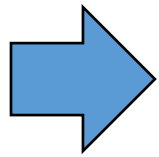


# Introduction to Bayesian Statistics - 6

*Edoardo Milotti*

Università di Trieste and INFN-Sezione di Trieste

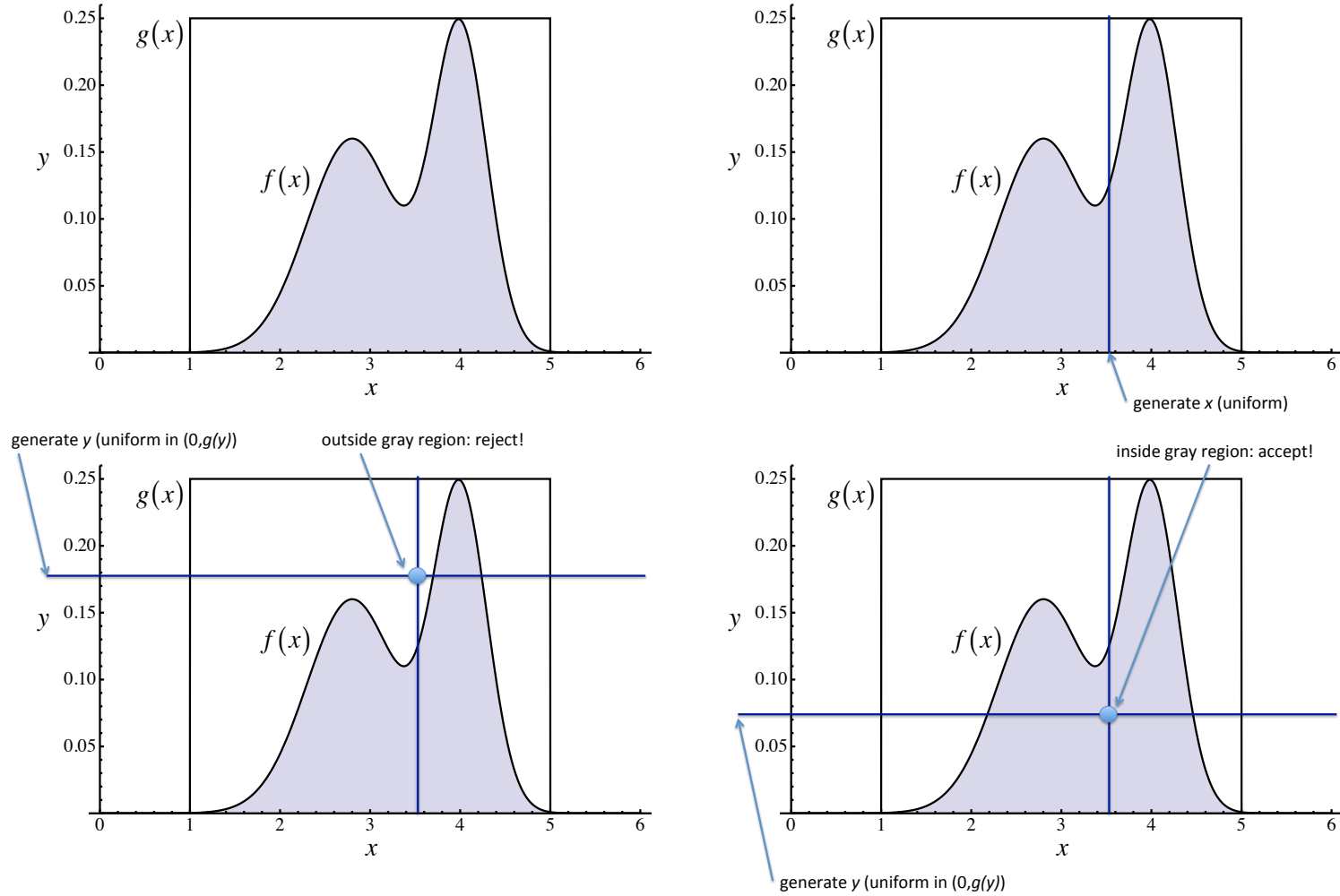
Bayesian estimates often require the evaluation of complex integrals. Usually these integrals can only be evaluated with numerical methods.



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 the Metropolis algorithm
6. Markov Chain Monte Carlo (MCMC)

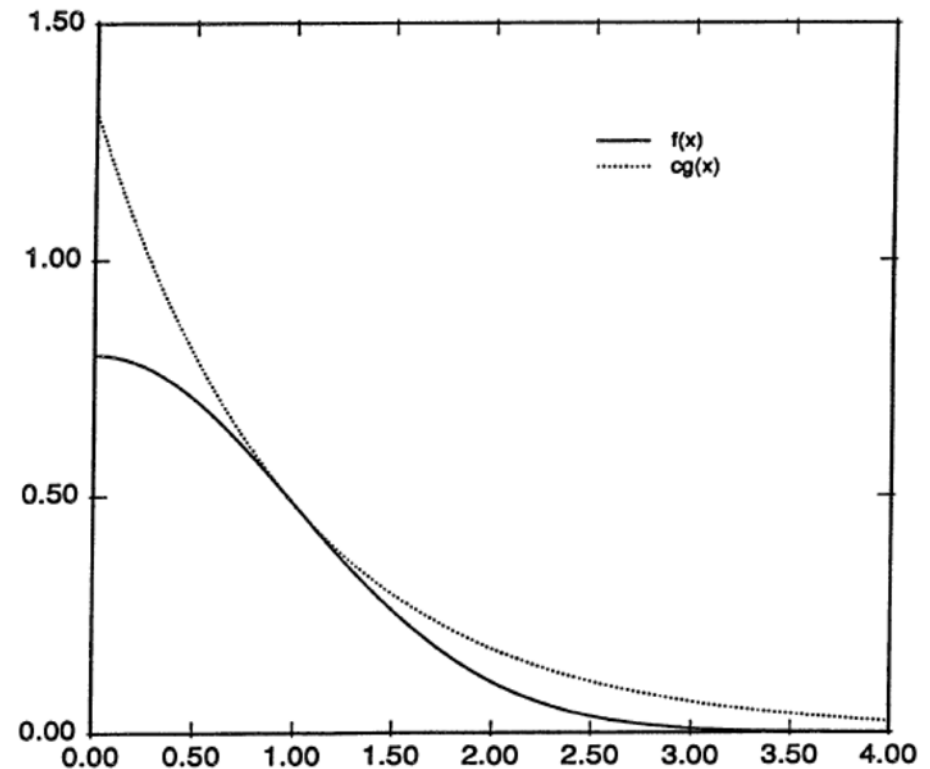
# 1. The acceptance rejection method



Example: random numbers with semi-Gaussian distribution from exponentially distributed random numbers.

$$f(x) = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2}\right) \quad x \geq 0$$

$$g(x) = \exp(-x)$$



## Definition of contact point (to maximize efficiency)

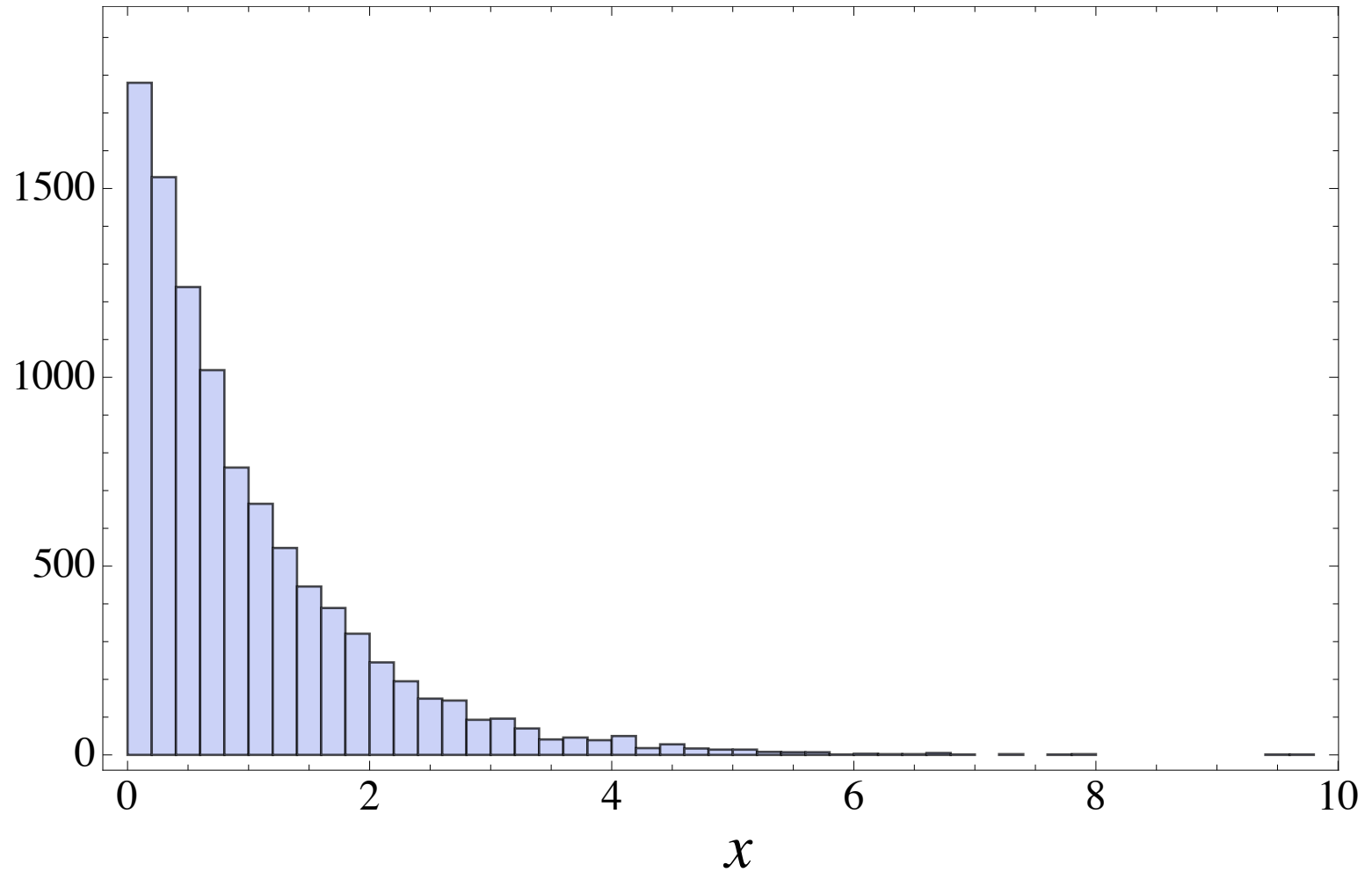
$$f(x) = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2}\right) \quad x \geq 0$$

$$g(x) = \exp(-x)$$

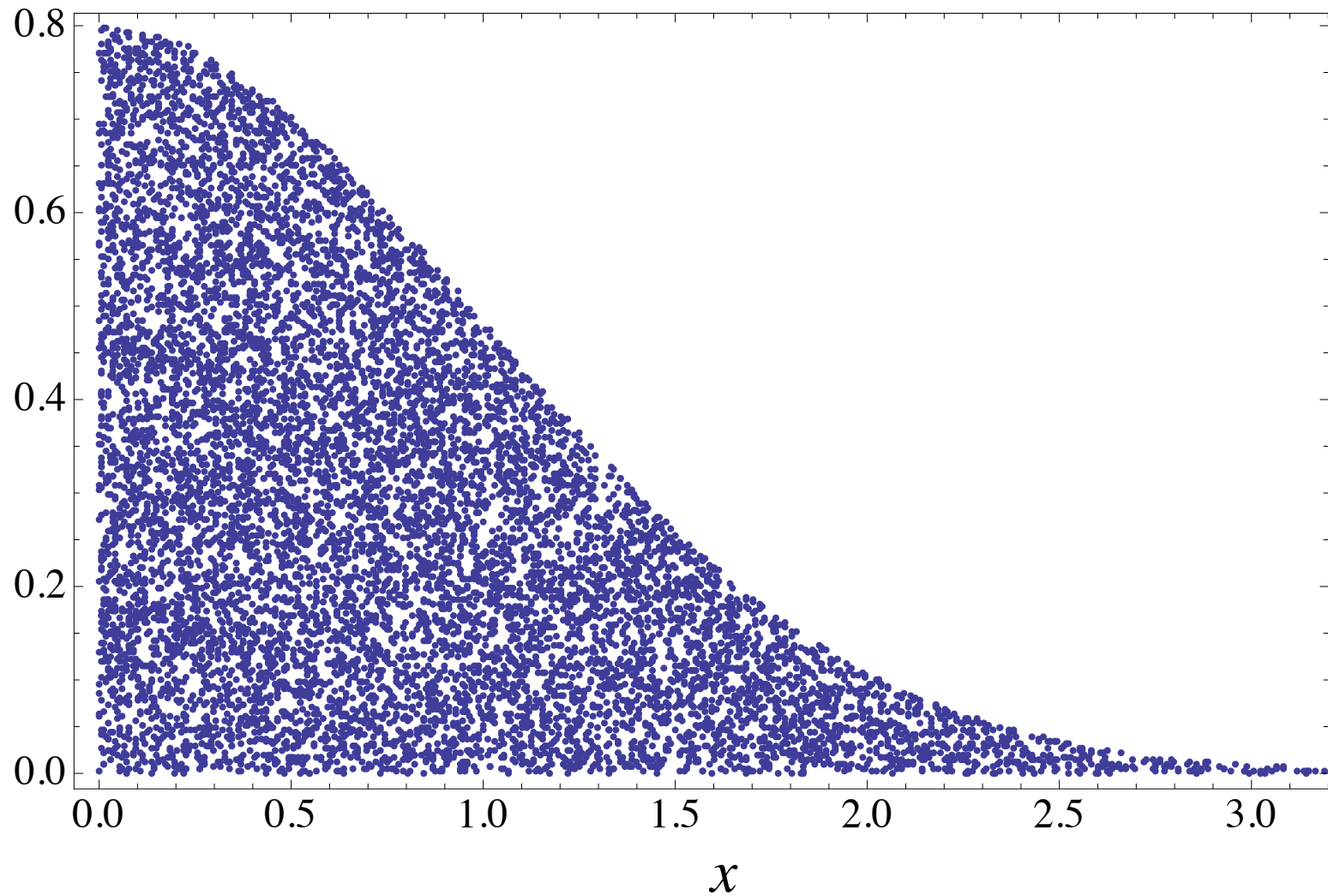
$$\Rightarrow \begin{cases} f(x) = cg(x) \\ f'(x) = cg'(x) \end{cases} \Rightarrow \begin{cases} \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2}\right) = c \exp(-x) \\ x \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2}\right) = c \exp(-x) \end{cases}$$

$$\Rightarrow x = 1; \quad c = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2} + x\right) \approx 1.31549$$

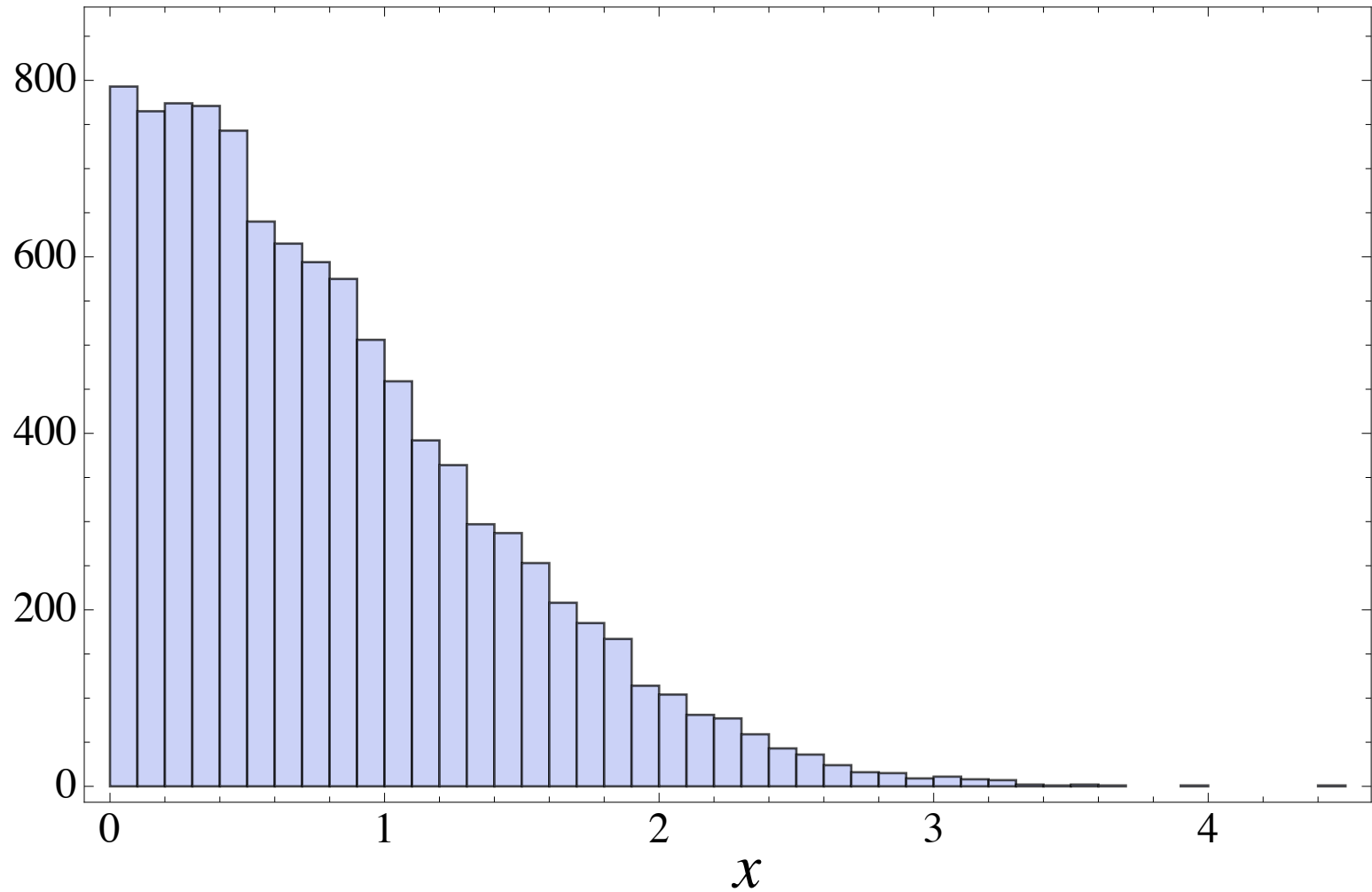
# Exponentially distributed values



# A/R accepted values (10000 accepted sample pairs)

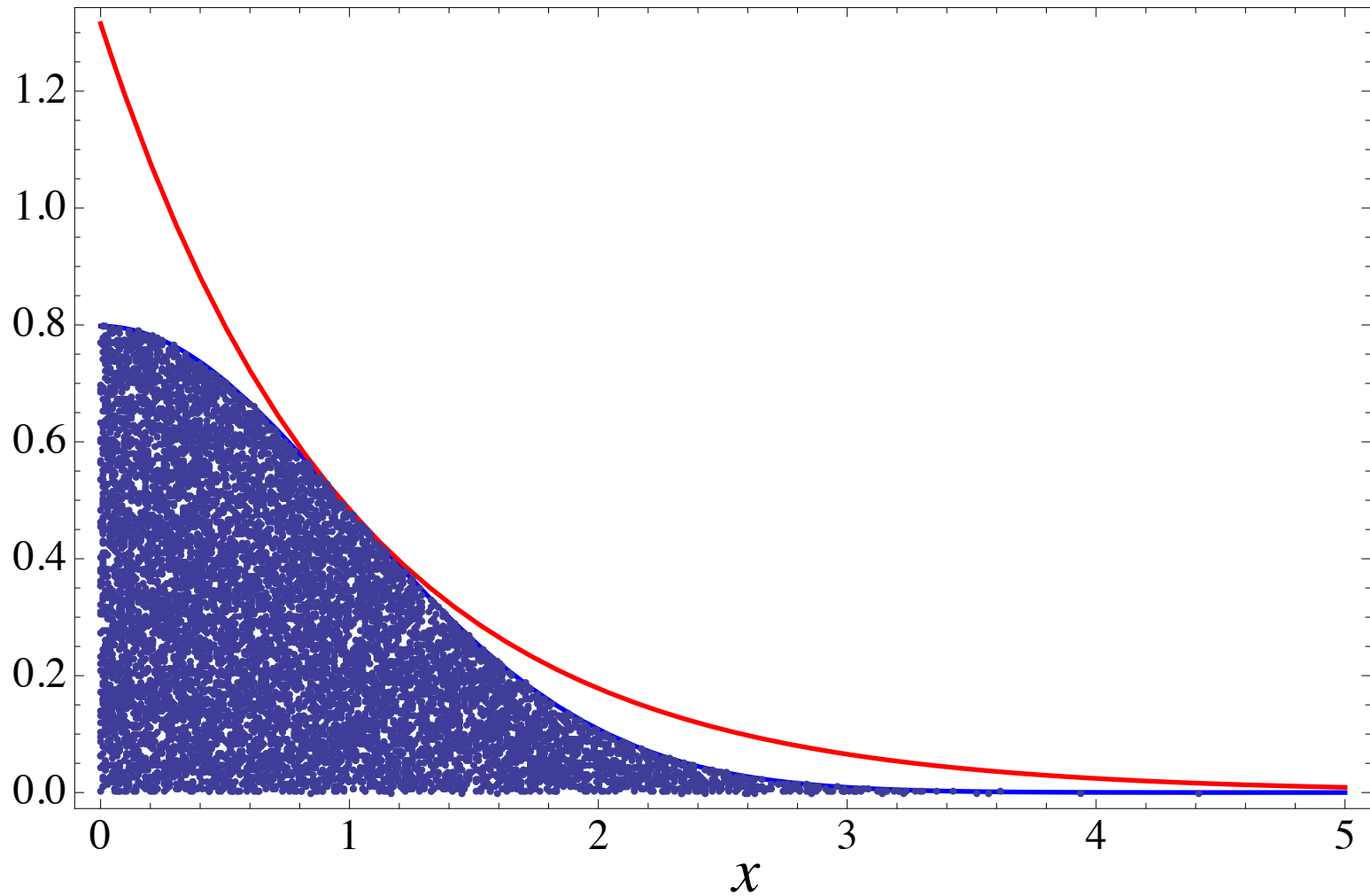


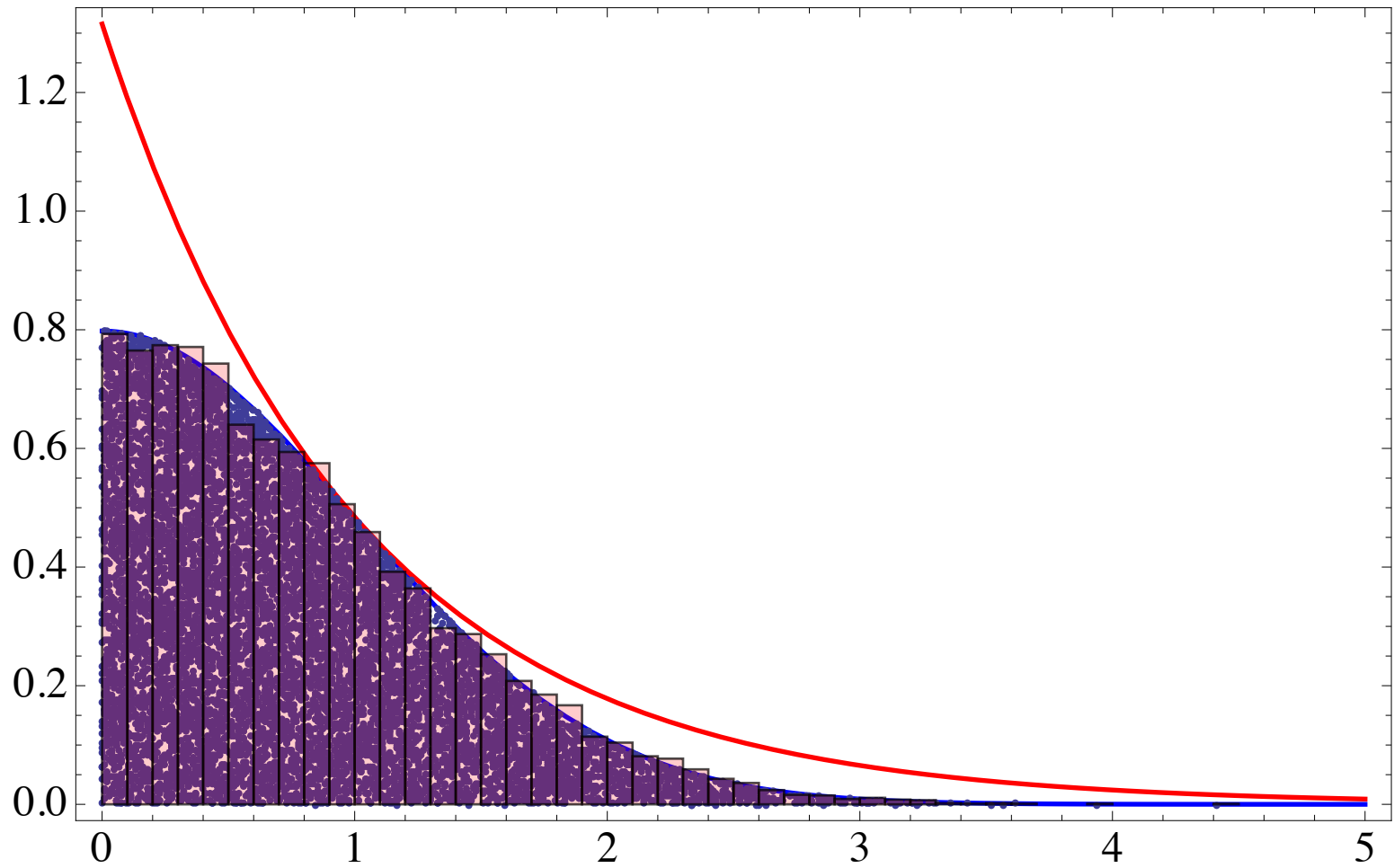
# Histogram of accepted $x$ values





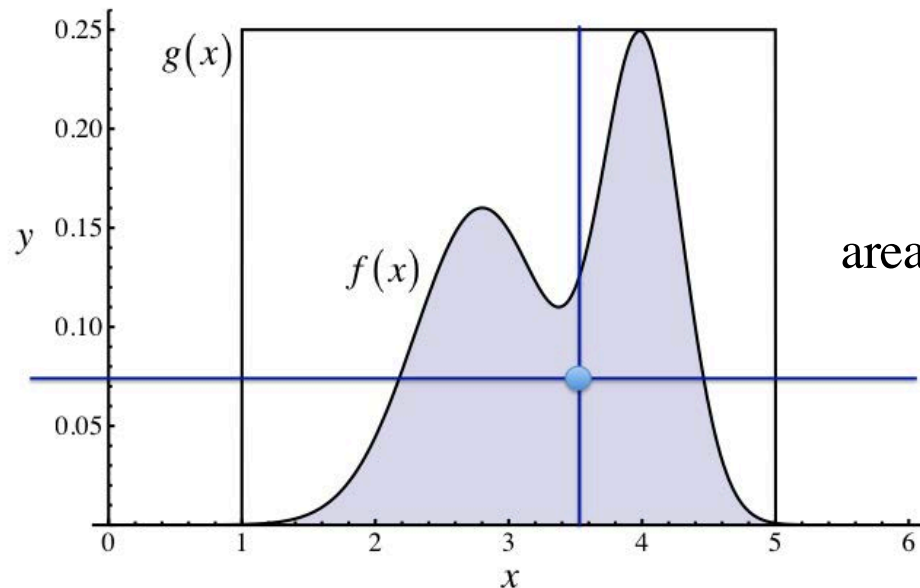
# Comparison with the original distributions





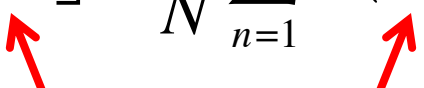
Now notice that in this method we generate pairs of real numbers  $(u, \theta)$  that are uniformly distributed between  $f(\theta)$  and the x-axis, therefore we can use these pairs to estimate the total area under the curve

(here the reference area is the area of the enclosing rectangle which corresponds to a uniform distribution)



$$\text{area} = \frac{\# \text{ of accepted pairs}}{\# \text{ of pairs}} \text{reference area}$$

In general, if  $h(x) = f(x)p(x)$ , where  $p$  is a pdf

$$\int_a^b h(x) dx = \int_a^b f(x)p(x) dx = E_p[f(x)] \approx \frac{1}{N} \sum_{n=1}^N f(x_n)$$


here the  $x$  are i.i.d with pdf  $p(x)$

and we find that the variance of this estimate of the integral is

$$\frac{1}{N} \left\{ \frac{1}{N-1} \sum_{n=1}^N [f(x_n) - E_p[f(x)]]^2 \right\}$$

**We encounter a problem with this method when we must sample functions that have many narrow peaks.**

## 2. Importance sampling

this pdf is troublesome ...

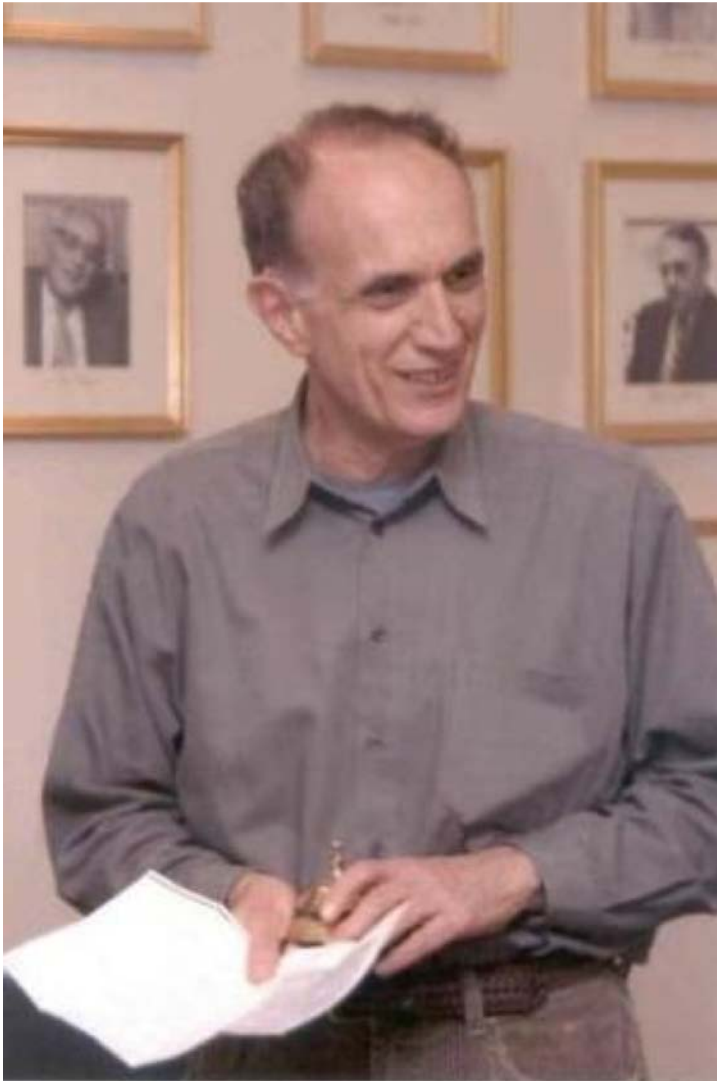
therefore we use this ...

$$\int_a^b h(x) dx = \int_a^b f(x) p(x) dx = \int_a^b \left[ f(x) \frac{p(x)}{q(x)} \right] q(x) dx$$
$$= E_q \left[ f(x) \frac{p(x)}{q(x)} \right] \approx \frac{1}{N} \sum_{n=1}^N f(x_n) \frac{p(x_n)}{q(x_n)}$$

here the  $x$  are i.i.d with pdf  $q(x)$

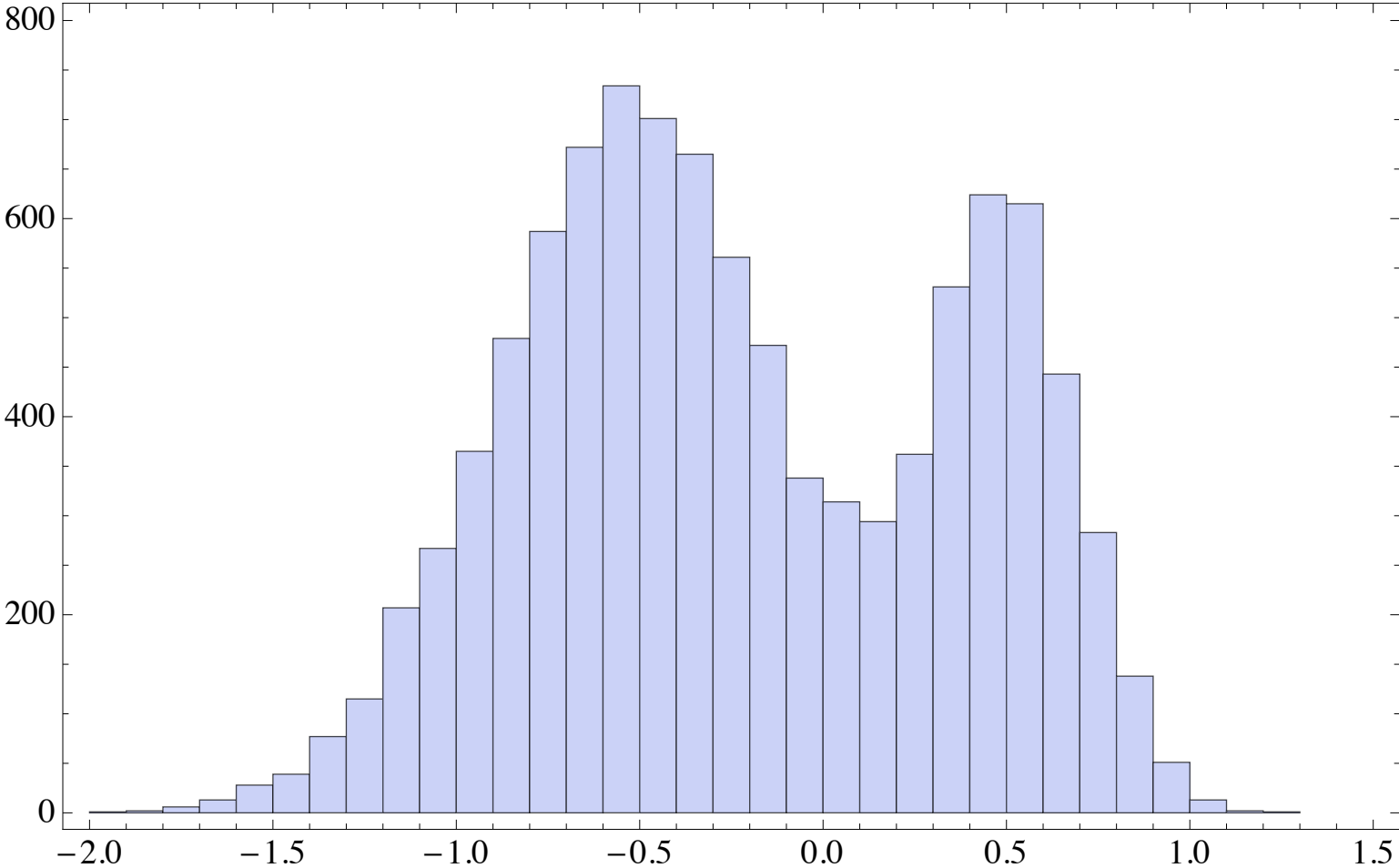
**These methods are still not very efficient and there is a better alternative, the Markov Chain Monte Carlo method**

### 3. *Bootstrap (B. Efron, 1977) and the importance of edf's*

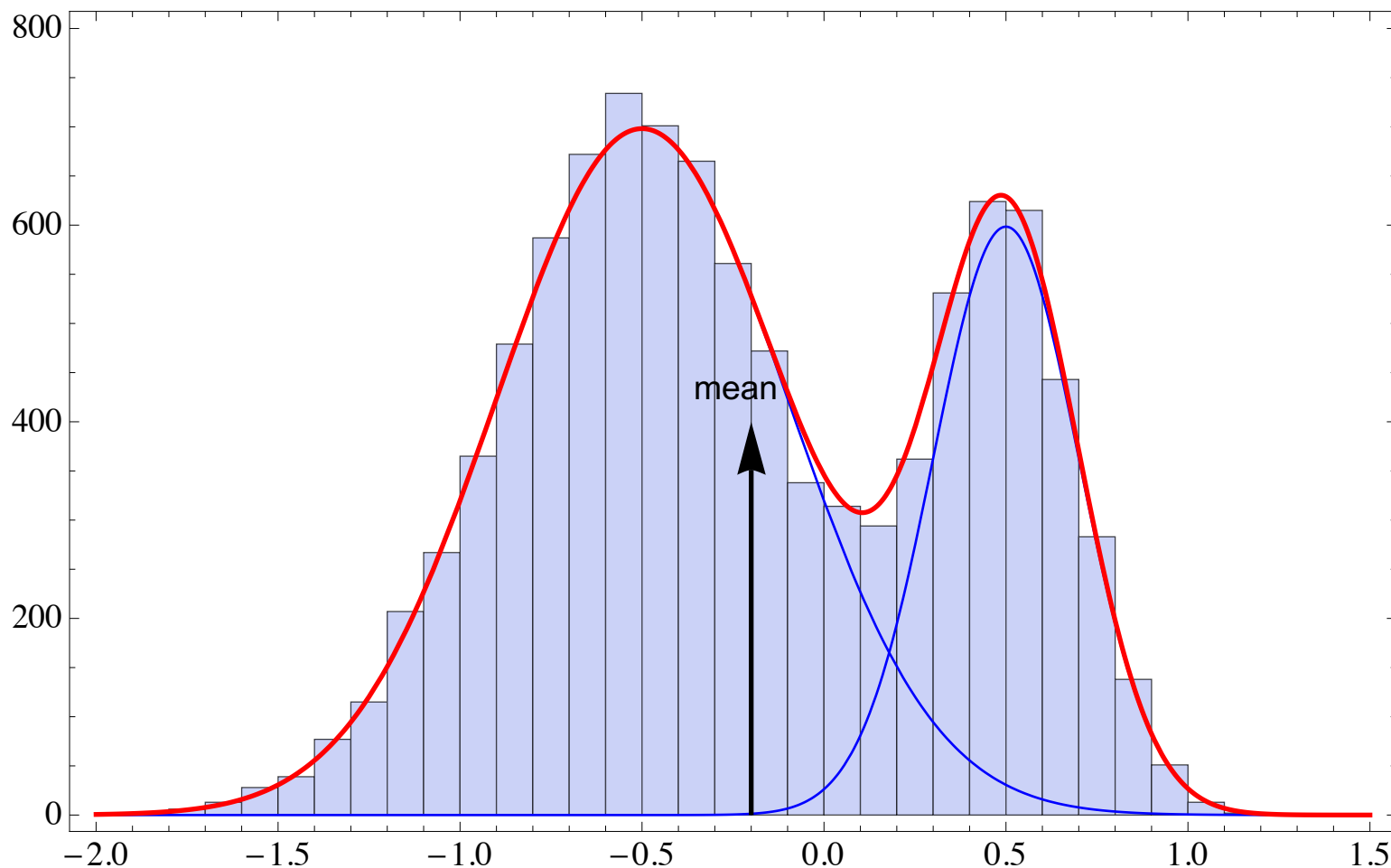


The bootstrap method is a resampling technique that helps calculate many statistical estimators

consider the distribution of a set of measurements

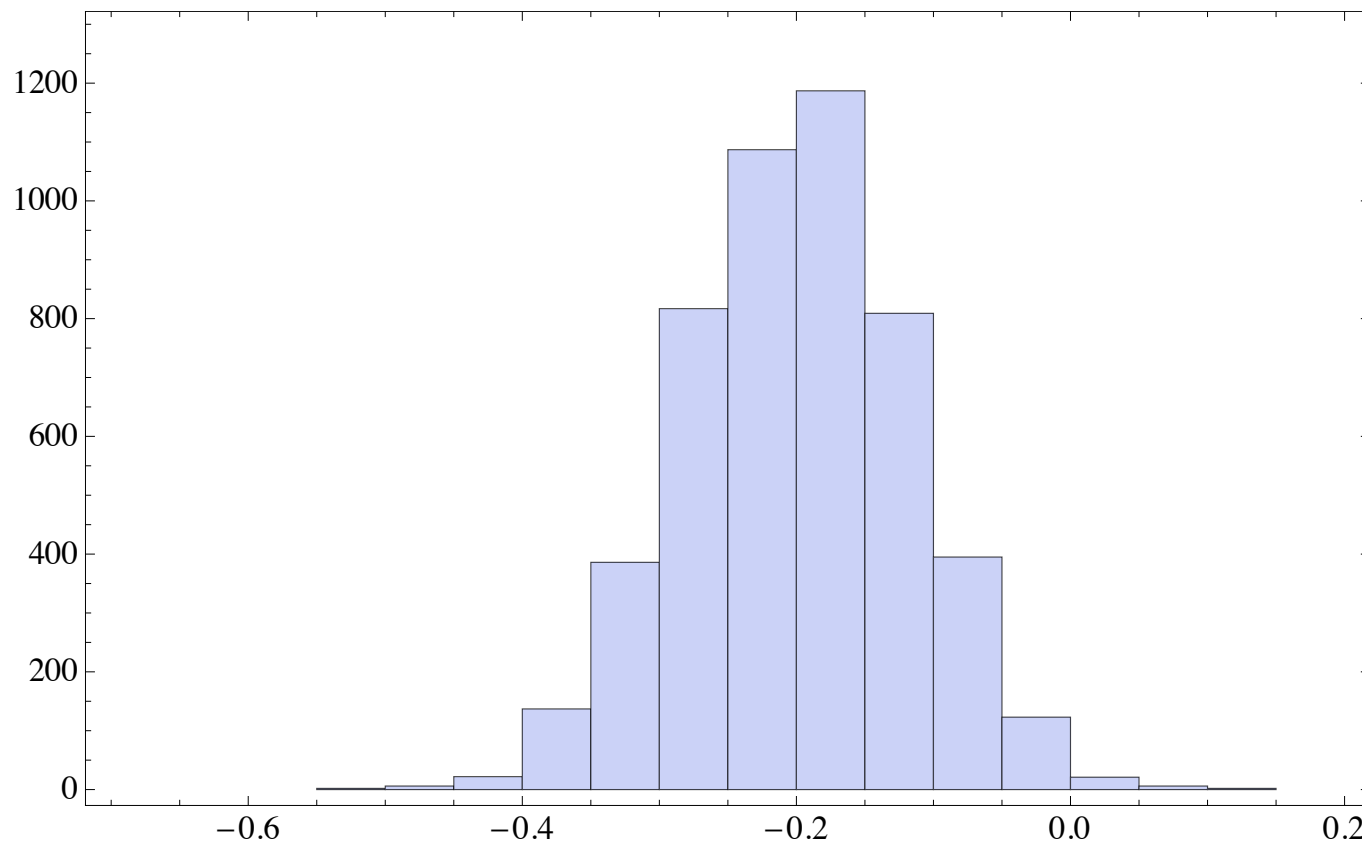


the distribution of data is an approximation of the “true” underlying distribution (in this case a mixture model)



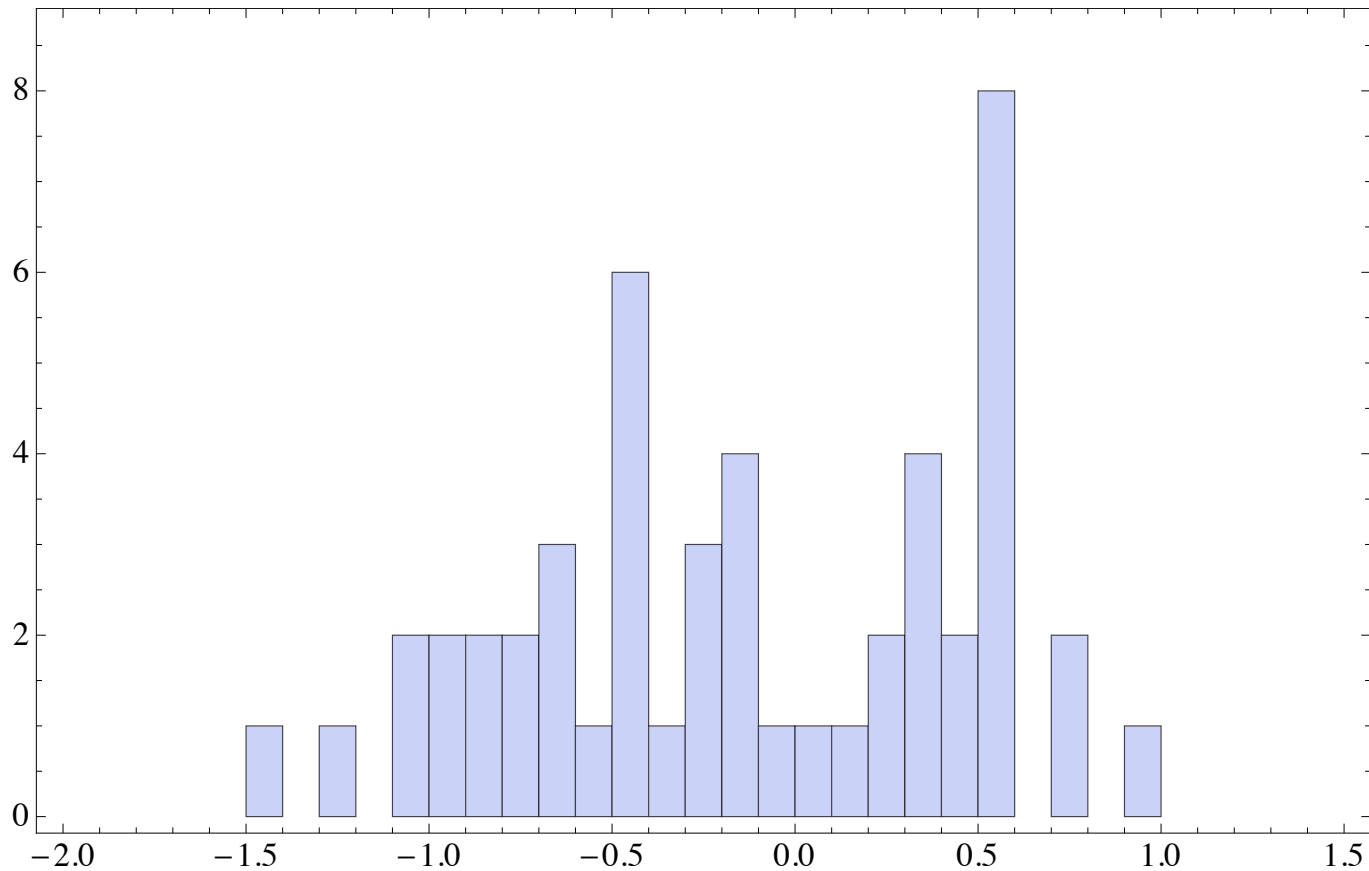


distribution of mean value obtained from 5000 sets of data (sample size = 50)



You can do this if you have large datasets ... but what if you have only a handful of measurements?

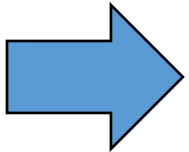
example: single dataset (same size as before, 50 measurements)



the distribution is a rough representation of the underlying distribution ... and yet it can be used just as before ...

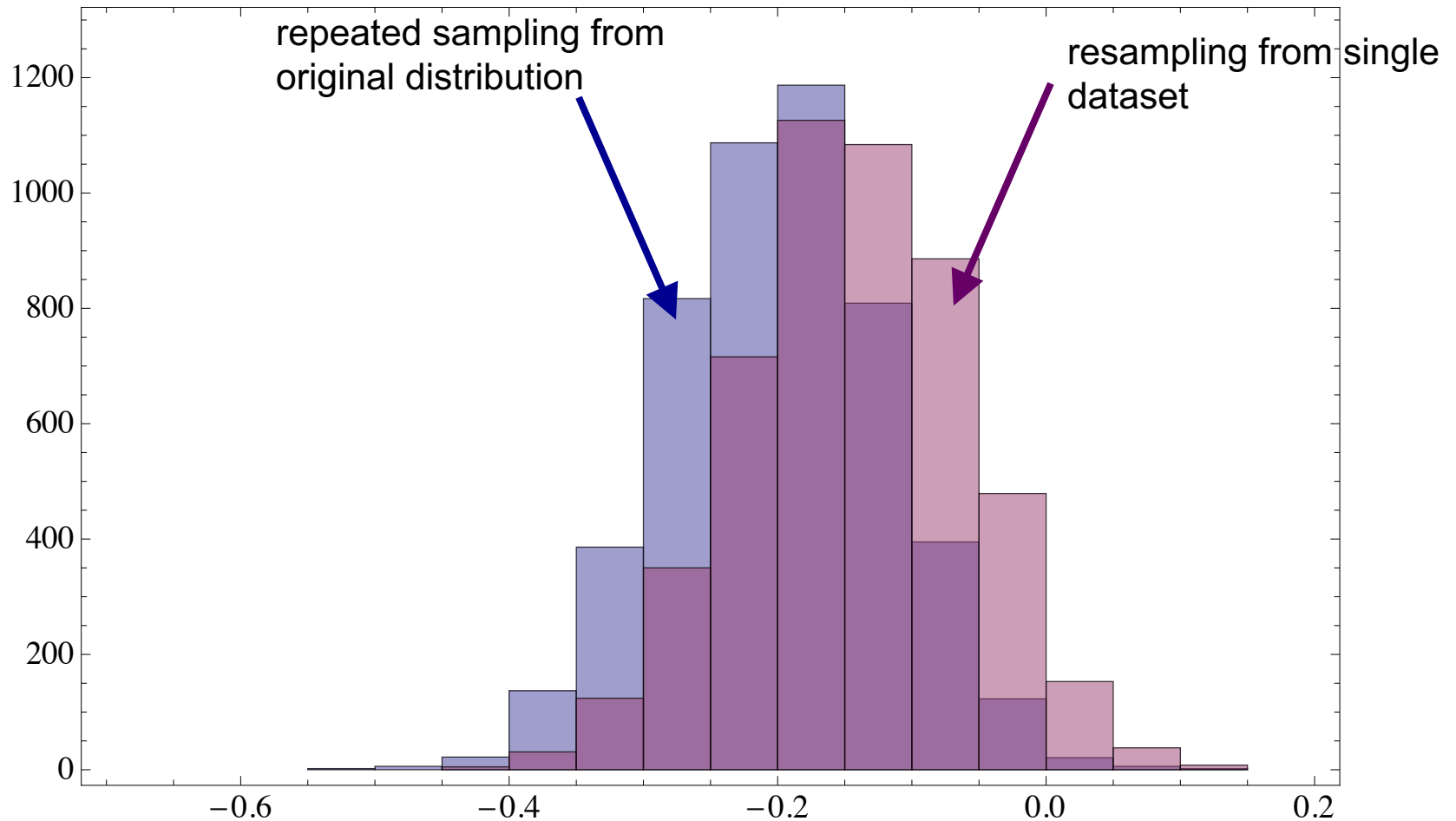
## Bootstrap recipe:

if you want to find the distribution of the mean (or any other statistical estimator) use the dataset itself to generate new datasets



resample from dataset (with replacement)

# distribution of mean value



true mean: -0.2

mean from repeated sampling (size = 250000):  $-0.200222 \pm 0.0813632$

mean from resampling dataset (size = 50):  $-0.142699 \pm 0.0838678$

counts of CD4 lymphocytes

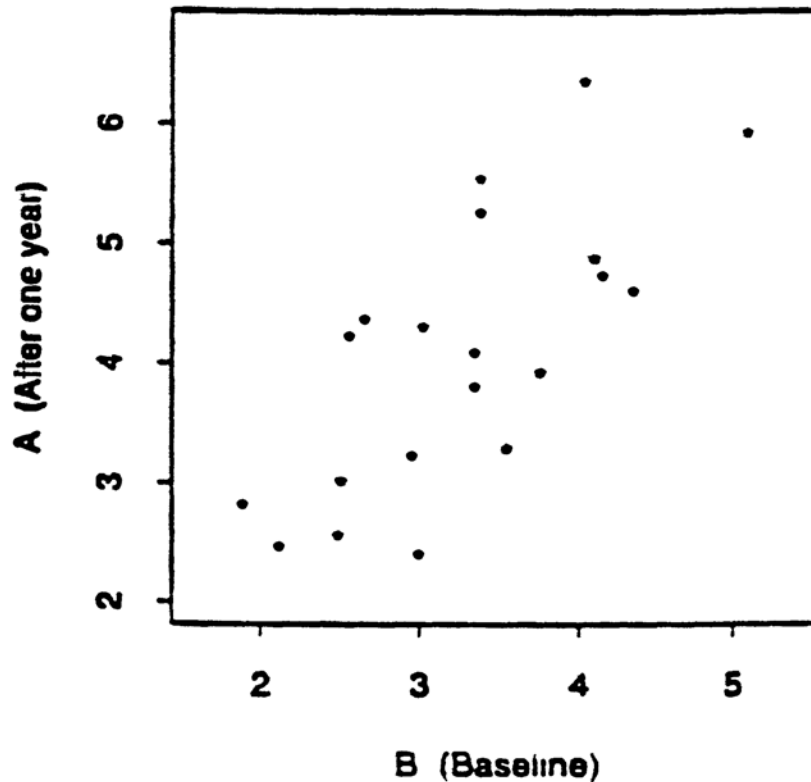


FIG. 1. *The cd4 data; cd4 counts in hundreds for 20 subjects, at baseline and after one year of treatment with an experimental anti-viral drug; numerical values appear in Table 1.*

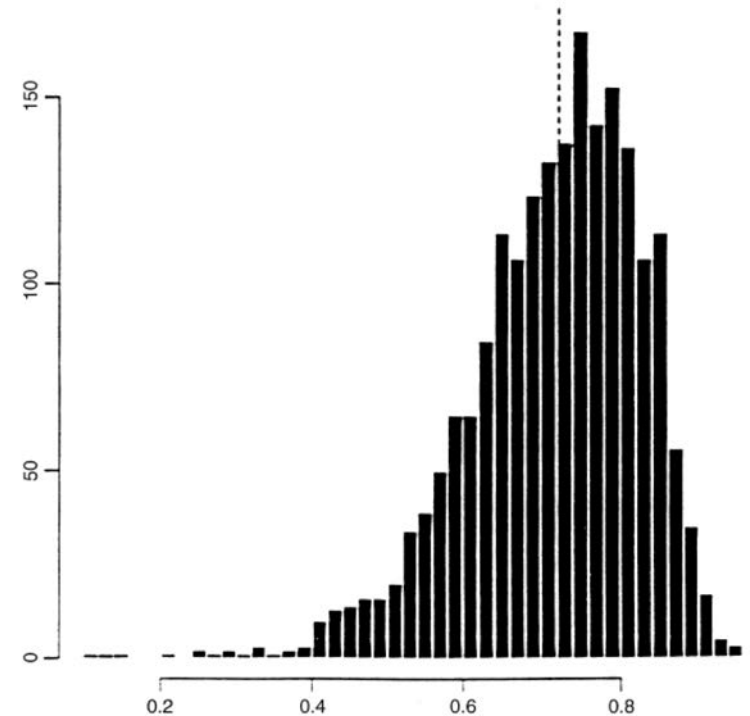


FIG. 3. *Histogram of 2,000 bootstrap correlation coefficients; bivariate normal sampling model.*

bootstrap estimate of correlation coefficient distribution

Example from Di Ciccio & Efron, *Statistics of Science* **11** (1996) 189 and Efron, *Statistics of Science* **13** (1998) 95

4. *Bayesian methods in a sampling-resampling perspective*  
(Smith & Gelfand, 1992)

# **Bayesian Statistics Without Tears: A Sampling–Resampling Perspective**

A. F. M. SMITH and A. E. GELFAND\*

---

Even to the initiated, statistical calculations based on Bayes's Theorem can be daunting because of the numerical integrations required in all but the simplest applications. Moreover, from a teaching perspective, introductions to Bayesian statistics—if they are given at all—are circumscribed by these apparent calculational difficulties. Here we offer a straightforward sampling–resampling perspective on Bayesian inference, which has both pedagogic appeal and suggests easily implemented calculation strategies.

*In Bayesian methods we have to evaluate many integrals, like, e.g.,*

$$p(\theta|x) = \frac{l(\theta; x)p(\theta)}{\int l(\theta; x)p(\theta) d\theta} \leftarrow \text{normalization (evidence)}$$

$$p(\phi|x) = \int p(\phi, \psi|x) d\psi. \leftarrow \text{marginalization}$$

$$E[m(\theta)|x] = \int m(\theta)p(\theta|x) d\theta \leftarrow \text{averages (statistical estimators)}$$

except in simple cases, explicit evaluation of such integrals will rarely be possible, and realistic choices of likelihood and prior will necessitate the use of sophisticated numerical integration or analytic approximation techniques (see, for example, Smith et al. 1985, 1987; Tierney and Kadane, 1986). This can pose problems for the applied practitioner seeking routine, easily implemented procedures. For the student, who may already be puzzled and discomforted by the intrusion of too much calculus into what ought surely to be a simple, intuitive, statistical learning process, this can be totally off-putting.



# Bayesian learning as a resampling procedure

$$p(\theta|x) = \frac{l(\theta;x)}{\int l(\theta;x)p(\theta)d\theta} p(\theta)$$

1. prior distribution defined by initial samples

2. Bayes factor distorts the distribution of initial samples

3. posterior distribution corresponds to a resampling of initial samples

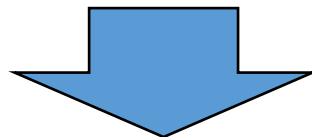
Example (McCullagh & Nelder): take two sets of binomially distributed independent random variables  $X_{i1}$  and  $X_{i2}$  ( $i=1,2,3$ )

$$X_{i1} = \text{Binomial}(n_{i1}, \theta_1)$$

$$X_{i2} = \text{Binomial}(n_{i2}, \theta_2)$$

The observed random variables are the sums

$$Y_i = X_{i1} + X_{i2}$$



$$\text{likelihood} = \prod_{i=1}^3 \sum_{j_i} \binom{n_{i1}}{j_i} \binom{n_{i2}}{y_i - j_i} \theta_1^{j_i} (1 - \theta_1)^{n_{i1} - j_i} \theta_2^{y_i - j_i} (1 - \theta_2)^{n_{i2} - (y_i - j_i)}$$

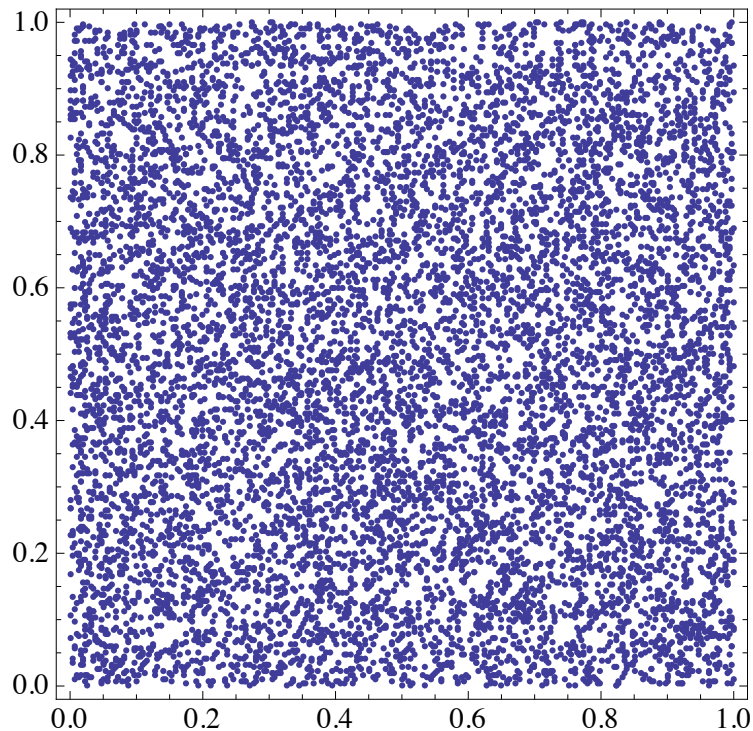
$$\max(0, y_i - n_{i2}) \leq j_i \leq \min(n_{i1}, y_i)$$

# Sample data

	<b>1</b>	<b>2</b>	<b>3</b>
$n_{i1}$	5	6	4
$n_{i2}$	5	4	6
$y_i$	7	5	6

# Example of implementation in *Mathematica*

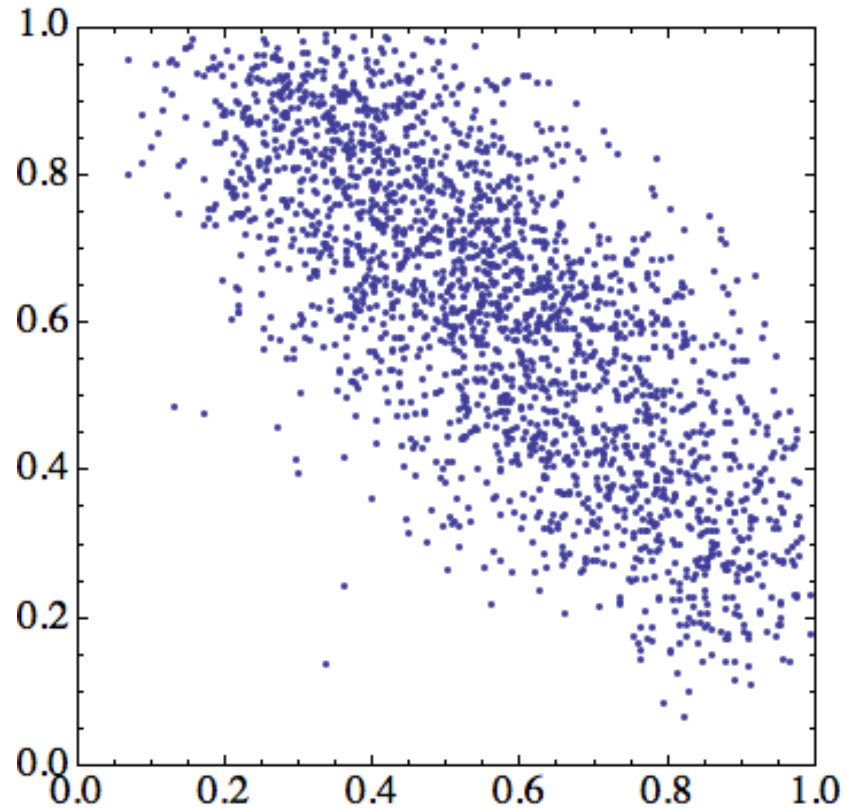
```
n1 = {5, 6, 4};  
n2 = {5, 4, 6};  
yi = {7, 5, 6};  
  
Clear[likelihood];  
likelihood[th1_, th2_] :=  
Product[Sum[Binomial[n1[[i]], j] * Binomial[n2[[i]], yi[[i]] - j] * th1^j * (1 - th1)^(n1[[i]] - j) *  
th2^(yi[[i]] - j) * (1 - th2)^(n2[[i]] - yi[[i]] + j), {j, Max[0, yi[[i]] - n2[[i]], Min[n1[[i]], yi[[i]]]}],  
{i, 1, 3}];  
  
ns = 10000;  
th = Table[{RandomReal[], RandomReal[]}, {ns}];
```



prior distribution (uniform in  
2D parameter space)

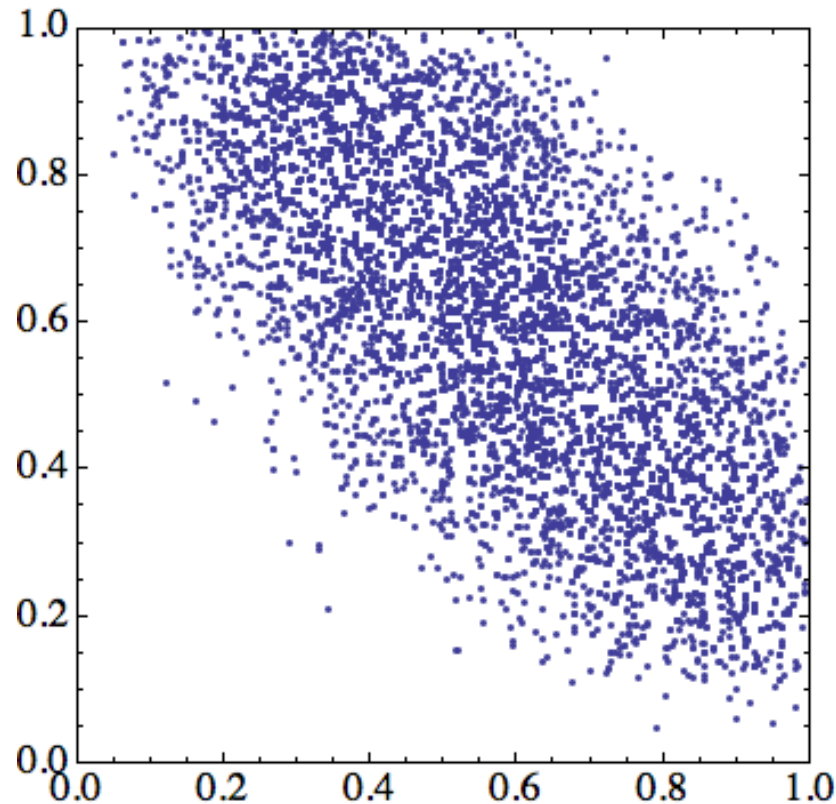
# Posterior as a resampled prior using acceptance-rejection

```
lt = Table[likelihood[th[[k, 1]], th[[k, 2]]], {k, 1, ns}];  
norm = Max[lt];  
w = lt / norm;  
  
thr = {}; ntot = 0;  
For[kn = 1, kn ≤ ns,  
  If[w[[kn]] > RandomReal[], ntot++; AppendTo[thr, th[[kn]]];  
  kn++]
```

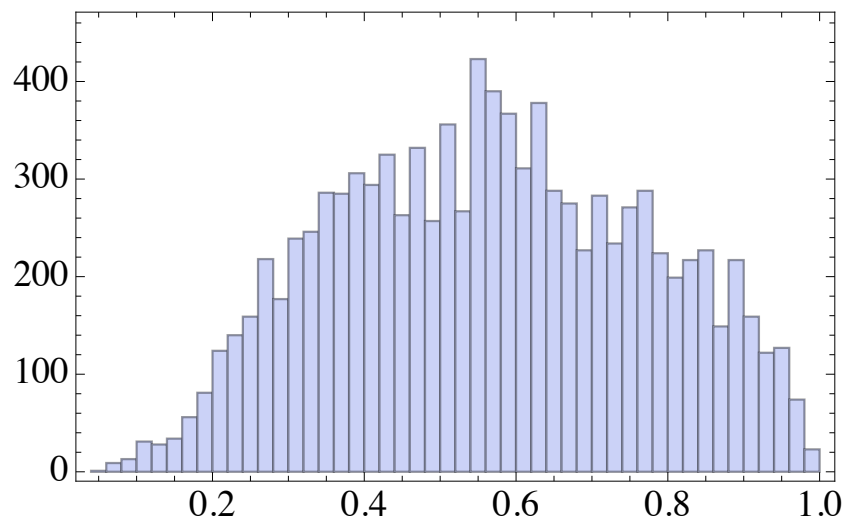


# Posterior as a resampled prior using weighted bootstrap

```
lt = Table[likelihood[th[[k, 1]], th[[k, 2]]], {k, 1, ns}];  
sum = Apply[Plus, lt];  
w = lt / sum;  
  
thr = Table[{0, 0}, {ns}];  
ntot = 0;  
While[ntot < ns,  
  kn = RandomInteger[{1, ns}];  
  If[RandomReal[] < w[[kn]], ntot++; thr[[ntot]] = th[[kn]]];  
]
```

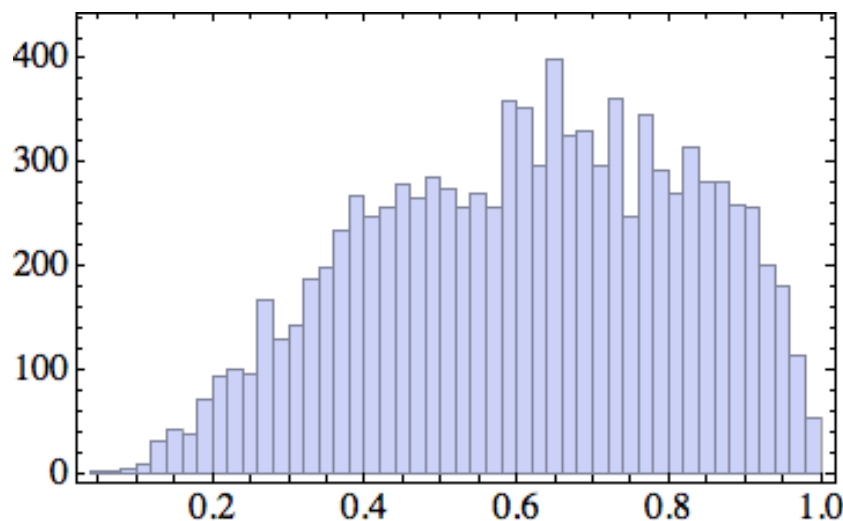


The resampled points are representative of the posterior distribution and can be used to evaluate any sample estimate



Marginalized distribution of  $\theta_1$

Sample mean:  $0.564 \pm 0.002$



Marginalized distribution of  $\theta_2$

Sample mean:  $0.613 \pm 0.002$

## 5. *Very short introduction to Markov chains*

Consider a system such that:

- the system can occupy a finite or countably infinite set of states  $S_n$ ;
- the system changes state randomly at discrete times  $t = 1, 2, \dots$ ;
- if the system is in state  $S_i$ , then the probability that the system goes into state  $S_j$  is

$$P\left(S_{t+1} = s_j \mid S_t = s_k, \dots, S_{t-n} = s_l, \dots\right)$$

i.e., this probability depends only on the previous state, and is independent of all previous states (**this is the Markov property**);

- the transition probabilities  $p_{ij}$  do not depend on time  $n$ .
- **Such a system is a special type of discrete time stochastic process, which is called Markov chain.**



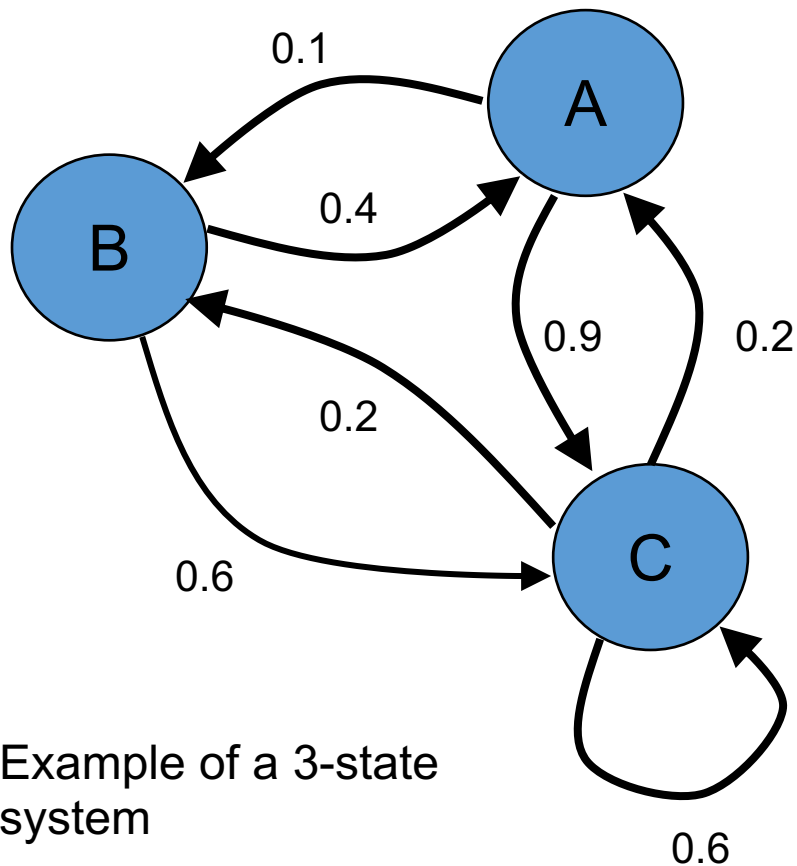
In general we should have

$$P\left(S_{t+1} = s_j \mid S_t = s_k, \dots, S_{t-n} = s_l, \dots\right)$$

however the Markov property tells us that only the previous state is important in determining the next state

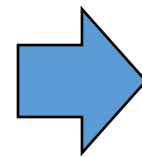
$$P\left(S_{t+1} = s_j \mid S_t = s_k, \dots, S_{t-n} = s_l, \dots\right) = P\left(S_{t+1} = s_j \mid S_t = s_k\right) = P(k \rightarrow j)$$

Now let  $\pi_j(t)$  be the probability that the system is in state  $j$  at time  $t$ :



Example of a 3-state system

$$\begin{aligned}\pi_j(t+1) &= \sum_k P(k \rightarrow j) \pi_k(t) \\ &= \sum_k T_{kj} \pi_k(t) \\ \boldsymbol{\pi}(t+1) &= \boldsymbol{\pi}(t) \mathbf{T}\end{aligned}$$



$$\mathbf{T} = \begin{pmatrix} 0 & 0.1 & 0.9 \\ 0.4 & 0 & 0.6 \\ 0.2 & 0.2 & 0.6 \end{pmatrix}$$

transition matrix; it belongs to the class of “stochastic matrices” (rows add to 1)

The following equation also holds

$$\boldsymbol{\pi}(t+1) = \boldsymbol{\pi}(t)\mathbf{T} = \boldsymbol{\pi}(t-1)\mathbf{T}^2$$

and more generally

$$\boldsymbol{\pi}(t+k) = \boldsymbol{\pi}(t)\mathbf{T}^k$$

Moreover we find

$$\boldsymbol{\pi}(t+k+m) = \boldsymbol{\pi}(t+k)\mathbf{T}^m = \boldsymbol{\pi}(t)\mathbf{T}^k\mathbf{T}^m = \boldsymbol{\pi}(t)\mathbf{T}^{k+m}$$

and therefore

discrete version of the Chapman-Kolmogorov eq.

$$\mathbf{T}^{k+m} = \mathbf{T}^k\mathbf{T}^m \quad \rightarrow$$

$$T_{ij}^{(k+m)} = \sum_n T_{in}^{(k)} T_{nj}^{(m)}$$

It can be shown that Markov chains have a stationary distribution

$$\boldsymbol{\pi}^* = \boldsymbol{\pi}^* \mathbf{T}$$

such that the *detailed balance* also holds

$$\pi_i P(i \rightarrow k) = \pi_k P(k \rightarrow i) \quad \text{i.e.} \quad \pi_i T_{ik} = \pi_k T_{ki}$$

Indeed we see that

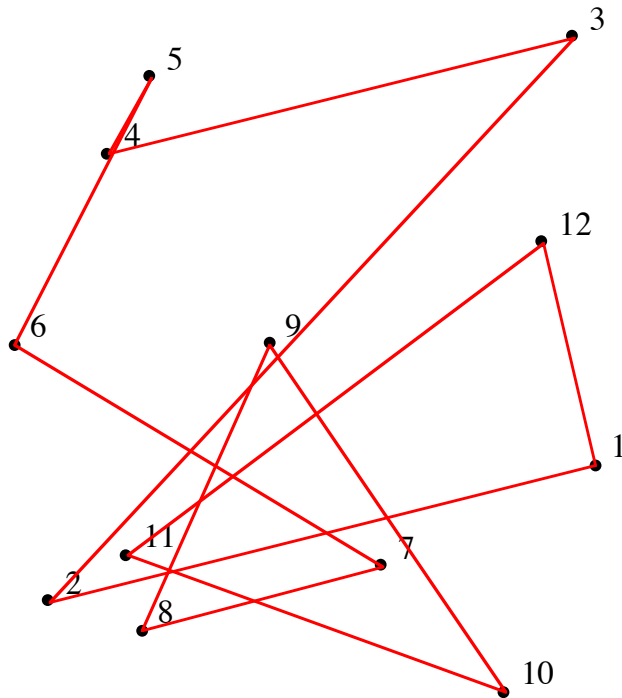
$$\pi_i(t+1) = \sum_k T_{ki} \pi_k(t) = \sum_k T_{ik} \pi_i(t) = \pi_i(t) \sum_k T_{ik} = \pi_i(t)$$

and therefore the distribution is stationary.

*Detailed balance holds if and only if the distribution is stationary.*

Now we consider the *Traveling Salesman Problem* (TSP), where we want to find the shortest closed path that connects  $N$  cities.

The problem was first stated by the Viennese mathematician Karl Menger and is one of the most studied problem in combinatorial optimization.



12 “cities” randomly distributed in the  $(0,1)$  square: the path corresponds to a random permutation of the sequence of cities.

(path length  $L=1.93834$ )

Paths are enumerated by permutations of “city names”, e.g., {9, 2, 7, 8, 1, 12, 4, 5, 3, 10, 11, 6} (start at 9, step to 2, and so on until you reach 6 and then return to 9).

The total number of configurations (undirected paths) is

$$\frac{1}{2}(n-1)!$$

The problem belongs to the class of NP-complete problems (Non-Polynomial complexity, a class of particularly hard problems)

In such cases there is only one known solution: the full enumeration of all paths.

# Approximate solution of the TSP with the Simulated Annealing algorithm

path length  energy of the system

exploration of the configuration space with the *Metropolis algorithm* (Metropolis, Rosenbluth Rosenbluth ,Teller and Teller, 1953)

THE JOURNAL OF CHEMICAL PHYSICS

VOLUME 21, NUMBER 6

JUNE, 1953

## Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,  
*Los Alamos Scientific Laboratory, Los Alamos, New Mexico*

AND

EDWARD TELLER,\* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

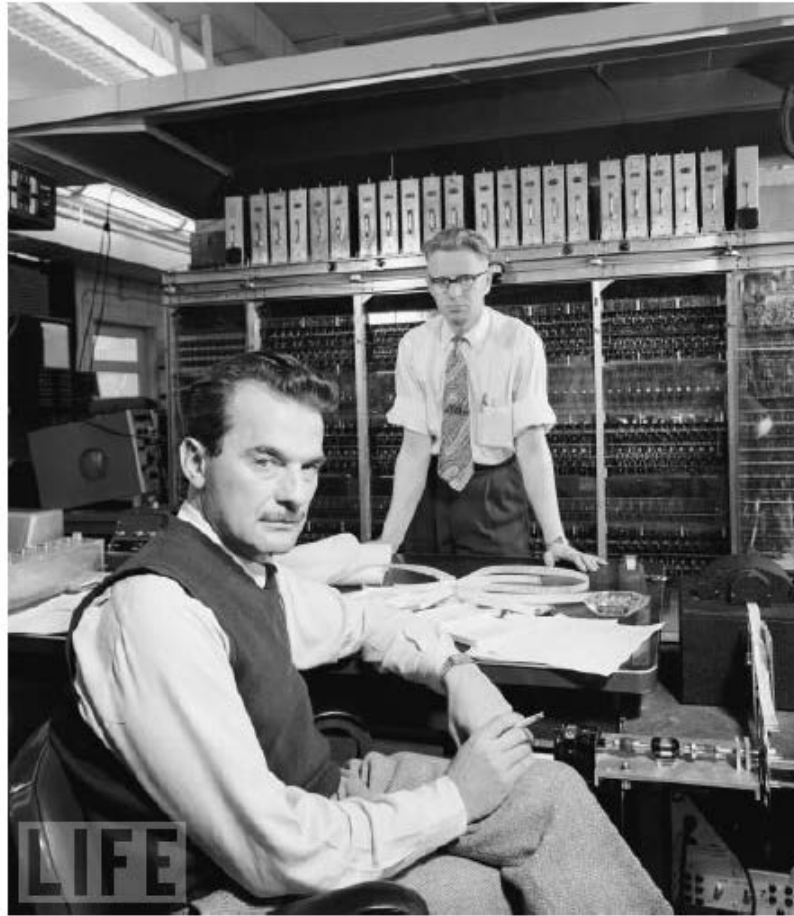


Figure 8.14: Portrait of American computer scientists Nicholas Metropolis (1915 - 1999) (seated) and James Henry Richardson (1918 - 1996) at Los Alamos National Laboratory, Los Alamos, New Mexico, November 1953 (from <http://www.life.com>).

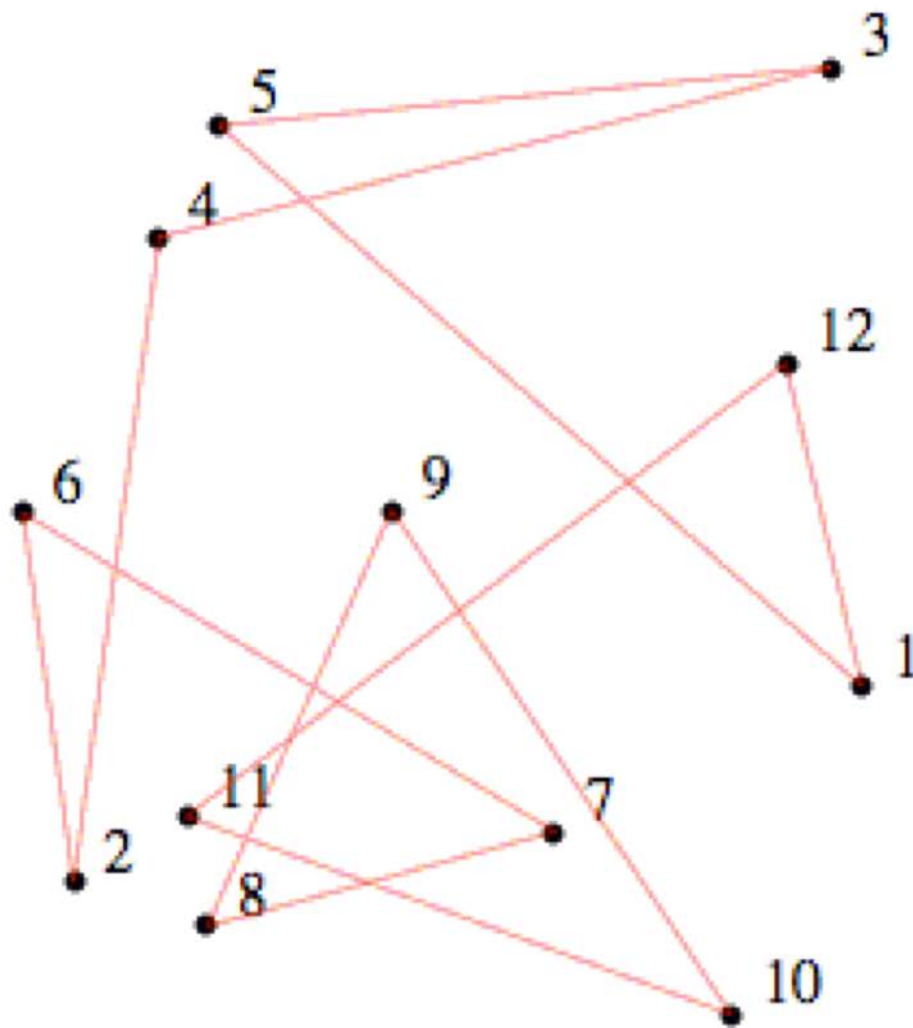


1. We generate a new configuration  $C'$  from the present configuration  $C$
2. We compute the energy of the new configuration,  $E'$
3. We compute the energy difference  $\Delta E = E' - E$
4. The new configuration is accepted with probability  $p$

$$\begin{cases} p = 1 & \Delta E < 0 \\ p = \exp\left(-\frac{\Delta E}{kT}\right) & \Delta E \geq 0 \end{cases}$$

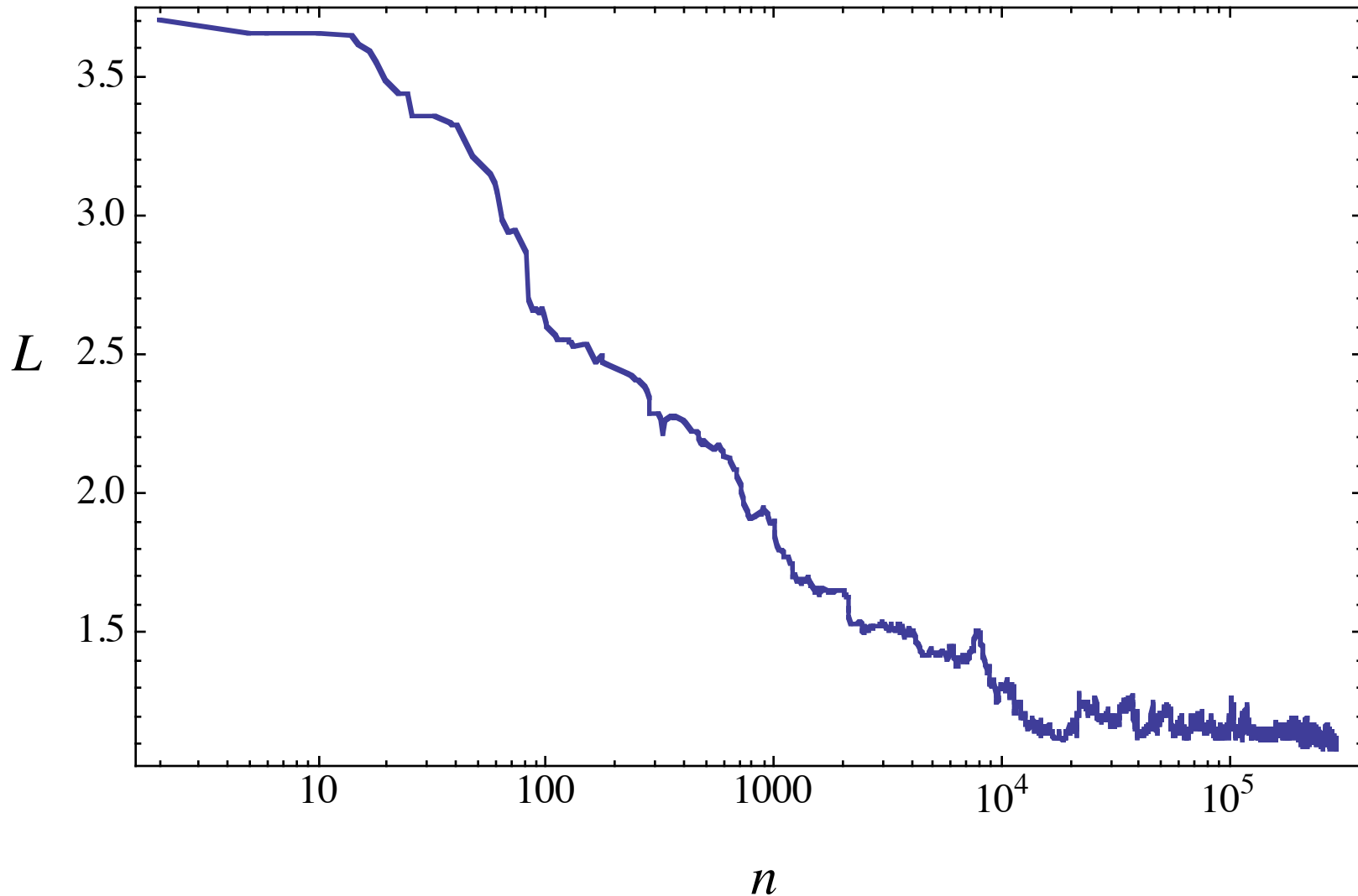
### Additional details

- the algorithm needs a slow cooling (it is common to choose an exponential cooling schedule)
- if cooling is not gradual, the system can get stuck into a local minimum
- simple exchanges of pairs of cities are the individual moves in the SA solution of the TSP
- the individual steps from one configuration to the next can be described by a Markov chain



$k = 1$   
 $T = 0.05$   
 $L = 1.84655$

# Decrease of total path length in a realization of the SA solution of the 50-cities problem



Here we note that the transition probability can be written as follows

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

Moreover, the algorithm preserves detailed balance

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

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

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

$$\frac{P(C')}{P(C)} = \exp \left( -\frac{(E' - E)}{kT} \right) \quad \leftarrow \text{ Boltzmann's distribution}$$

Moreover

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

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

## 6. MCMC – definition of the Metropolis-Hastings (M-H) algorithm (1970)


- we define the transition probability

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

and the target density

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

- we take state


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

- we choose randomly another state  $\mathbf{y}$  and we accept it ( $\mathbf{y} \rightarrow \mathbf{x}_{n+1}$ ) with probability

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

If the transition probability is symmetrical, then the acceptance probability takes on the simpler form

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

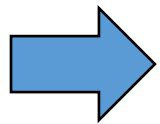
and it depends on the target density only.

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

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

• case  $\frac{\pi(\mathbf{y}) q(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x}) q(\mathbf{x}, \mathbf{y})} \geq 1$

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



$$\pi(\mathbf{x}) P(\mathbf{x} \rightarrow \mathbf{y}) = \pi(\mathbf{x}) q(\mathbf{x}, \mathbf{y})$$

$$\pi(\mathbf{y}) P(\mathbf{y} \rightarrow \mathbf{x}) = \pi(\mathbf{y}) q(\mathbf{y}, \mathbf{x}) \frac{\pi(\mathbf{x}) q(\mathbf{x}, \mathbf{y})}{\pi(\mathbf{y}) q(\mathbf{y}, \mathbf{x})} = \pi(\mathbf{x}) q(\mathbf{x}, \mathbf{y})$$



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

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

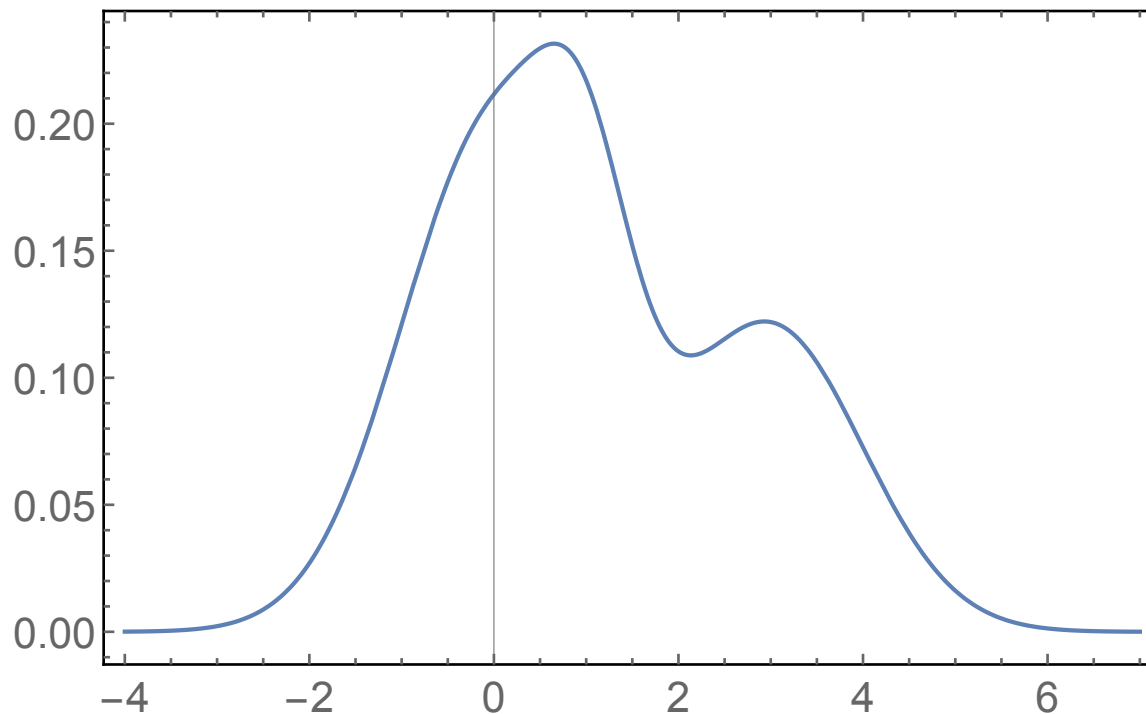
$$\Rightarrow \begin{aligned} \pi(\mathbf{x})P(\mathbf{x} \rightarrow \mathbf{y}) &= \pi(\mathbf{x})q(\mathbf{x},\mathbf{y}) \frac{\pi(\mathbf{y})q(\mathbf{y},\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x},\mathbf{y})} = \pi(\mathbf{y})q(\mathbf{y},\mathbf{x}) \\ \pi(\mathbf{y})P(\mathbf{y} \rightarrow \mathbf{x})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

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

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

(a three-component mixture model)



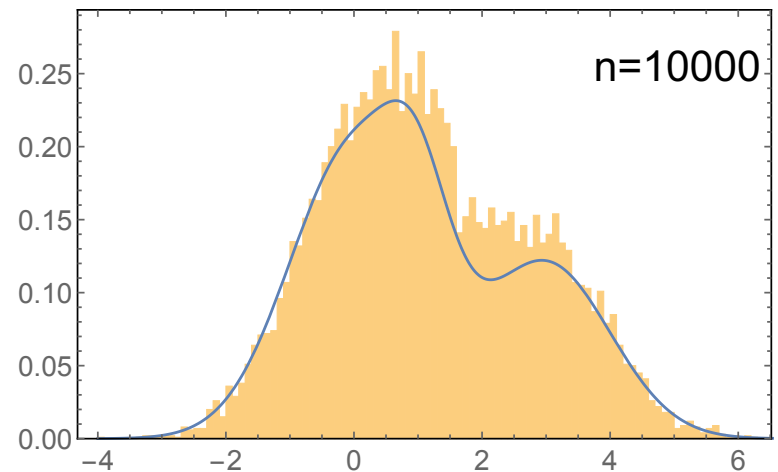
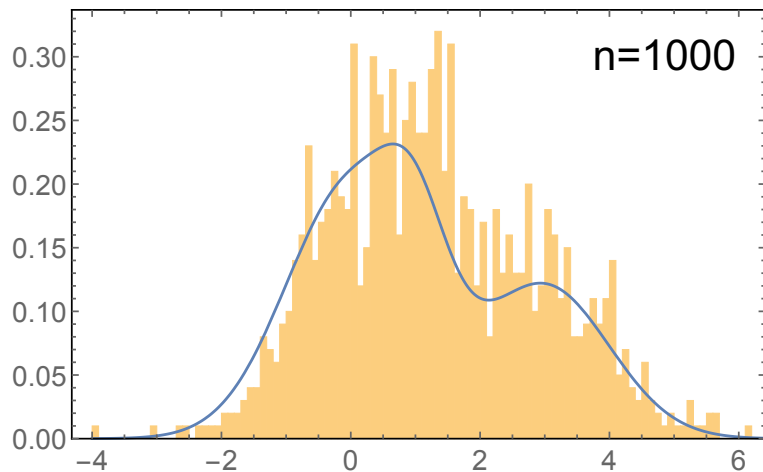
```

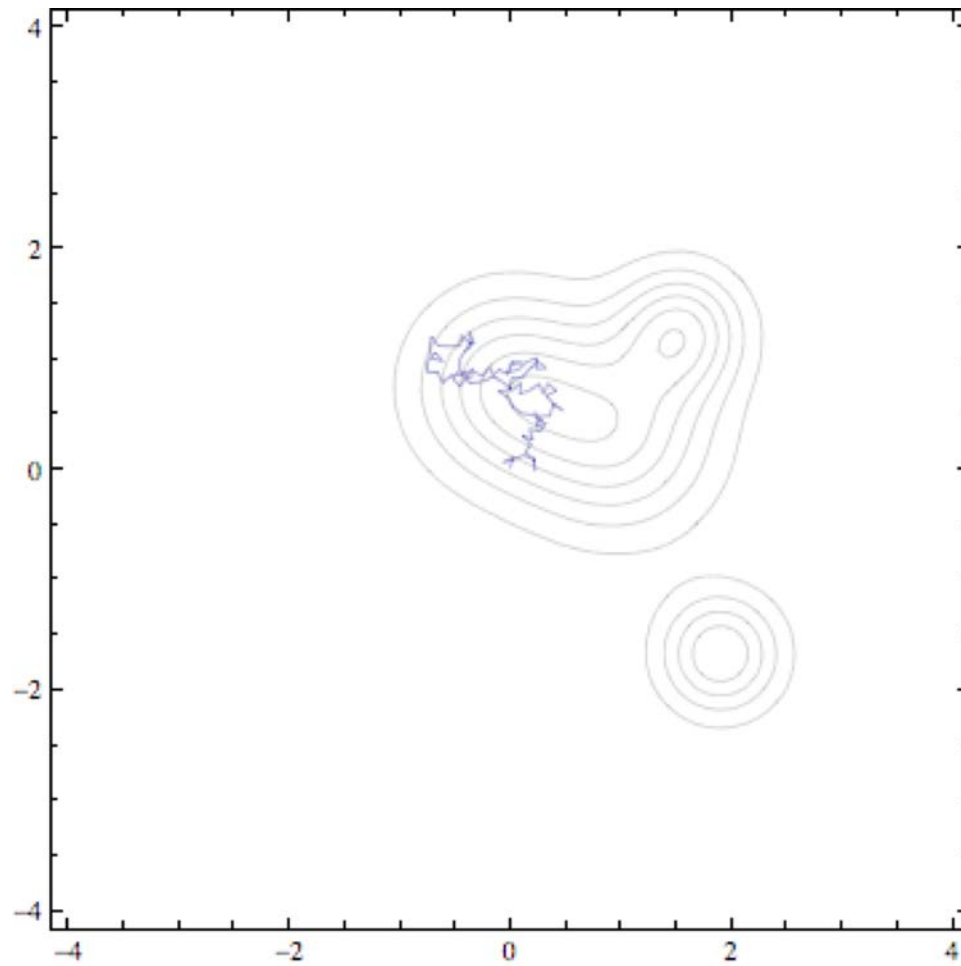
nrmax = 40 000;

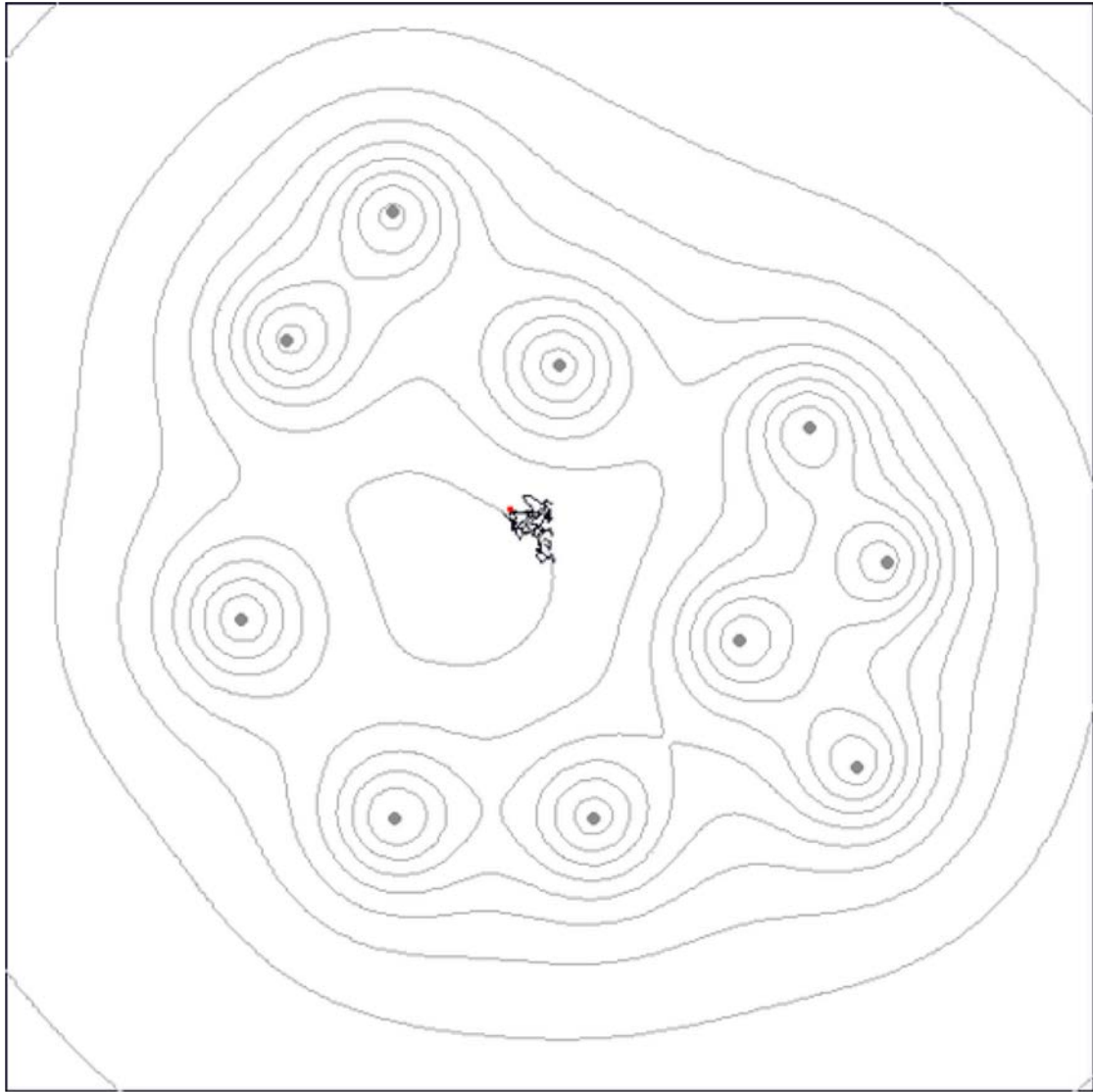
xr = Table[0, {nrmax}];
xr[[1]] = -4;

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

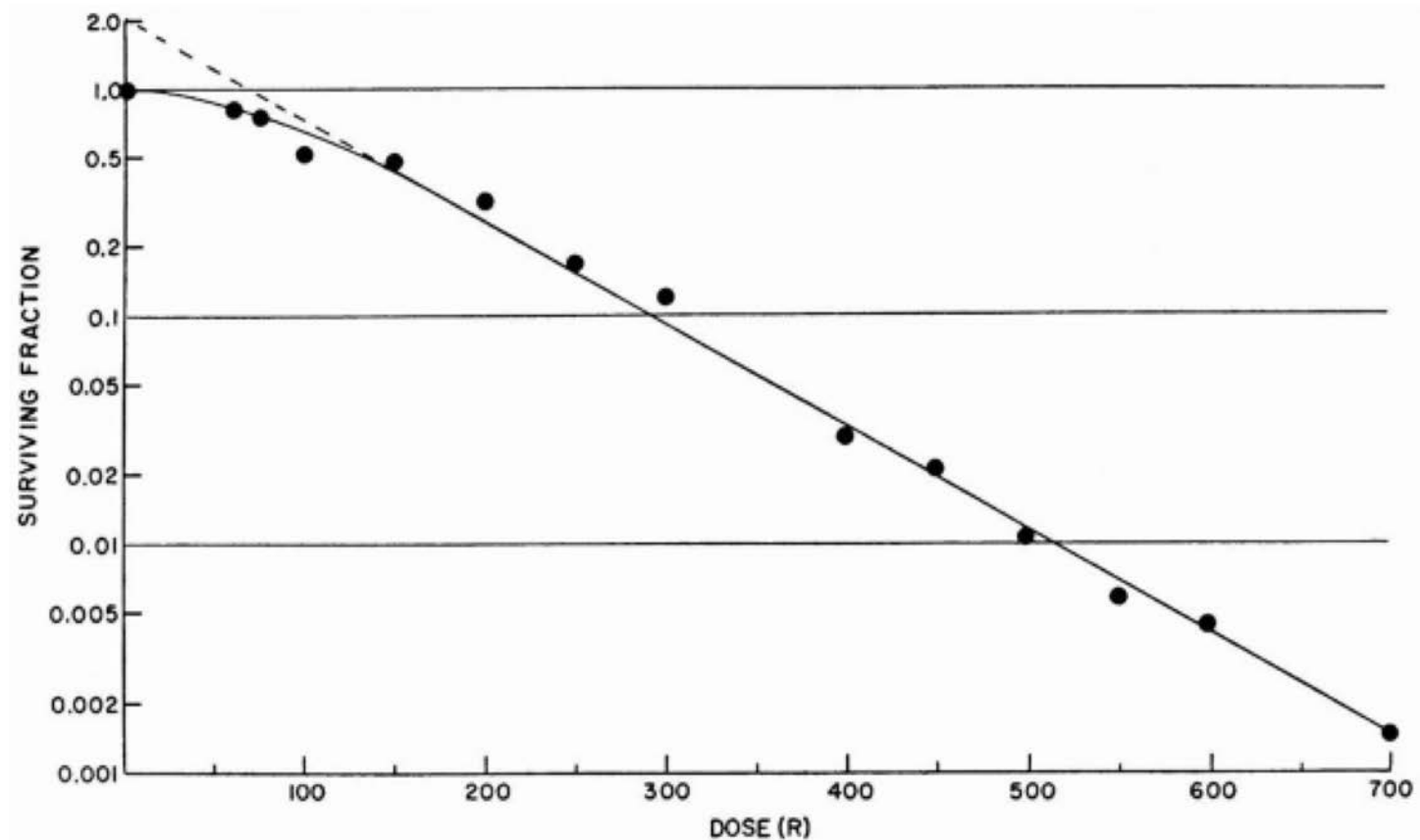
```







# 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

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

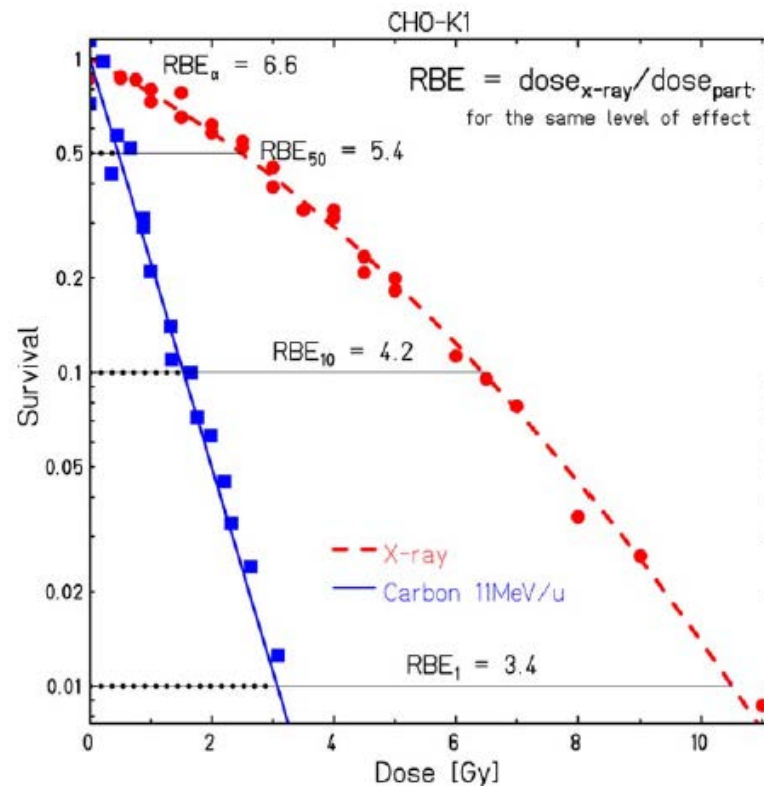


Fig. 1. Clonogenic survival curves illustrating the higher efficiency of the carbon ions compared with X-rays [10] (courtesy of the author, dr. Wilma K. Weyrather).

# Target theory

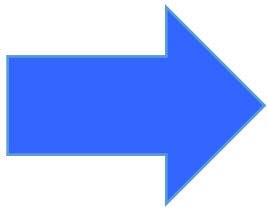
## Simple Poisson model:

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:  $a = D/D_0$



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



## 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

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

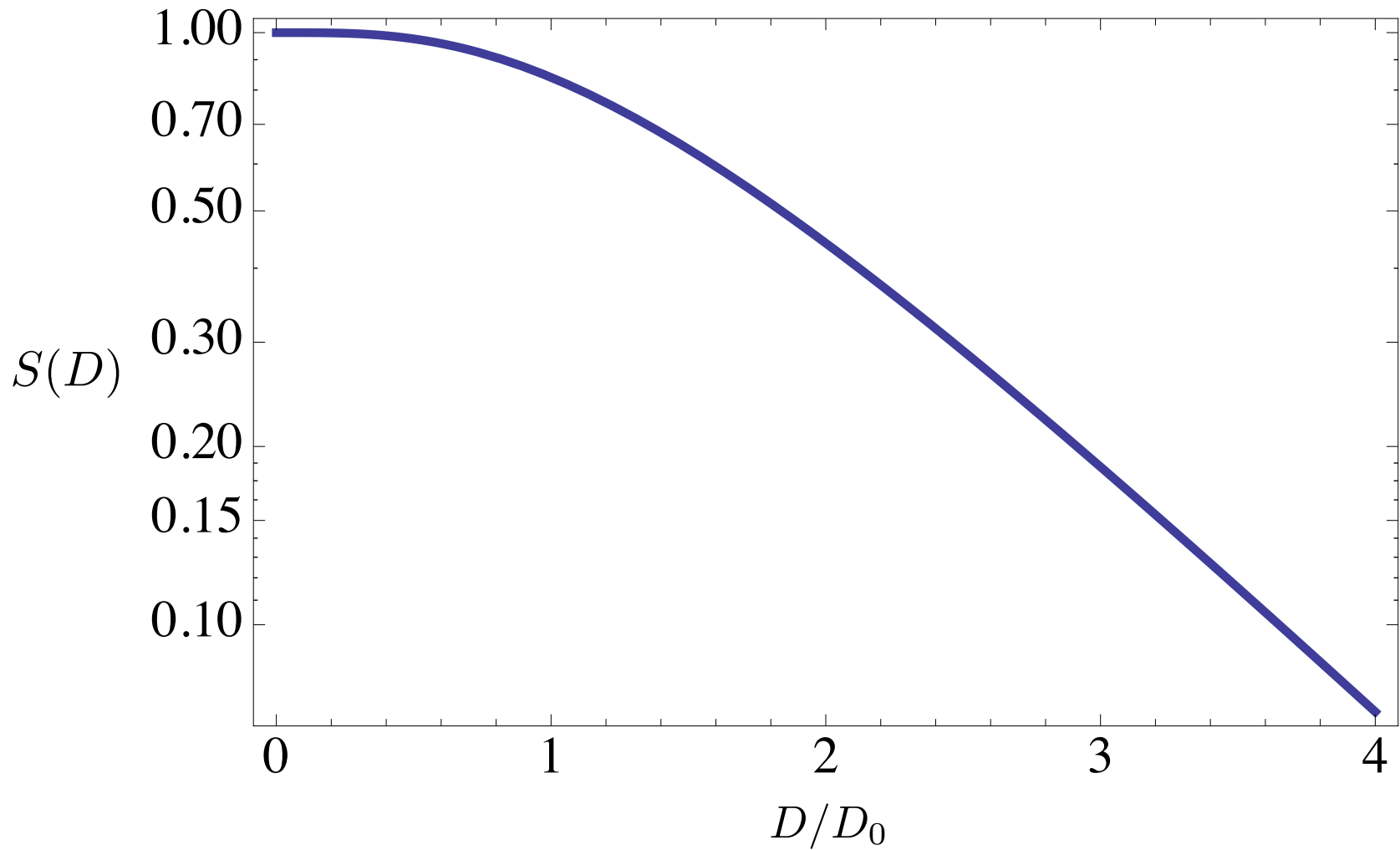
then, for large dose

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

i.e.,

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

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



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

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

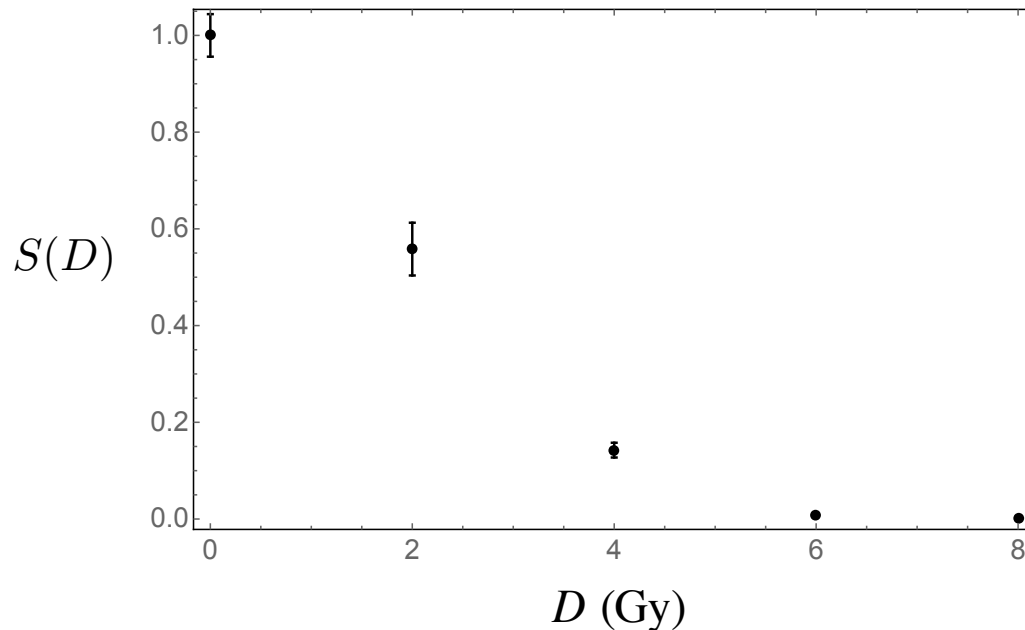
The derivatives differ in the origin, and the multitarget model fails to reproduce the observed linear-quadratic law.

# 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

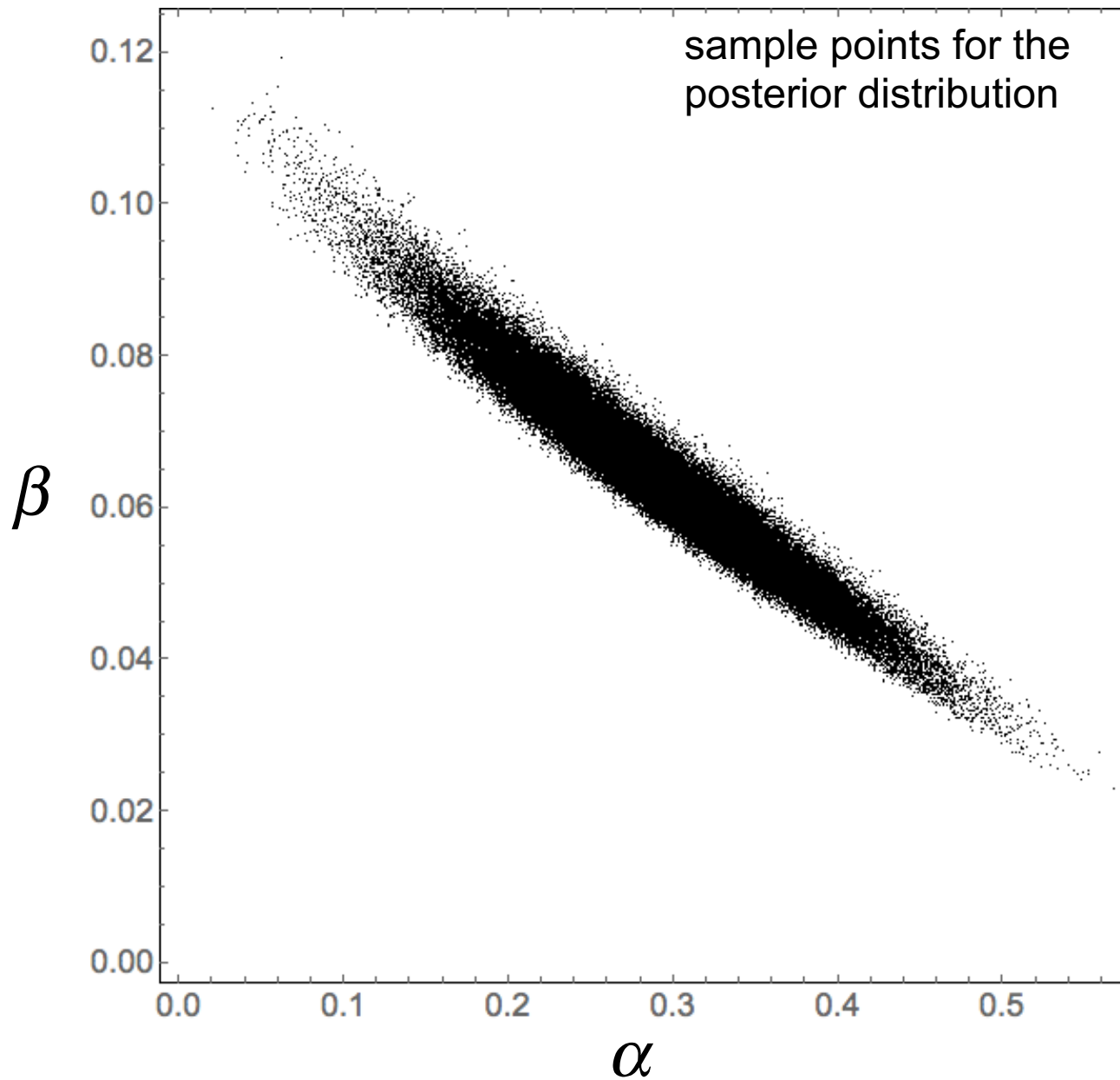
$$L(\alpha, \beta) = \prod_k \exp\left(-\frac{(S_k - S(\alpha, \beta))^2}{2\sigma_k^2}\right)$$

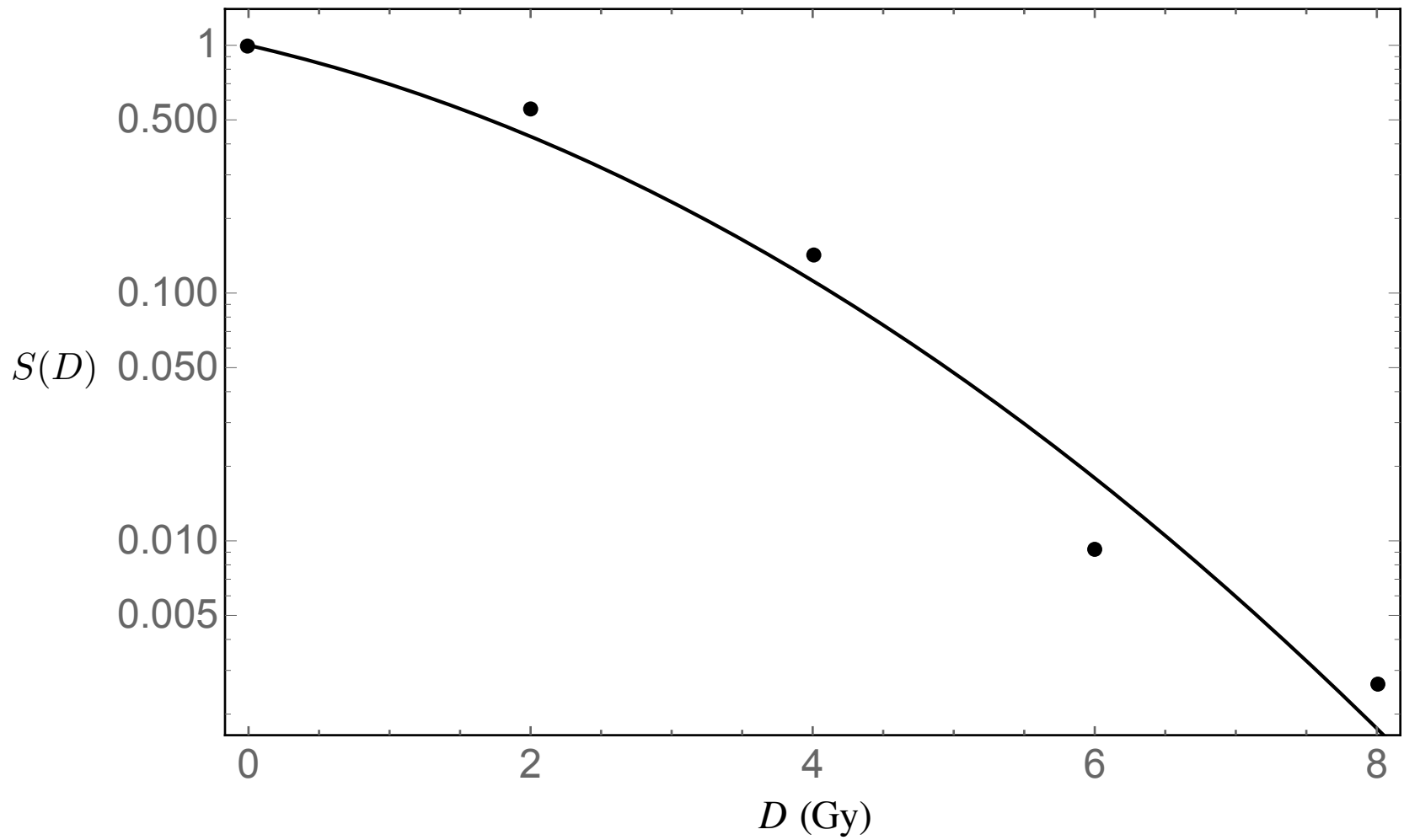
## 1b. Chose exponential priors for the parameters

## 1c. Complete posterior pdf

$$p(\alpha, \beta | \{S_k\}, I) = \left[ \prod_k \exp\left(-\frac{(S_k - S(\alpha, \beta))^2}{2\sigma_k^2}\right) \right] \exp(-0.1\alpha) \exp(-0.1\beta)$$

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





## 2a. Simple Gaussian likelihood for the RCR model

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

## 2b. Chose exponential priors for the parameters

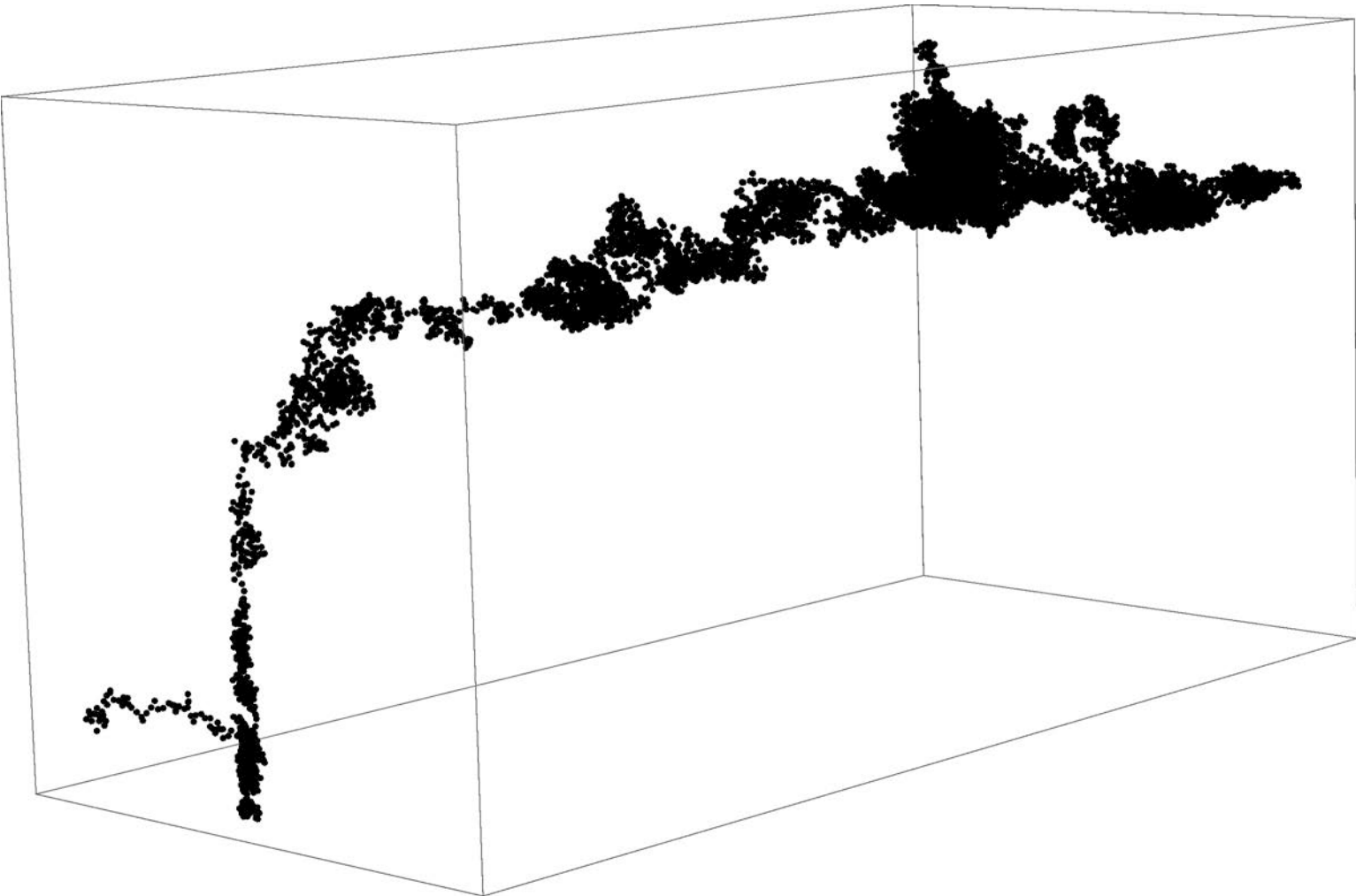
## 2c. Complete posterior pdf

$$p(a,b,c|\{S_k\},I) = \left[ \prod_k \exp\left(-\frac{(S_k - S(a,b,c))^2}{2\sigma_k^2}\right) \right] e^{-0.2a} e^{-0.2b} e^{-0.2c}$$

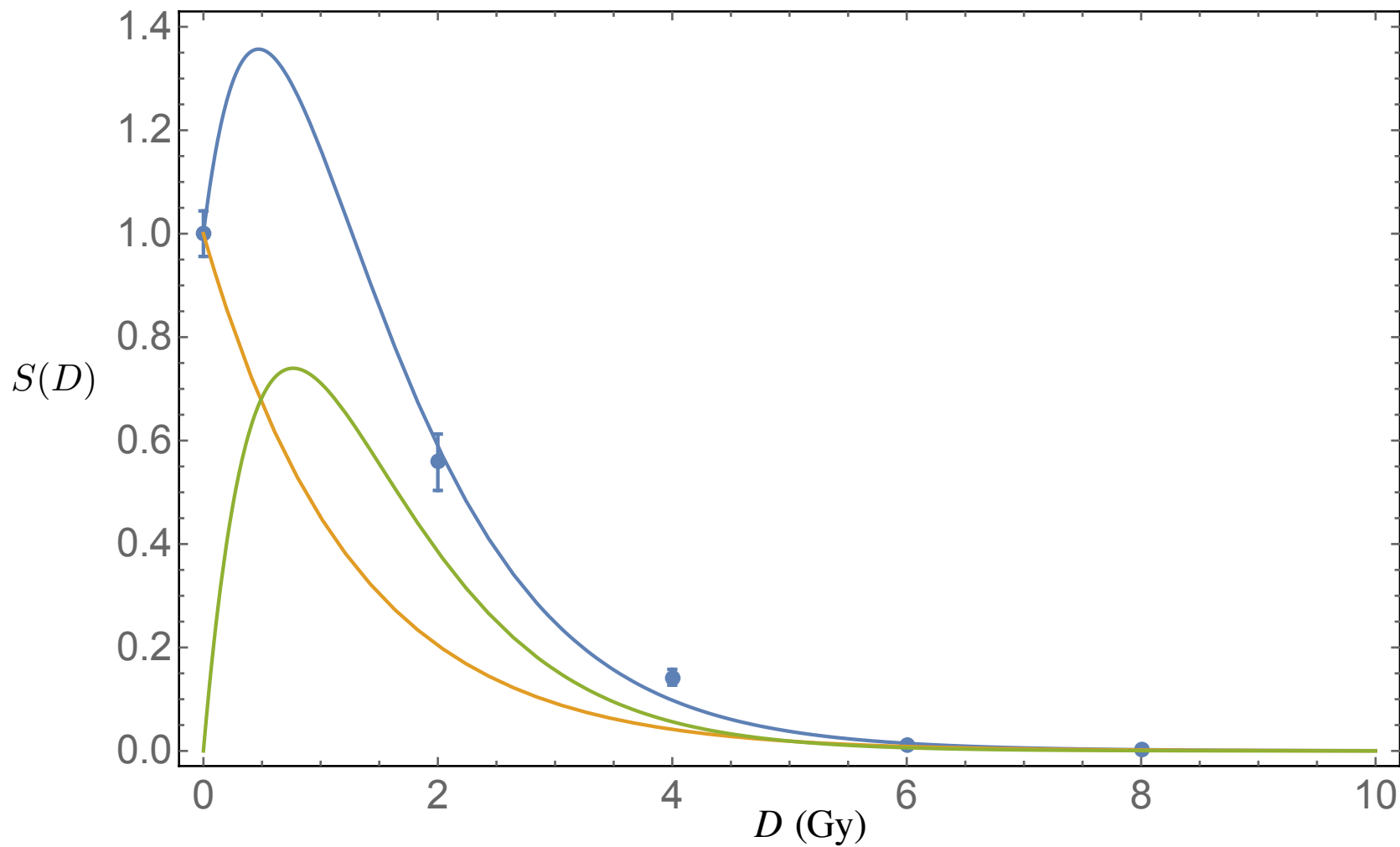
## 2d. Use MCMC to find the MAP estimate (and any moment of the pdf)



# Path in (a,b,c) space

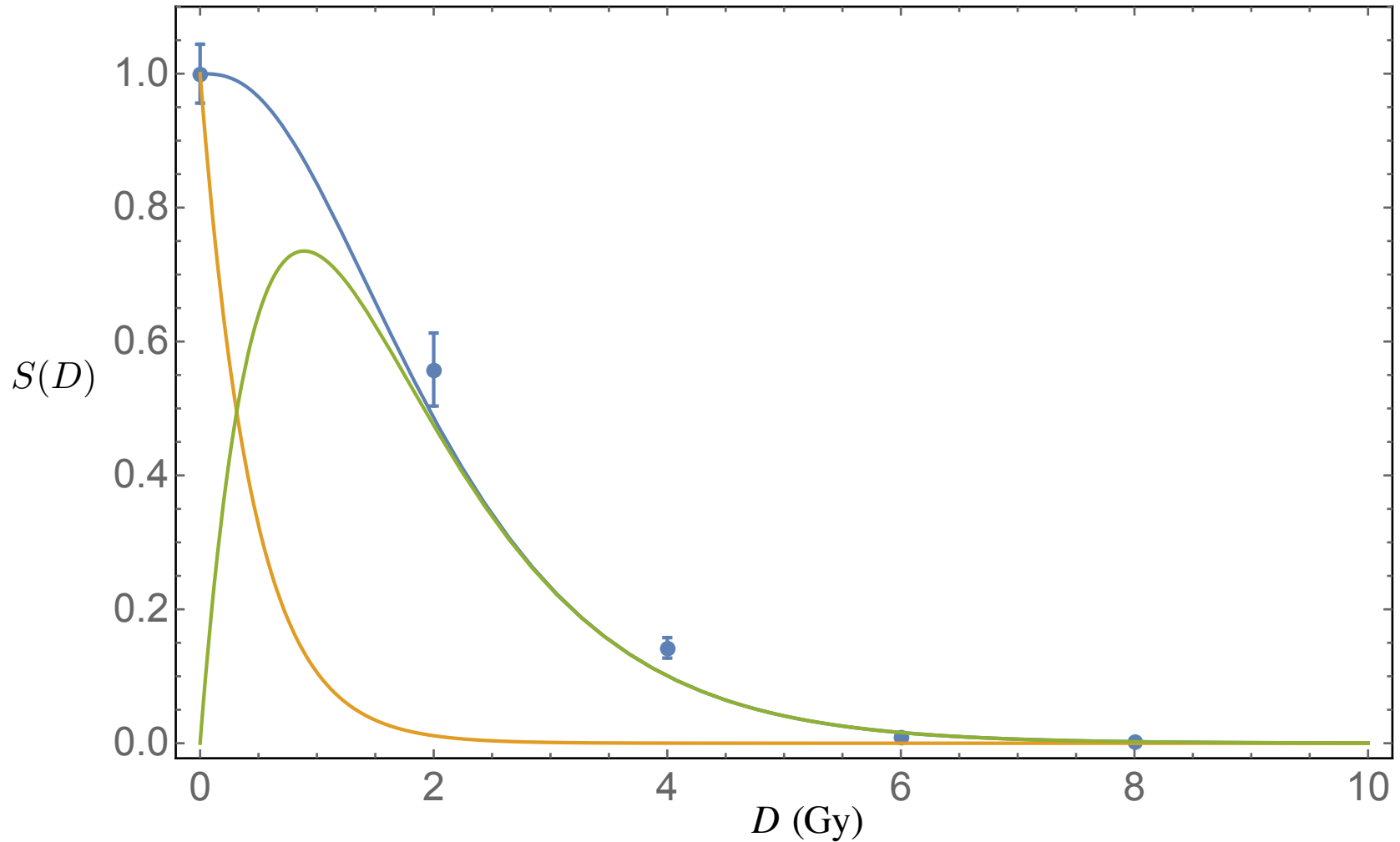


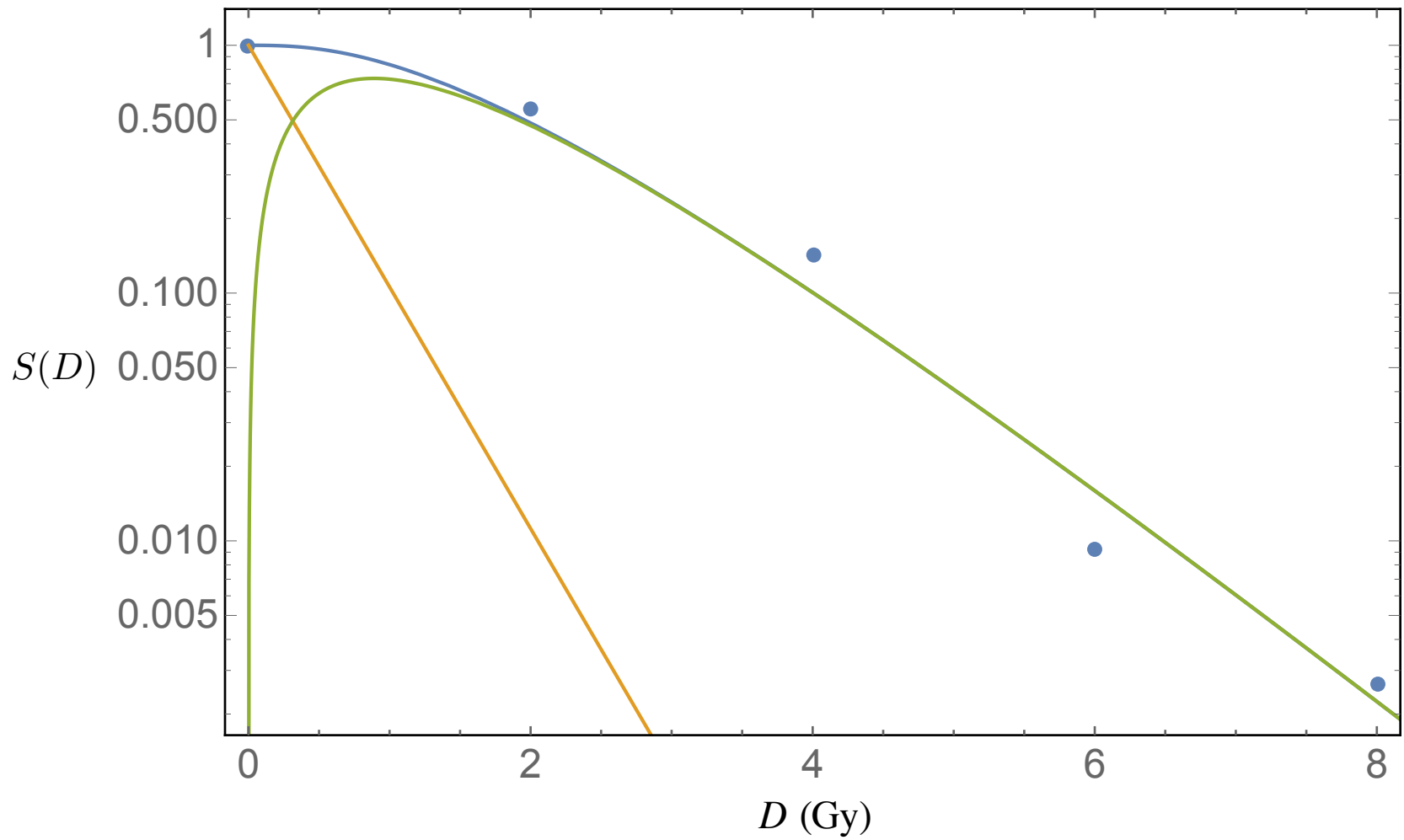
# Fit showing individual components: unsatisfactory result



## Revise priors to include constraint on derivative

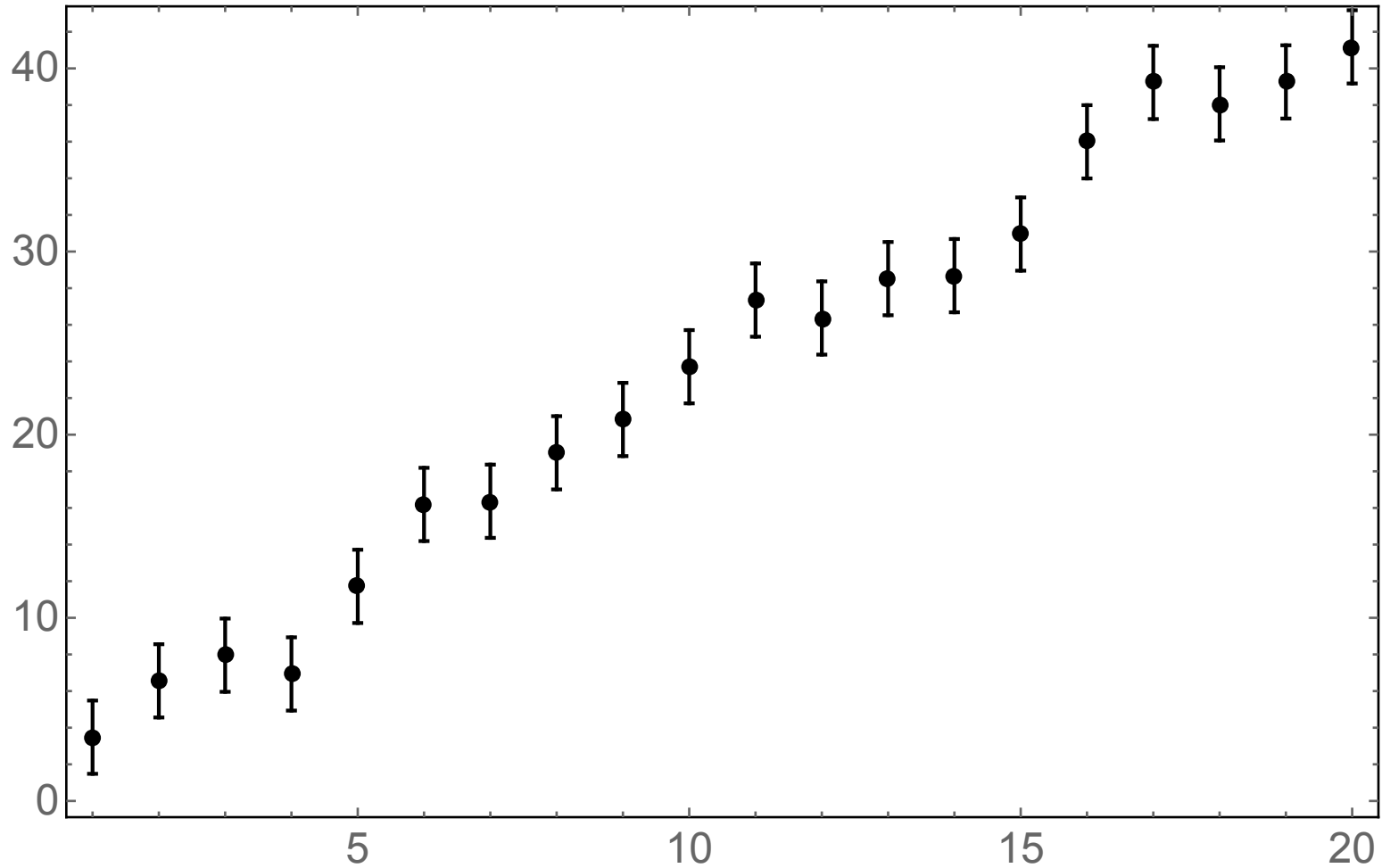
(priors vanish where derivative in the origin is negative)



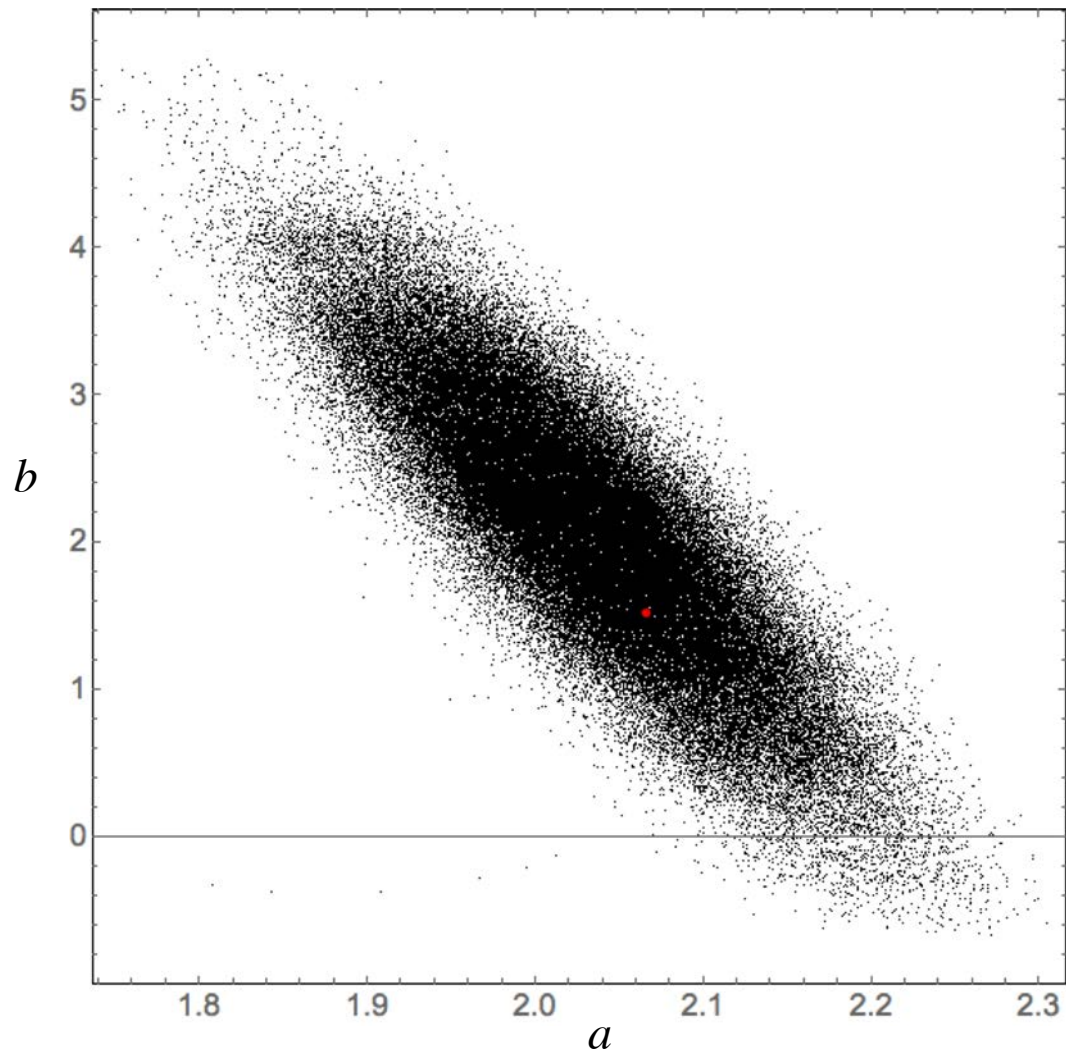


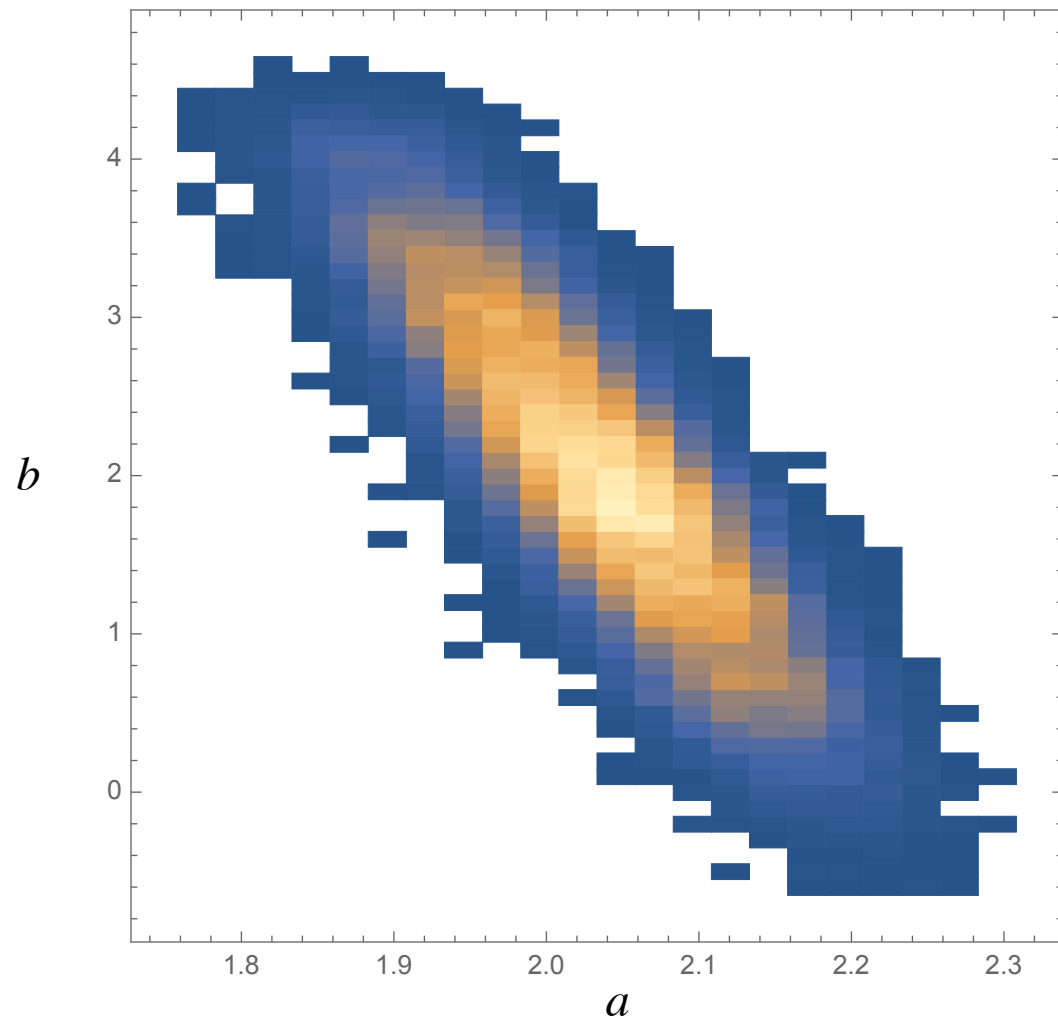
# “Straight line fit” with the MCMC

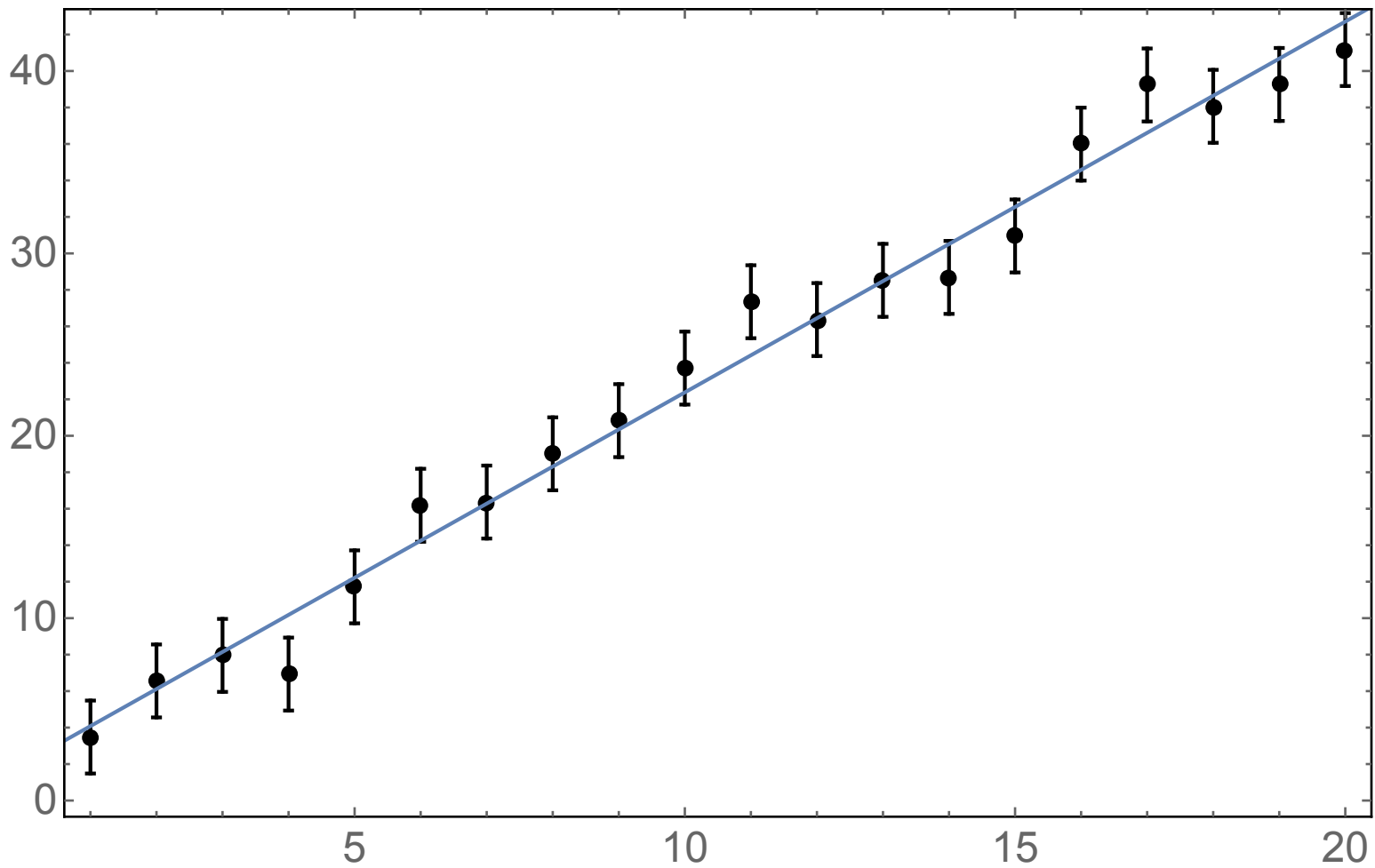
An example with Gaussian errors and exponential priors.



model  $y = ax + b$









## •References:

- D. Flury, “ Acceptance-Rejection Sampling Made Easy ” SIAM Review **32** (1990) 474
- H. Varian, “Bootstrap Tutorial”, The Mathematica Journal **9** (2005) 768
- A. F. M. Smith and A. E. Gelfand: “Bayesian Statistics Without Tears: A Sampling-Resampling Perspective”, Am. Stat. **46** (1992) 84
- B. Walsh: “Markov Chain Monte Carlo and Gibbs Sampling”,  
<http://web.mit.edu/~wingated/www/introductions/mcmc-gibbs-intro.pdf>
- S. P. Brooks: “Markov Chain Monte Carlo and Its Application”, The Statistician **47** (1998) 69