showing individual components: unsatisfactory result



Revise priors to include constraint on derivative

(priors vanish where derivative in the origin is positive)





$S = \exp(-aD) + bD\exp(-cD)$
1
0.500
0.100
$S(D)$ 0.050
0.010
0.005
0
2
4
6
8
D (Gy)