Monte Carlo methods

outline

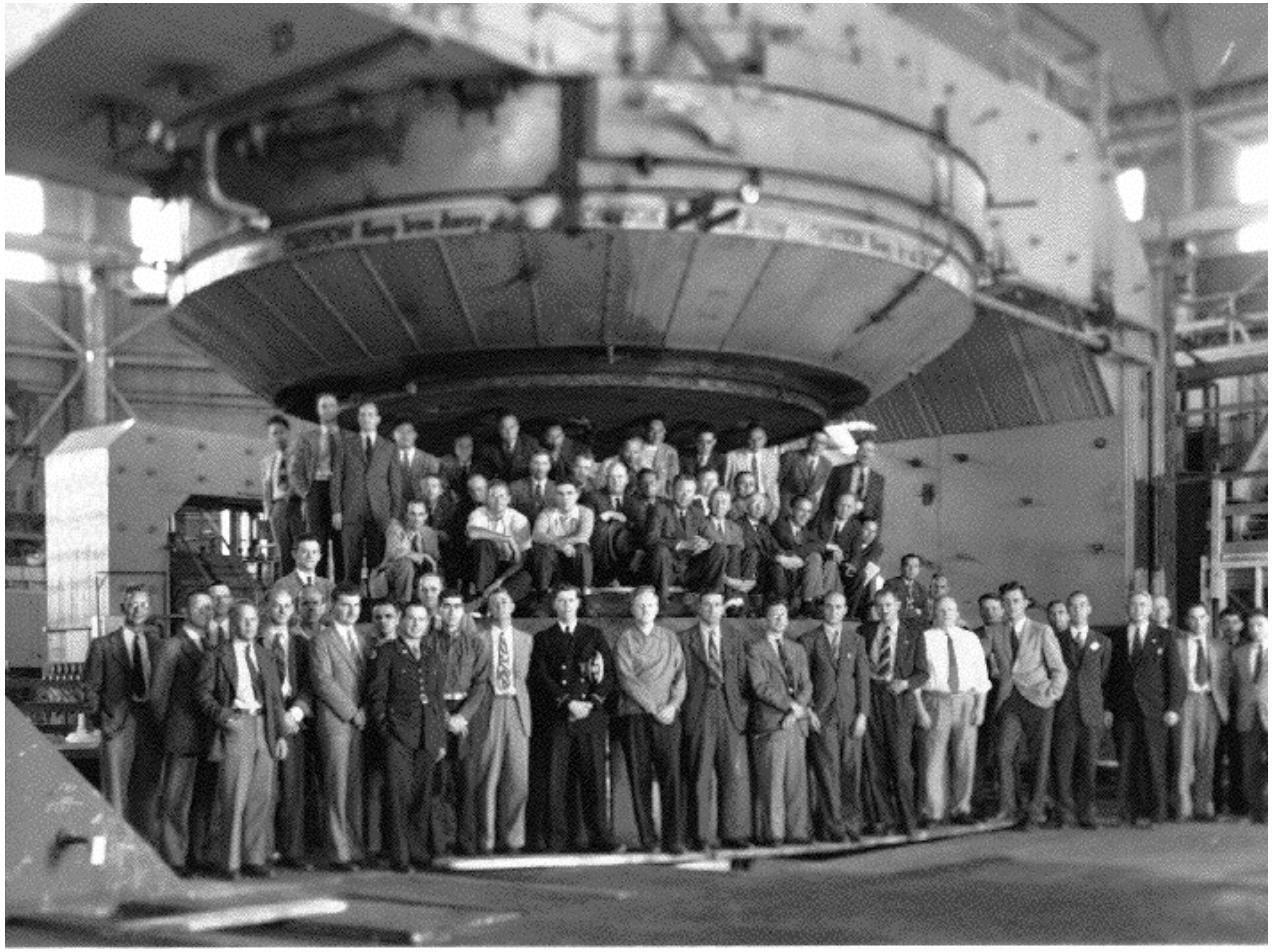
Stocastic Processes in Science Inference

- History of Monte Carlo Methods
- Application of MC to probabilistic inference
- A simple MC simulation
- MC simulations applications in Urban Science - Traffic flow, Resque
- Rejection & Importance Sampling

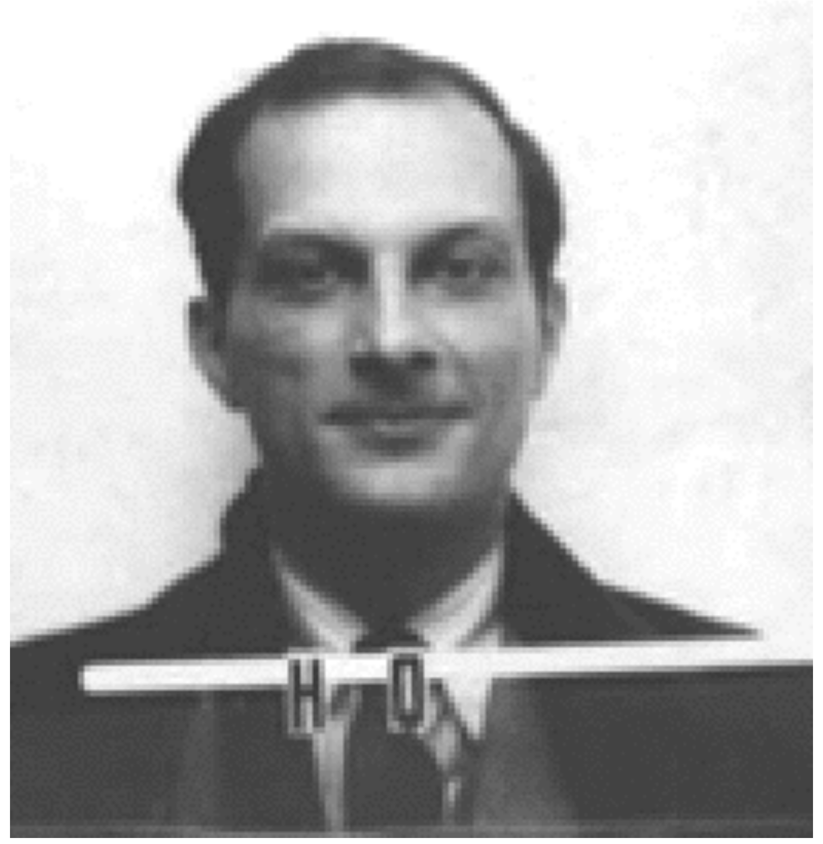
Markov Chain Monte Carlo

- Markovian Processes and Markov chains
- Bayes theorem and the posterior distribution
- Metropolis Hasting (and Gibbs sampling) MCMC
- Affine Invariant MCMC
- convergence criteria

The Manhattan Project



MC - history



Stanislav Ulam

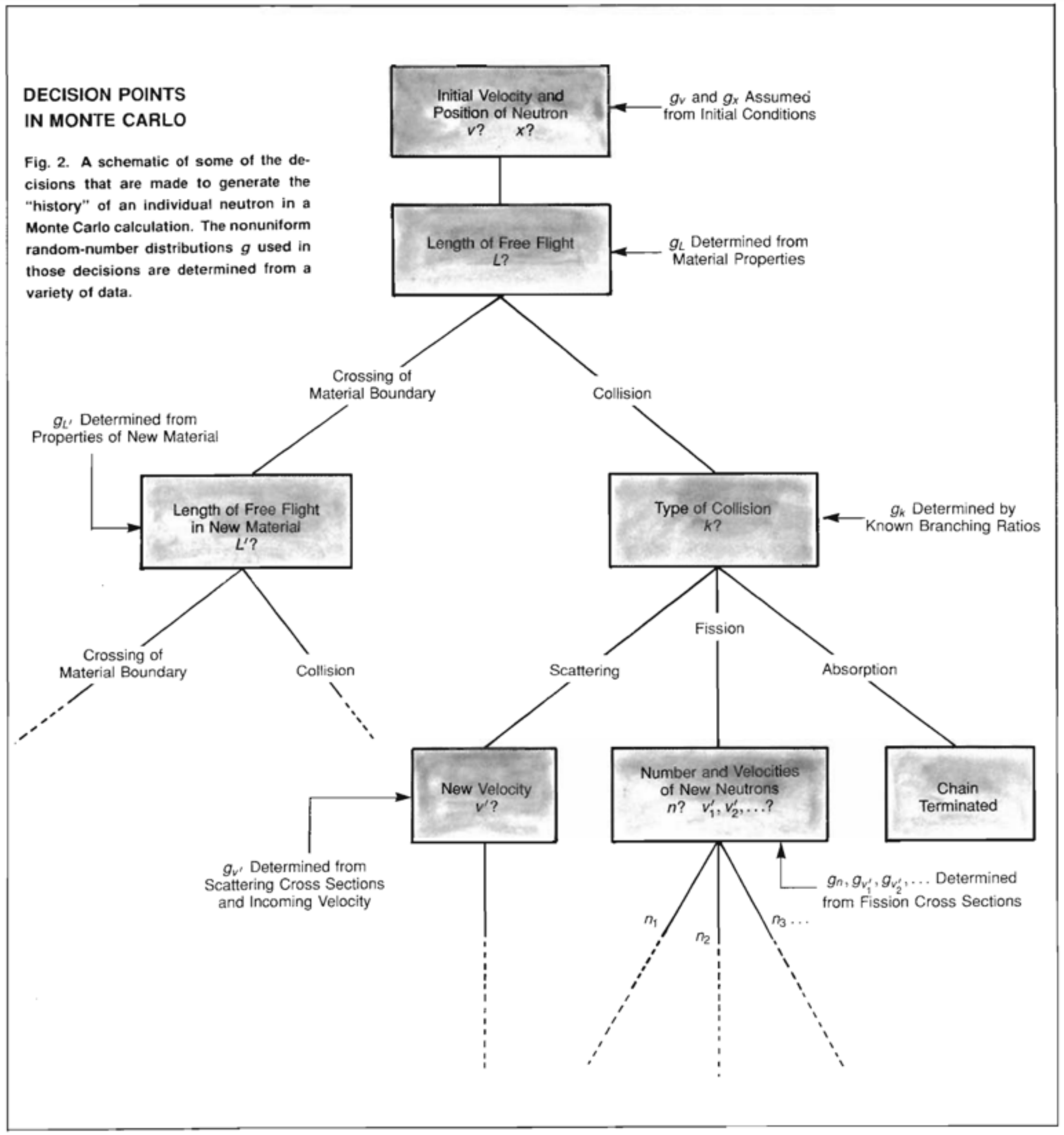
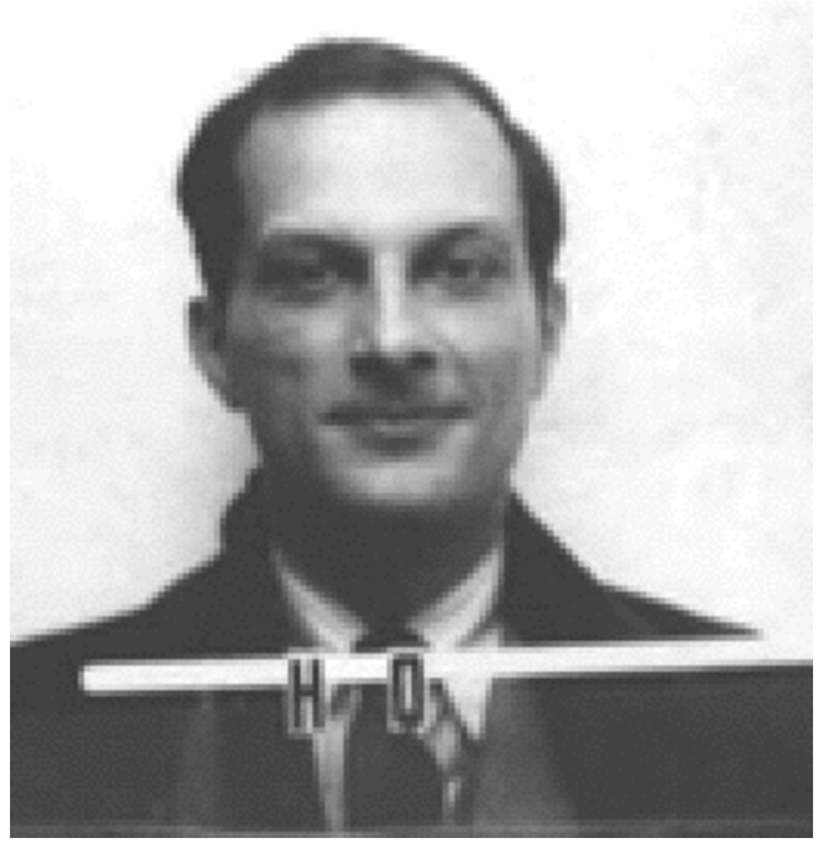


Fig. 2. A schematic of some of the decisions that are made to generate the "history" of an individual neutron in a Monte Carlo calculation. The nonuniform random-number distributions g used in those decisions are determined from a variety of data.

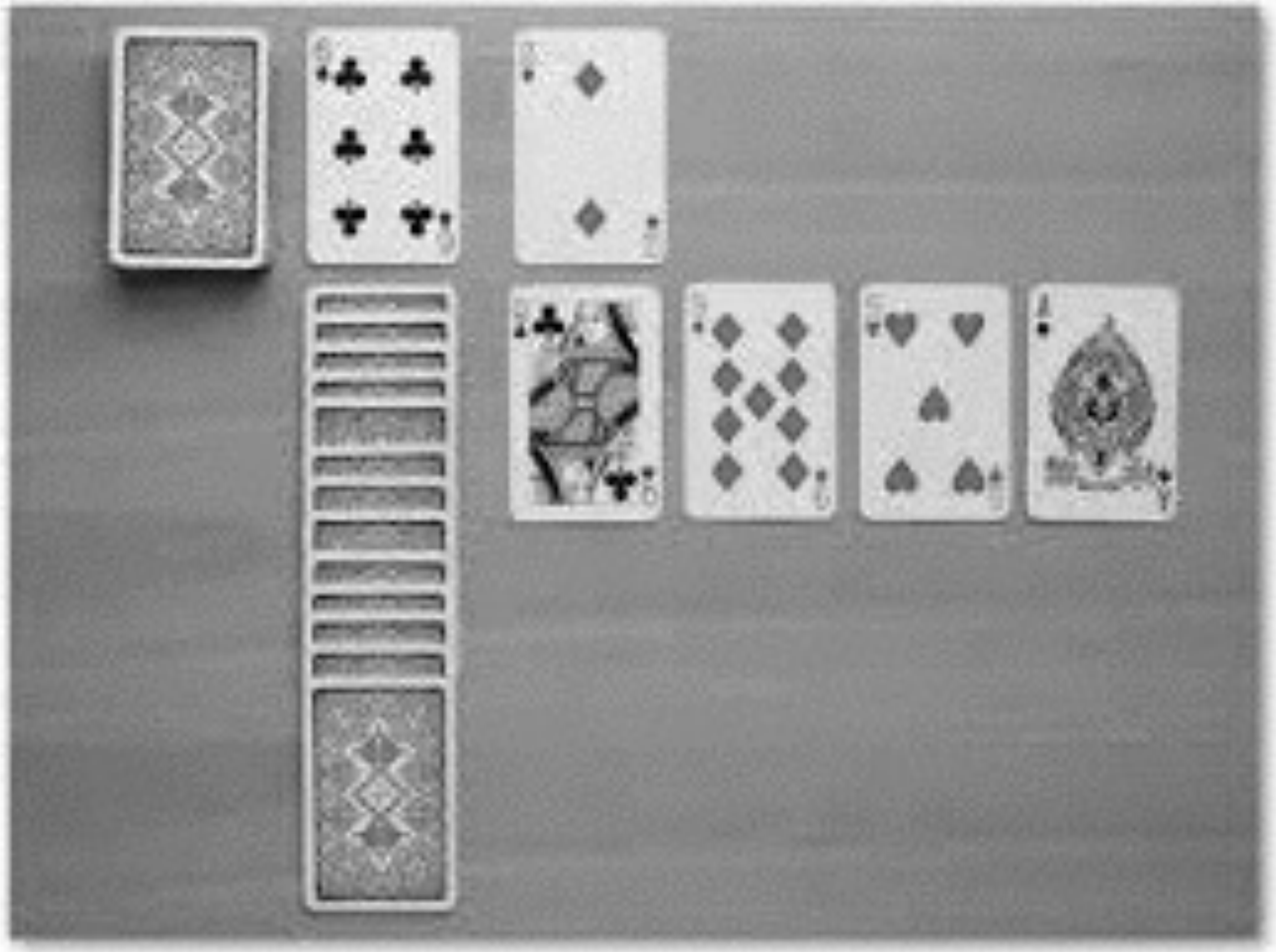


Stanislav Ulam

What are the chances that a Canfield solitaire laid out with 52 cards will come out successfully?

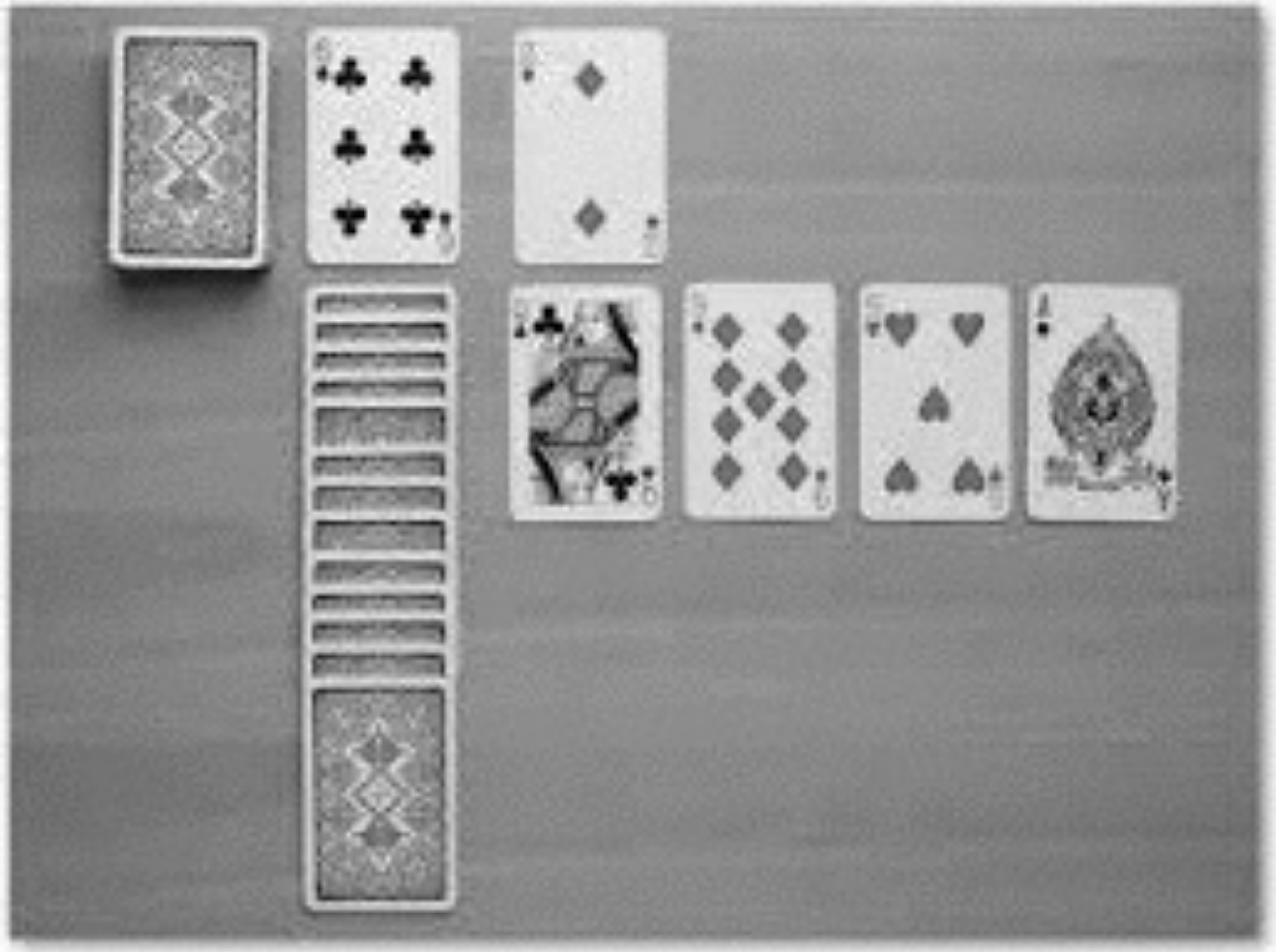
The number of different games is
$52! = 52x51x52…x3x2x1 \sim 8x10^{67}$

Canfield Solitaire



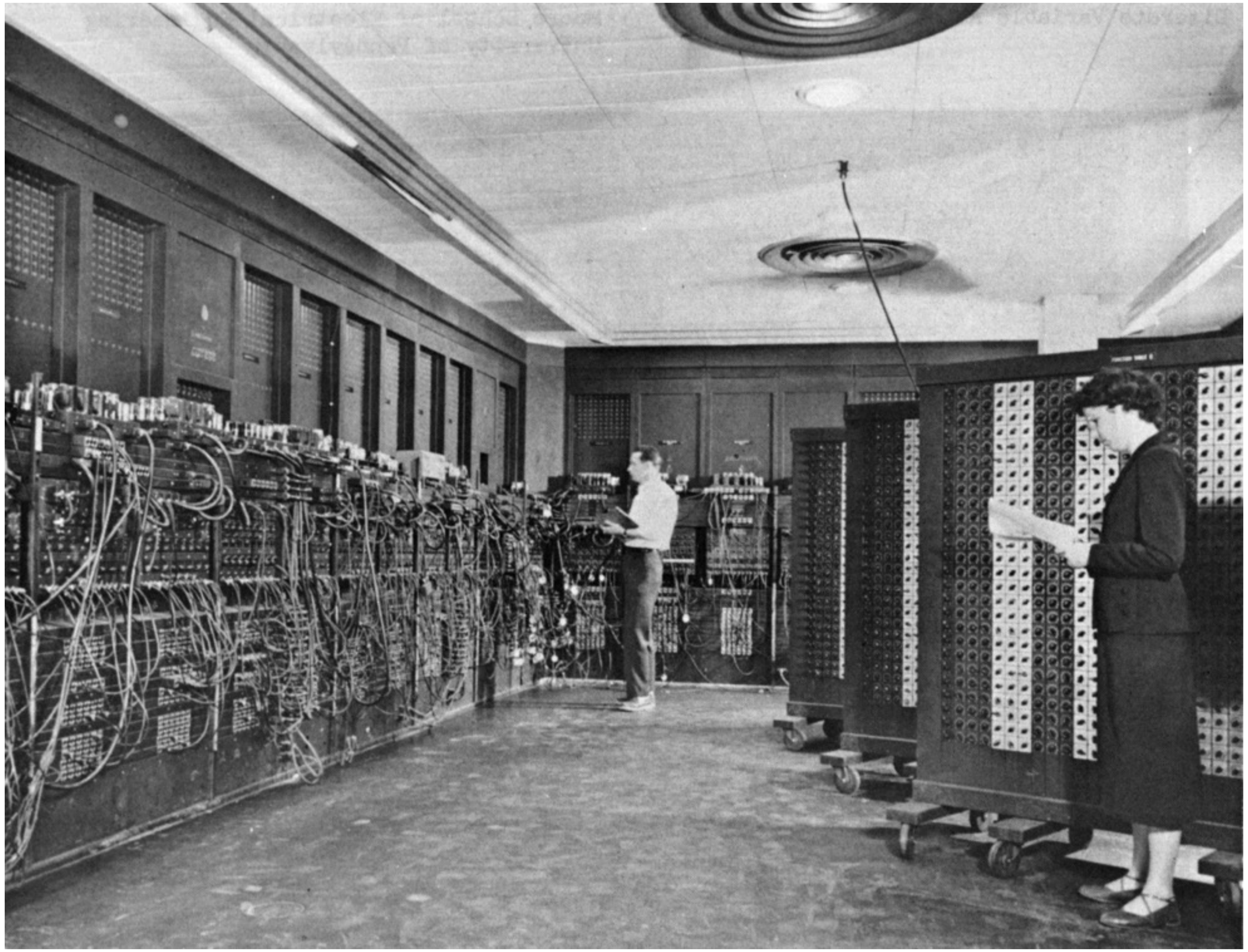
MC - history

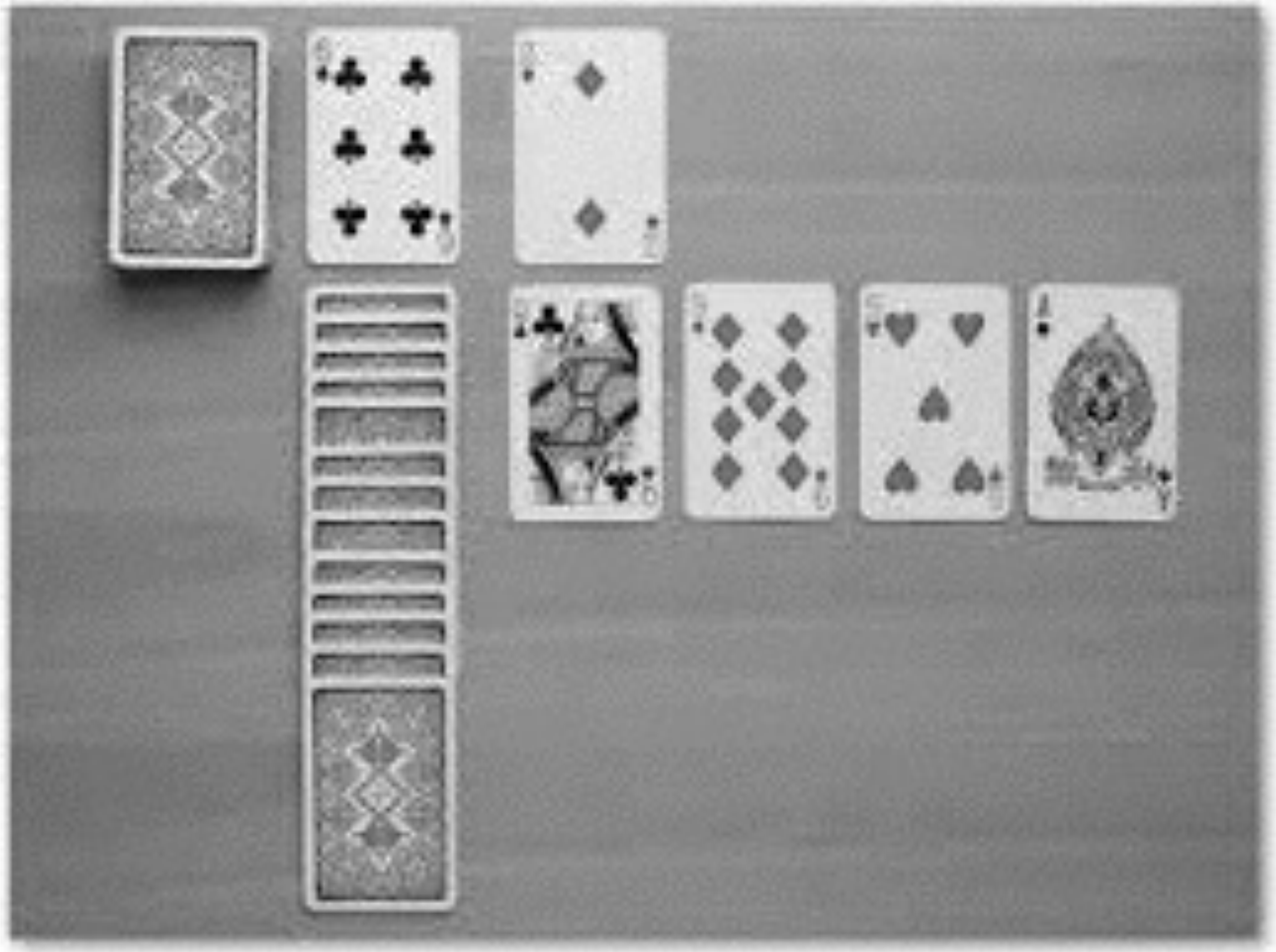
"What are the chances that a Canfield solitaire laid out with 52 cards will come out successfully? After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether **a more practical method than *abstract thinking* might not be to lay it out say one hundred times and simply observe and count the number of successful play"**



http://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-88-9068

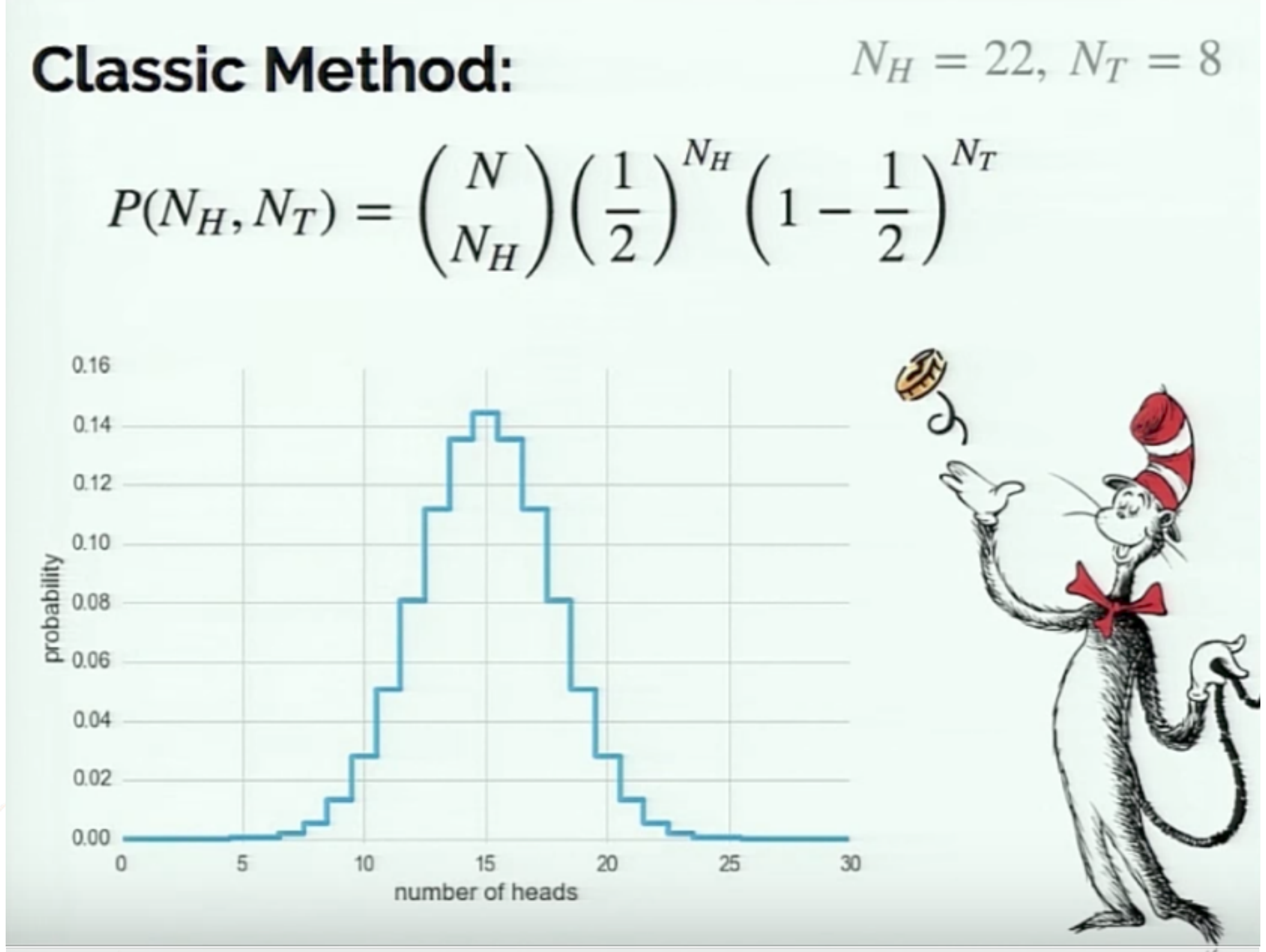
MC - history

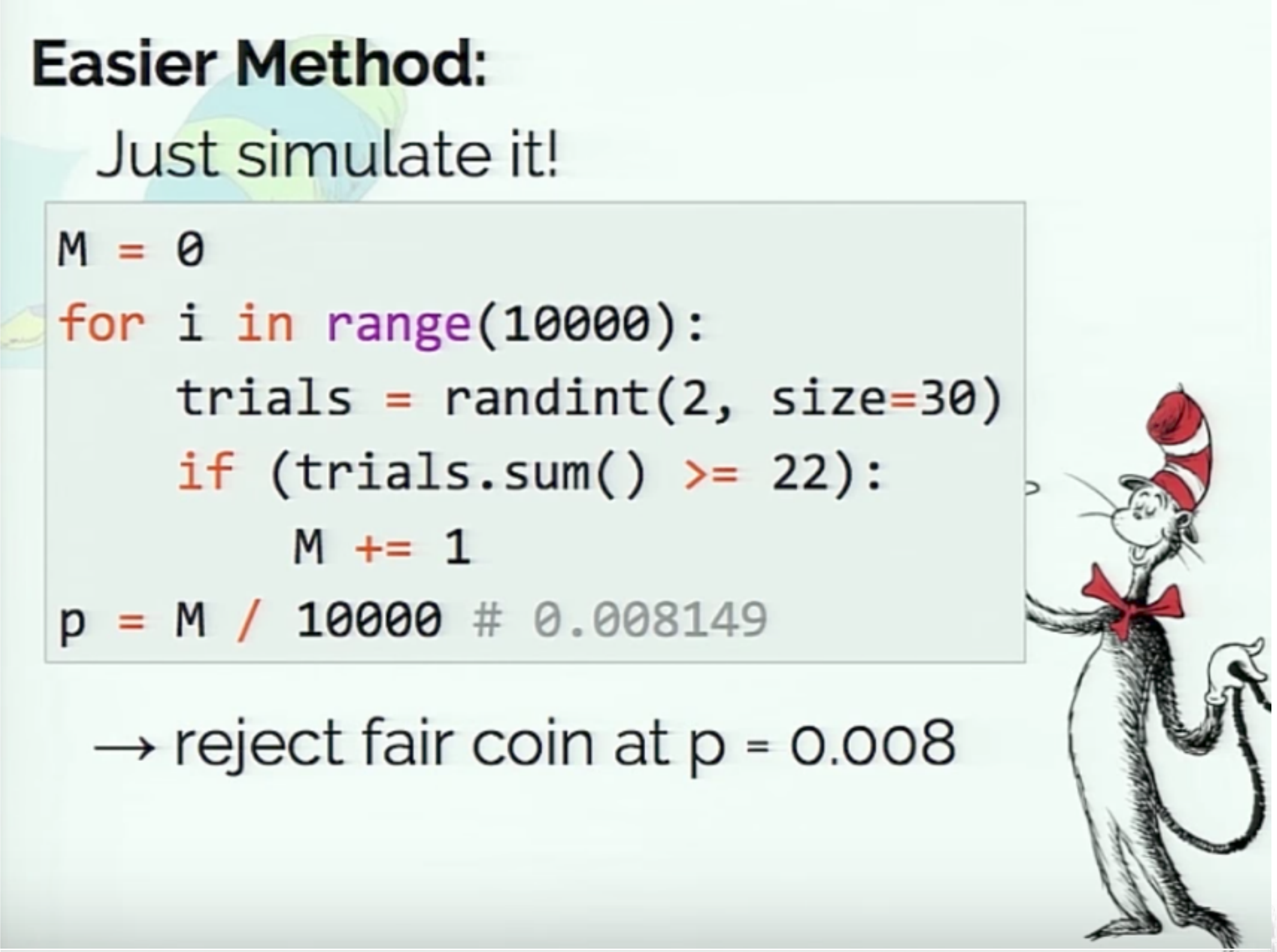




ENIAC It weighed more than 30 short tons (27 t), was roughly 2.4 m × 0.9 m × 30 m (8 × 3 × 100 feet) in size, occupied 167 m^2 (1,800 ft^2), consumed 150 kW of electricity.
500FLOPS vs today's Macbook pro ~1TeraFLOP

MC - history

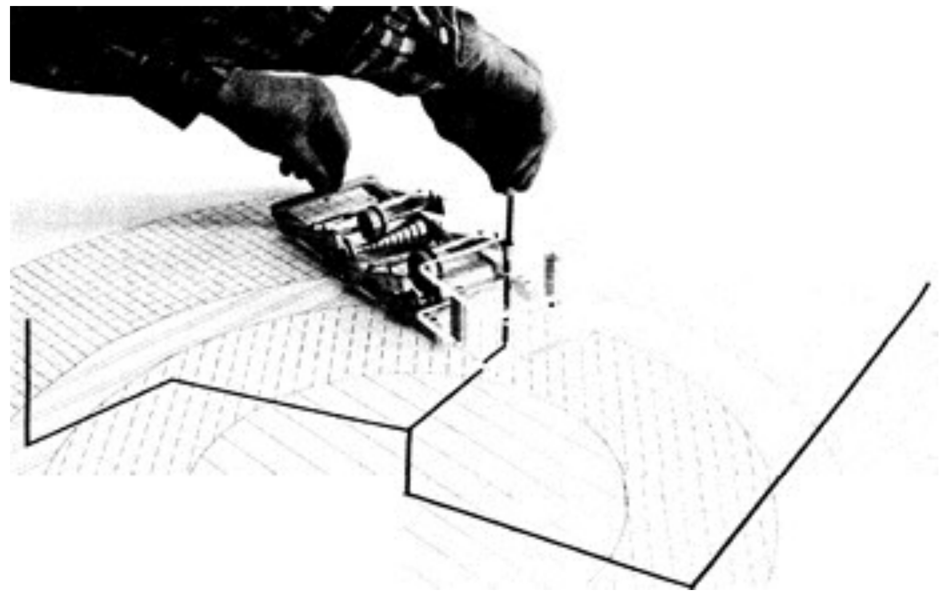


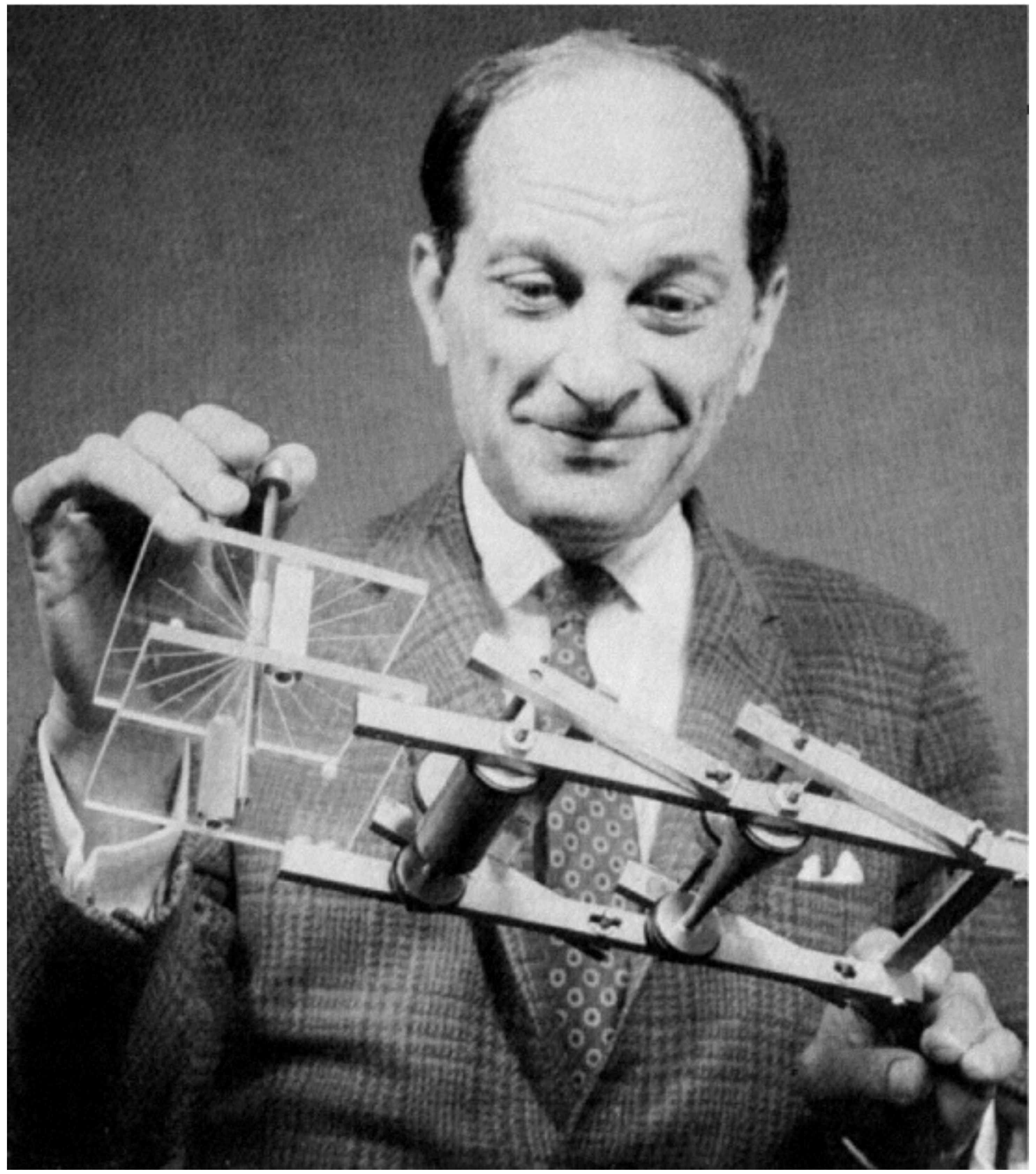


[Statistics for Hackers, Jake Vanderplas PyCon16](https://www.youtube.com/watch?v=Iq9DzN6mvYA)
https://www.youtube.com/watch?v=Iq9DzN6mvYA

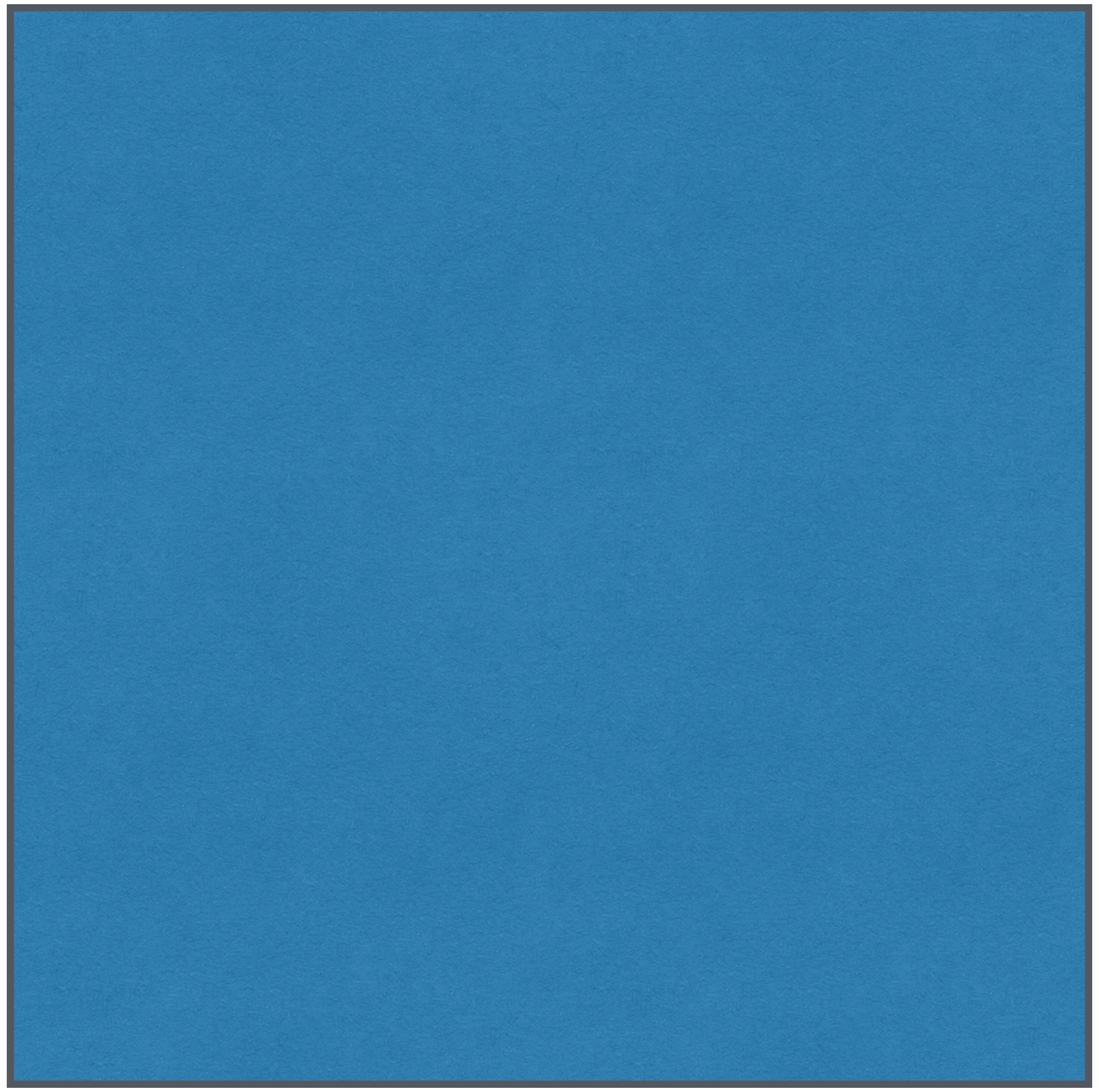
The Fermiac

Enrico Fermi looked really smart with his predictions…



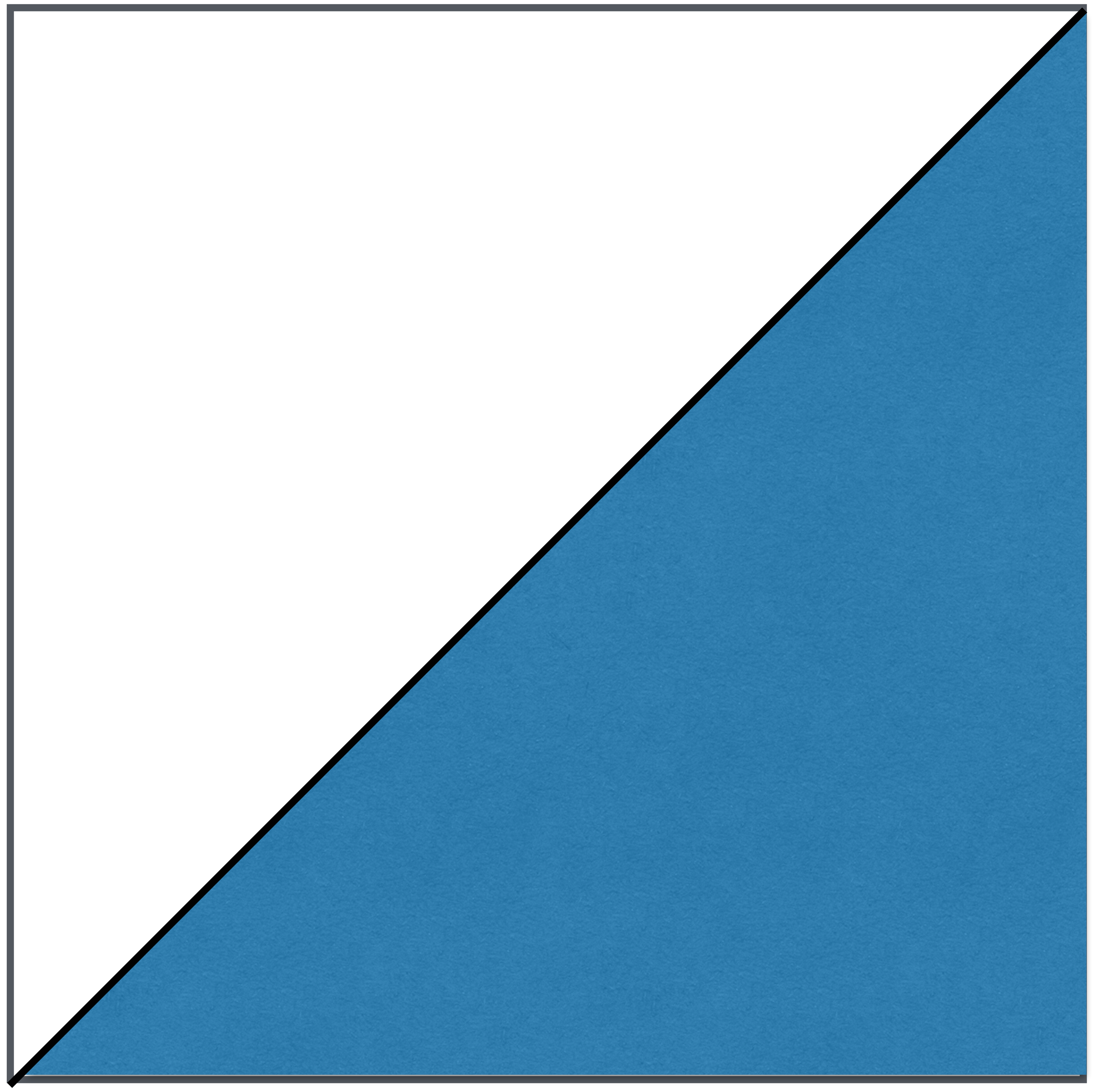


MC - motivation: expectations



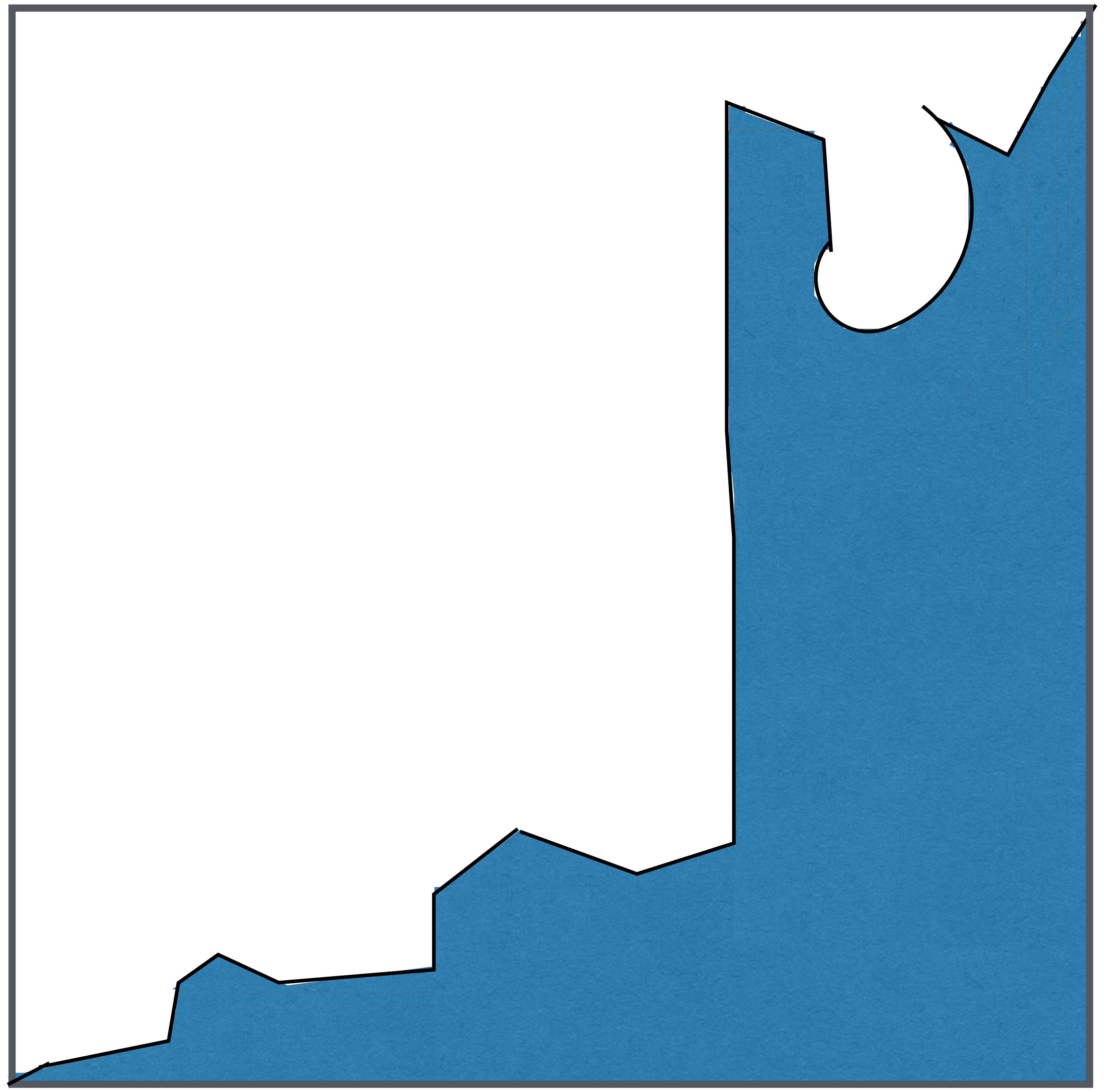
Area: base x height

MC - motivation: expectations



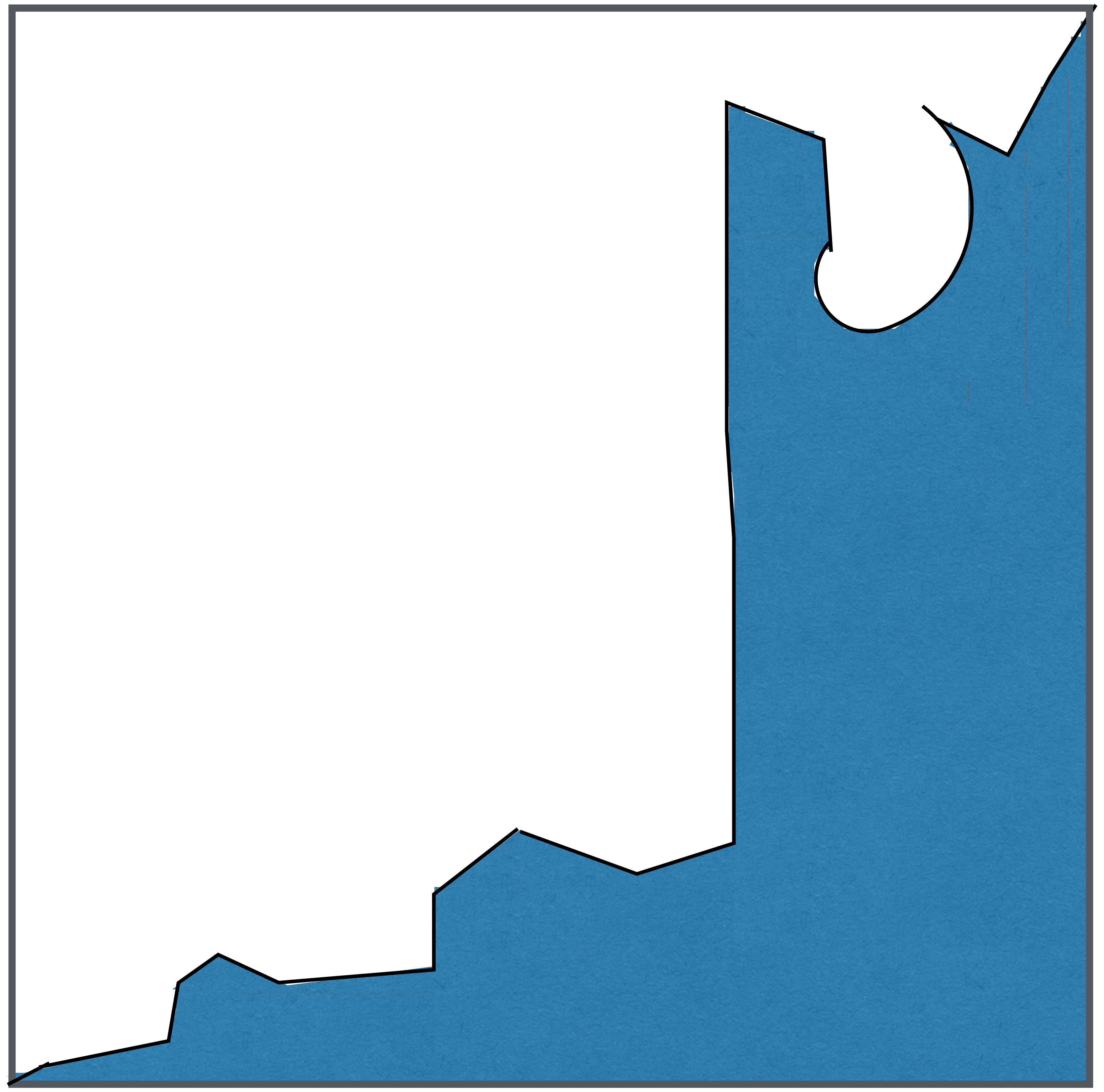
Area: $\frac{\text{base x height}}{2}$

MC - motivation: expectations



Area: ??????

MC - motivation: expectations



Area: ??????

MCArea.ipynb

MC - motivation: expectations

Why am I bothering with areas? - Expectation values are related to areas

Mean $$< \vec{x} >= \frac{1}{N} \sum_{i=1}^{N} N(x_i)$$

$\vec{x}$ = [0, 2, 6, 15, 2]

<x> = 25 / 5 = 5

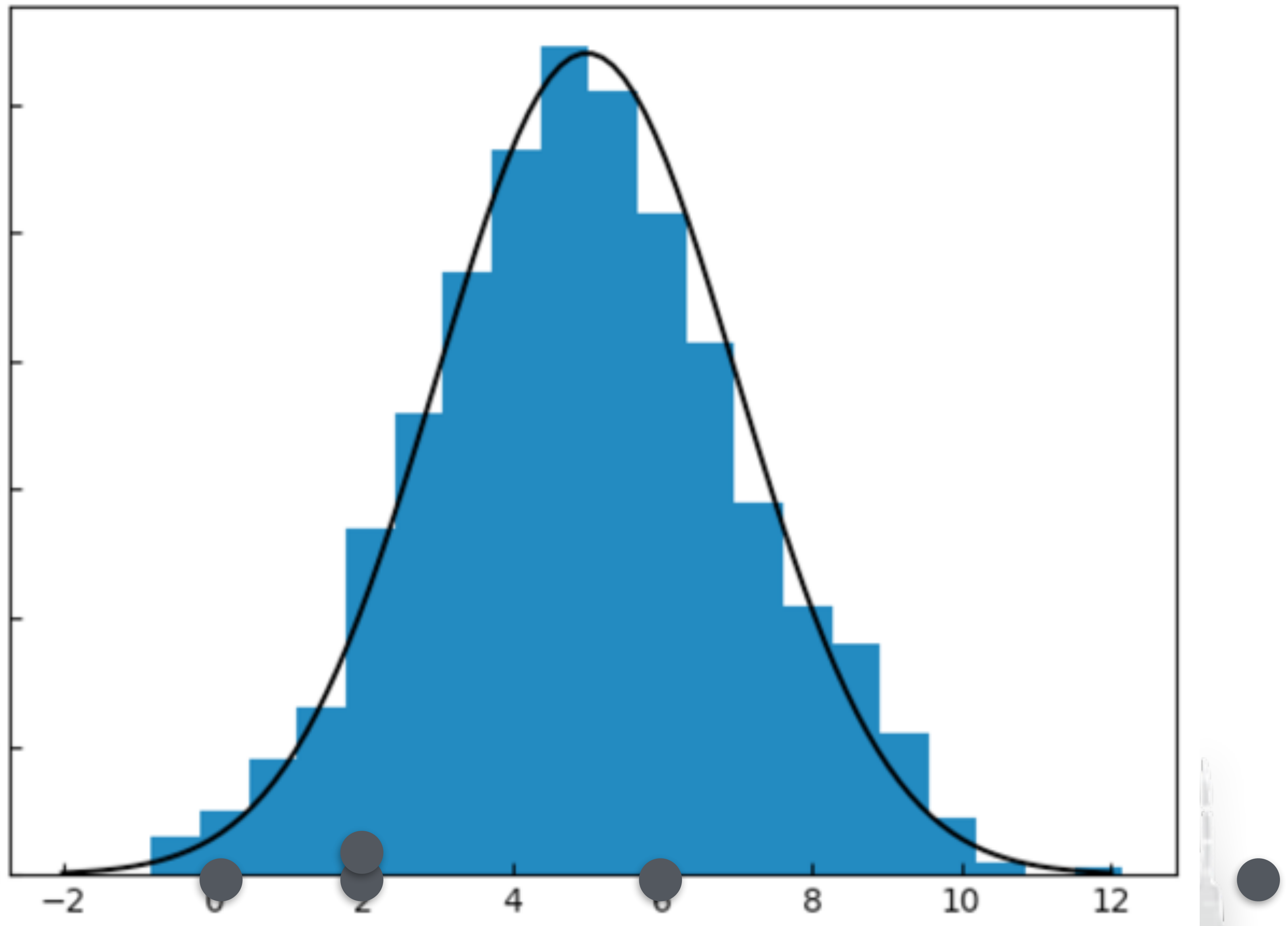
MC - motivation: expectations

Why am I bothering with areas? - Expectation values are related to areas

Mean of a sample

$$< \vec{x} >= \frac{1}{N} \sum_{i=1}^{N} N(x_i)$$

$\vec{x}$ = [0, 2, 6, 15, 2]

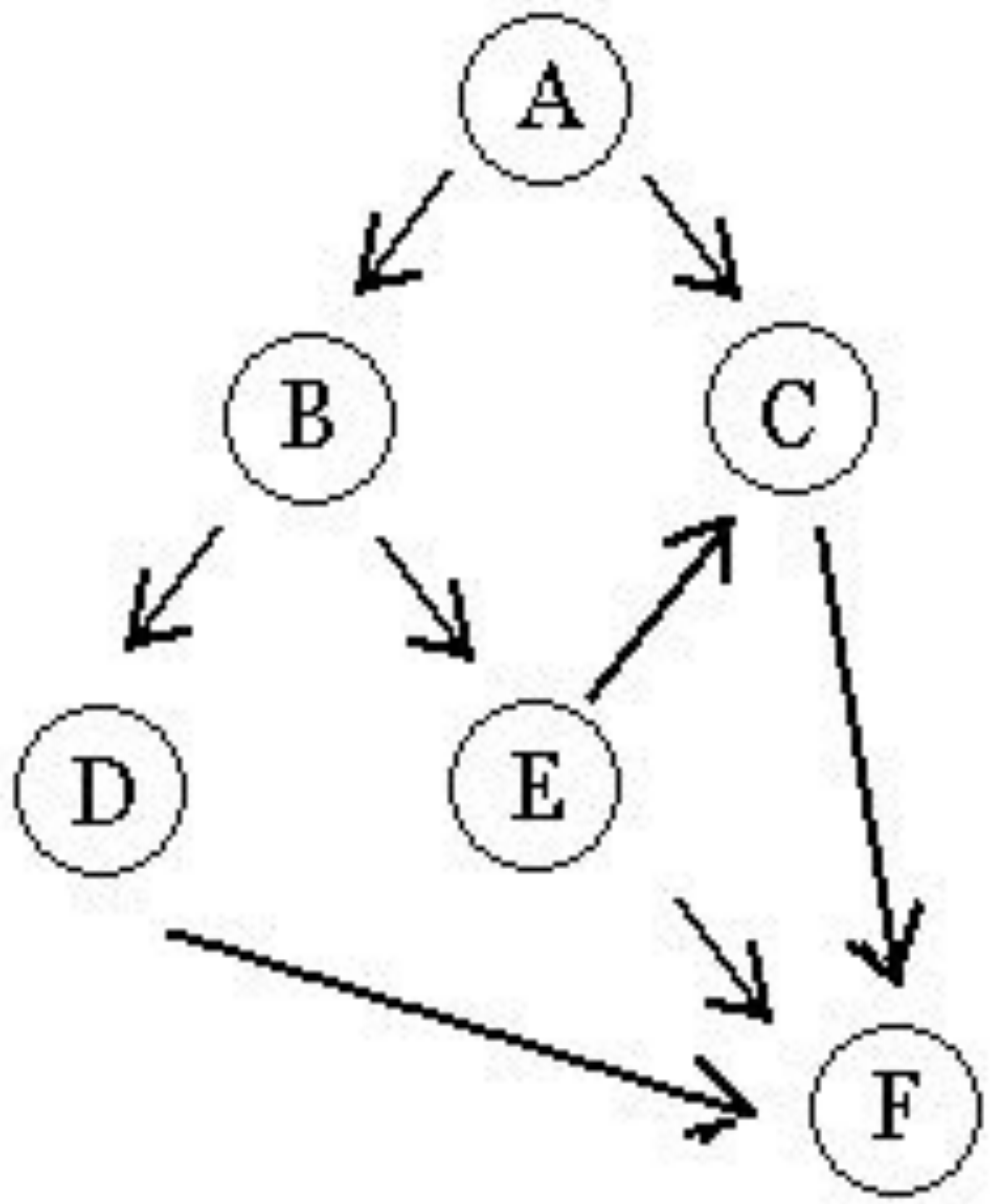


Mean of a continuous distribution

$$mean(X) = E[X] = \int X f(X) dX$$

$$Var(X) = E[X^2] - (E[X])^2.$$

MC - motivation: simulations



Sample

$A \sim P(A)$

$B \sim P(B|A)$

$C \sim P(C|A,E)$

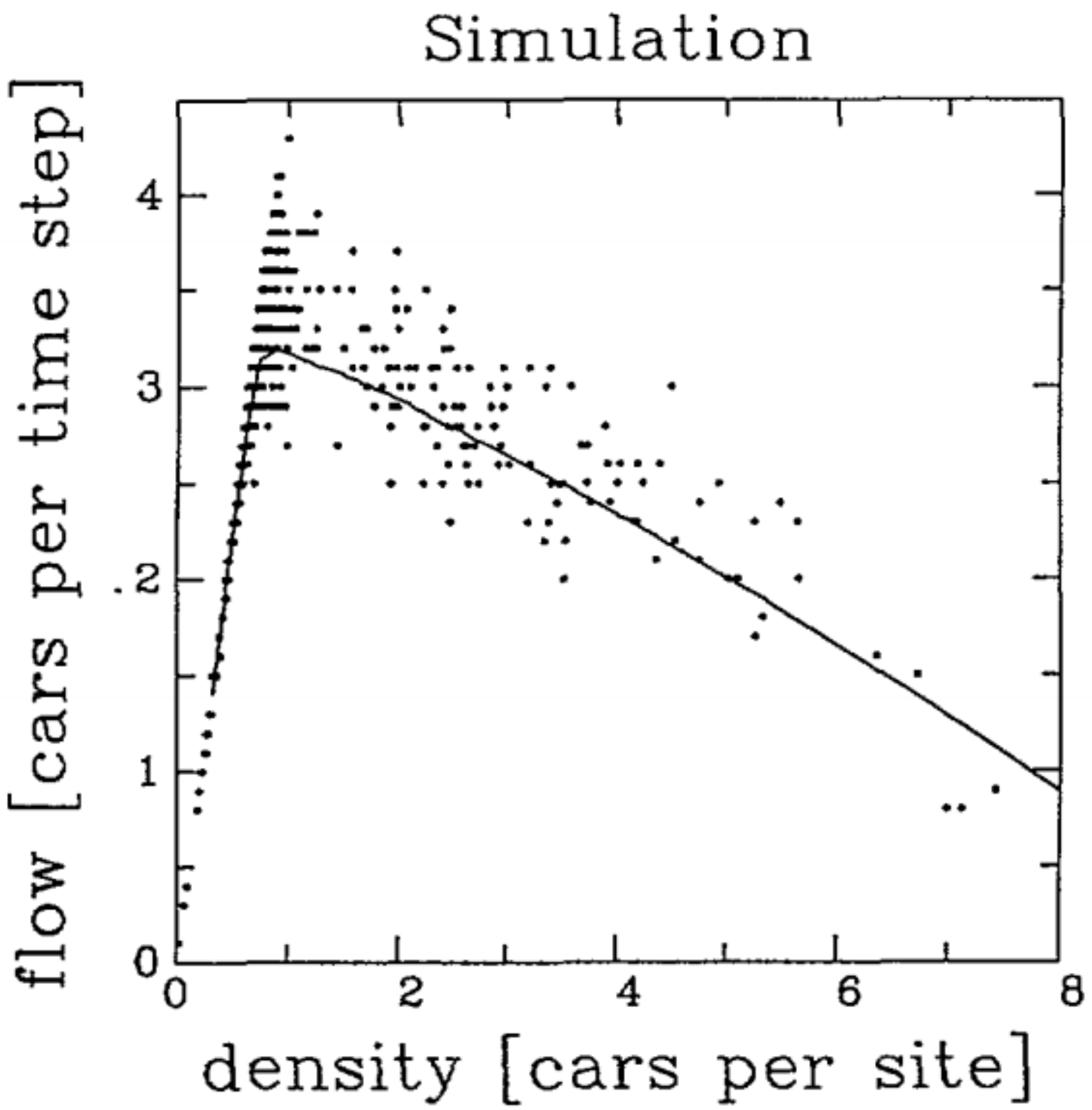
$D \sim P(D|B)$

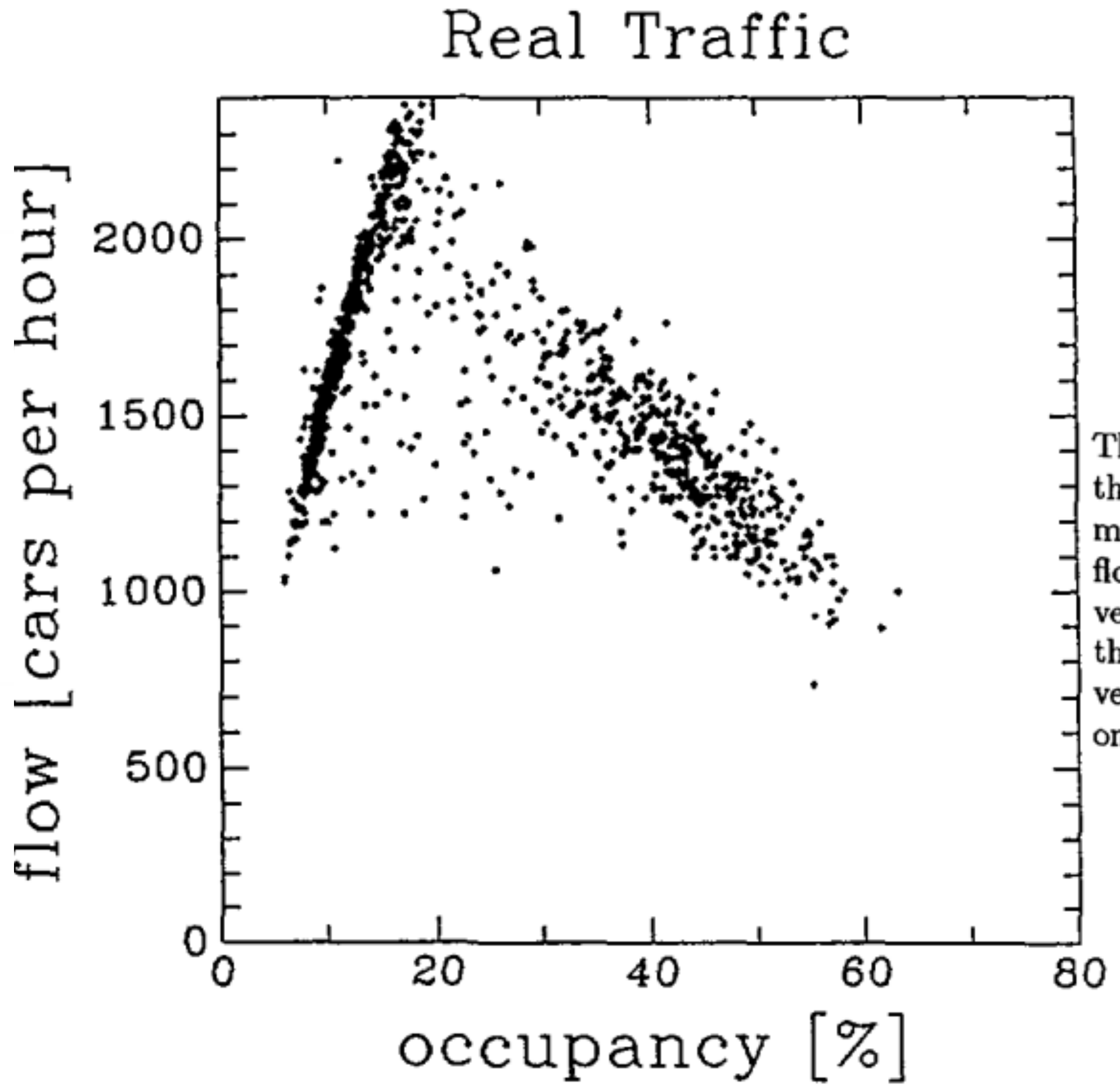
$E \sim P(E|B)$

$F \sim P(F|C,D,E)$

A long history of MC simulation in traffic flow analysis

<u>A cellular automaton model for freeway traffic</u> Nagel & Schreckenber 1992

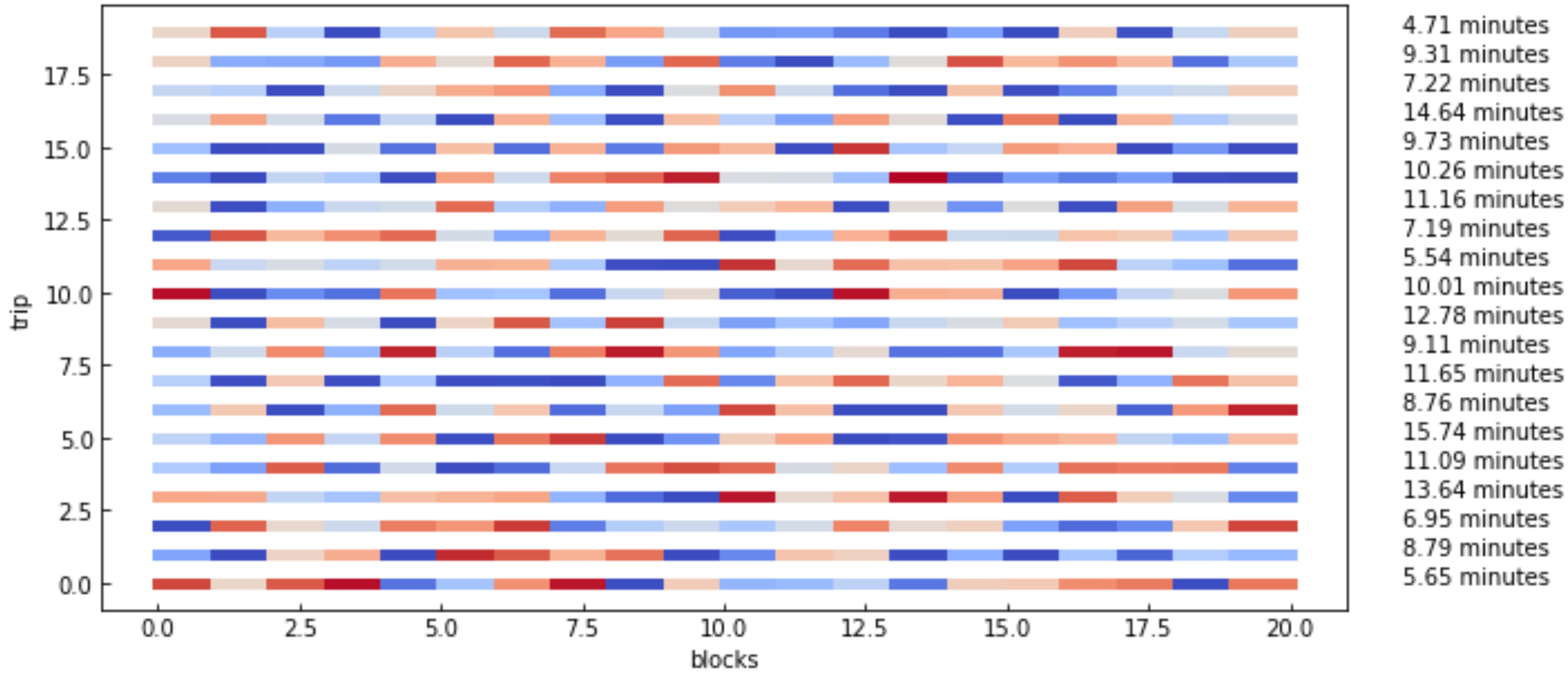




1) **Acceleration:** if the velocity v of a vehicle is lower than $v_{\max}$ and if the distance to the next car ahead is larger than $v + 1$, the speed is advanced by one $[v \to v + 1]$.
2) **Slowing down (due to other cars):** if a vehicle at site i sees the next vehicle at site $i + j$ (with $j \leq v$), it reduces its speed to $j - 1$ $[v \to j - 1]$.
3) **Randomization:** with probability p, the velocity of each vehicle (if greater than zero) is decreased by one $[v \to v - 1]$.
4) **Car motion:** each vehicle is advanced v sites.

Through the steps one to four very general properties of single lane traffic are modelled on the basis of integer valued probabilistic cellular automaton rules [9, 10]. Already this simple model shows nontrivial and realistic behavior. Step 3 is essential in simulating realistic traffic flow since otherwise the dynamics is completely deterministic. It takes into account natural velocity fluctuations due to human behavior or due to varying external conditions. Without this randomness, every initial configuration of vehicles and corresponding velocities reaches very quickly a stationary pattern which is shifted backwards (i.e. opposite the vehicle motion) one site per time step.

MC - Urban Applications



MCstreetLight.ipynb

MC - Urban Applications

Sample

A person’s depth

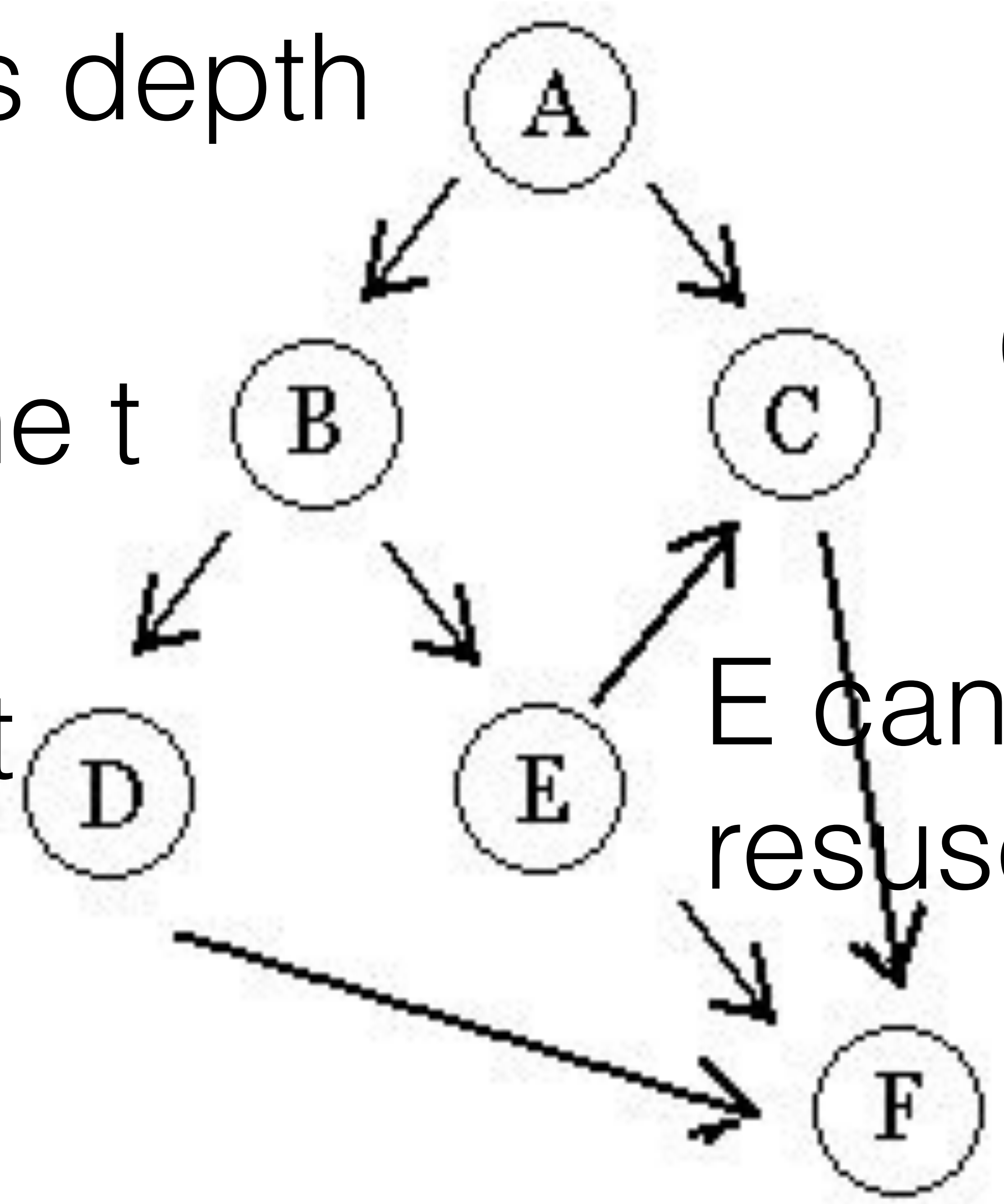
B prob to find them at time t

C they survive the avalange

D they are still alive at t

E can be resuscitated at time t

F person survives



A concept for optimizing avalanche rescue strategies using a Monte Carlo simulation approach

Ingrid Reiweger, Manuel Genswein, Peter Paal, Jürg Schweizer

Published: May 3, 2017 • https://doi.org/10.1371/journal.pone.0175877

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0175877

MC - motivation: sampling

SetUp 1:

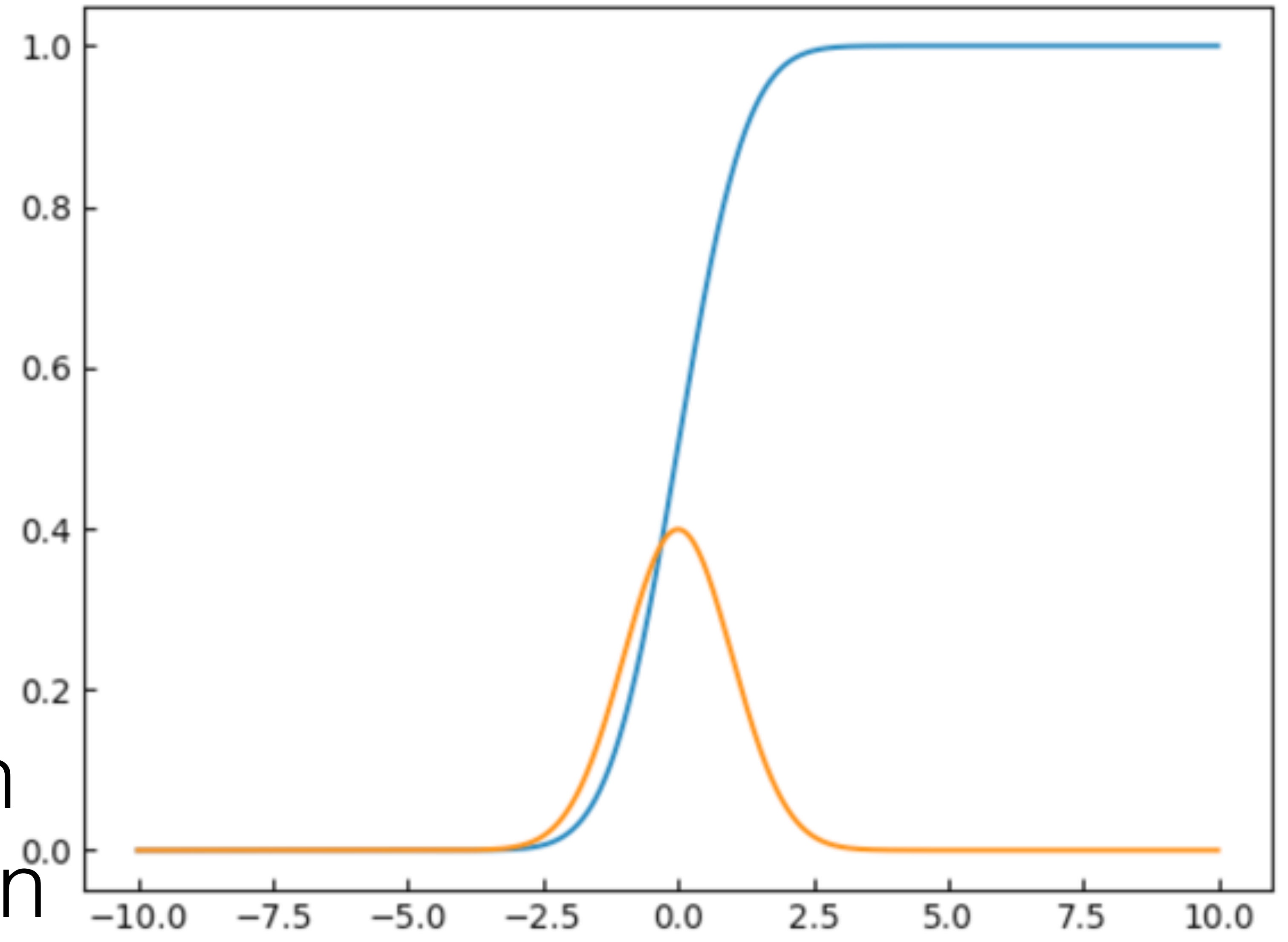
1. I have a distribution described by some formula ***P(x)*** (its PDF)

2. The function can be integrated : *e.g. Gaussian*

$$P(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \int P(x)dx=\frac{1}{2}\left[1+erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

3. If I can *take the integral* of the PDF ***P(x)*** I can calculate the CDF ***F(x)***

4. If I know *and can invert* ***F(x)*** *(i.e. calculate* $\boldsymbol{F^{-1}(u)}$*)* I know at which percentile a value is and I can directly sample from the distribution



MC - motivation: sampling

SetUp 1:

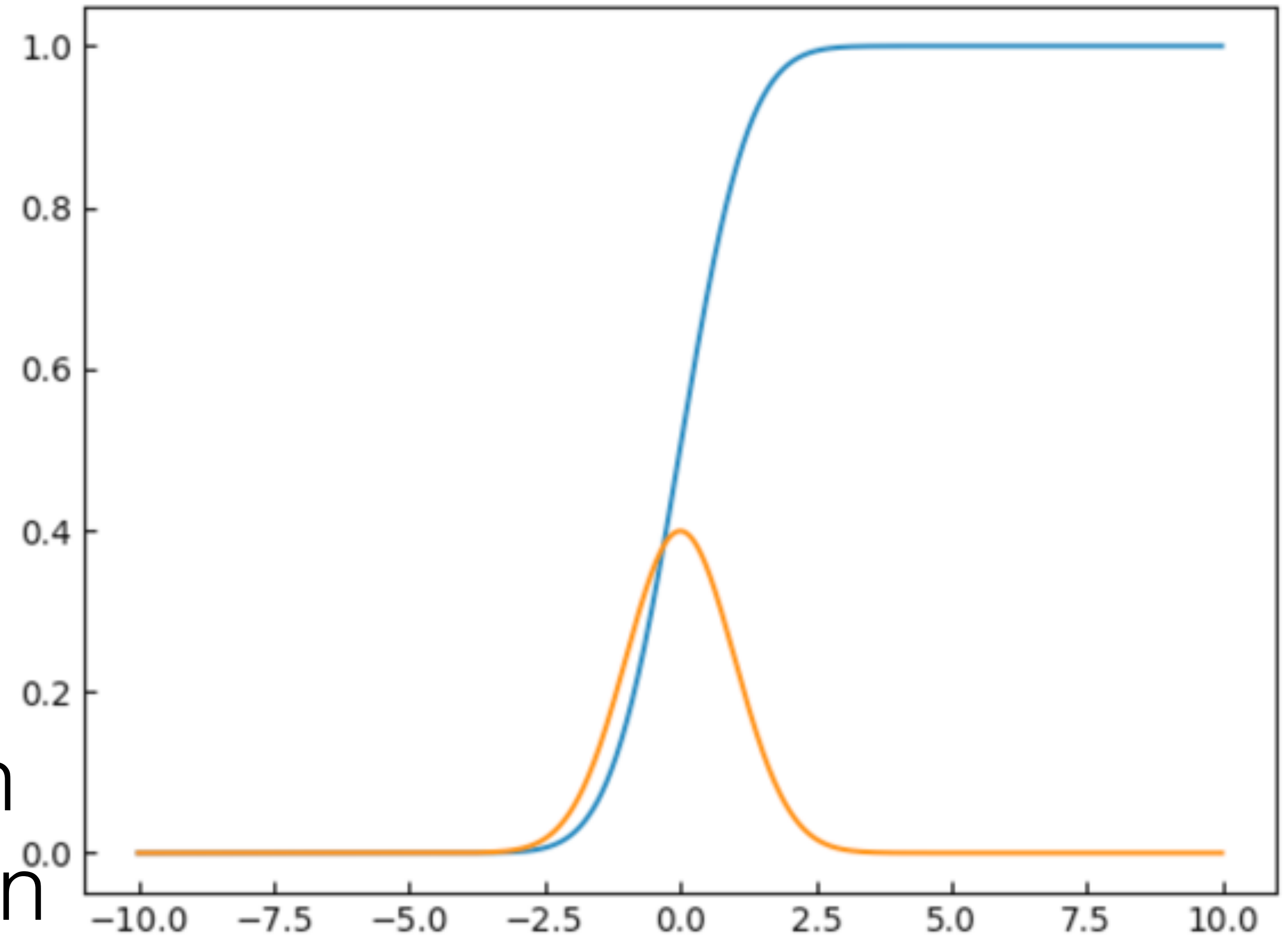
1. I have a distribution described by some formula ***P(x)***

2. The function can be integrated : *e.g. Gaussian*

$$P(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad \int P(x)dx=\frac{1}{2}\left[1+erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

3. If I can *take the integral* of the PDF ***P(x)*** I can calculate the CDF ***F(x)***

4. If I know *and can invert* ***F(x)*** *(i.e. calculate* ***F^{-1}(u)****)* I know at which percentile a value is and I can directly sample from the distribution



WHILE convergence: // ***P(x)*** is filled in

draw a uniform random number ***u ~ Uniform[0,1]***

calculate ***x*** = ***F^{-1}(u)*** *//x is a sample from* ***P***

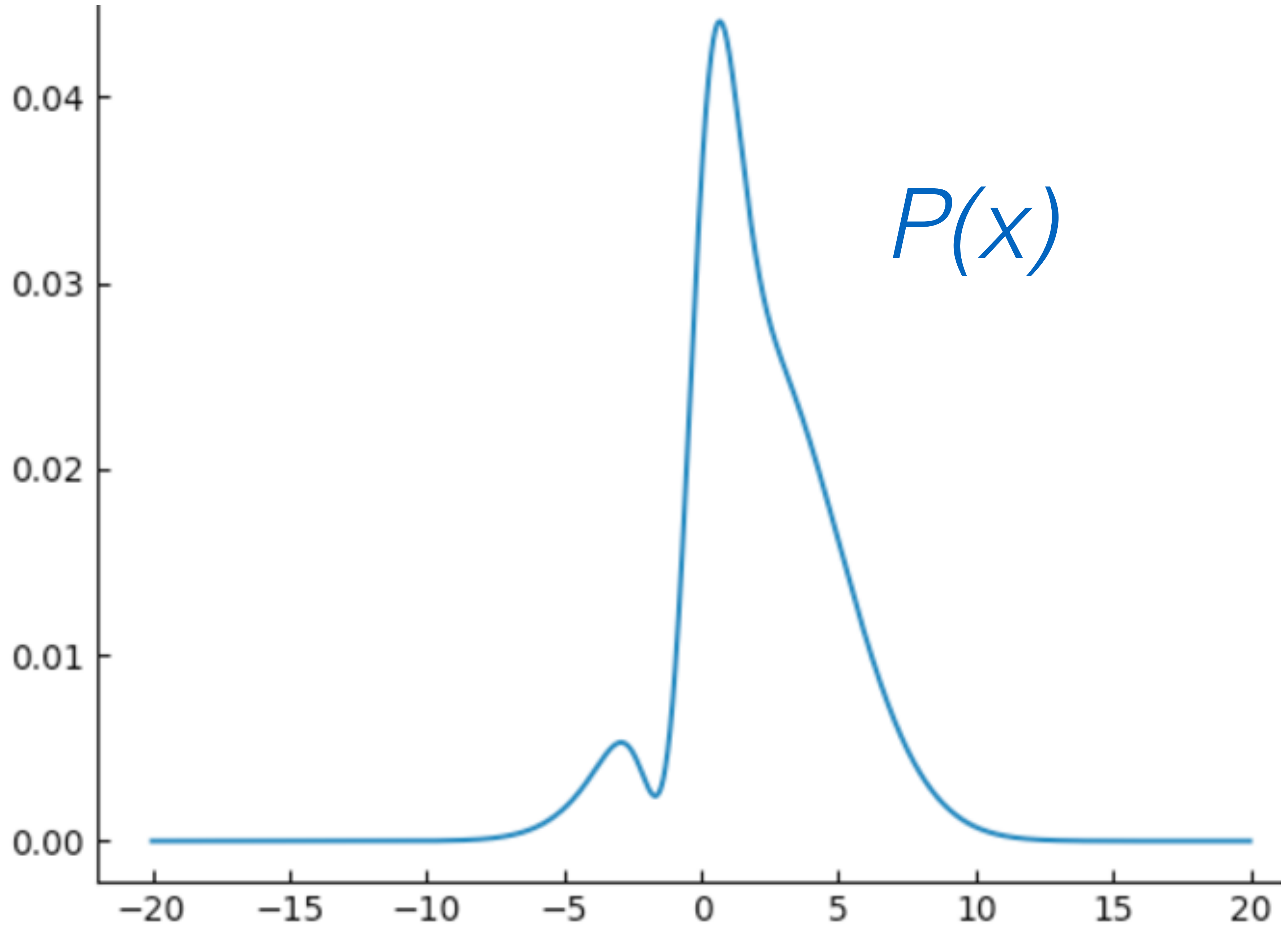
<u>**Slides on sampling from distributions**</u>
Paul E. Johnson 2015

MC - Rejection Sampling

SetUp 2:

1. I have a distribution described by some formula ***P(x)***

2. The function *cannot* be (easily) integrated :
   I dont know how to draw samples
   but I can calculate its value at every x



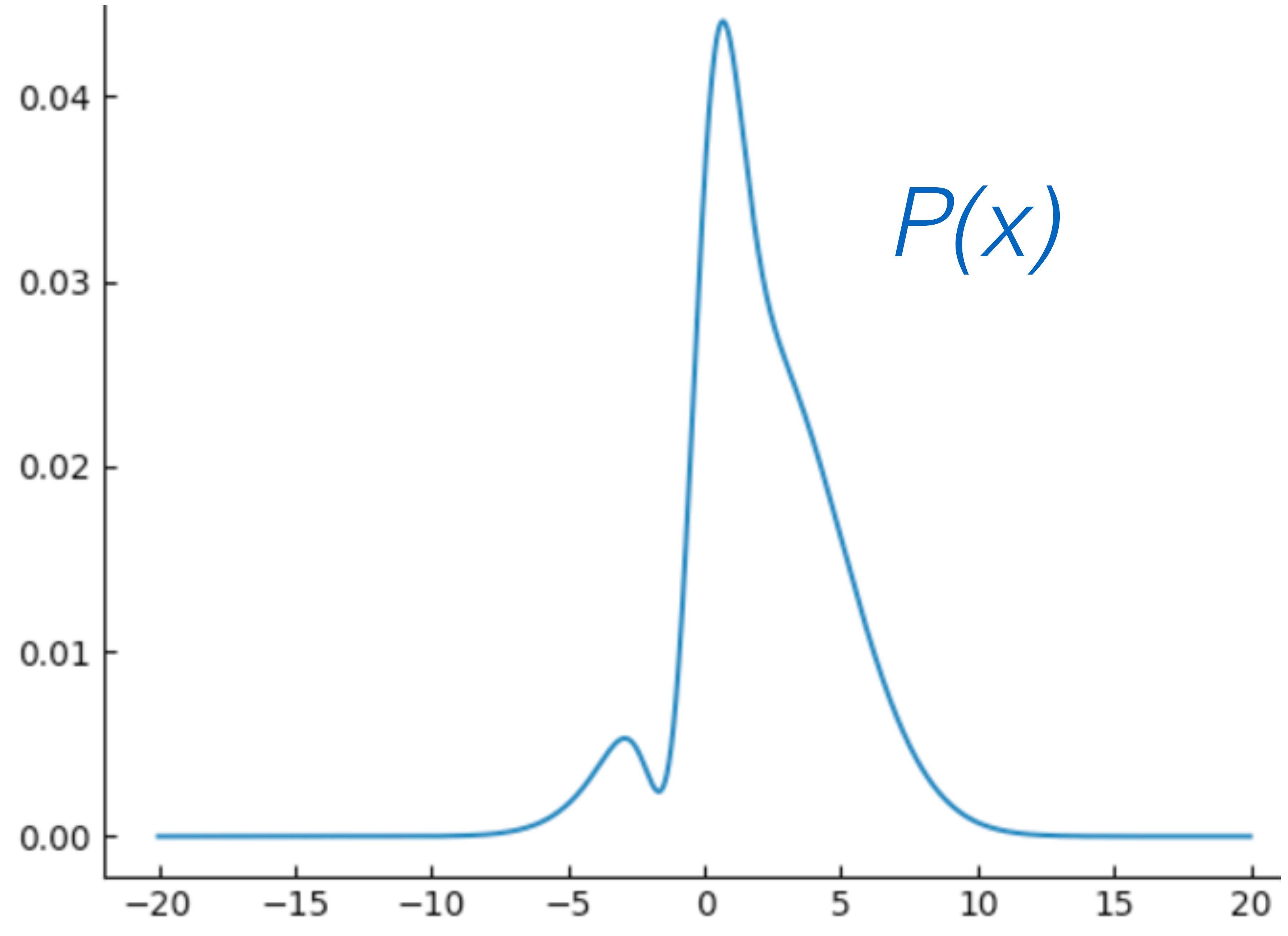
MC - Rejection Sampling

SetUp 2:

1. I have a distribution described by some formula ***P(x)***

2. The function *cannot* be (easily) integrated :
   I dont know how to draw samples
   but I can calculate its value at every x

3. There exist distributions - Q(x) - that are higher than the P(x) at every x



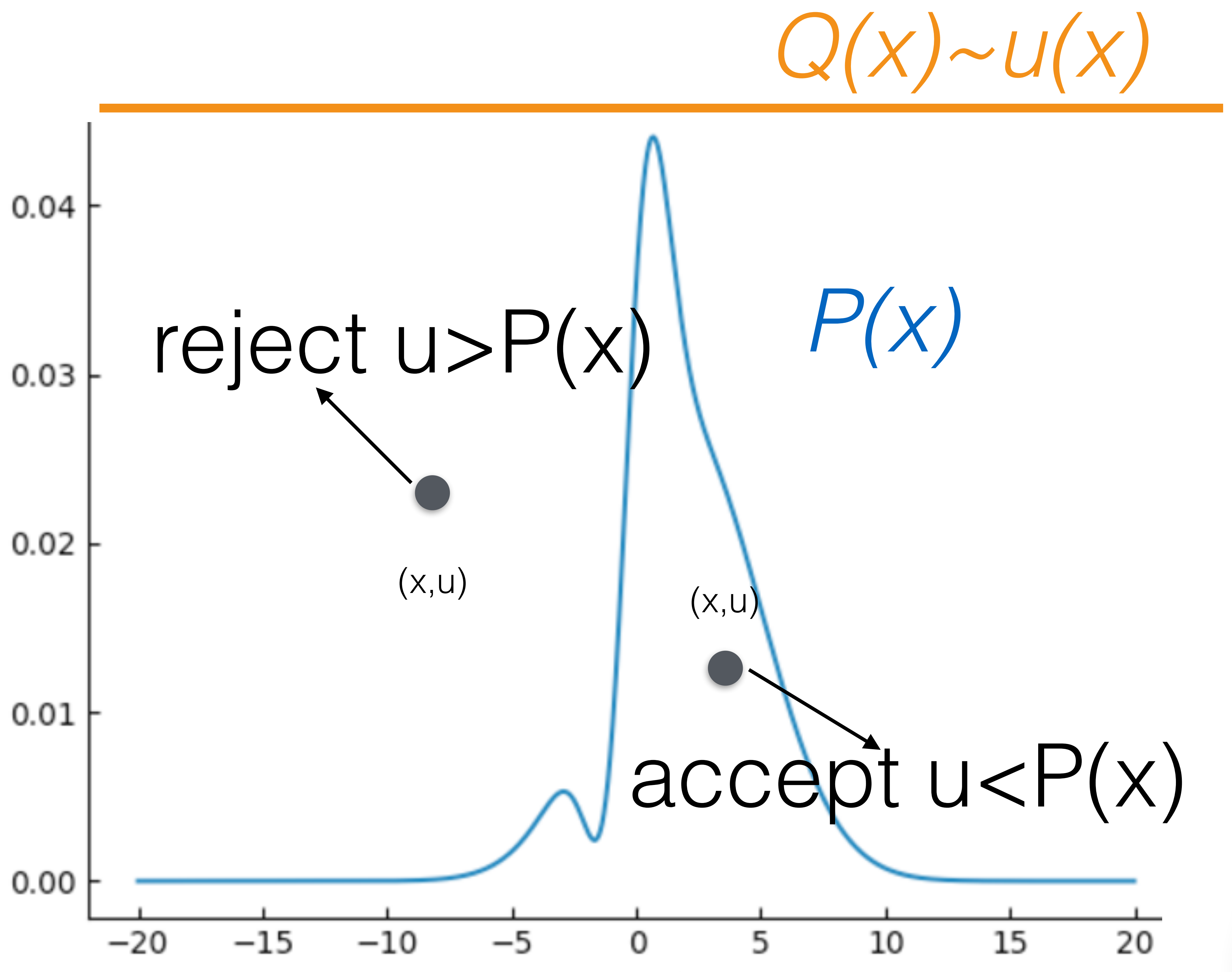
MC - Rejection Sampling

SetUp 2:

1. I have a distribution described by some formula ***P(x)***

2. The function *cannot* be (easily) integrated :
   I dont know how to draw samples
   but I can calculate its value at every x

3. There exist distributions - Q(x) - that are higher than the P(x) at every x : e.g. *Uniform distribution!*



MC - Rejection Sampling

SetUp 2:

1. I have a distribution described by some formula ***P(x)***

2. The function *cannot* be (easily) integrated :
   I dont know how to draw samples
   but I can calculate its value at every x

3. There exist distributions - Q(x) - that are higher than the P(x) at every x : e.g. *Uniform distribution!*

WHILE convergence: // ***P(x)*** is filled in

  draw a point ***x*** from ***Q(x)***

  calculate ***P(x)***

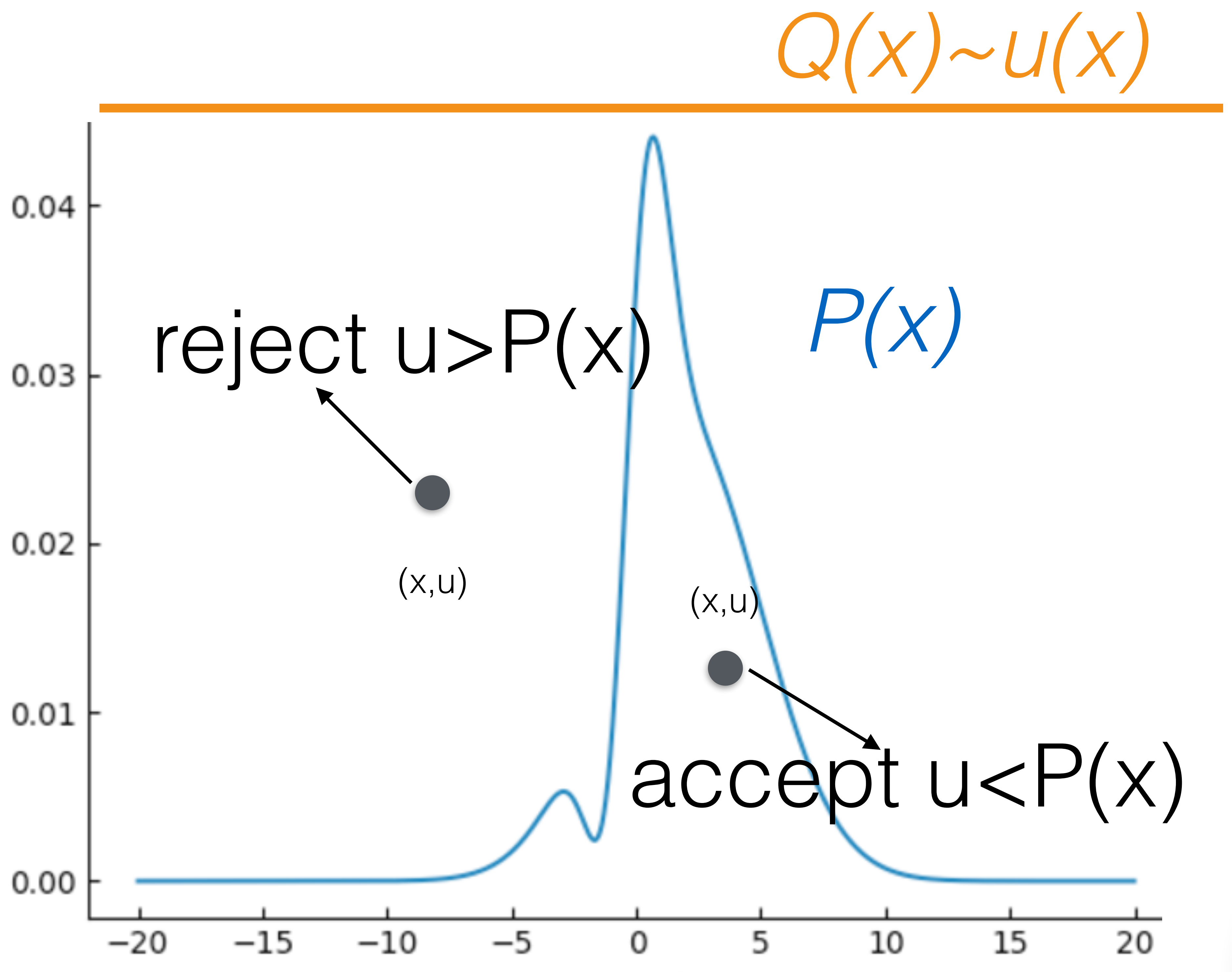
  draw a height ***u ~ Uniform[0,Q(x)]***

  IF : ***u <= P(x)***

    accept // point is sample of ***P(x)***

  ELSE :

    reject



MC - Rejection Sampling

SetUp 2:

1. I have a distribution described by some formula ***P(x)***

2. The function *cannot* be (easily) integrated :
   I dont know how to draw samples
   but I can calculate its value at every x

3. There exist distributions - Q(x) - that are higher than the P(x) at every x : e.g. *Gaussian distribution!*

WHILE convergence: // ***P(x)*** is filled in

draw a point ***x*** from ***Q(x)***

calculate ***P(x)***

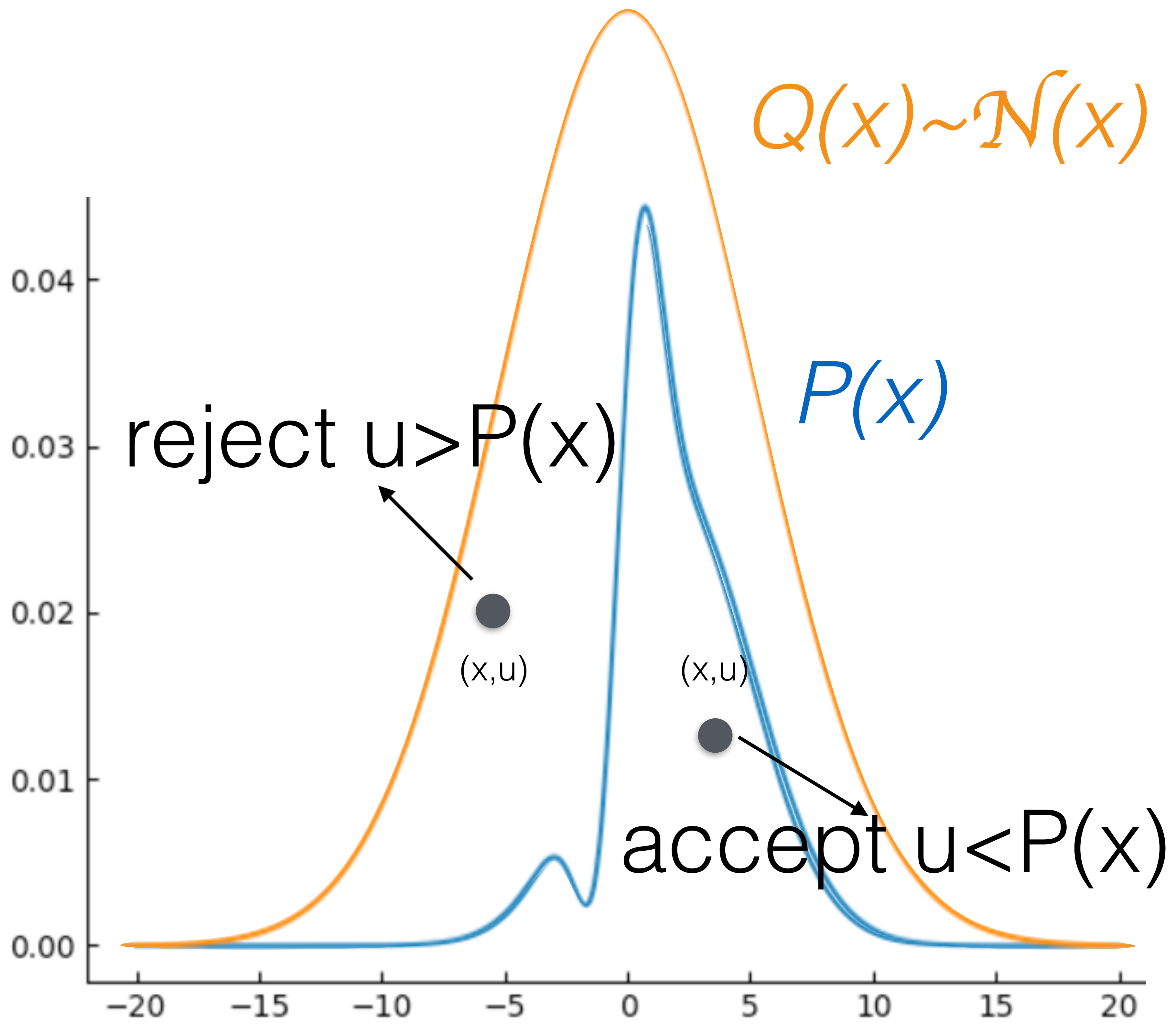
draw a height ***u ~ Uniform[0,Q(x)]***

IF : ***u <= P(x)***

accept // point is sample of ***P(x)***

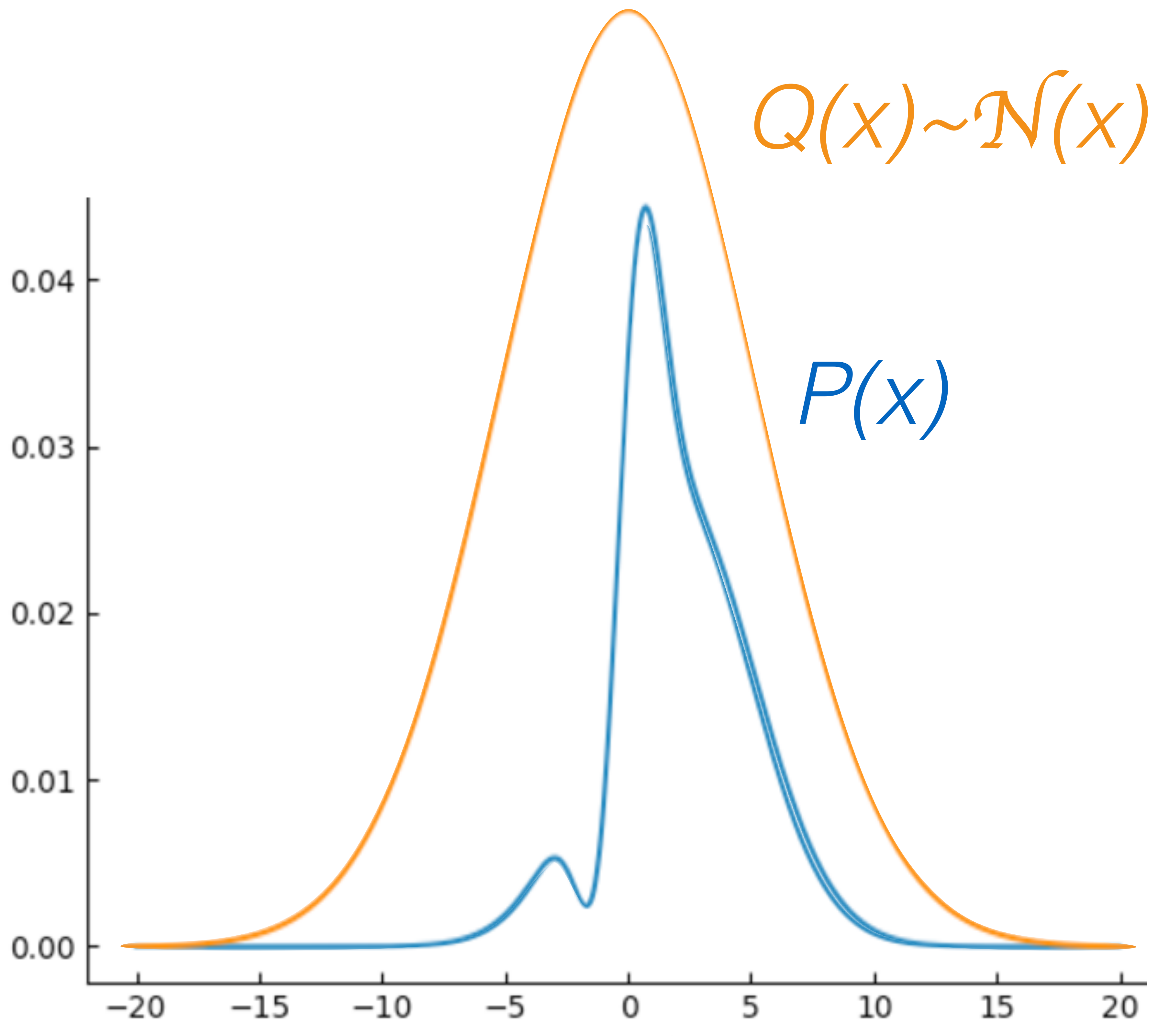
ELSE :

reject



SetUp 2:

1. I have a distribution described by some formula ***P(x)***
2. The function *cannot* be (easily) integrated :
   I dont know how to draw samples
   but I can calculate its value at every x
3. There exist distributions - Q(x)



$$\int f(x)P(x)dx = \int f(x)\frac{P(x)}{Q(x)}Q(x)dx, \quad (Q(x)>0 \ \ if P(x)>0)$$
$$\approx \frac{1}{S}\sum_{s=1}^{S} f(x_s)\frac{P(x_s)}{Q(x_s)}, \qquad x(s) \sim Q(x)$$

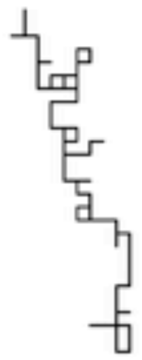
Q(x) <=> P(x) guarantees that the integral does not diverge
chose Q(x) s.t. Q(x) is large where f(x) P(x) is large

Markov Chain Monte Carlo

Markov Chain

Markov Chain

memory-less stockastic process:
make predictions for the future of the process based solely on its present state independently from the previous history;
i.e. the next state of the process is based on a chosen distribution (e.g. gaussian) with parameters that depend only on the current state (e.g. with mean at the current state)



Markov Chain

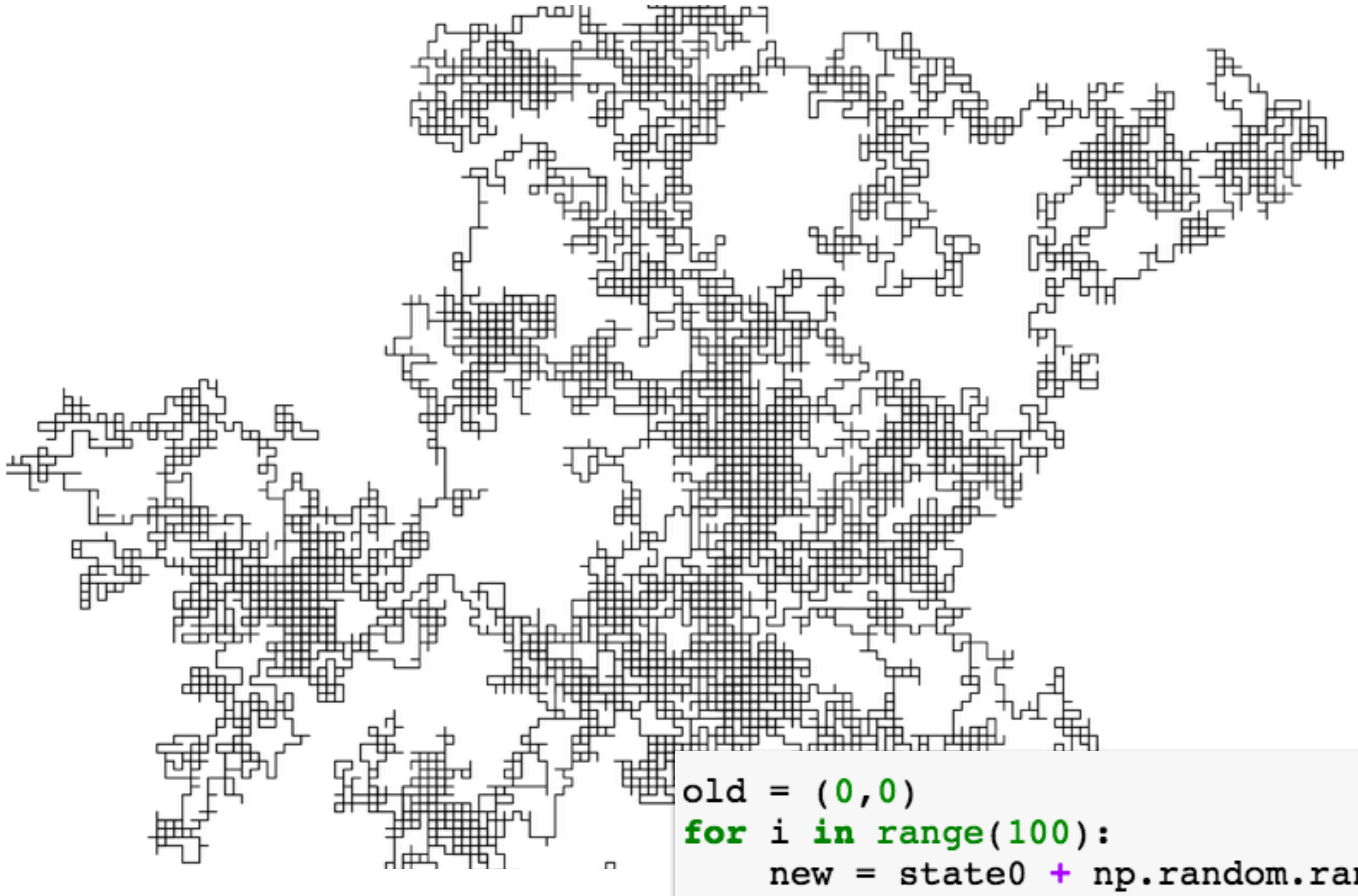
memory-less stockastic process:

e.g.:

Random Walk -> *next position is a stockastic perturbation over current position*

Gamblers' ruin

Markov processes



```
old = (0,0)
for i in range(100):
    new = state0 + np.random.rand(2)
    old = new
```

Markov Chain

memory-less stockastic process:

e.g.:

Random Walk -> choose next position as a gaussian perturbation over the current

Gamblers' ruin

Markov Chain Monte Carlo

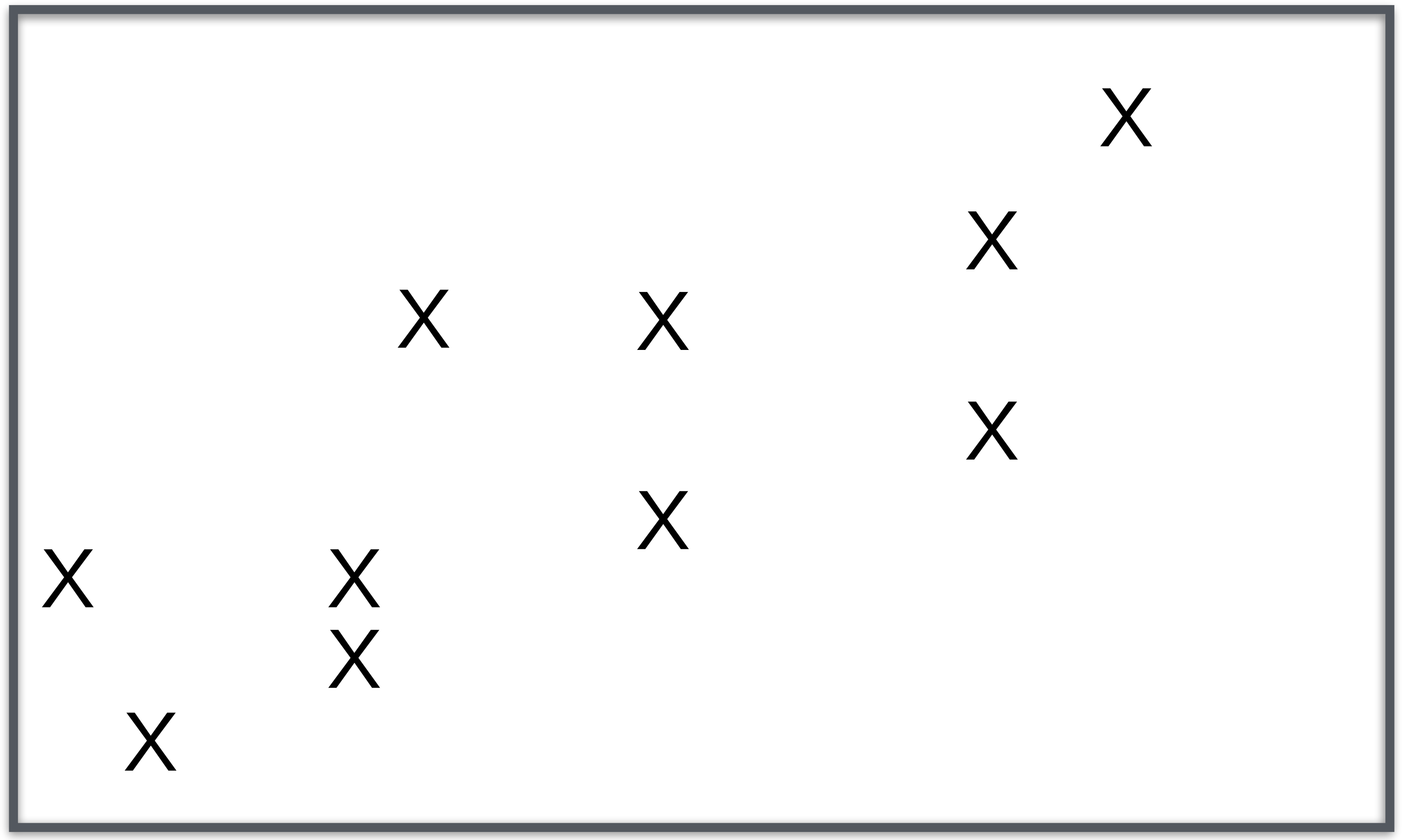
MCMC - motivation

I have a model and I want to find the best parameters to describe my data

Data ***D***

Model - some function *f(**θ**)*

Paremeters ***θ***



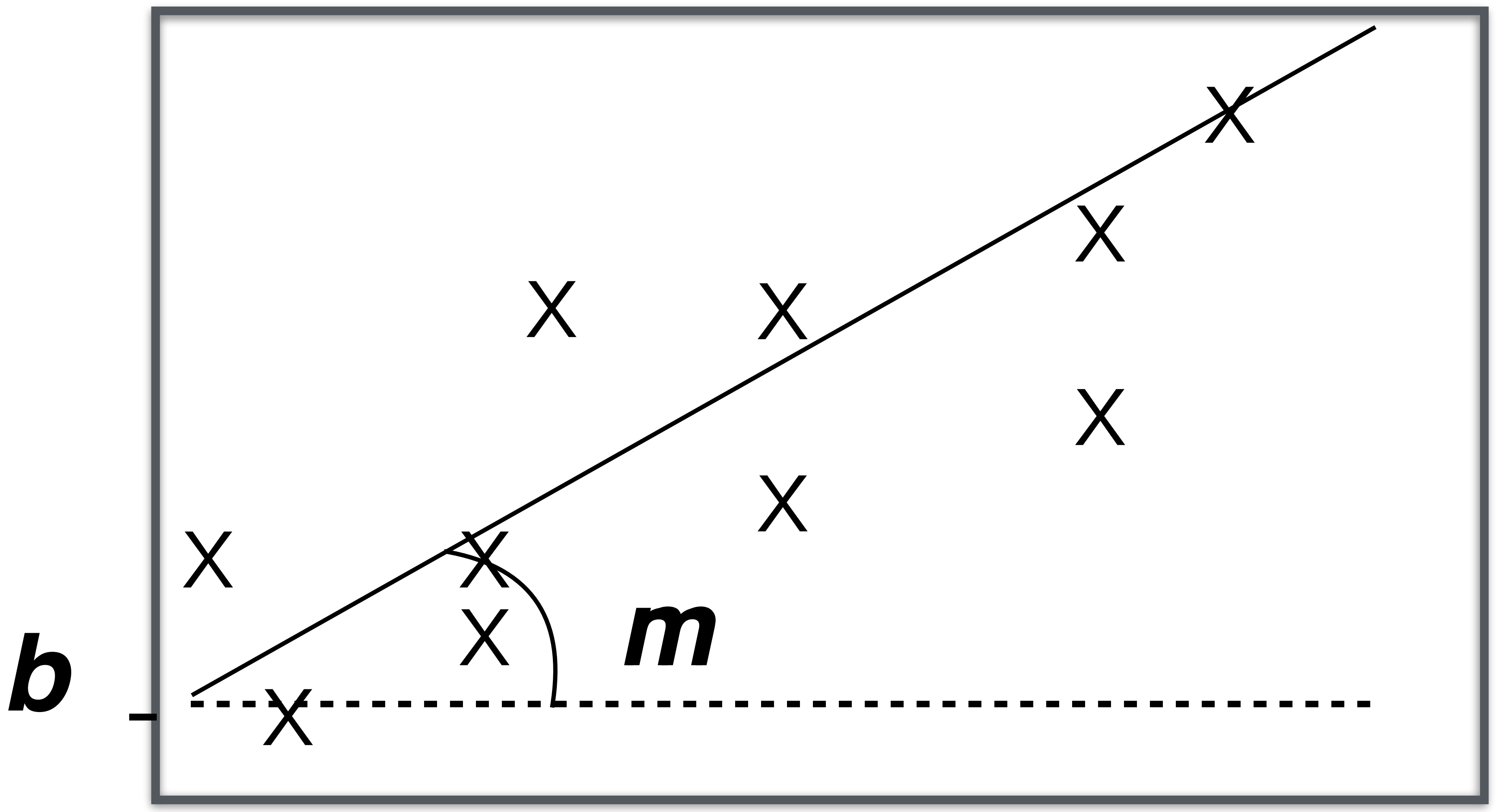
MCMC - motivation

I have a model and I want to find the best parameters to describe my data

Data ***D***

Model $f(\mathbf{m}, \mathbf{b}) = mx + b$

Paremeters $\boldsymbol{\theta} = (\mathbf{m},\mathbf{b})$



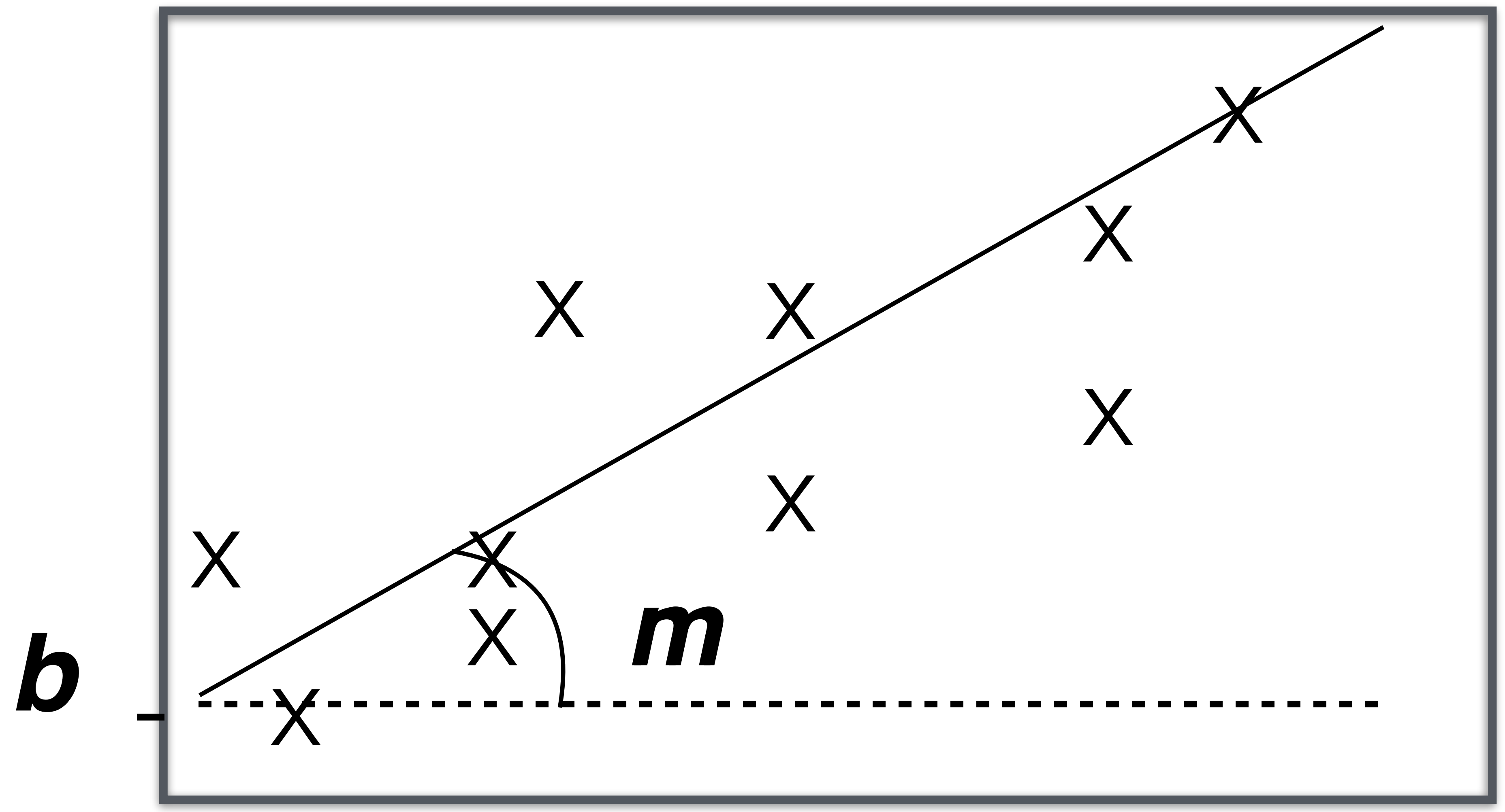
MCMC - motivation

I have a model and I want to find the best parameters to describe my data

Data ***D***

Model $f(\mathbf{m}, \mathbf{b}) = mx + b$

Paremeters $\boldsymbol{\theta} = (\mathbf{m,b})$



To find the best model parameters:
maximize likelihood: ***θ such that P(D|θ) is max***

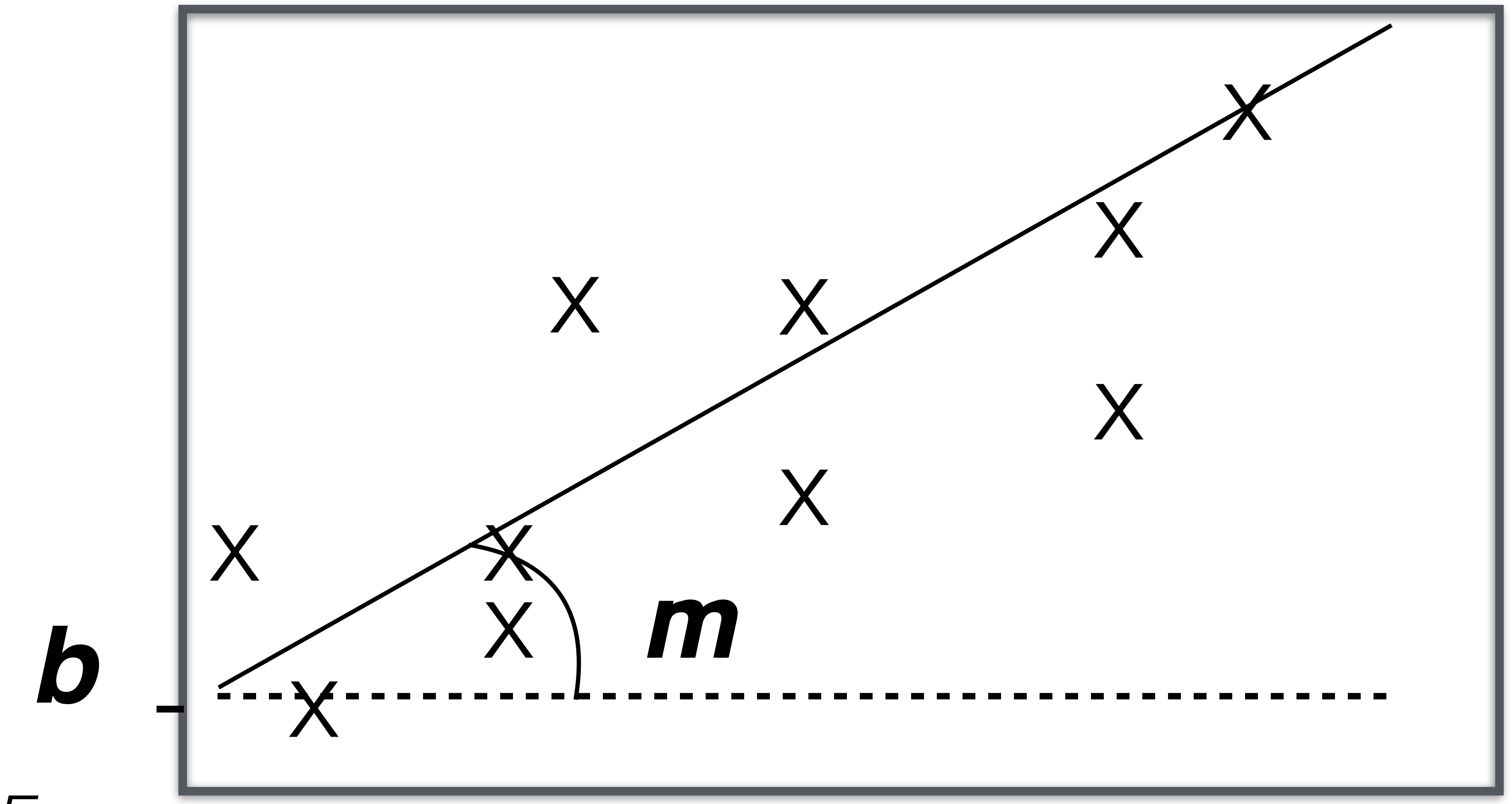
https://github.com/fedhere/PUI2016_fb55/blob/master/HW6_fb55/building_nrg_solution.ipynb

I have a model and I want to find the best parameters to describe my data

Data ***D***

Model *f*(***m*, *b***) = *mx* + *b*

Paremeters ***θ* = (*m*,*b*)**



Refresher about *likelihood:*

- *The likelihood of a distribution has the same form as its PDF*
- *the likelohood of a Gaussian distribution is:* $L \equiv P(D|N,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\ \sigma} \Pi_{i=1}^{N} exp\left(-\frac{(\mu-x)^2}{2\sigma^2}\right)$
- *generally we like to work in log space with likelihoods because they can be very large numbers and finding the maximum is equivalent to finding a 0 in log space*

$$log(L) = -\frac{1}{2}\sum_{i=1}^{N} log(2\pi) - \sum_{i=1}^{N} log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\frac{(\mu-x)^2}{\sigma^2} = C - \frac{1}{2}\chi^2$$

PUI2016 HW 6, lecture 6

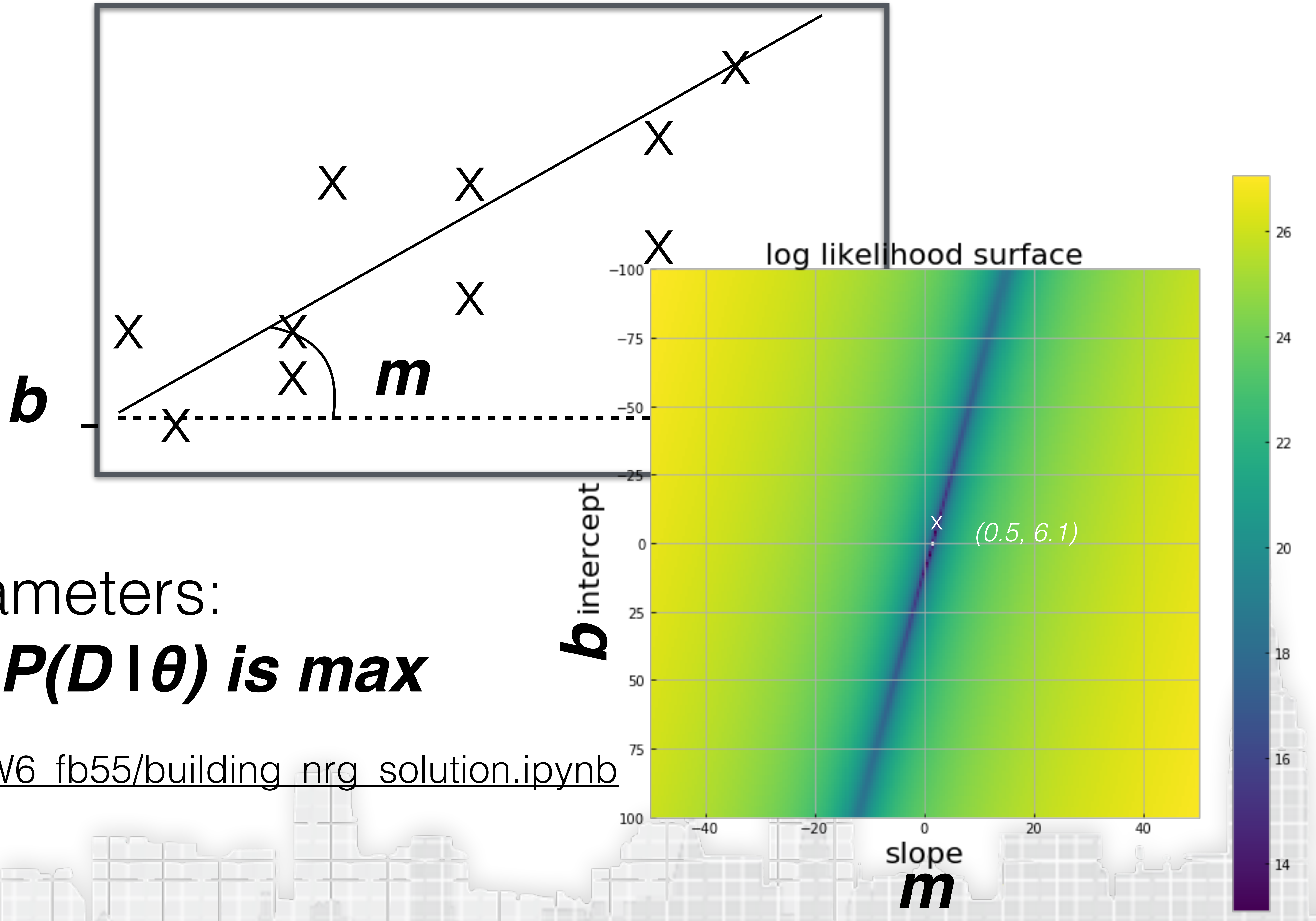
MCMC - motivation

I have a model and I want to find the best parameters to describe my data

Data ***D***

Model $f(\mathbf{m}, \mathbf{b}) = mx + b$

Paremeters $\boldsymbol{\theta = (m,b)}$



To find the best model parameters:
maximize likelihood: ***θ such that P(D|θ) is max***

https://github.com/fedhere/PUI2016_fb55/blob/master/HW6_fb55/building_nrg_solution.ipynb

Bayes theorem:

$$P(\theta|D)=\frac{P(D|\theta)\,P(\theta)}{P(D)}$$

Bayes theorem:

$$P(\theta|D) = \frac{P(D|\theta)\,P(\theta)}{P(D)}$$

Definitions:

posterior: joint probability distributin of a parameter set (*m*, *b*) condition upon some data *D* and a model hypothesys *f*

Bayes theorem: $$P(\theta|D,f) = \frac{P(D|\theta,f)\,P(\theta,f)}{P(D|f)}$$

Definitions: **posterior**

posterior: joint probability distributin of a parameter set (*m*, *b*) condition upon some data *D* and a model hypothesys *f*

Bayes theorem:

$$\underbrace{P(\theta|D,f)}_{\text{posterior}} = \frac{\overbrace{P(D|\theta,f)}^{\text{likelihood}}\,\overbrace{P(\theta,f)}^{\text{prior}}}{\underbrace{P(D|f)}_{\text{evidence}}}$$

Definitions:

posterior: joint probability distributin of a parameter set (*m*, *b*) condition upon some data *D* and a model hypothesys *f*

MCMC - motivation

Bayes theorem:

$$P(\theta|D,f) = \frac{P(D|\theta,f)\,P(\theta,f)}{P(D|f)}$$

prior

Definitions: posterior

posterior: joint probability distributin of a parameter set (*m*, *b*) condition upon some data *D* and a model hypothesys *f*

prior: “intellectual” knowledge about the model parameters

Bayes theorem: $$P(\theta|D,f) = \frac{P(D|\theta,f)\,P(\theta,f)}{P(D|f)}$$

prior

Definitions: **posterior**

posterior: joint probability distributin of a parameter set (*m*, *b*) condition upon some data *D* and a model hypothesys *f*

prior: "intellectual" knowledge about the model parameters

e.g.: energy consumption increaded w number of units: m>0

Bayes theorem:

$$\underbrace{P(\theta|D,f)}_{\text{posterior}} = \frac{P(D|\theta,f)\,\overbrace{P(\theta,f)}^{\text{prior}}}{\underbrace{P(D|f)}_{\text{evidence}}}$$

Definitions:

posterior: joint probability distributin of a parameter set (*m*, *b*) condition upon some data *D* and a model hypothesys *f*

prior: “intellectual” knowledge about the model parameters

evidence: marginal likelihood of data under the model

$$P(D|f) = \int P(D|\theta,f)P(\theta|f)d\theta$$

Bayes theorem: $$\underbrace{P(\theta|D,f)}_{\text{posterior}} = \frac{P(D|\theta,f)\overbrace{P(\theta,f)}^{\text{prior}}}{\cancel{\underbrace{P(D|f)}_{\text{evidence}}}}$$

Definitions: posterior prior evidence

posterior: joint probability distributin of a parameter set (*m*, *b*) condition upon some data *D* and a model hypothesys *f*

prior: "intellectual" knowledge about the model parameters

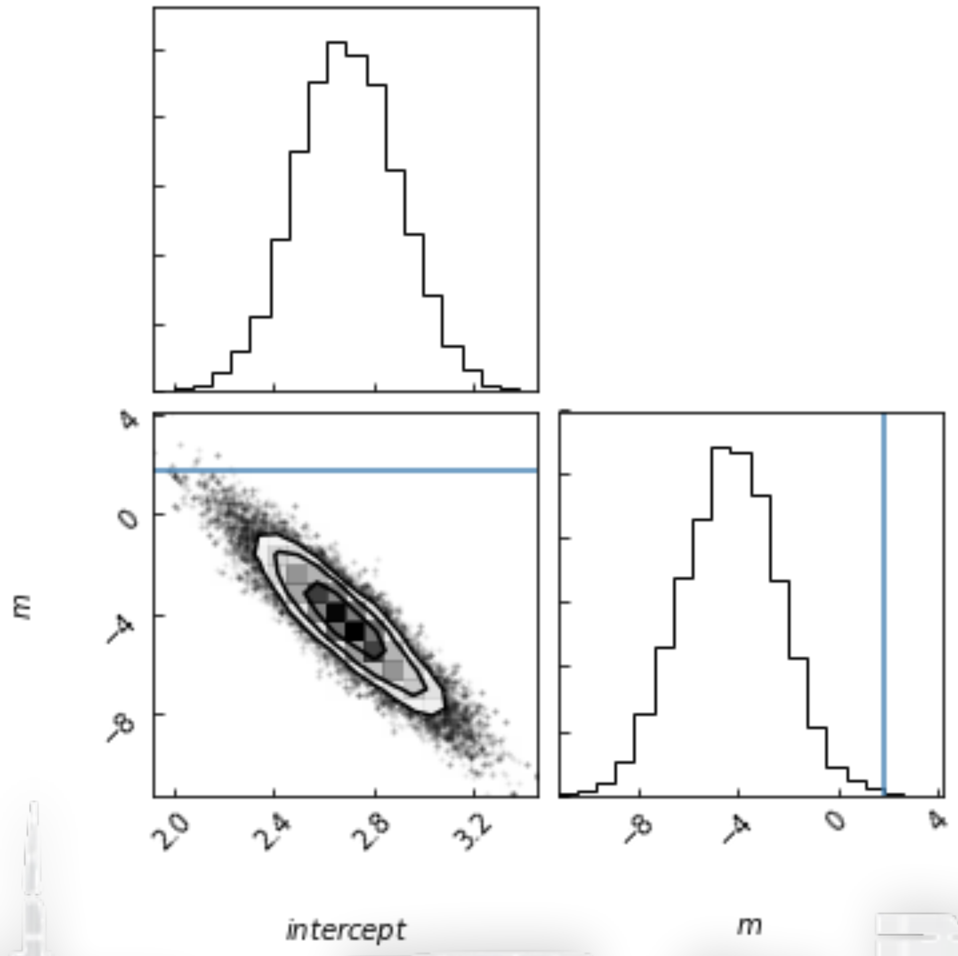
evidence: marginal likelihood of data under the model
its constant in θ so we can ignore it $P(D|f)=\int P(D|\theta,f)P(\theta|f)d\theta$

Bayes theorem: $$P(\theta|D,f) \propto P(D|\theta,f)\,P(\theta,f)$$

Definitions: **posterior**

posterior: joint probability distributin of a parameter set (*m*, *b*) condition upon some data *D* and a model hypothesys *f*

triangle plot



MCMC - Metropolis Hastings algorithm

Bayes theorem: $$P(\theta|D,f) \propto P(D|\theta,f)\,P(\theta,f)$$

Goal: sample the posterior distribution

A nice tutorial on MCMC by Thomas Wiecki (Quantopian)

While My MCMC Gently Samples

Bayesian modeling, Computational Psychiatry, and Python

choose a starting point **current** = $\boldsymbol{\theta_0 = (m,b)}$

WHILE convergence criterion is met:

calculate current posterior **post_curr =** $\boldsymbol{P(D|\theta,f)}$

/*proposal*/
chose a new set of parameters **new =** $\boldsymbol{\theta_{new} = (m,b)}$

calculate the new posterior **post_new=**$\boldsymbol{P(D|\theta_{new},f)}$

IF ***post_new > post_curr:***
current = new
ELSE**:**
/*accept with probability $\boldsymbol{P(D|\theta_{new},f) / P(D|\theta,f)}$ */
r = random uniform number [0,1]
IF **r <** ***post_new / post_orig:***
current = new
ELSE:
pass //do nothing

MCMC - Metropolis Hastings algorithm

Bayes theorem: $$P(\theta|D,f) \propto P(D|\theta,f)\,P(\theta,f)$$

Goal: sample the posterior distribution

Questions:

1. how do I choose the next point?

choose a starting point **current** = ***θ₀= (m,b)***

WHILE convergence criterion is met:

calculate current posterior **post_curr =** ***P(D|θ,f)***

/*proposal*/
chose a new set of parameters **new =** ***θ_new= (m,b)***

calculate the new posterior **post_new=*P(D|θ_new,f)***

IF ***post_new > post_curr:***
current = new
ELSE***:***
/*accept with probability ***P(D|θ_new,f) / P(D|θ,f)*** */
r = random uniform number [0,1]
IF **r <** ***post_new / post_orig:***
current = new
ELSE:
pass //do nothing

MCMC - Metropolis Hastings algorithm

Bayes theorem: $P(\theta|D,f) \propto P(D|\theta,f)P(\theta,f)$

Goal: sample the posterior distribution

Questions:

1. how do I choose the next point?

   Any *Markovian* process

```
choose a starting point current = θ0= (m,b)
WHILE convergence criterion is met:
        calculate current posterior post_curr = P(D|θ,f)
        /*proposal*/
        chose a new set of parameters new = θnew= (m,b)
        calculate the new posterior post_new=P(D|θnew,f)
        IF post_new > post_curr:
            current = new
        ELSE:
            /*accept with probability P(D|θnew,f) / P(D|θ,f) */
            r = random uniform number [0,1]
            IF r < post_new / post_orig:
                current = new
            ELSE:
                pass  //do nothing
```

MCMC - Metropolis Hastings algorithm

Bayes theorem: $P(\theta|D,f) \propto P(D|\theta,f)\,P(\theta,f)$

Goal: sample the posterior distribution

Questions:

1. how do I choose the next point?

   Any *Markovian* process

   Any *ergodic* process
   (with enough time all locations will be explored)

```
choose a starting point current = θ0= (m,b)
WHILE convergence criterion is met:
        calculate current posterior post_curr = P(D|θ,f)
        /*proposal*/
        chose a new set of parameters new = θnew= (m,b)
        calculate the new posterior post_new=P(D|θnew,f)
        IF post_new > post_curr:
            current = new
        ELSE:
            /*accept with probability P(D|θnew,f) / P(D|θ,f) */
            r = random uniform number [0,1]
            IF r < post_new / post_orig:
                current = new
            ELSE:
                pass  //do nothing
```

MCMC - Metropolis Hastings algorithm

Bayes theorem: $P(\theta|D,f) \propto P(D|\theta,f)P(\theta,f)$

Goal: sample the posterior distribution

```
choose a starting point current = θ0= (m,b)
WHILE convergence criterion is met:
        calculate current posterior post_curr = P(D|θ,f)
        /*proposal*/
        chose a new set of parameters new = θnew= (m,b)
        calculate the new posterior post_new=P(D|θnew,f)
        IF post_new > post_curr:
            current = new
        ELSE:
            /*accept with probability P(D|θnew,f) / P(D|θ,f) */
            r = random uniform number [0,1]
            IF r < post_new / post_orig:
                current = new
            ELSE:
                pass  //do nothing
```

Questions:

1. how do I choose the next point?

   Any *Markovian* process

   Any *ergodic* process
   CN: detailed balance

detailed balance $\pi(x_1)p(x_2|x_1)=\pi(x_2)p(x_1|x_2)$

Metropolis Rosenbluth Rosenbluth Teller 1953 - Hastings 1970

MCMC - Metropolis Hastings algorithm

Bayes theorem: $P(\theta|D,f) \propto P(D|\theta,f)\,P(\theta,f)$

Goal: sample the posterior distribution

choose a starting point **current** = ***θ_0= (m,b)***

WHILE convergence criterion is met:

calculate current posterior **post_curr = *P(D|θ,f)***

*/*proposal*/*
chose a new set of parameters **new = *θ_{new}= (m,b)***

calculate the new posterior **post_new=*P(D|θ_{new},f)***

IF ***post_new > post_curr:***
current = new
ELSE***:***
/*accept with probability ***P(D|θ_{new},f) / P(D|θ,f)*** */
r = random uniform number [0,1]
IF **r *< post_new / post_orig:***
current = new
ELSE:
pass //do nothing

DYI_MCMC.ipynb

MCMC - Metropolis Hastings algorithm

Bayes theorem: $P(\theta|D,f) \propto P(D|\theta,f)\,P(\theta,f)$

Goal: sample the posterior distribution

```
choose a starting point current = θ0= (m,b)
WHILE convergence criterion is met:
        calculate current posterior post_curr = P(D|θ,f)
        /*proposal*/
        chose a new set of parameters new = θnew= (m,b)
        calculate the new posterior post_new=P(D|θnew,f)
        IF post_new > post_curr:
            current = new
        ELSE:
            /*accept with probability P(D|θnew,f) / P(D|θ,f) */
            r = random uniform number [0,1]
            IF r < post_new / post_orig:
                current = new
            ELSE:
                pass  //do nothing
```

Questions:

1. how do I choose the next point?

   Any *Markovian* process

   Any *ergodic* process
   CN: detailed balance

detailed balance $\pi(x_1)p(x_2|x_1)=\pi(x_2)p(x_1|x_2)$

Metropolis Rosenbluth Rosenbluth Teller 1953 - Hastings 1970

MCMC - Gibbs sampler

Bayes theorem: $P(\theta|D,f) \propto P(D|\theta,f)\,P(\theta,f)$

Goal: sample the posterior distribution

Questions:

1. how do I choose the next point?

   Gibbs sampling:
   Metropolis Hastings proposal distribution with change *along a single direction at a time => always accept*
   must know the integral P(D|f) along that direction

```
choose a starting point current = θ0= (m,b)
WHILE convergence criterion is met:
        calculate current posterior post_curr = P(D|θ,f)
        /*proposal*/
        chose a new set of parameters new = θnew= (m,b)
        calculate the new posterior post_new=P(D|θnew,f)
        IF post_new > post_curr:
            current = new
        ELSE:
            /*accept with probability P(D|θnew,f) / P(D|θ,f) */
            r = random uniform number [0,1]
            IF r < post_new / post_orig:
                current = new
            ELSE:
                pass  //do nothing
```

detailed balance $\pi(x_1)p(x_2|x_1)=\pi(x_2)p(x_1|x_2)$

Metropolis Rosenbluth Rosenbluth Teller 1953 - Hastings 1970

MCMC - Gibbs sampler

Bayes theorem: $P(\theta|D,f) \propto P(D|\theta,f)\,P(\theta,f)$

Goal: sample the posterior distribution

```
choose a starting point current = θ0= (m,b)
WHILE convergence criterion is met:
        calculate current posterior post_curr = P(D|θ,f)
        /*proposal*/
        chose a new set of parameters new = θnew= (m,b)
        calculate the new posterior post_new=P(D|θnew,f)
        IF post_new > post_curr:
            current = new
        ELSE:
            /*accept with probability P(D|θnew,f) / P(D|θ,f) */
            r = random uniform number [0,1]
            IF r < post_new / post_orig:
                current = new
            ELSE:
                pass  //do nothing
```

Questions:

1. how do I choose the next point?

Other options:
simulated annealing (good for multimodal)
parallel tempering (good for multimodal)
differential evoution (good for covariant spaces)

detailed balance $\pi(x_1)p(x_2|x_1)=\pi(x_2)p(x_1|x_2)$

Metropolis Rosenbluth Rosenbluth Teller 1953 - Hastings 1970

MCMC - Gibbs sampler

Bayes theorem: $P(\theta|D,f) \propto P(D|\theta,f)\,P(\theta,f)$

Goal: sample the posterior distribution

Questions:

1. how do I choose the next point?

Other options:
affine invariant : EMCEE package

```
choose a starting point current = θ0= (m,b)
WHILE convergence criterion is met:
        calculate current posterior post_curr = P(D|θ,f)
        /*proposal*/
        chose a new set of parameters new = θnew= (m,b)
        calculate the new posterior post_new=P(D|θnew,f)
        IF post_new > post_curr:
            current = new
        ELSE:
            /*accept with probability P(D|θnew,f) / P(D|θ,f) */
            r = random uniform number [0,1]
            IF r < post_new / post_orig:
                current = new
            ELSE:
                pass  //do nothing
```

detailed balance $\pi(x_1)p(x_2|x_1)=\pi(x_2)p(x_1|x_2)$

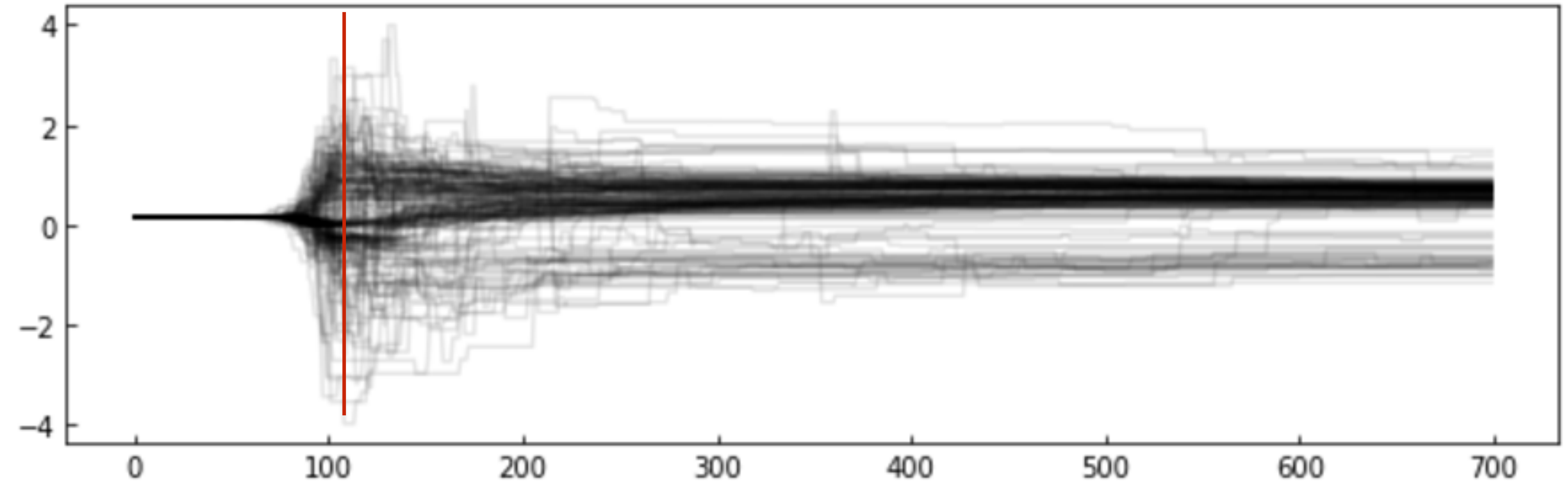
Metropolis Rosenbluth Rosenbluth Teller 1953 - Hastings 1970

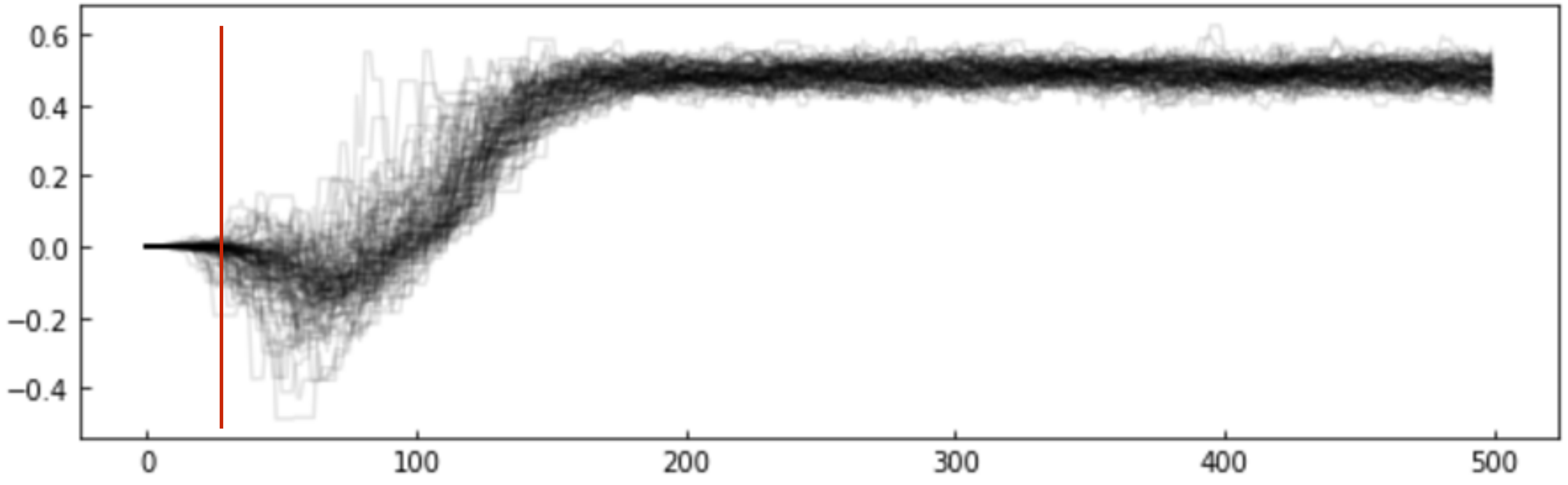
MCMC - EMCEE

MCMC - convergence

Bayes theorem: $$P(\theta|D,f) \propto P(D|\theta,f)\,P(\theta,f)$$

Goal: sample the posterior distribution





Questions:

1. how do I choose the next point?
2. when have I sampled the posterior adequatly?
   has your chain *burned-in* ?

Bayes theorem: $P(\theta|D,f) \propto P(D|\theta,f)P(\theta,f)$

Goal: sample the posterior distribution

Questions:

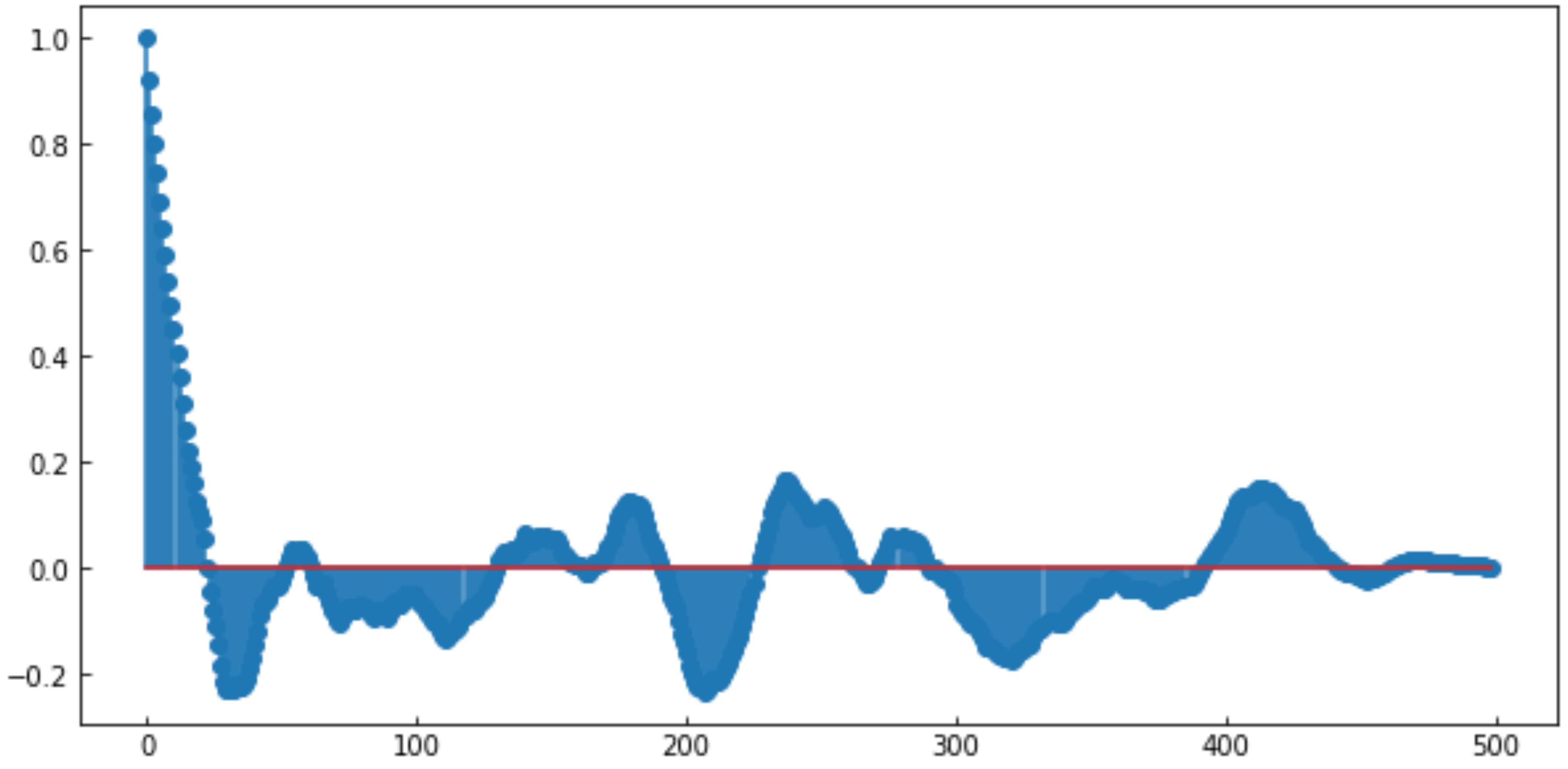
1. how do I choose the next point?

2. when have I sampled the posterior adequatly?
   has your MCMC *converged* ?

MCMC - convergence

Bayes theorem: $P(\theta|D,f) \propto P(D|\theta,f)\,P(\theta,f)$

Goal: sample the posterior distribution



Questions:

1. how do I choose the next point?
2. when have I sampled the posterior adequatly?
   has your MCMC *converged* ?

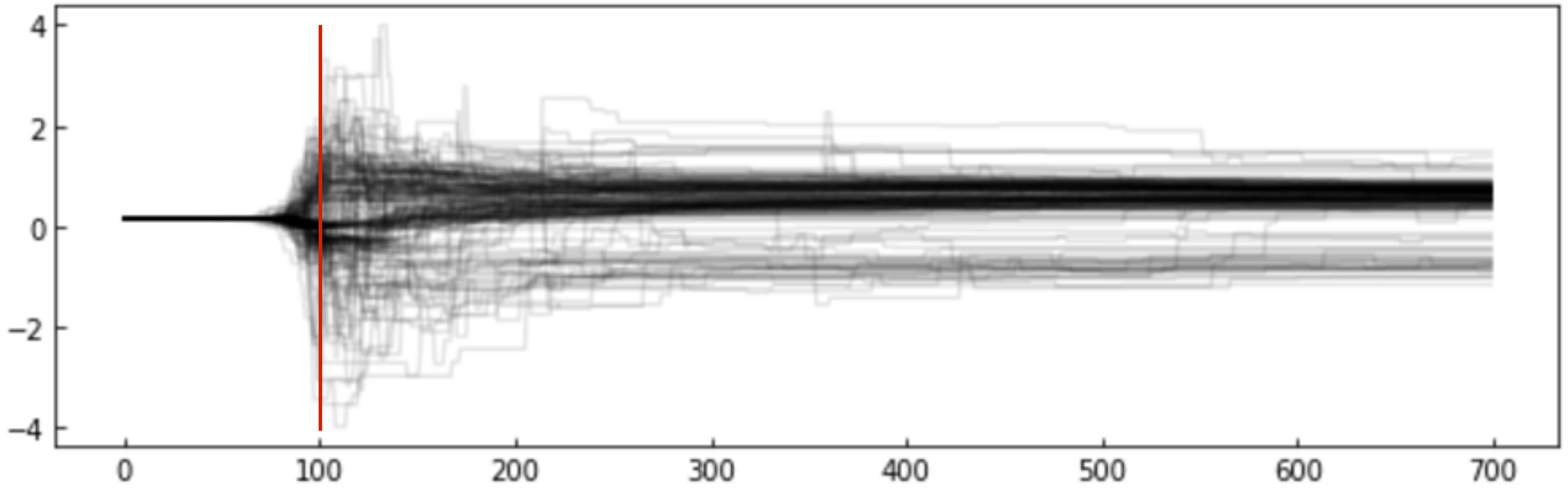
   a. **check autocorrelation within a chain (*Raftery*)**
   b. check that all chains coverged to same region (a stationary distribution *GelmanRubin*)
   c. mean at beginning = mean at end (*Geweke*)
   d. check that entire chain reached stationary distribution (or a final fraction of the chain, *Heidelberg-Welch* using Cramer-von-Mises statistic)

MCMC - convergence

Bayes theorem: $$P(\theta|D,f) \propto P(D|\theta,f)\,P(\theta,f)$$

Goal: sample the posterior distribution

Questions:



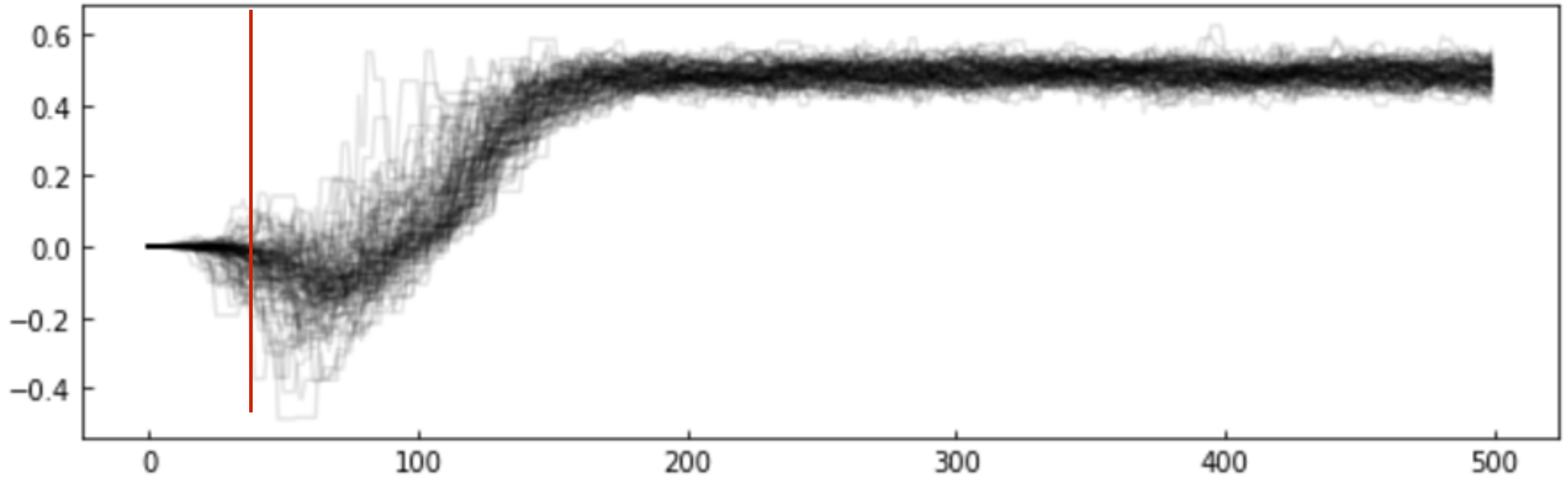
1. how do I choose the next point?

2. when have I sampled the posterior adequatly?
   has your MCMC *converged* ?
   a. check autocorrelation within a chain (*Raftery*)
   b. check that all chains coverged to same region (a stationary distribution *GelmanRubin*)
   c. mean at beginning = mean at end (*Geweke*)
   d. check that entire chain reached stationary distribution (or a final fraction of the chain, *Heidelberg-Welch* using Cramer-von-Mises statistic)

MCMC - convergence

Bayes theorem: $$P(\theta|D,f) \propto P(D|\theta,f)\,P(\theta,f)$$

Goal: sample the posterior distribution



Questions:

1. how do I choose the next point?

2. when have I sampled the posterior adequatly?
   has your MCMC *converged* ?
   a. check autocorrelation within a chain (*Raftery*)
   b. check that all chains coverged to same region (a stationary distribution *GelmanRubin*)
   c. **mean at beginning = mean at end (*Geweke*)**
   d. **check that entire chain reached stationary distribution (or a final fraction of the chain, *Heidelberg-Welch* using Cramer-von-Mises statistic)**

MCMC - Metropolis Hastings algorithm

Bayes theorem: $P(\theta|D,f) \propto P(D|\theta,f)P(\theta,f)$

Goal: sample the posterior distribution

Questions:

1. how do I choose the next point?
2. when have I sampled the posterior adequatly?
3. how can it be-the samples are *not independent!*

good point!…

choose a starting point **current** = $\boldsymbol{\theta_0}$**= (m,b)**

WHILE convergence criterion is met:

calculate current posterior **post_curr =** ***P(D|θ,f)***

*/*proposal*/*
chose a new set of parameters **new =** $\boldsymbol{\theta_{new}}$**= (m,b)**

calculate the new posterior **post_new=*P(D|***$\boldsymbol{\theta_{new}}$***,f)***

IF ***post_new > post_curr:***
current = new
ELSE***:***
/*accept with probability ***P(D|***$\boldsymbol{\theta_{new}}$***,f) / P(D|θ,f)*** */
r = random uniform number [0,1]
IF **r <** ***post_new / post_orig:***
current = new
ELSE:
pass //do nothing

Resources Markov Chain Monte Carlo

Information Theory, Inference, and Learning Algorithms
David J.C. MacKay, 2003

Numerical Recipes
Bill Press+ 1992 (+)

Ensemble samplers with affine invariance
Jonathan Goodman and Jonathan Weare 2010

Resources Markov Chain Monte Carlo

Slides on sampling from distributions
Paul E. Johnson 2015

EMCEE readme
D. Foerman-Mackey, D. Hogg, D. Lang, J. Goodman+ 2012

Quick Glossary

- **Stocastic**: random, following any distribution
- **PDF**: probability distribution function *P(x)* describes the *relative* likelihood of sample *x* compared
- **CDF**: cumulative distribution function - the probability that a value drawn from a distribution will be smaller than *x* $F(x)=\int_{-}^{x} P(x)$
- **Marginalize**: integrate along a dimension
- **Gaussian distribution**: a distribution with PDF $N(\mu,\sigma)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{\sigma^2}}$
- **Chi Squared χ^2:** a model fitting method based on the provable fact that (under proper assumption) the function follows a χ^2 distribution $\sum_{i=1}^{N}\frac{(M-D)^2}{\sigma^2} \sim \chi^2_{DOF}$
- **Likelihood**: in Bayes theorem: the term indicating the probability of the data under the model for a choice of parameters. More generally it can be thought of the probability of the parameters given the data
- **Posterior**: the probability of data given model calculated by Beyes theorem as likelihood * prior / evidence
- **Evidence**: the probability of the data given a model marginalized over all parameters
- **Prior**: prior, or otherwise obtained, knowledge about the problem which indicates how likeli the model parameter are for any value
- **Markovian process** - a process whose next stage depends stocastically on the current state only