# Monte Carlo methods

# outline

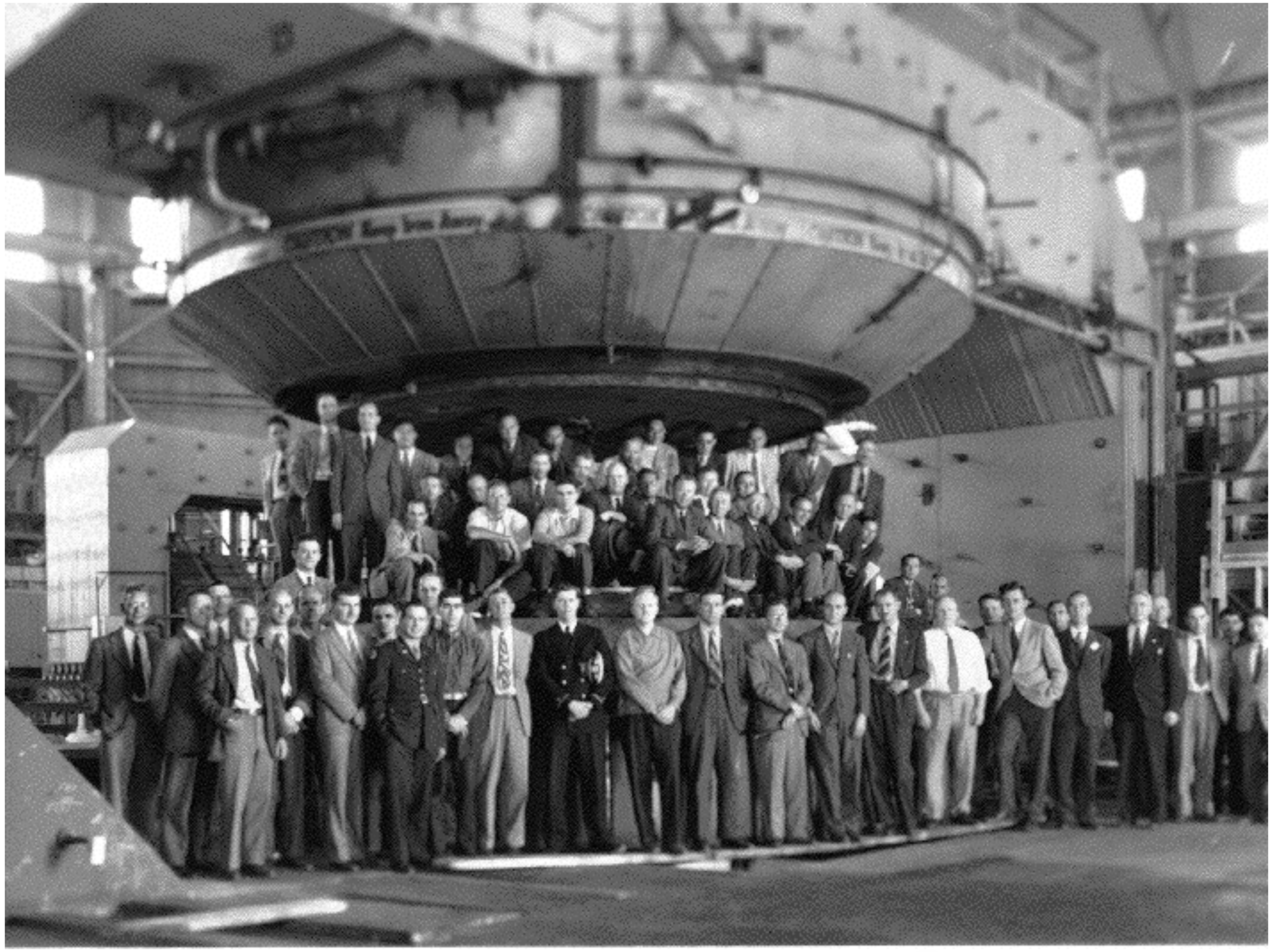
Stocastic Processes in Science Inference

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

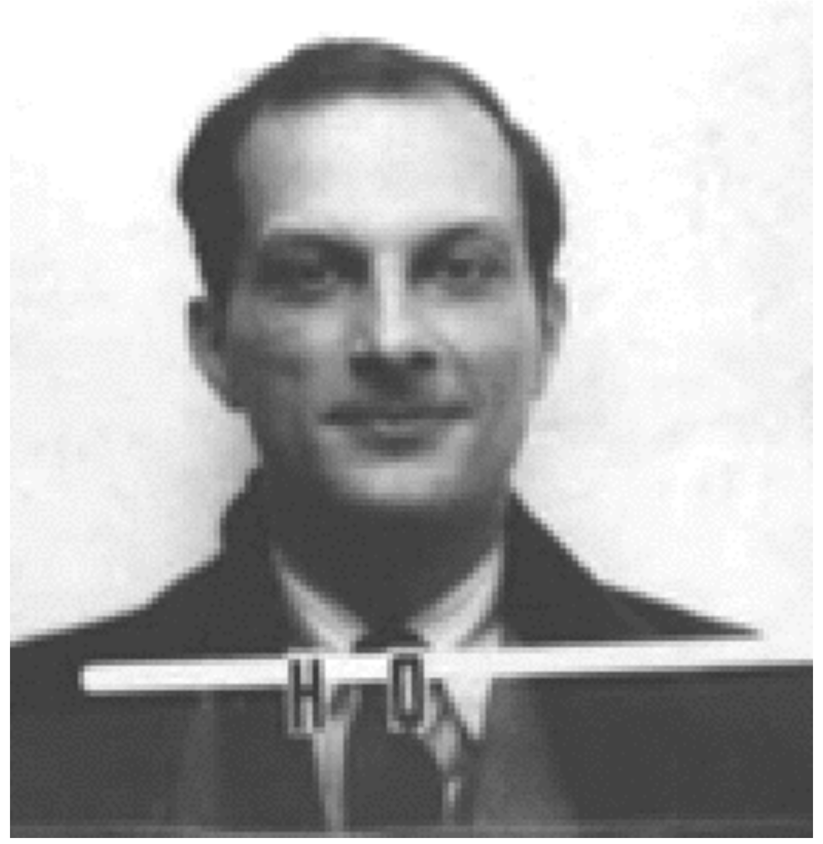
Markov Chain Monte Carlo

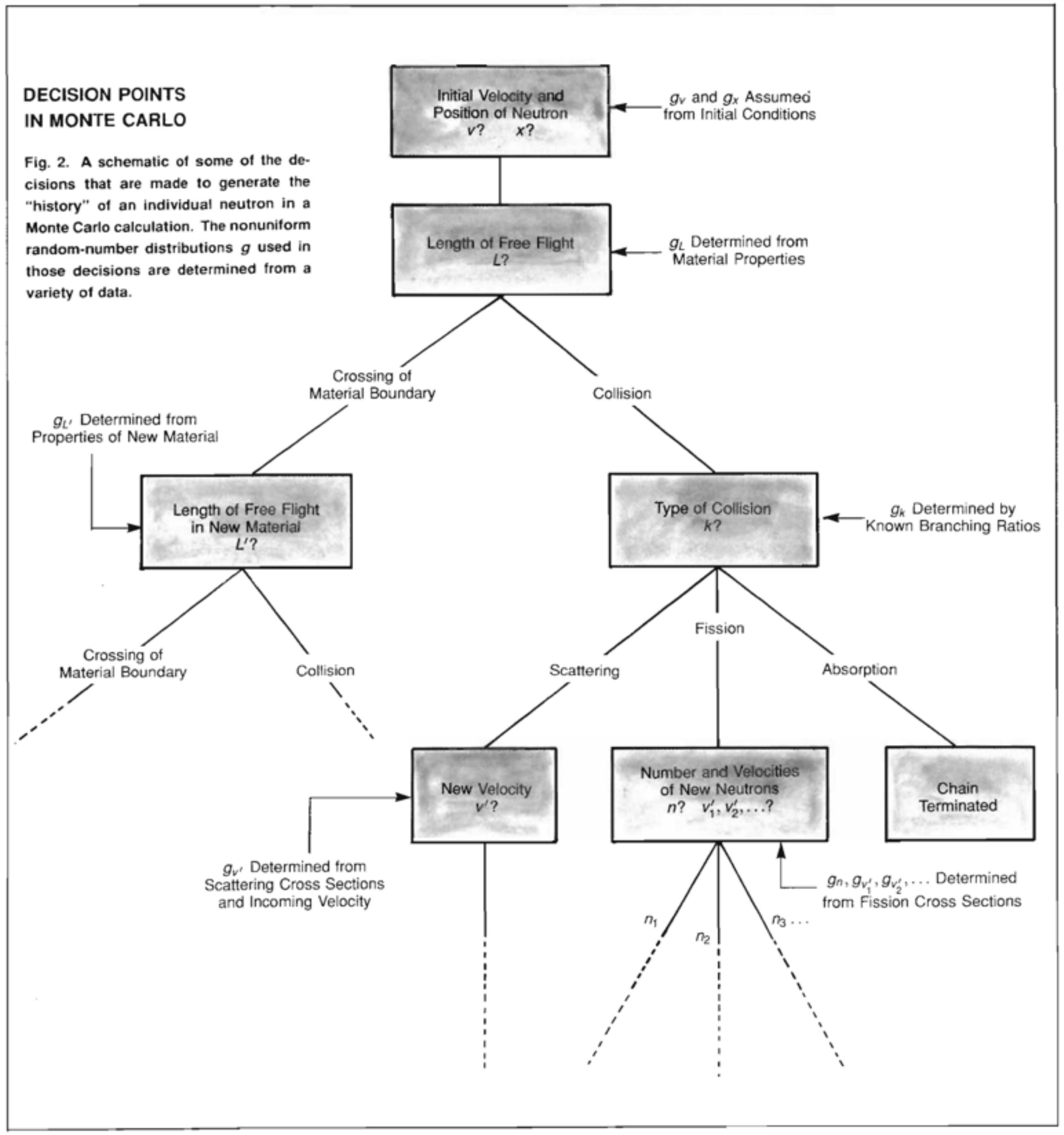
- Markovian Processes and Markov chains
- Bayes theorem and the posterior distribution
- Metropolis Hasting (and Gibbs sampling) MCMC
- Affine Invariant MCMC
- convergence criteria

The Manhattan Project



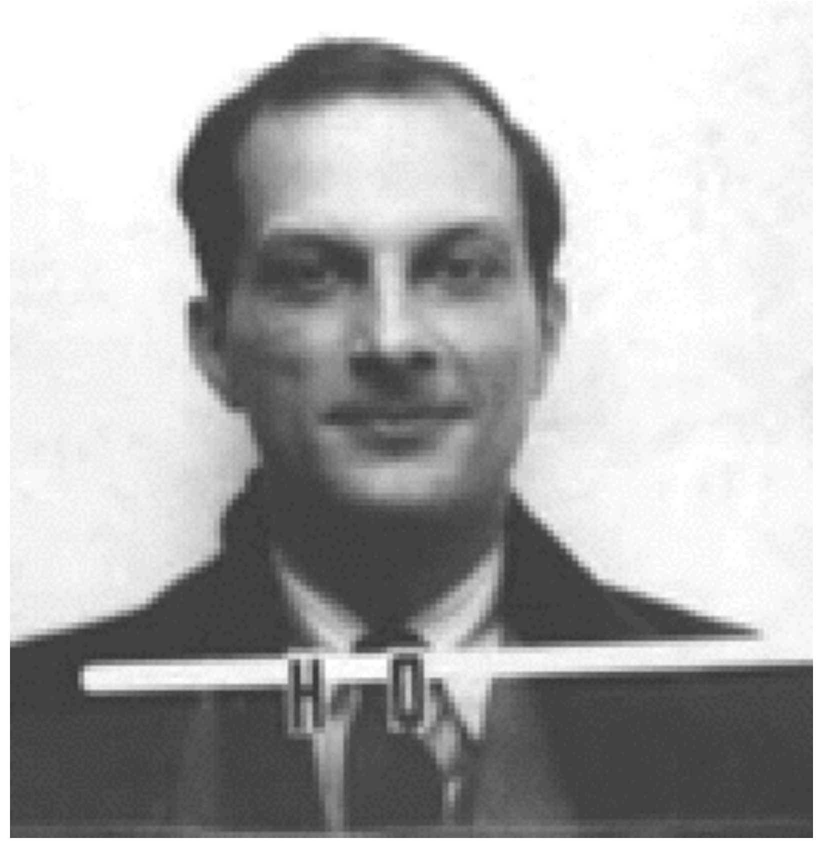
# history



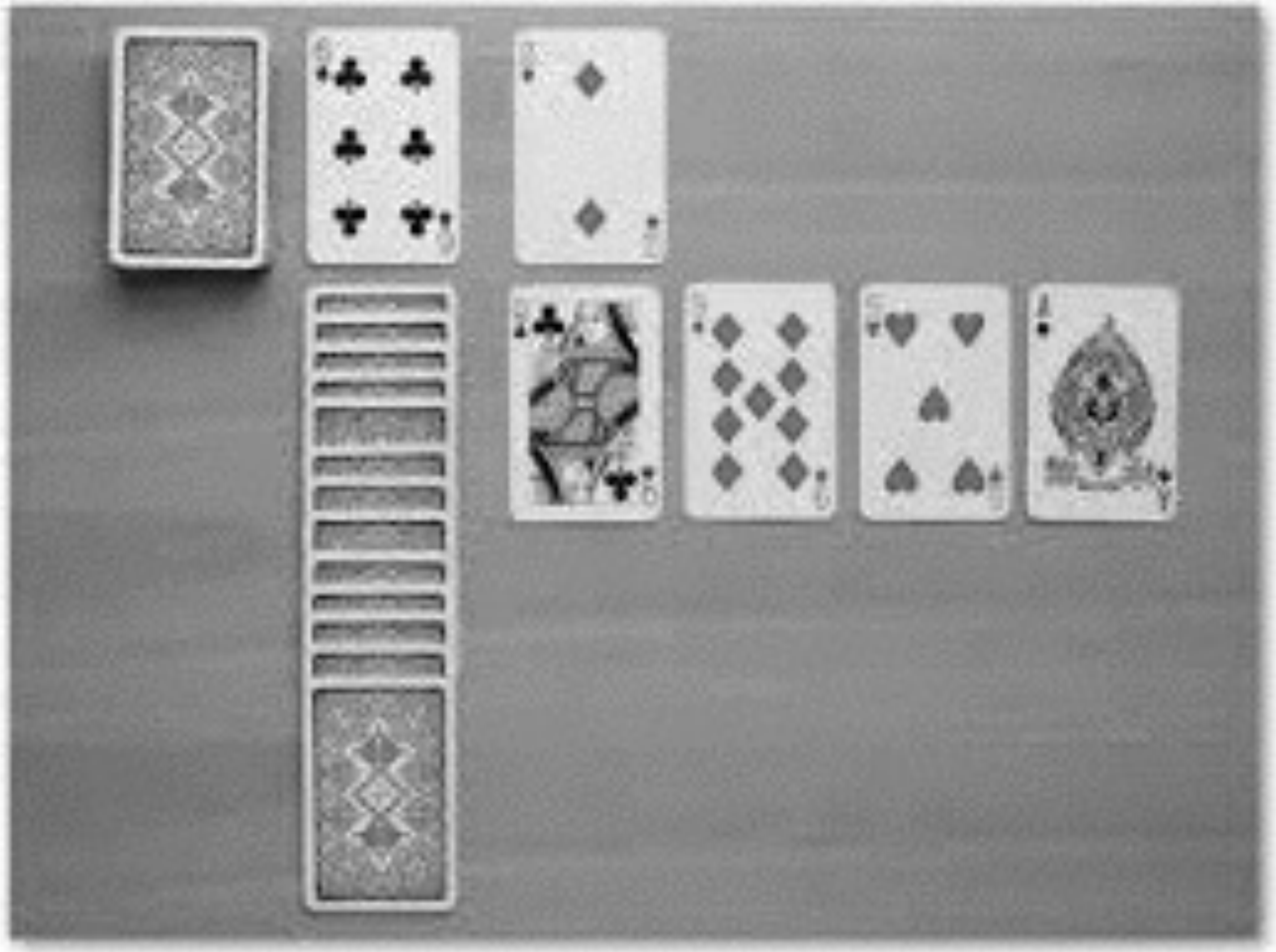


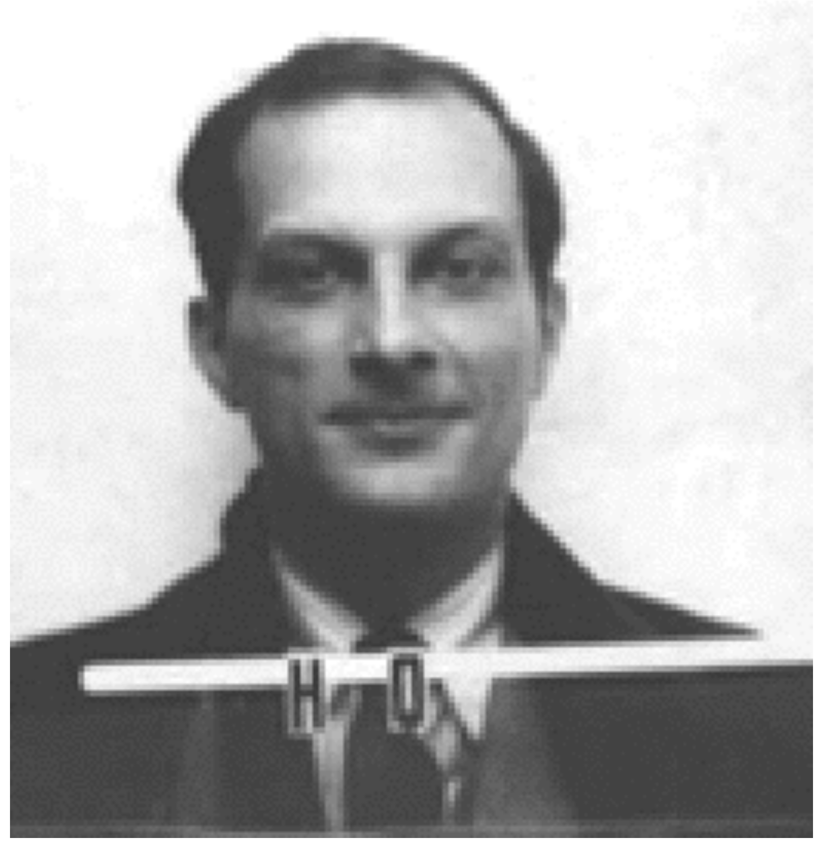

**DECISION POINTS IN MONTE CARLO**

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.



Canfield Solitaire

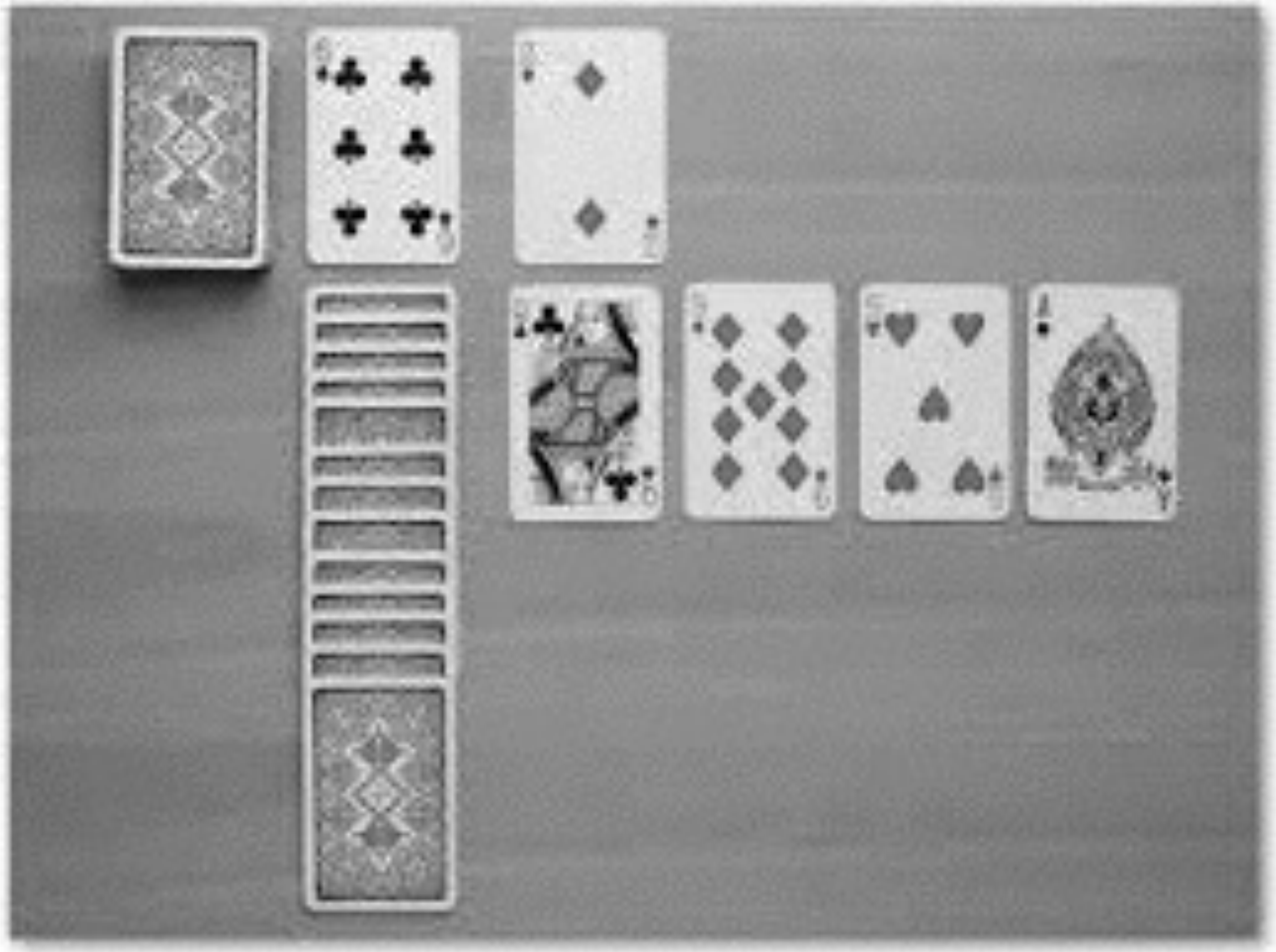




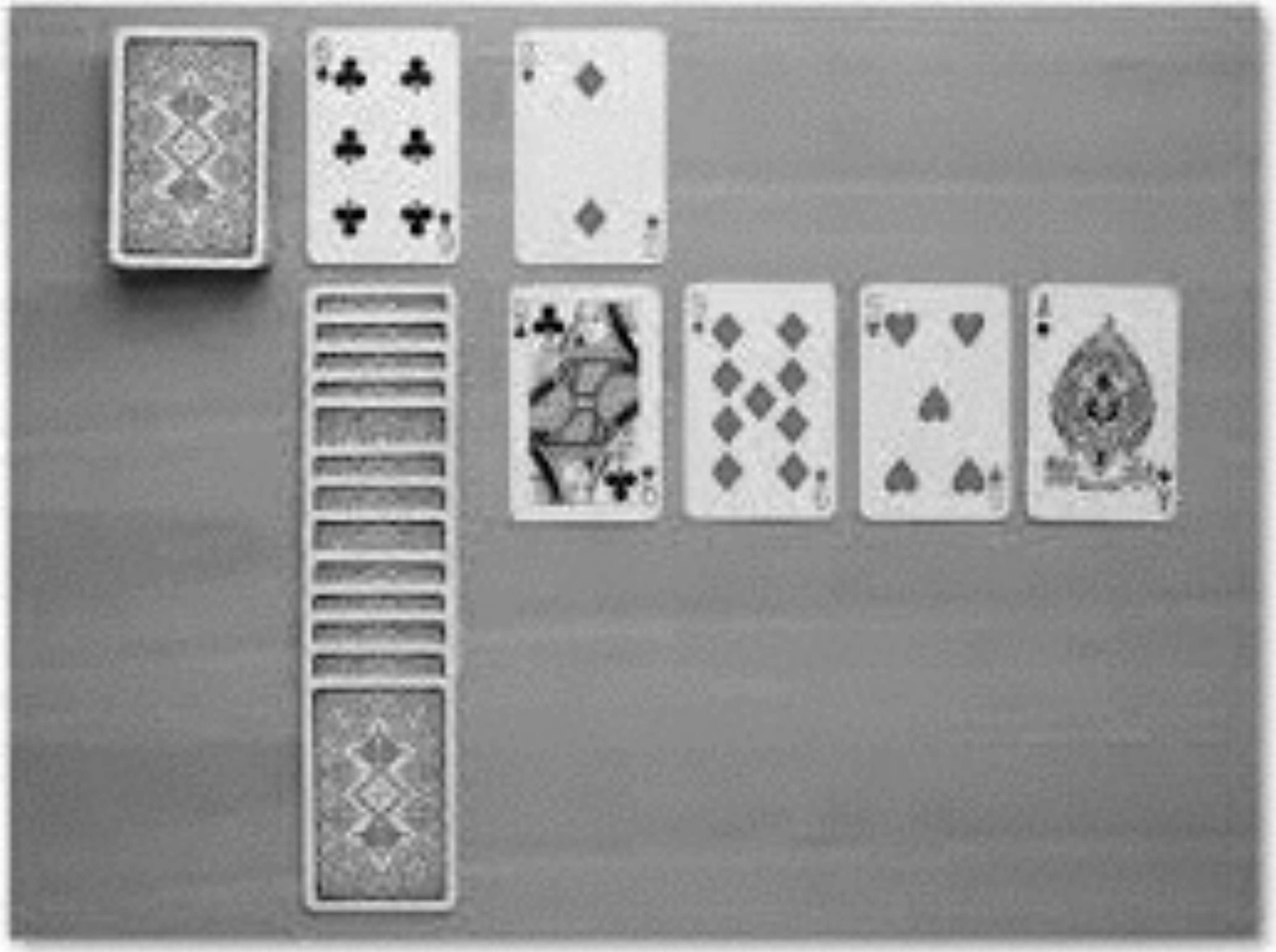

What are the chances that a Canfield solitaire laid out with 52 cards will come out successfully?

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

Canfield Solitaire

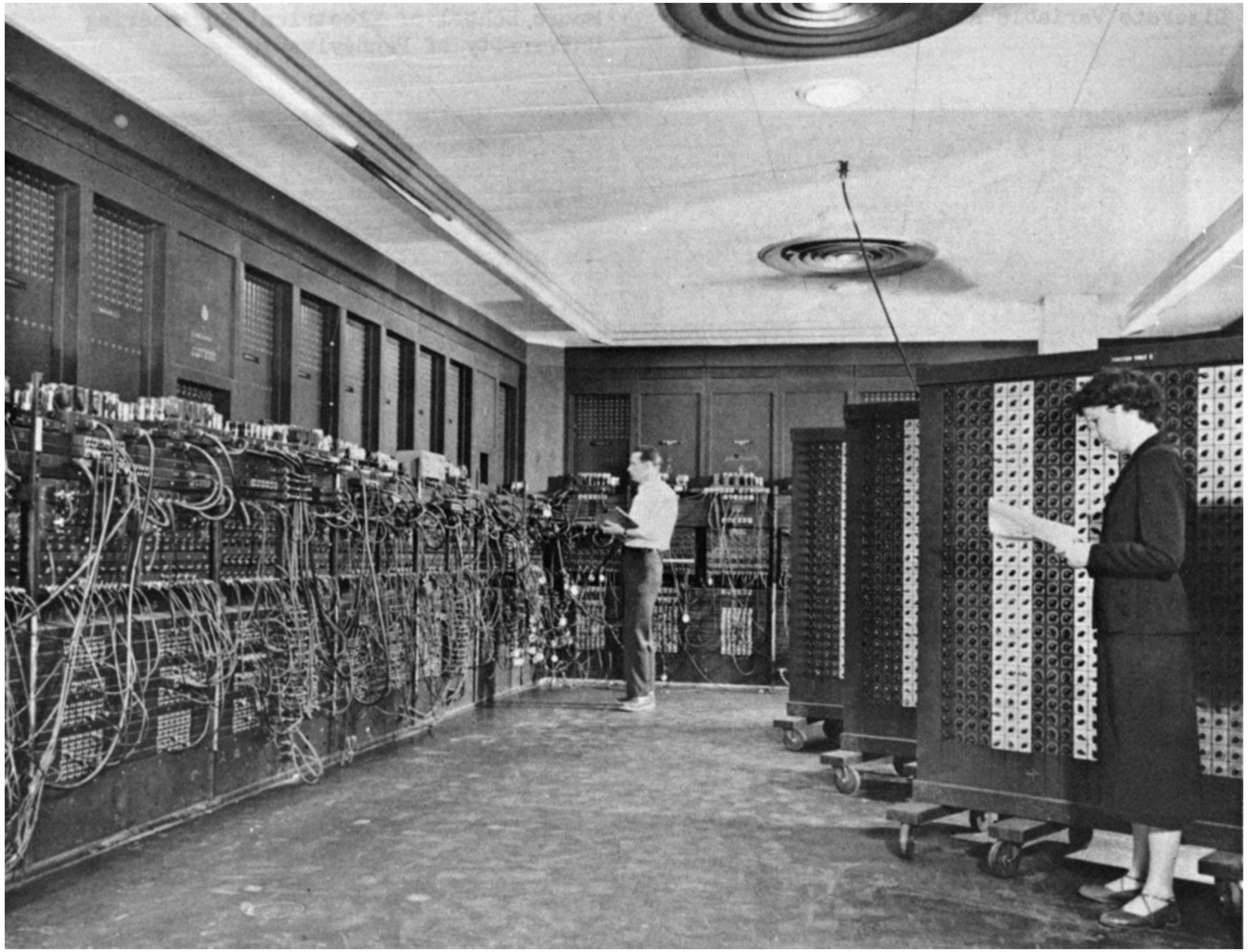


What are the chances that a Canfield solitaire laid out with 52 cards will come out successfully? After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether **a more practical method than "abstract thinking" might not be to lay it out say one hundred times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

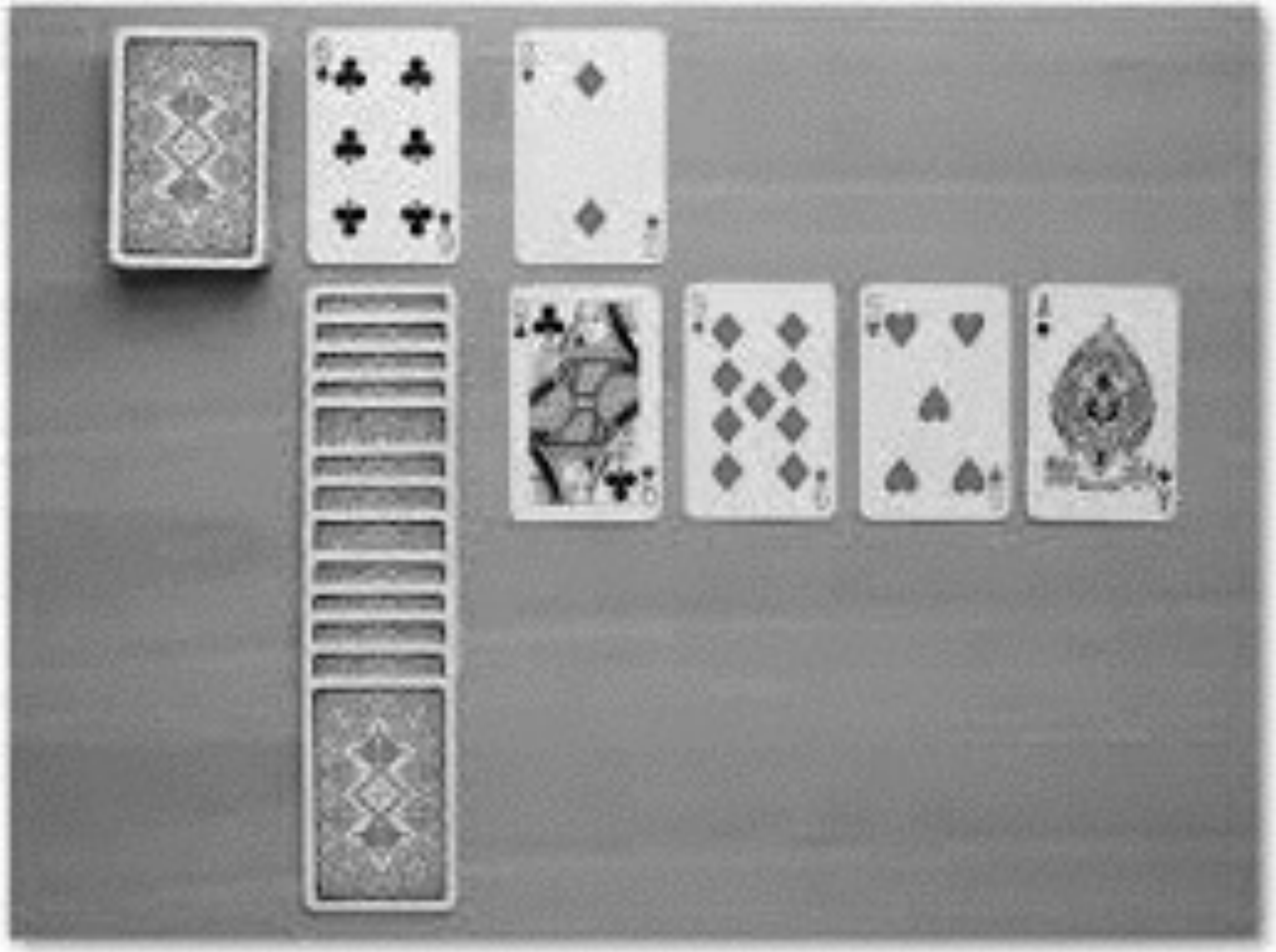


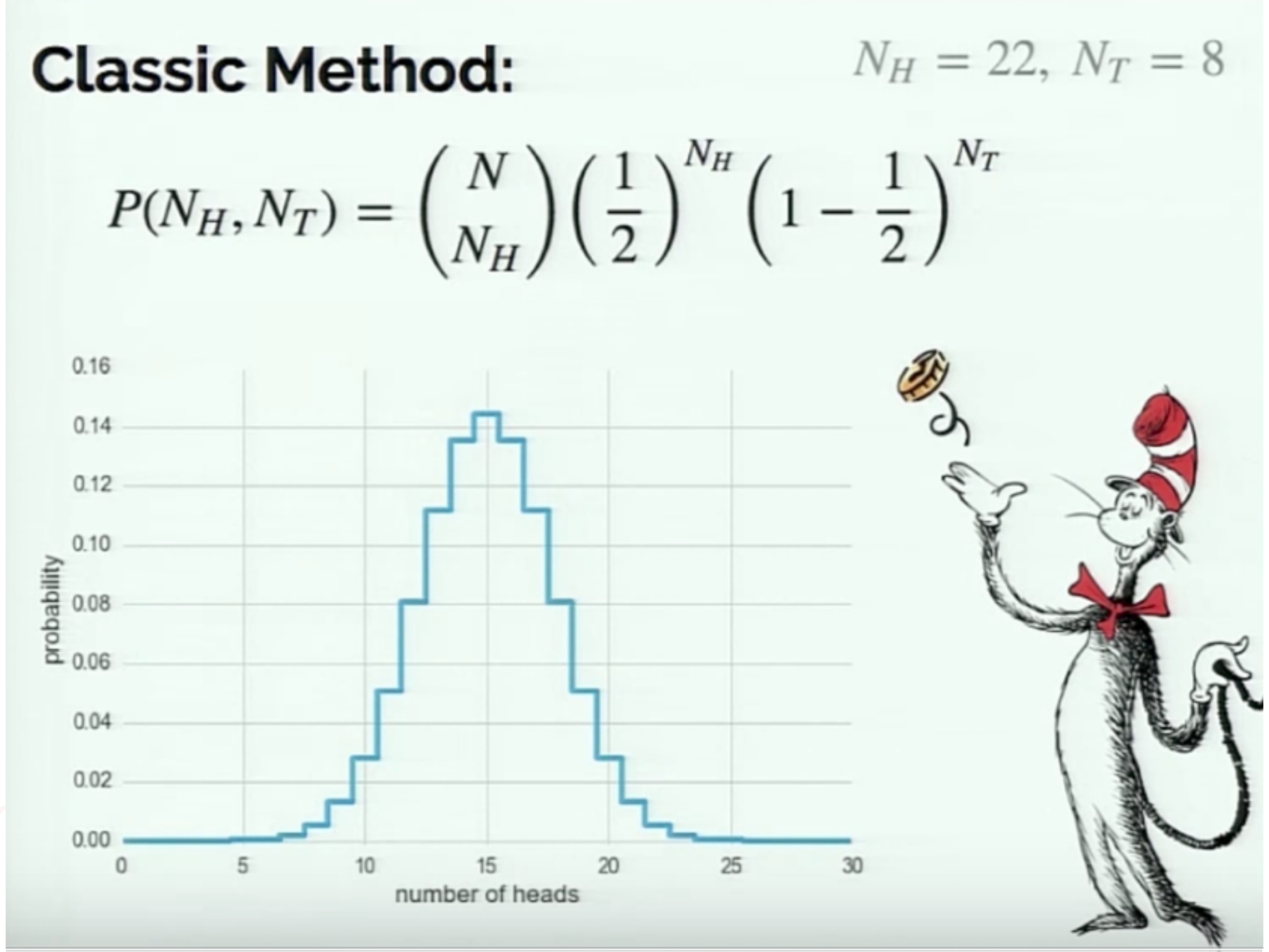
http://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-88-9068

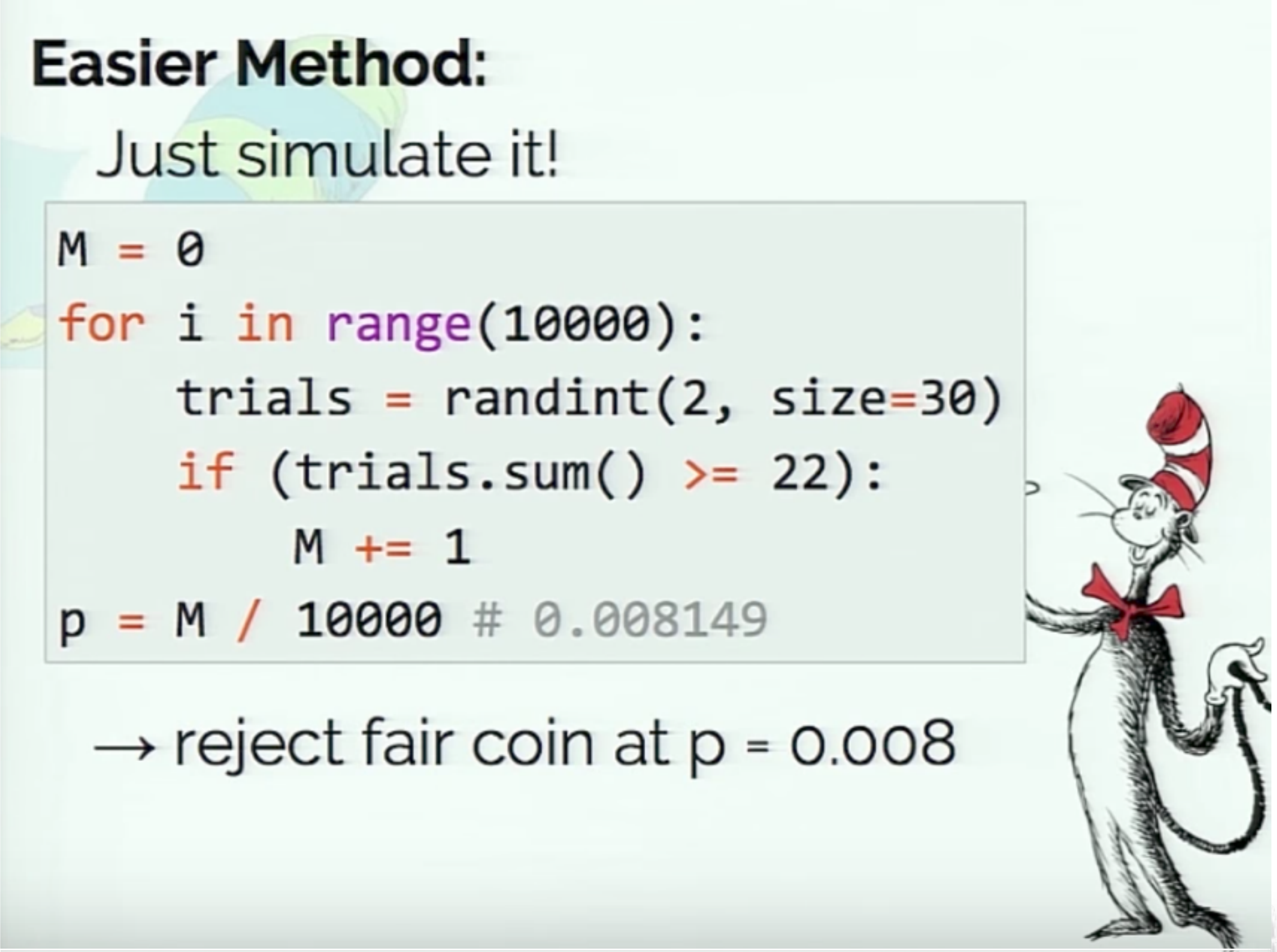
# history



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 m2 (1,800 ft2), consumed 150 kW of electricity.
500FLOPS vs today's Macbook pro~1TeraFLOP





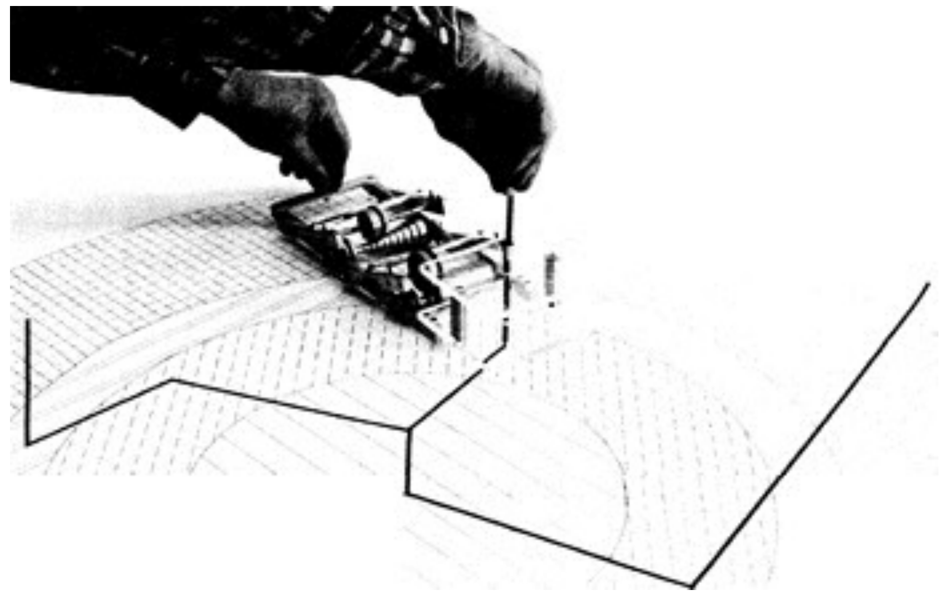



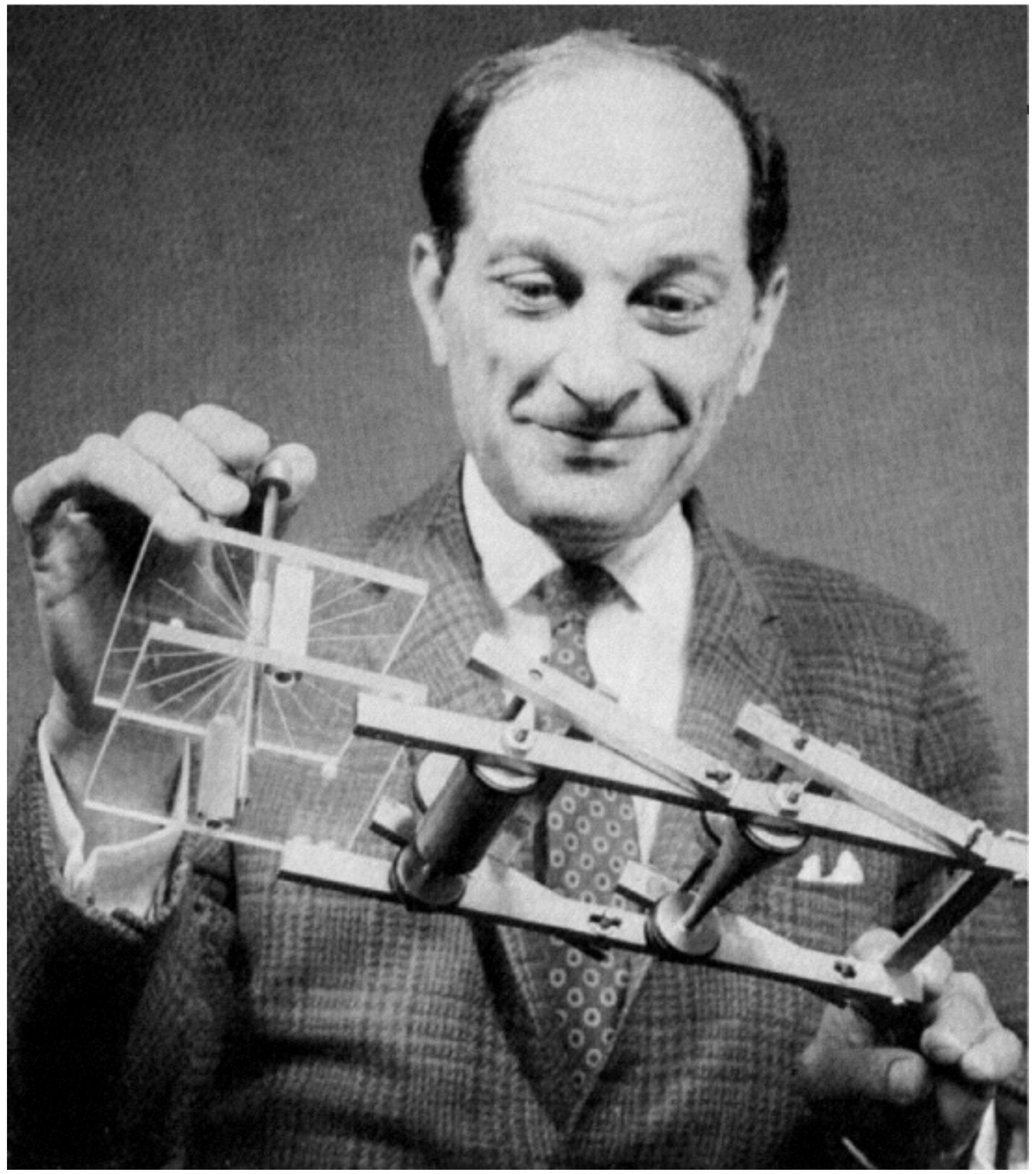

[Statistics for Hackers, Jake Vanderplas PyCon16](https://www.youtube.com/watch?v=Iq9DzN6mvYA)
https://www.youtube.com/watch?v=Iq9DzN6mvYA

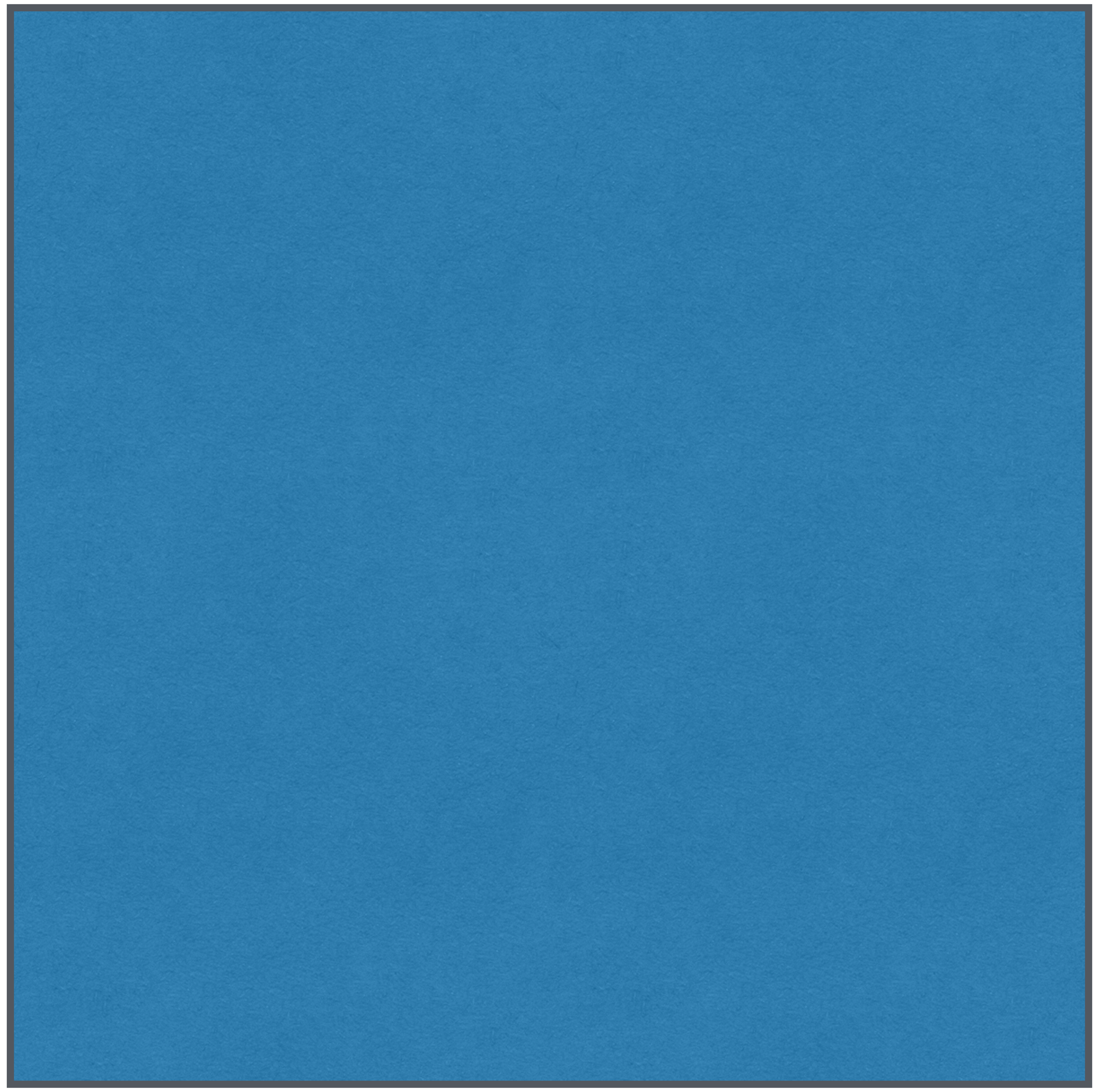
# history

The Fermiac

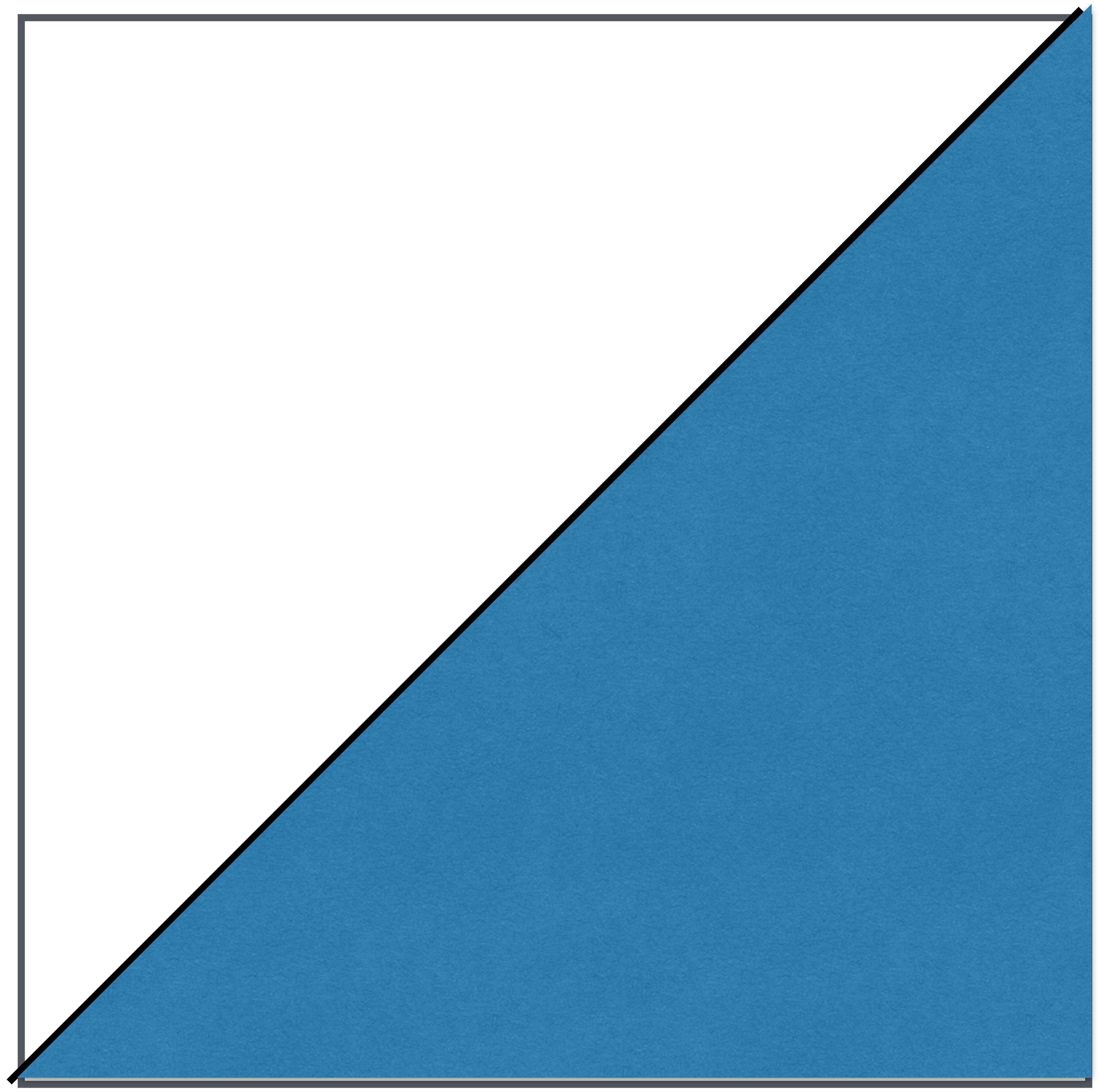
Enrico Fermi looked really smart with his predictions…





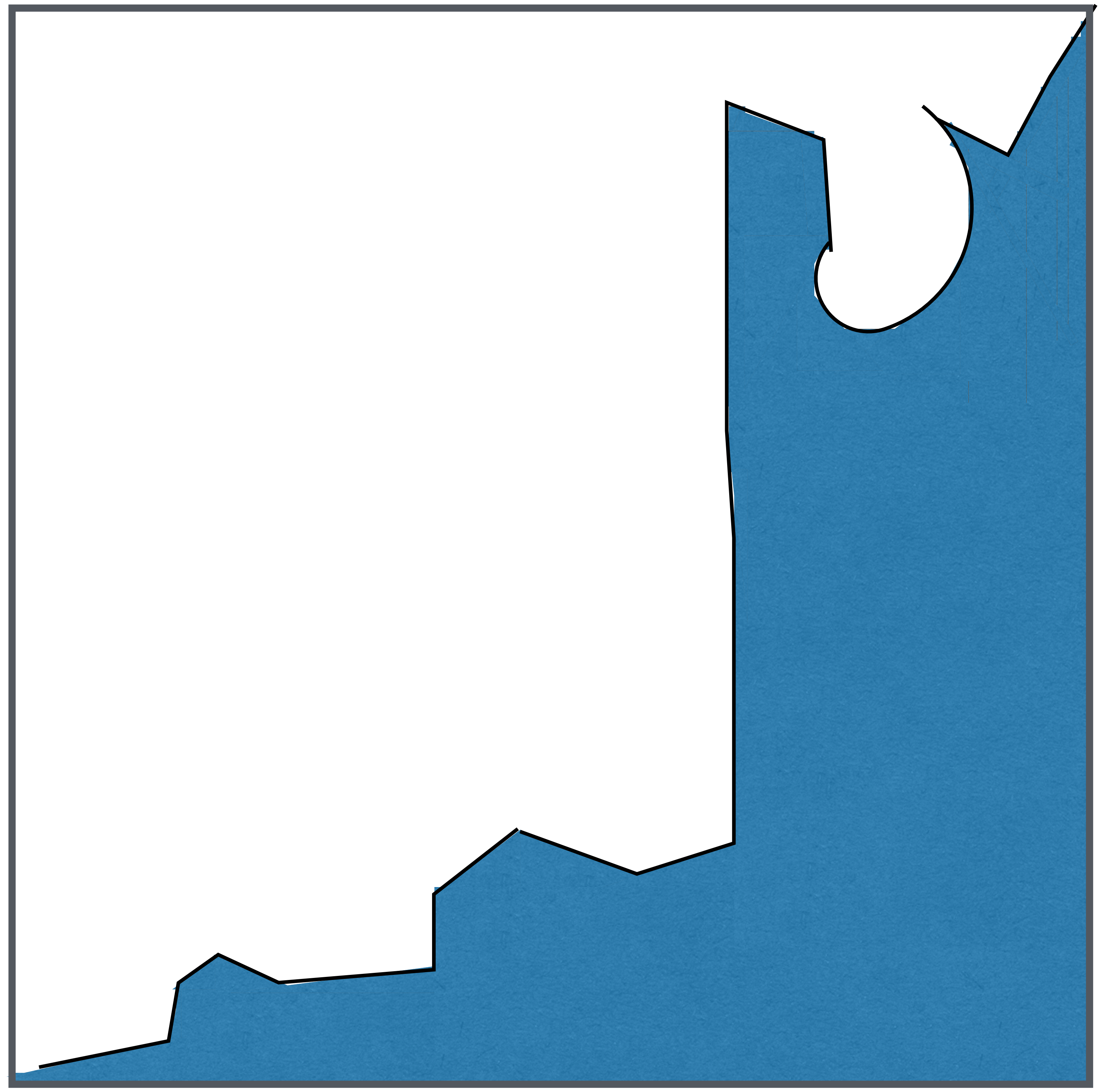


Area: base x height



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

# MC - motivation



Area: ??????

*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

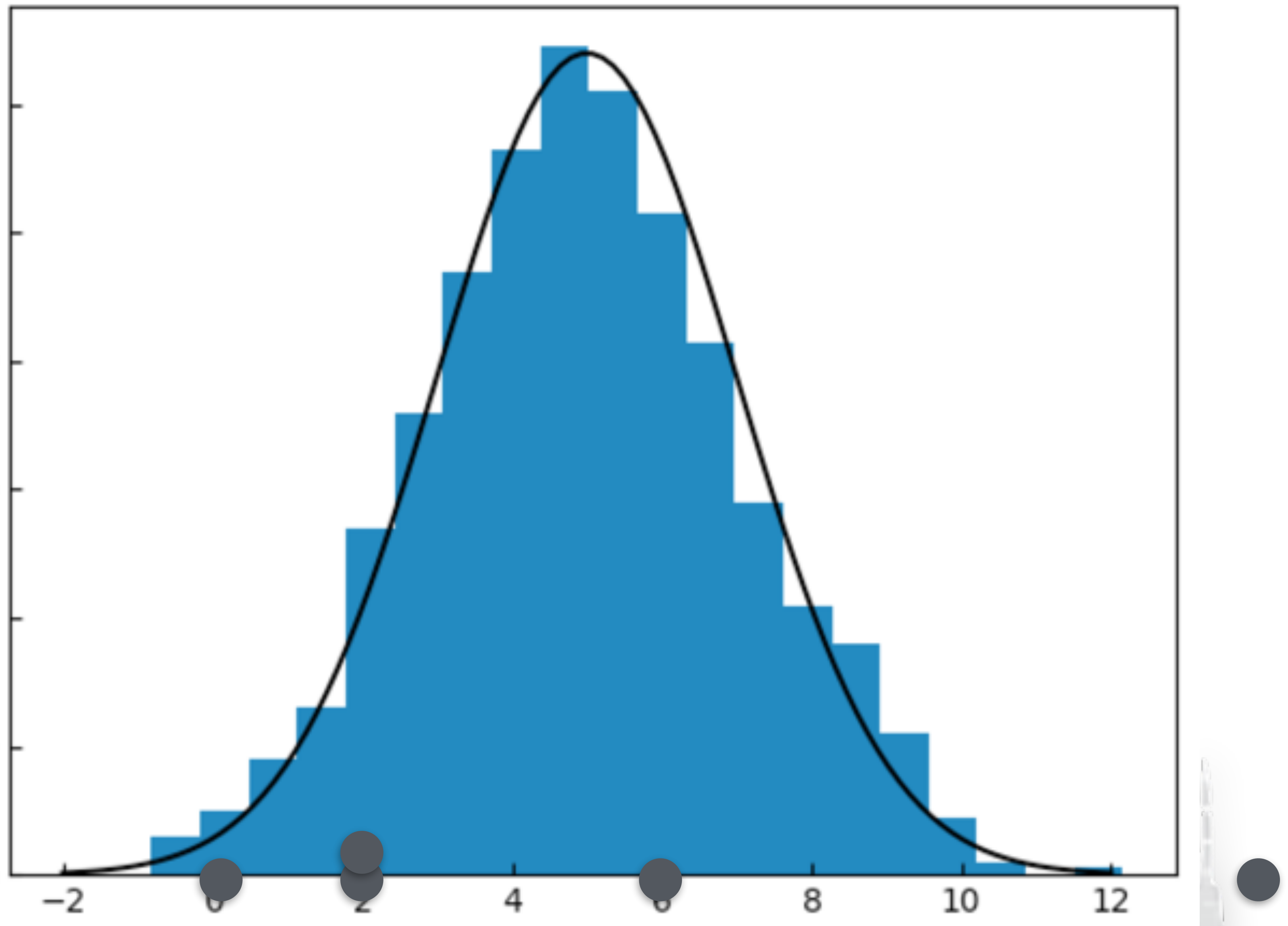
# *MC* - motivation

*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]



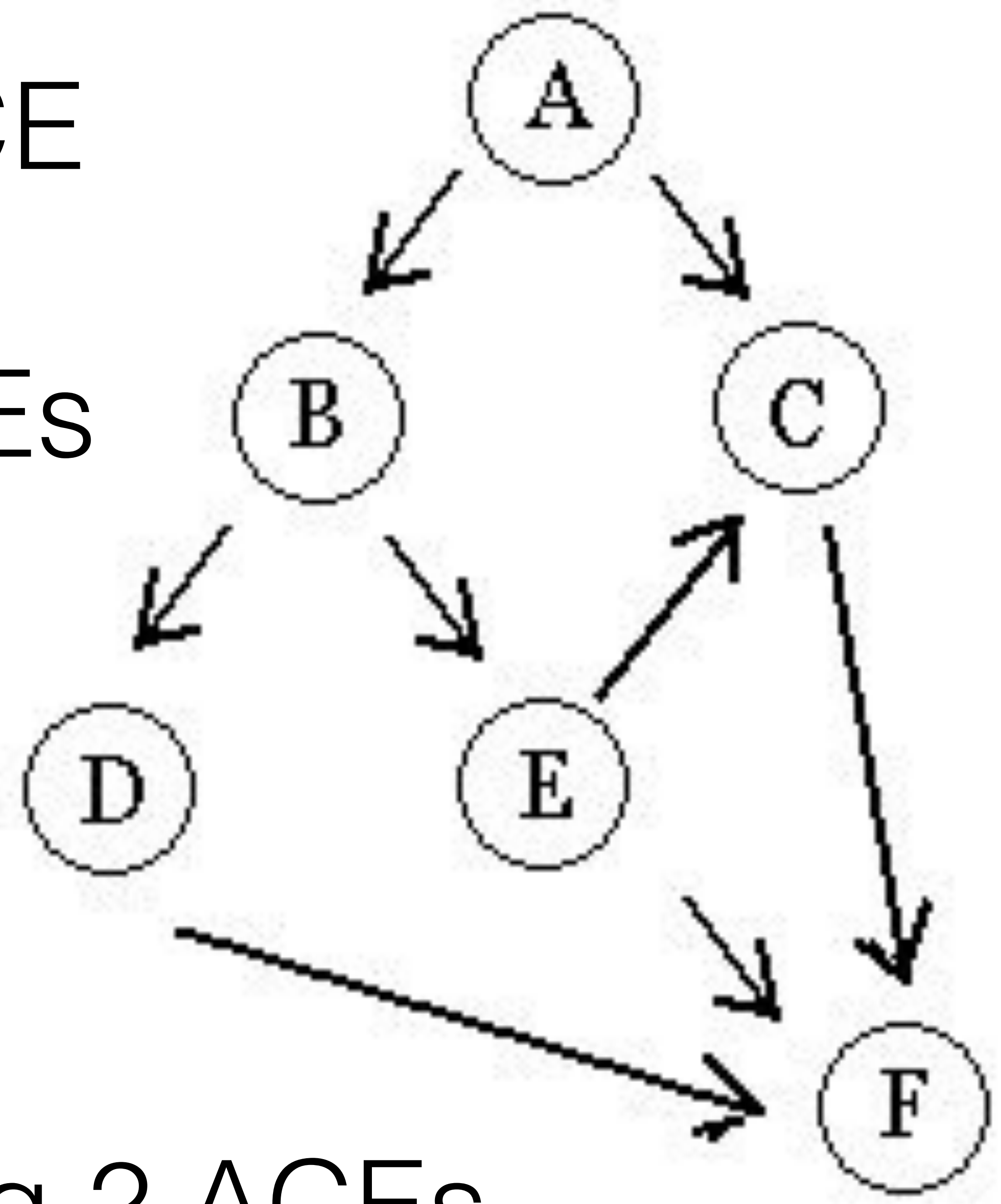

Mean of a continuous distribution

$$mean(X) = \int X f(X) dX$$

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

A P having ACE

B P having 2 ACEs




A P having 2 ACEs

**Sample**

$A \sim P(A)$

$B \sim P(B|A)$

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

$D \sim P(D|B)$

$E \sim P(E|B)$

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

**SetUp 1:**

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

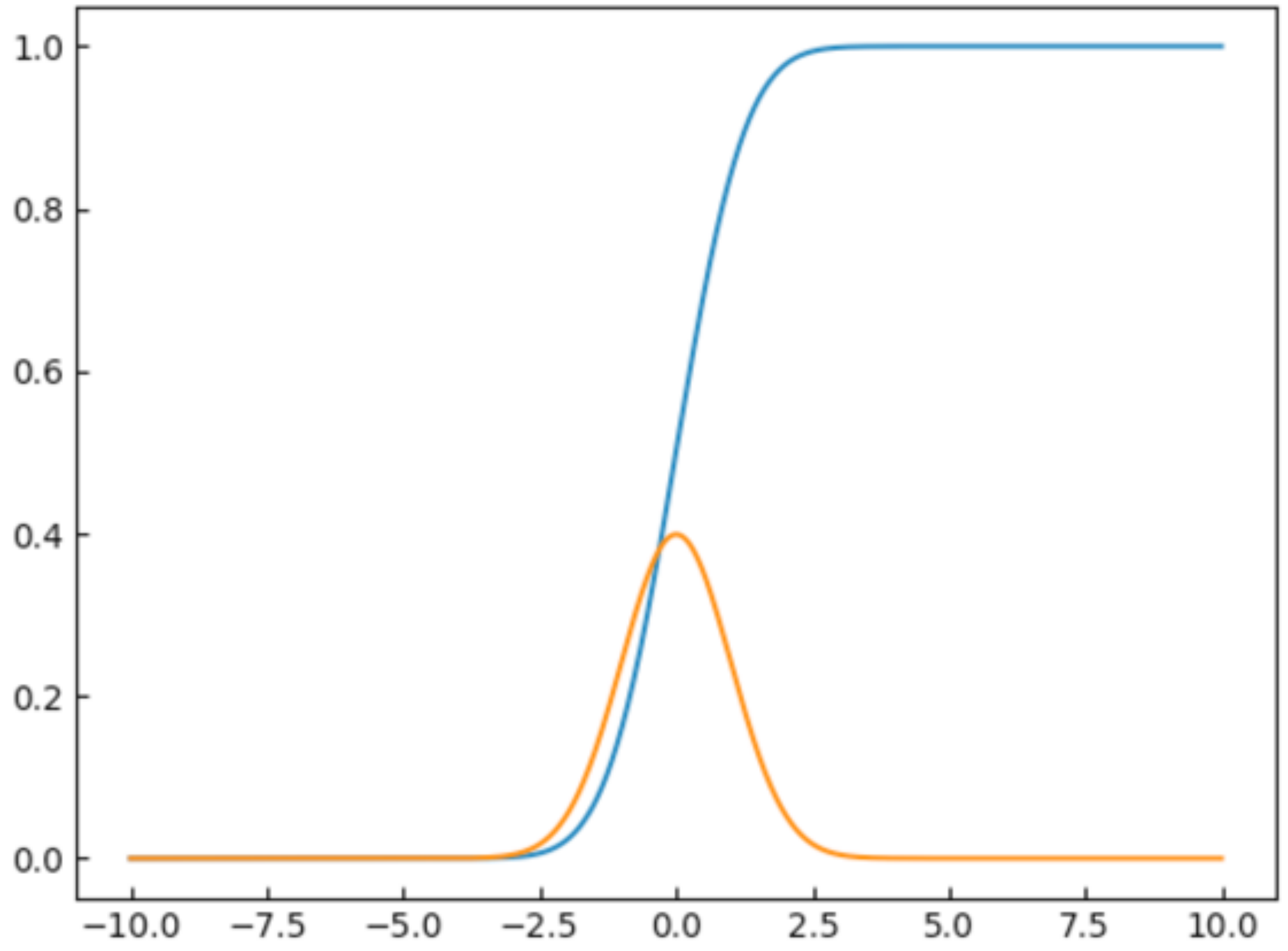
2. The function can be integrated : e.g. Gaussian

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

3. If I can take the integral of the PDF I can calculate the CDF

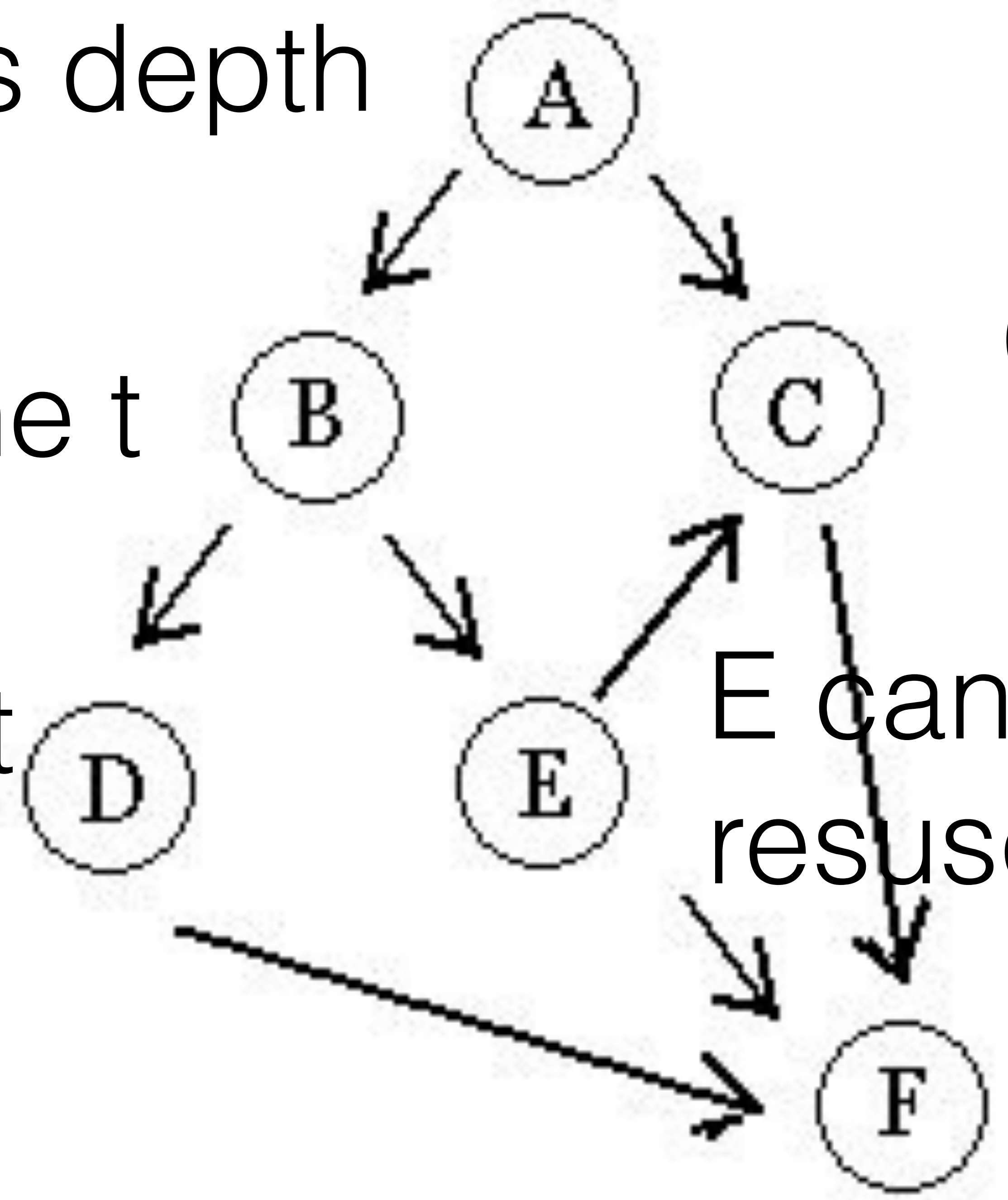
4. If I know the CDF I know at which percentile each value is

*I can directly samplefrom the distribution*



# *MC* - Rejection Sampling

**Sample**




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

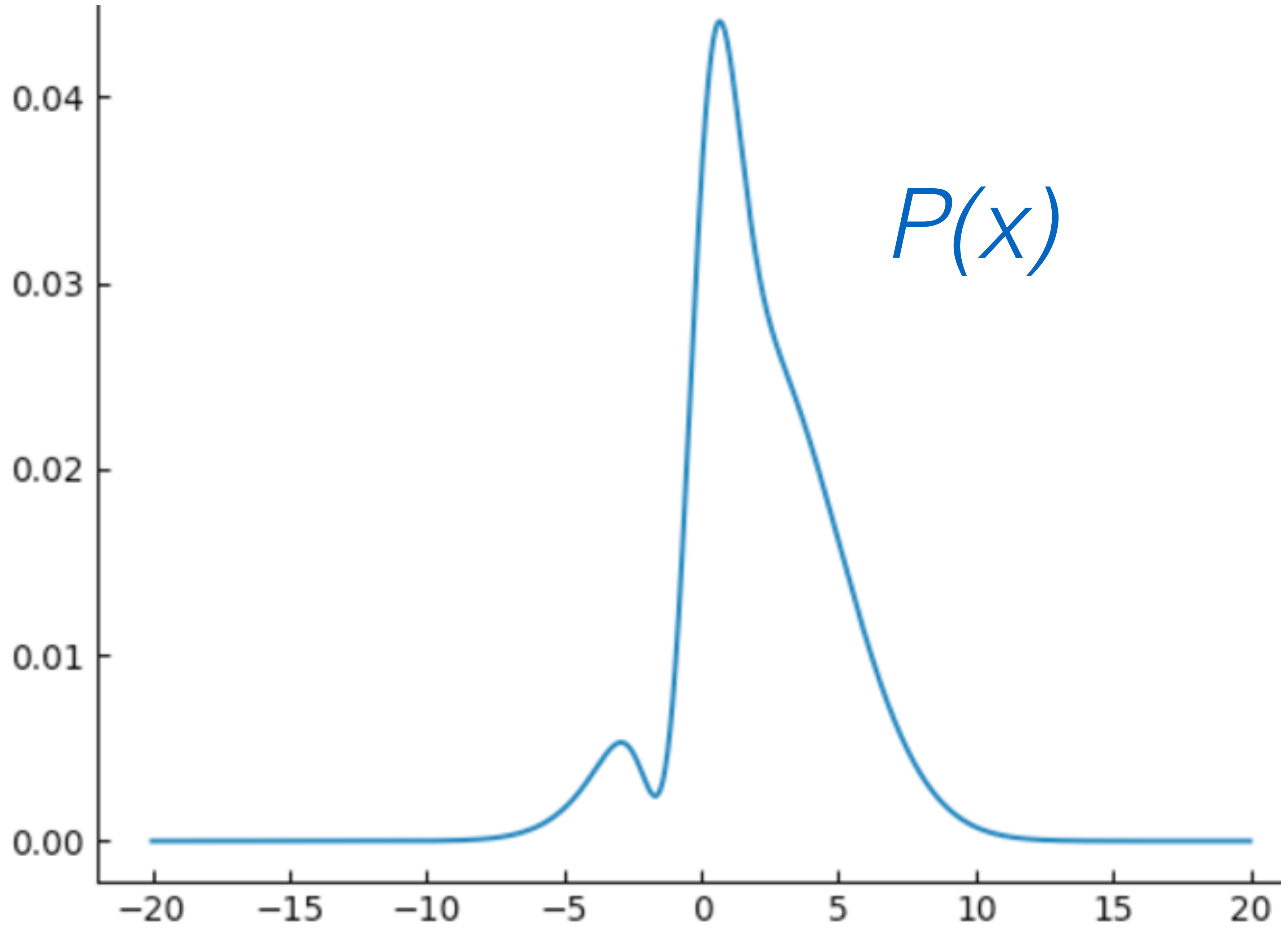
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0175877

# *MC* - 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***

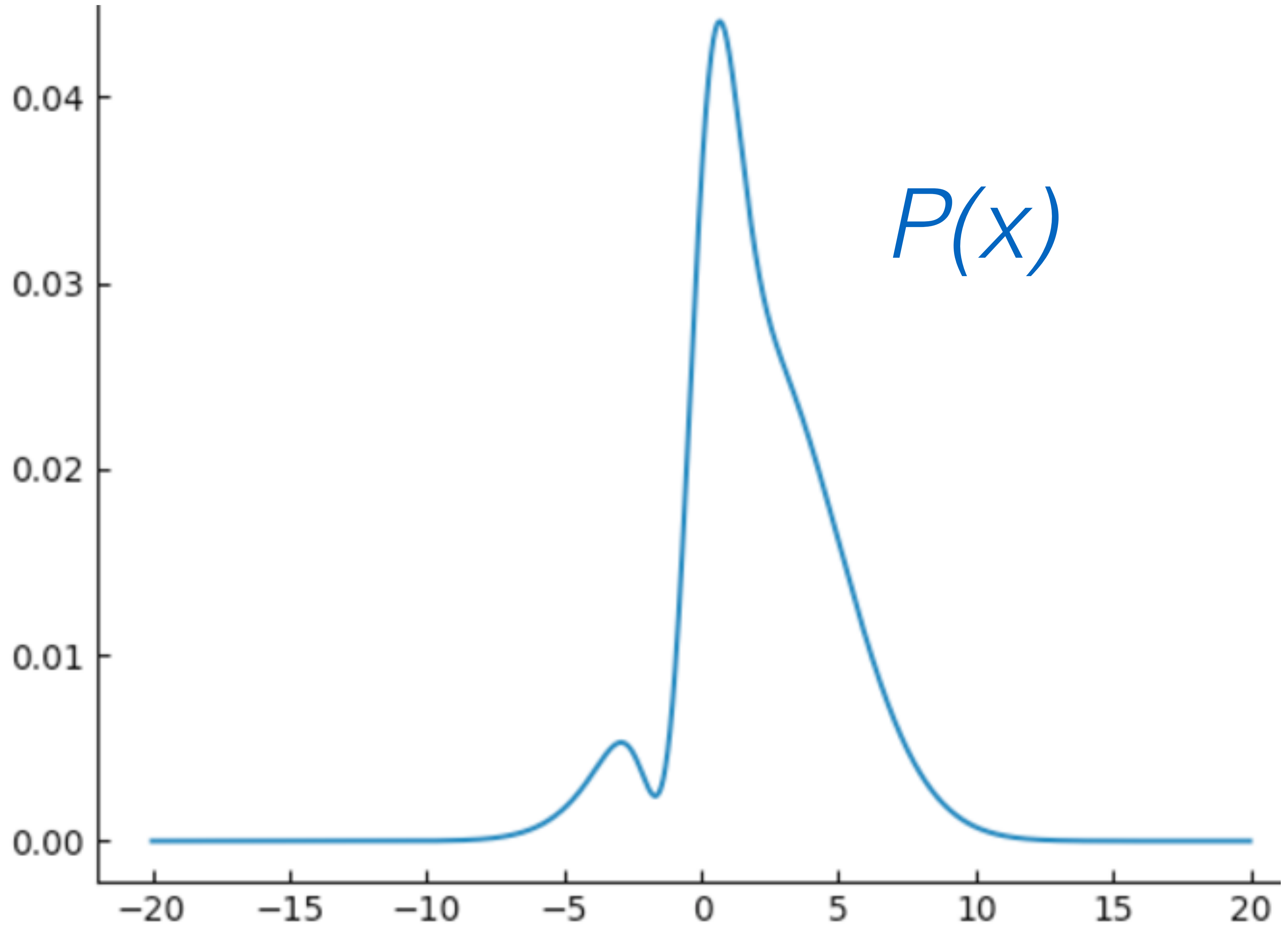


**SetUp 2:**

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 :



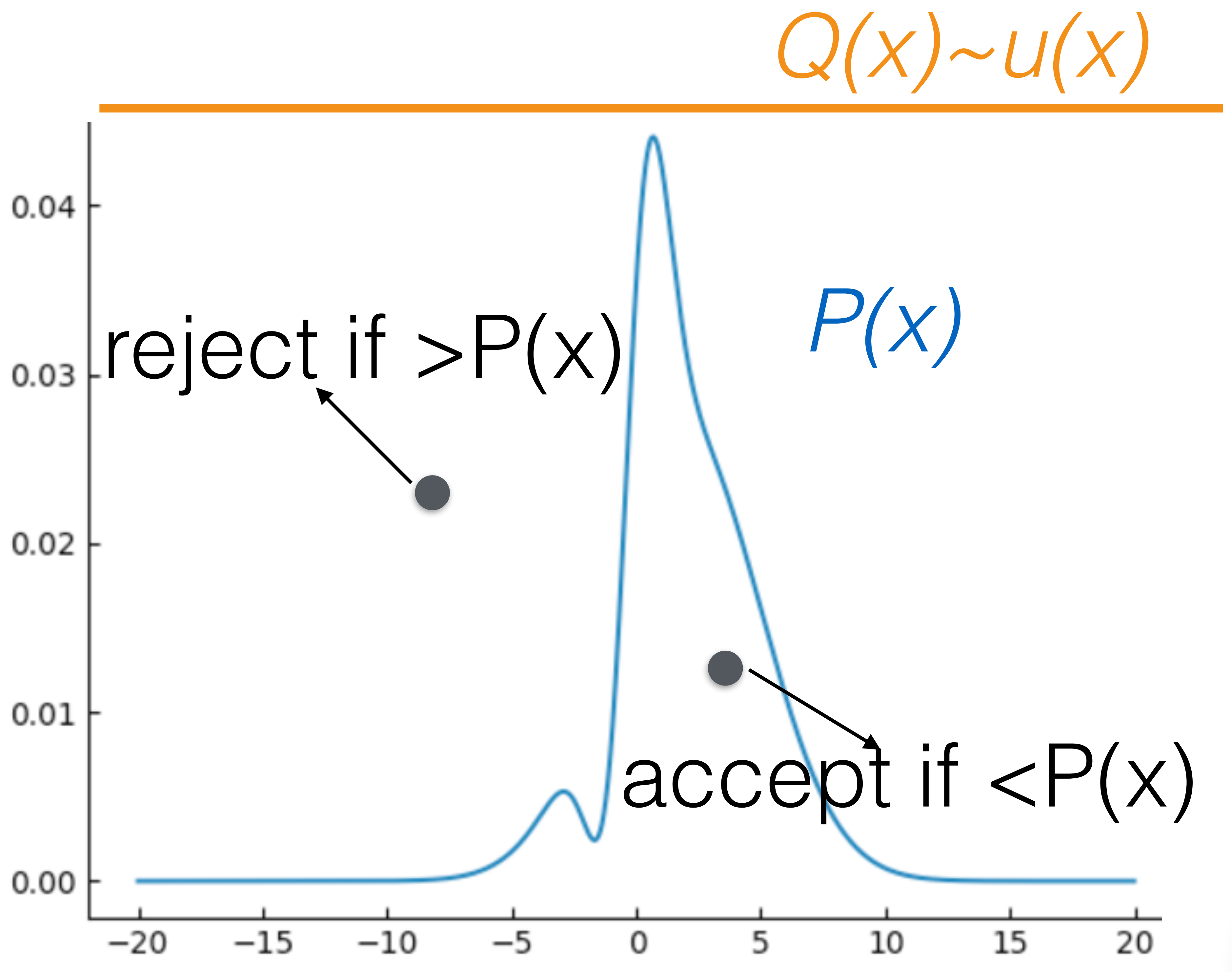
# *MC* - Rejection Sampling

**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)***

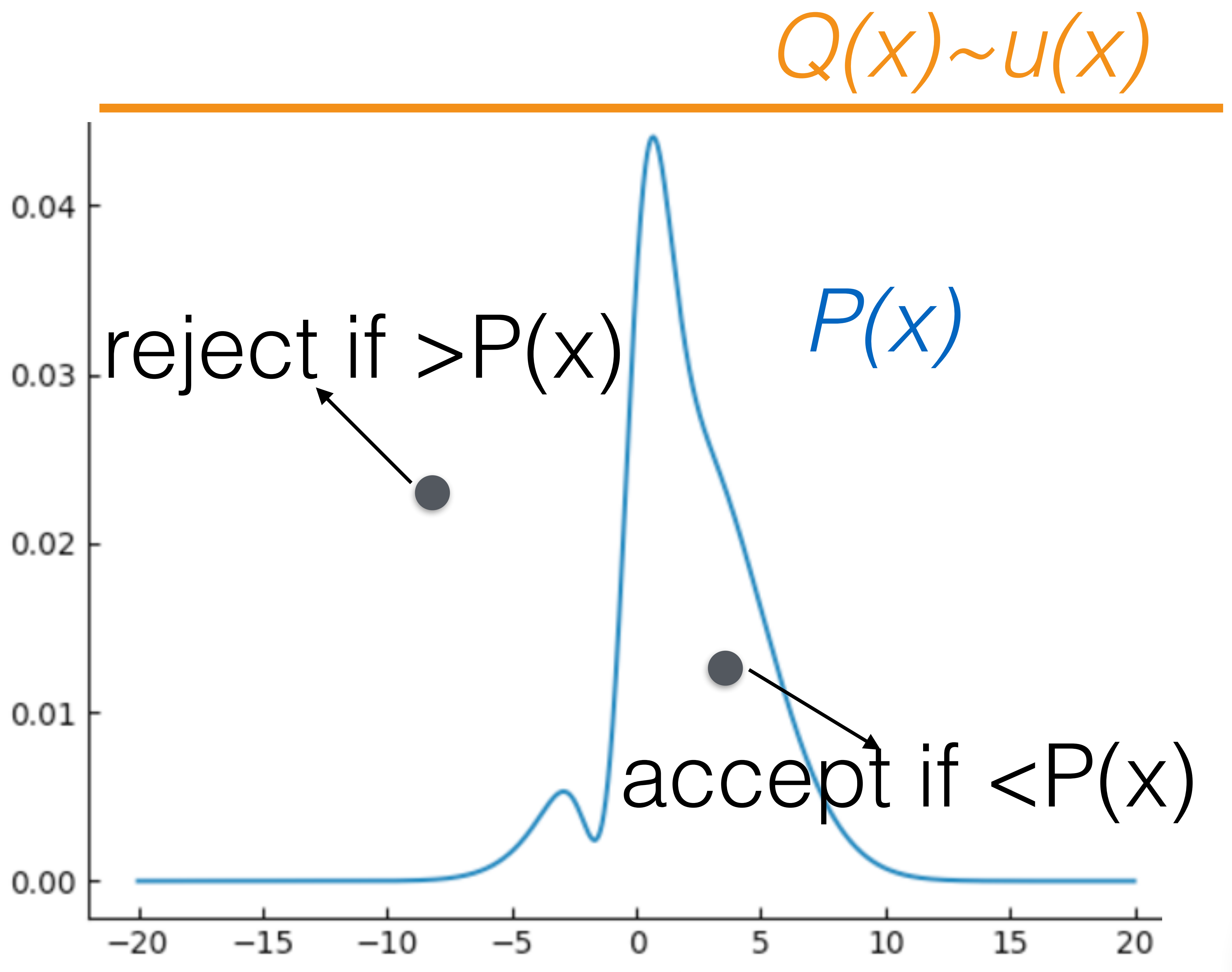
    draw a height ***u ~ Uniform[0,Q(x)]***

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

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

    ELSE :

        reject



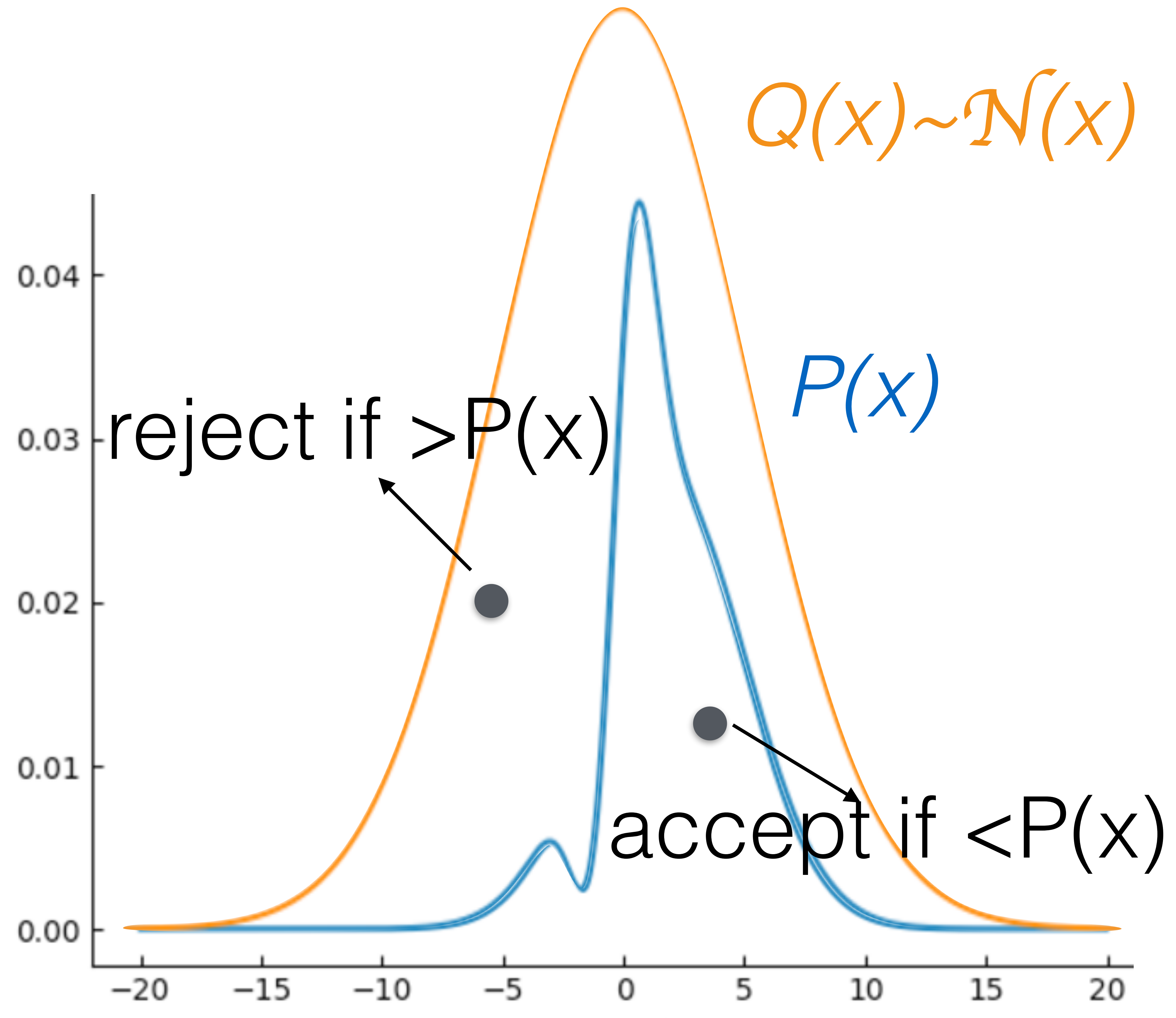
# *MC* - Rejection Sampling

**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

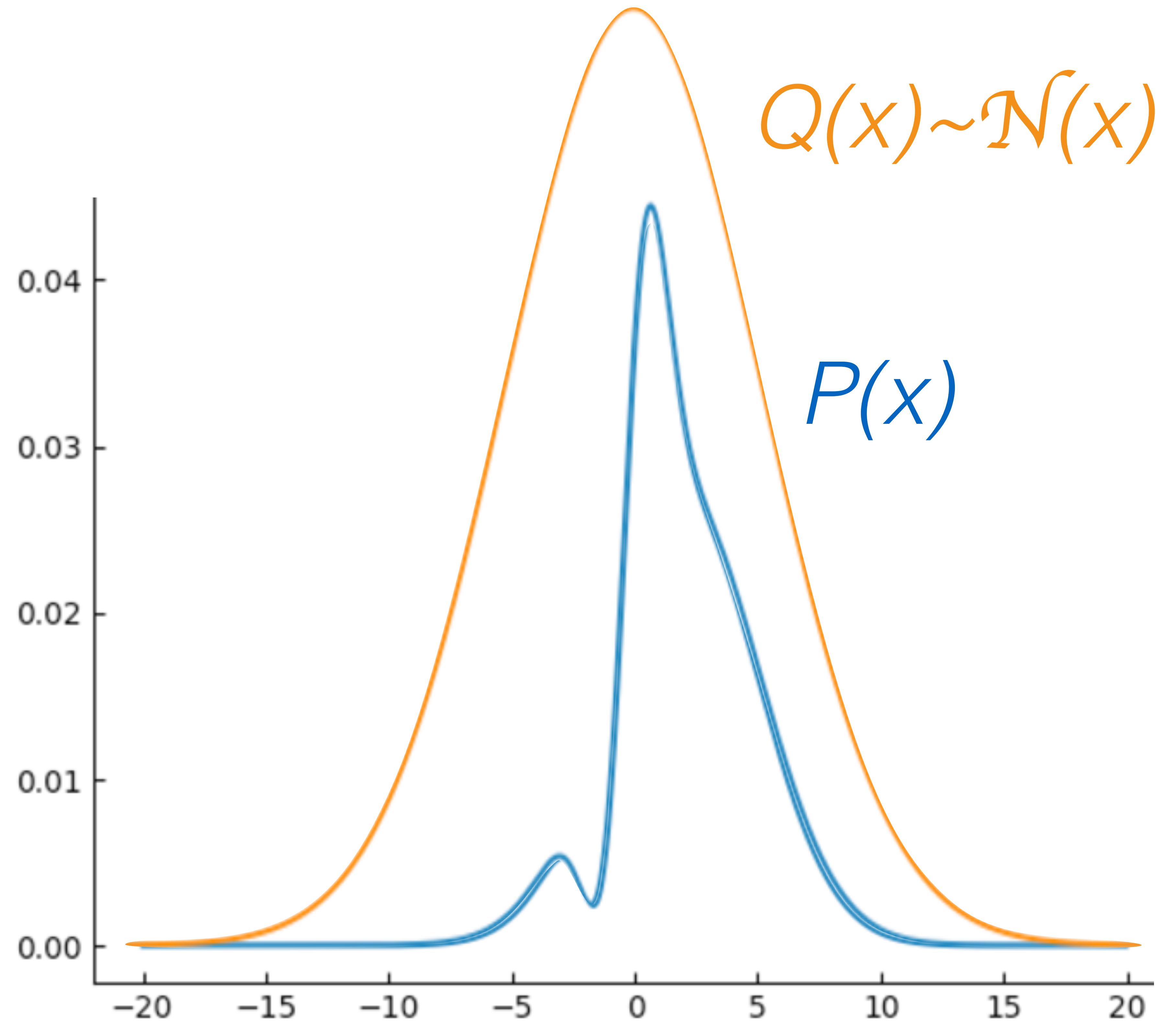
# *MC* - Importance Sampling

**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)




$$\int f(x)P(x)dx = \int f(x)\frac{P(x)}{Q(x)}Q(x)dx, \quad (Q(x)>0 \;\; if\, P(x)>0)$$

$$\approx \frac{1}{S}\sum_{s=1}^{S} f(x_s)\frac{P(x_s)}{Q(x_s)}, \qquad x(s)\sim Q(x)$$

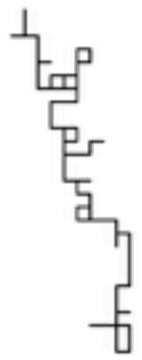
chose Q(x) s.t. Q(x) is large where f(x) P(x) is large

# Markov Chain Monte Carlo

## 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 processes*



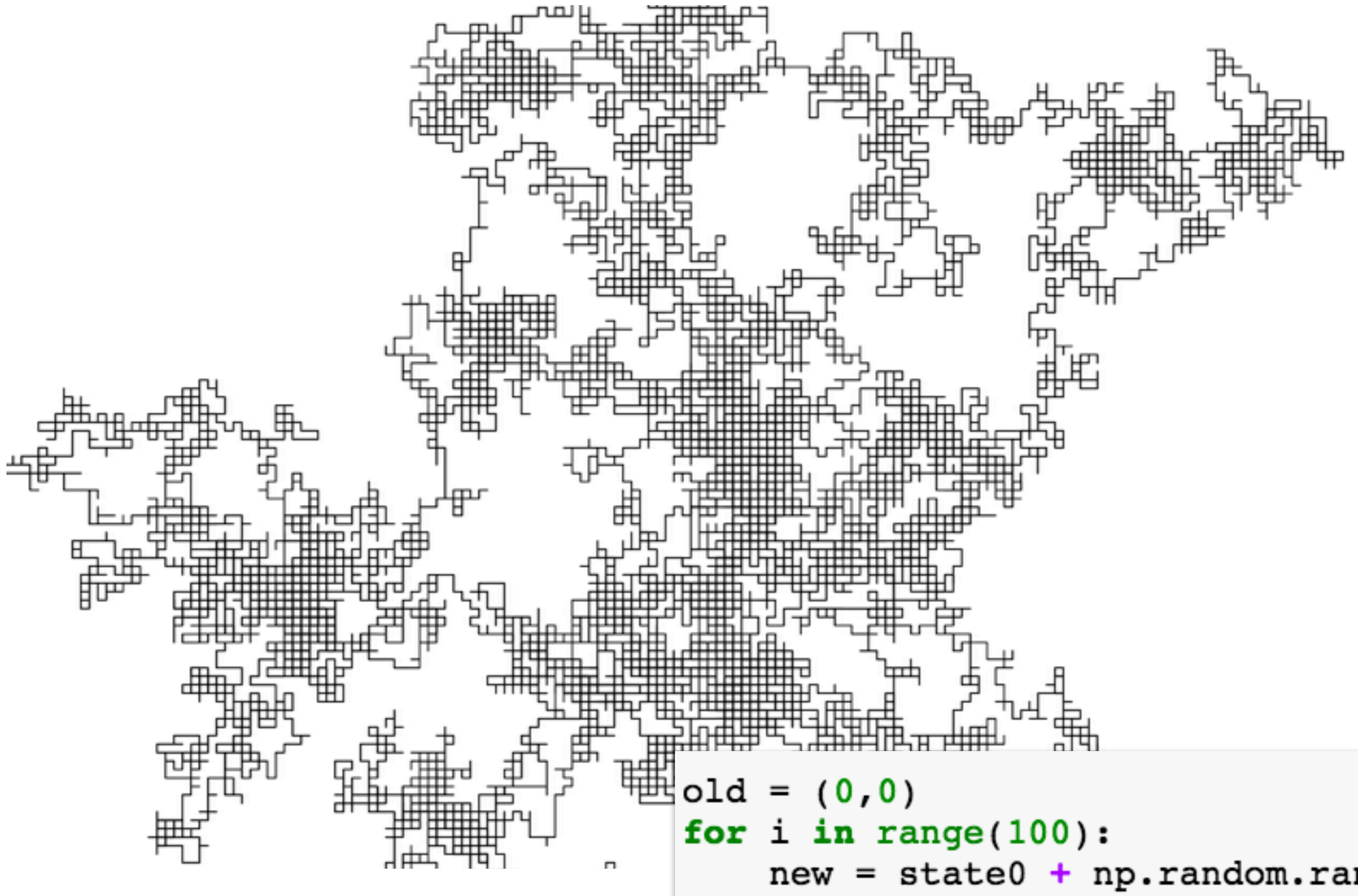
## Markov Chain

*memory-less stockastic process:*

e.g.:

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

Gamblers' ruin



```
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 processes

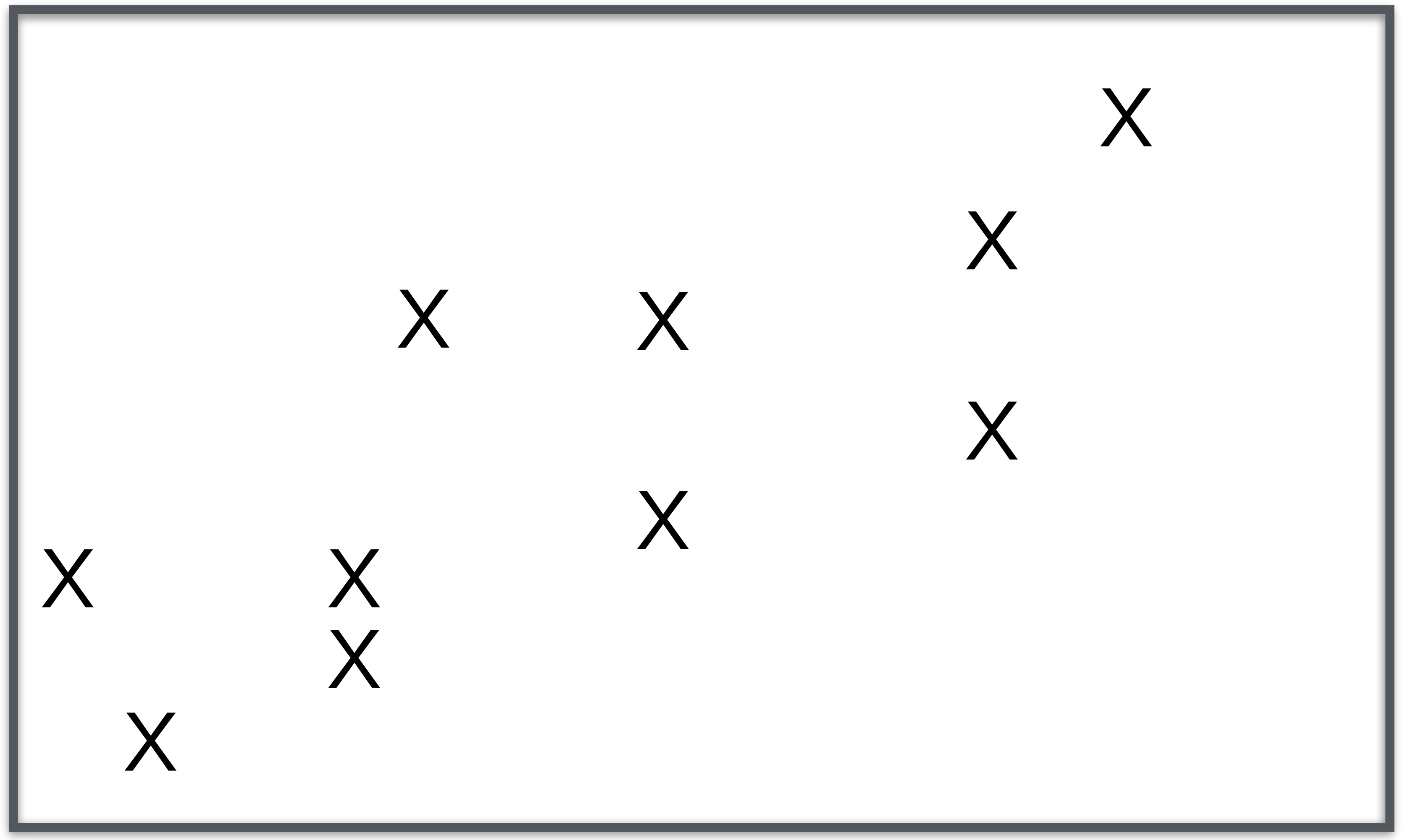
# *MCMC* - motivation

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***

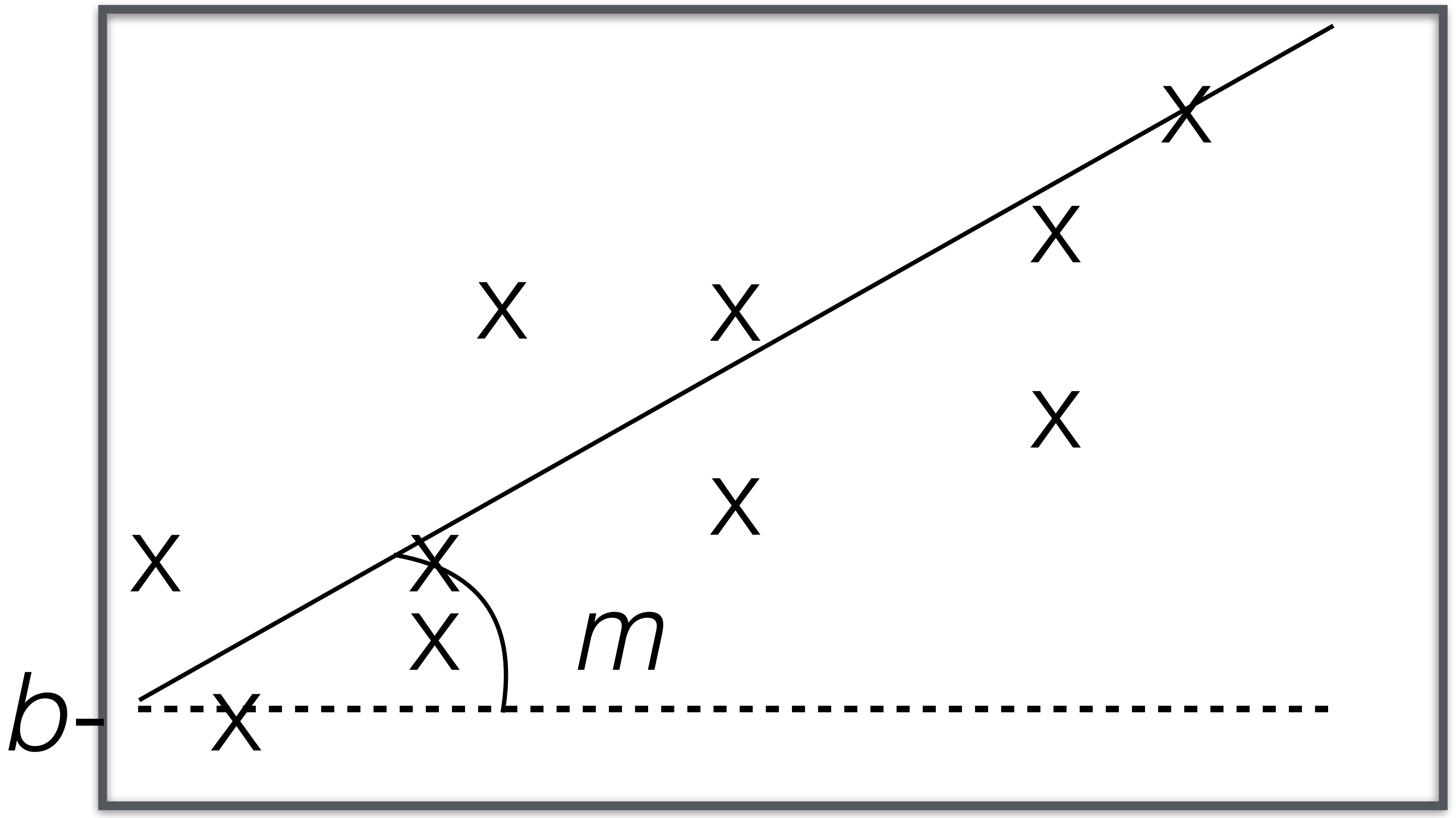


I have a model and I want to find the best parameters to describe my data

Data $\boldsymbol{D}$

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

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



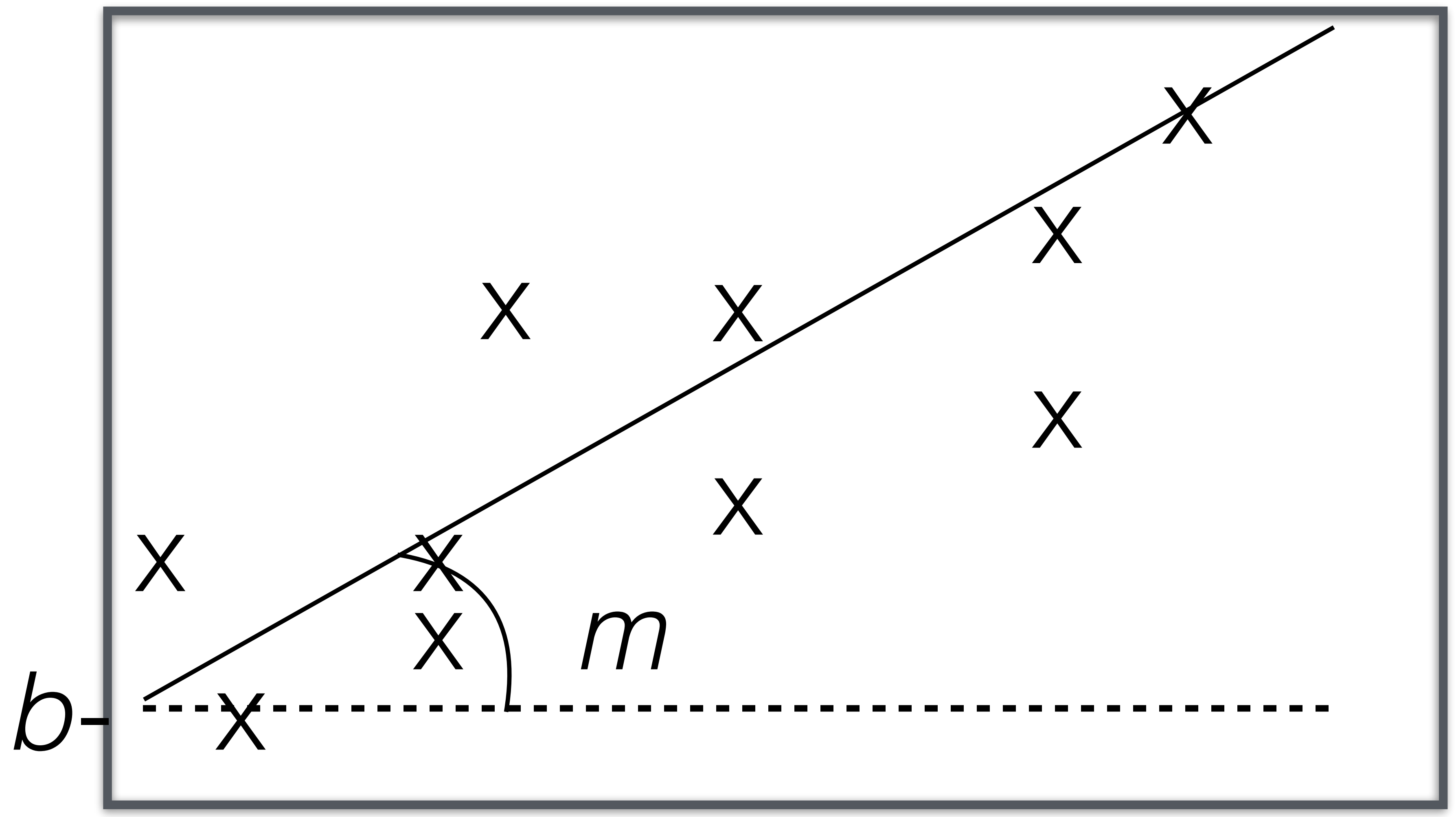
# *MCMC* - motivation

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




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

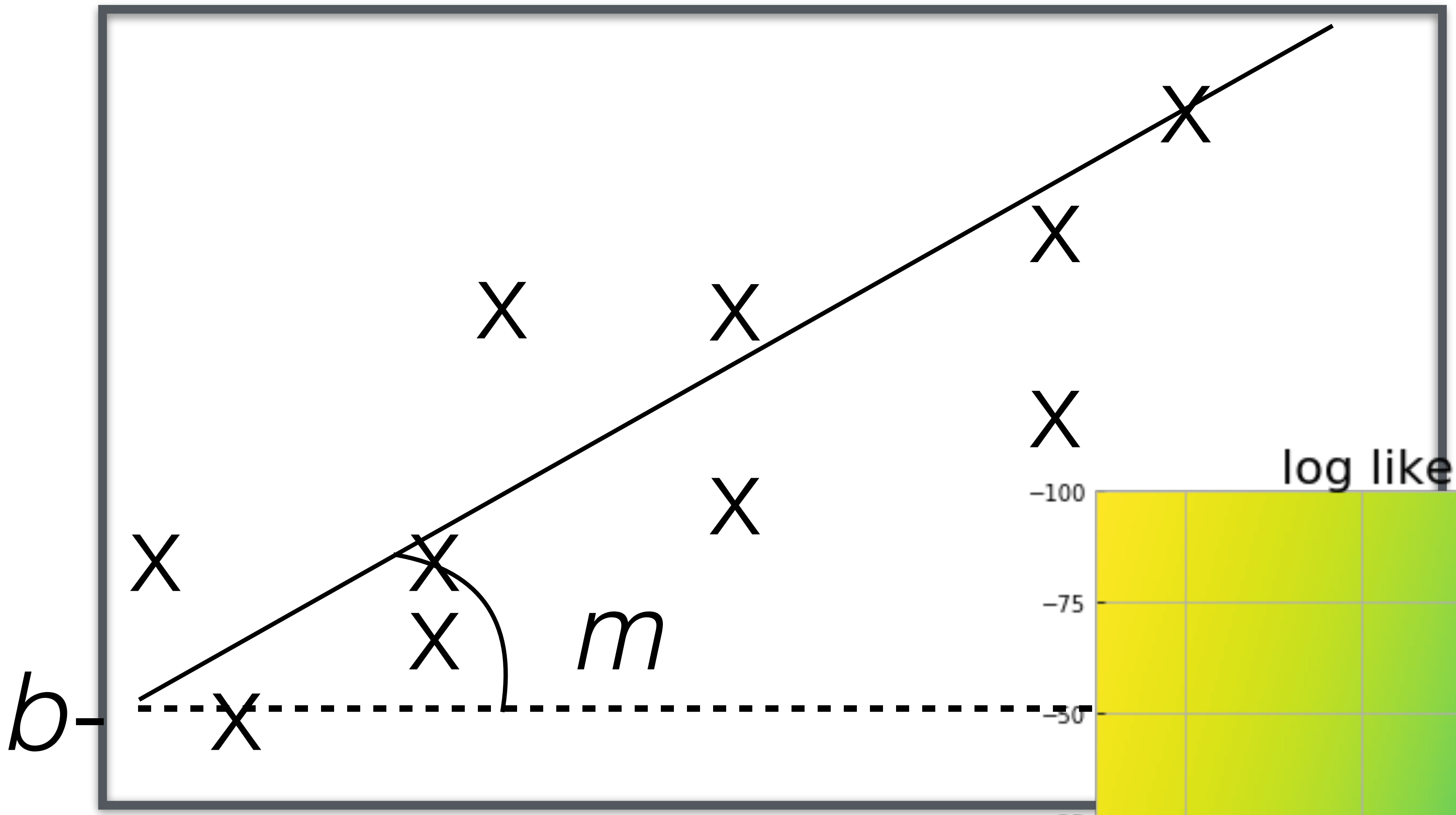
https://github.com/fedhere/PUI2016_fb55/blob/master/HW6_fb55/building_nrg_solution.ipynb

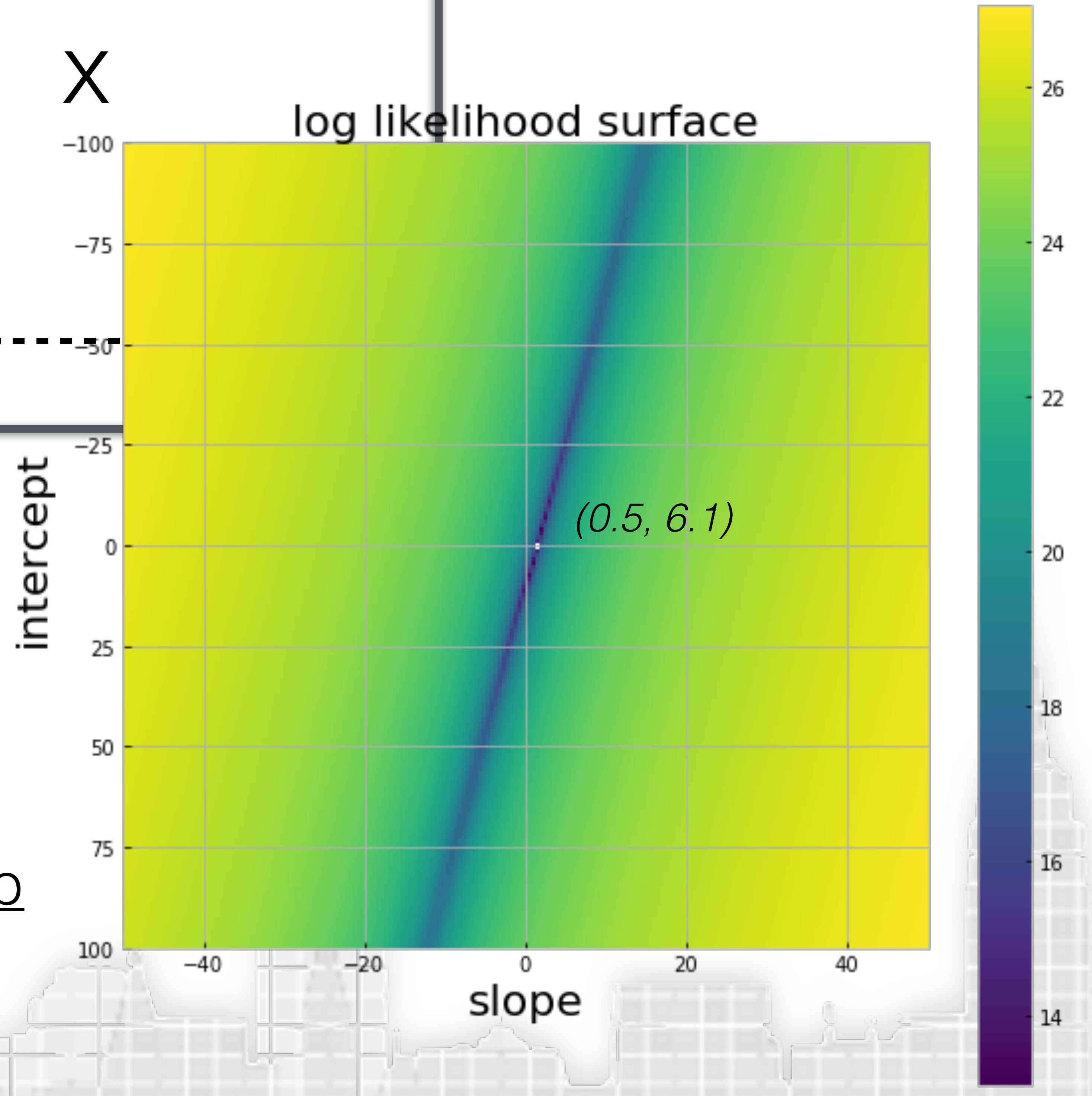
I have a model and I want to find the best parameters to describe my data

Data ***D***

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

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

**Bayes theorem**:

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

likelihood

prior

evidence

**Definitions**: **posterior**

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

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

**prior**

**Definitions**: **posterior**

**posterior:** joint probsbility 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 probsbility 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**:

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

prior

**Definitions**: **posterior**

**evidence**

**posterior:** joint probsbility 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**: $$P(\theta|D,f) = \frac{P(D|\theta,f)\,P(\theta,f)}{\cancel{P(D|f)}}$$

prior

**Definitions**: posterior

evidence

**posterior:** joint probsbility 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

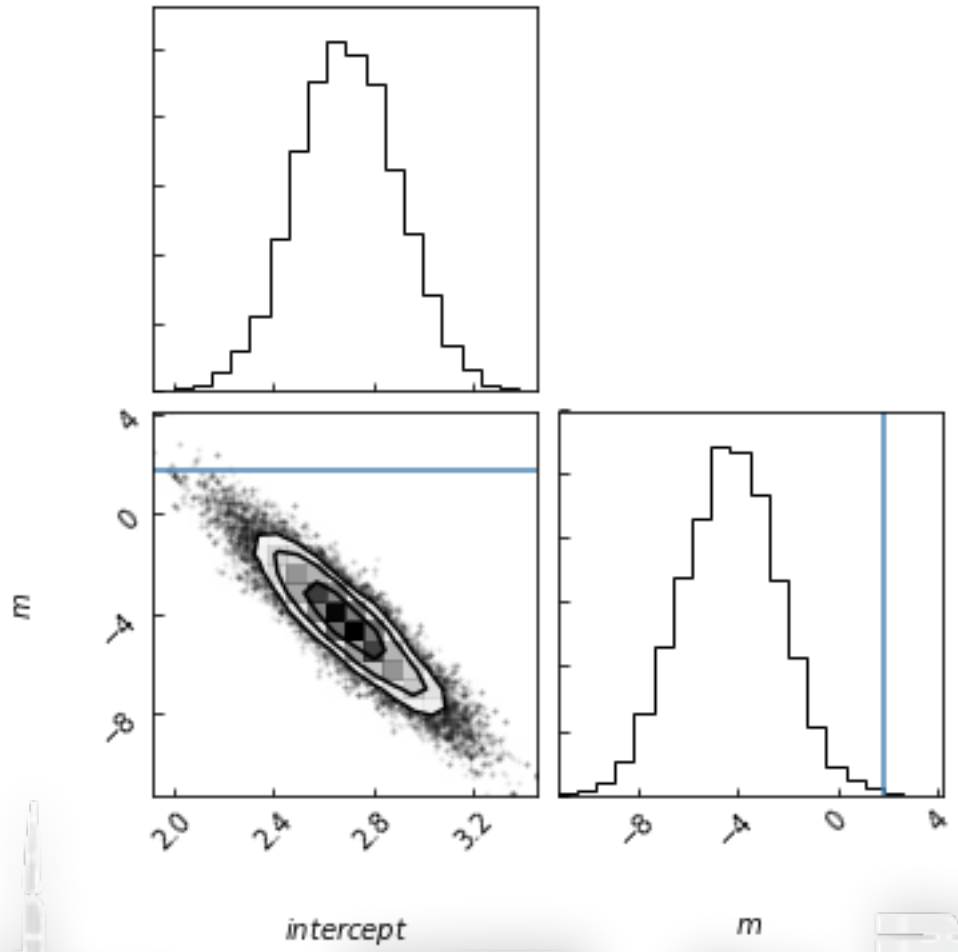
its constant in $\theta$ so we can ignore it $P(D|f) = \int P(D|\theta,f)P(\theta|f)d\theta$

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

**Definitions**: **posterior**

**posterior:** joint probsbility 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)


1. choose a starting point **current** = ***θ₀= (m,b)***
2. calculate current posterior **post_orig = *P(D|θ,f)***
3. WHILE convergence criterion is met:

```
        /*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_orig:
            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
```

4. GOTO point 2

**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?

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

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

3. WHILE convergence criterion is met:

```
        /*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_orig:
            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
```

4. GOTO point 2

**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

1. choose a starting point **current** = ***θ₀= (m,b)***
2. calculate current posterior **post_orig = *P(D|θ,f)***
3. WHILE convergence criterion is met:

   */*proposal*/*
   chose a new set of parameters **new = *θ<sub>new</sub>= (m,b)***

   calculate the new posterior **post_new=*P(D|θ<sub>new</sub>,f)***

   IF ***post_new > post_orig:***
   current = new
   ELSE***:***
   /*accept with probability ***P(D|θ<sub>new</sub>,f) / P(D|θ,f)*** */
   ***r =*** random uniform number [0,1]
   IF r < ***post_new / post_orig:***
   current = new
   ELSE:
   pass //do nothing
4. GOTO point 2

**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)

```
1. choose a starting point current = θ0= (m,b)
2. calculate current posterior post_orig = P(D|θ,f)
3. WHILE convergence criterion is met:
        /*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_orig:
            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
4. GOTO point 2
```

# *MCMC* - Metropolis Hastings algorithm

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

**Goal: sample the posterior distribution**

1. choose a starting point **current** = ***θ₀= (m,b)***
2. calculate current posterior **post_orig = *P(D|θ,f)***
3. WHILE convergence criterion is met:

    */*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_orig:***
        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
4. GOTO point 2

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

1. choose a starting point **current** = ***θ₀= (m,b)***
2. calculate current posterior **post_orig = *P(D|θ,f)***
3. WHILE convergence criterion is met:

    */*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_orig:***
        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
4. GOTO point 2

**Goal: sample the posterior distribution**

**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*

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

## Goal: sample the posterior distribution

1. choose a starting point **current** = ***θ₀= (m,b)***
2. calculate current posterior **post_orig = *P(D|θ,f)***
3. WHILE convergence criterion is met:

   */*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_orig:***
   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
4. GOTO point 2

**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*

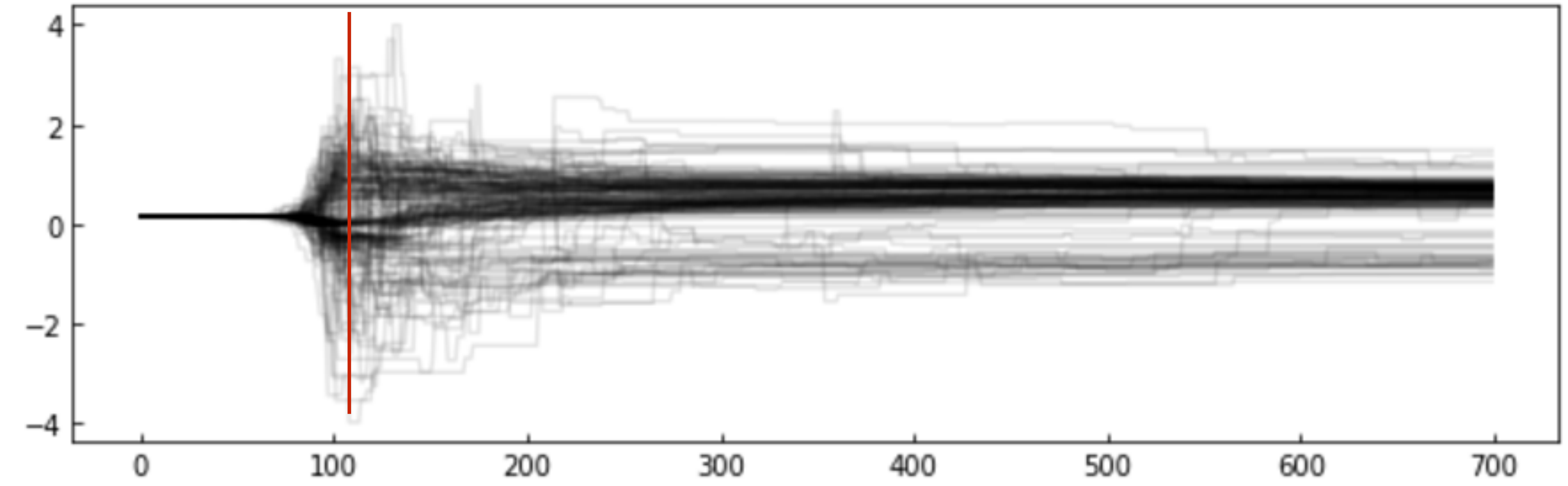
detailed balance $\pi(x_1)p(x_2|x_1)=\pi(x_2)p(x_1|x_2)$

*Metropolis Rosenbluth Rosenbluth Teller 1953 - Hastings 1970*

# MCMC - EMCEE

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

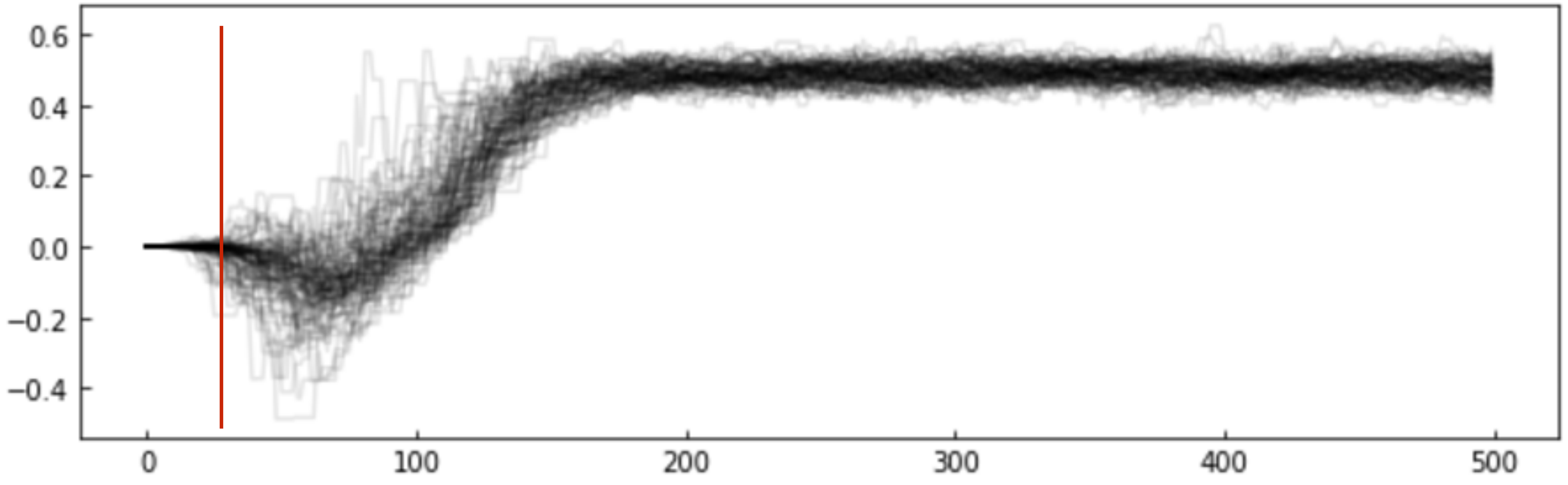
# Goal: sample the posterior distribution




**Questions:**

1. how do I choose the next point?

2. when have I sampled the posterior adequatly?
   has your chain *burned-in* ?



# *MCMC* - convergence

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

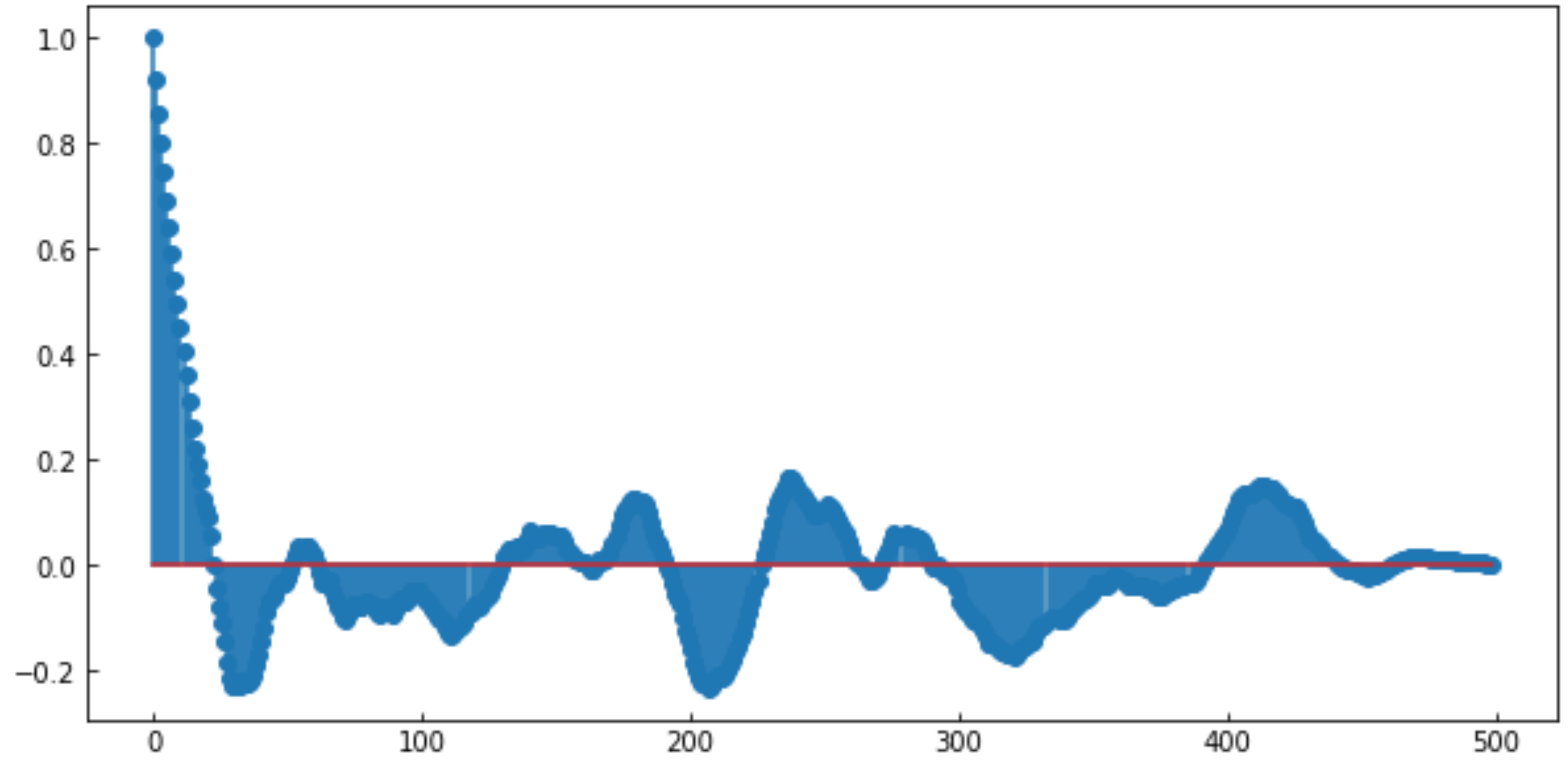
**Goal: sample the posterior distribution**

**Questions:**

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

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

## Goal: sample the posterior distribution



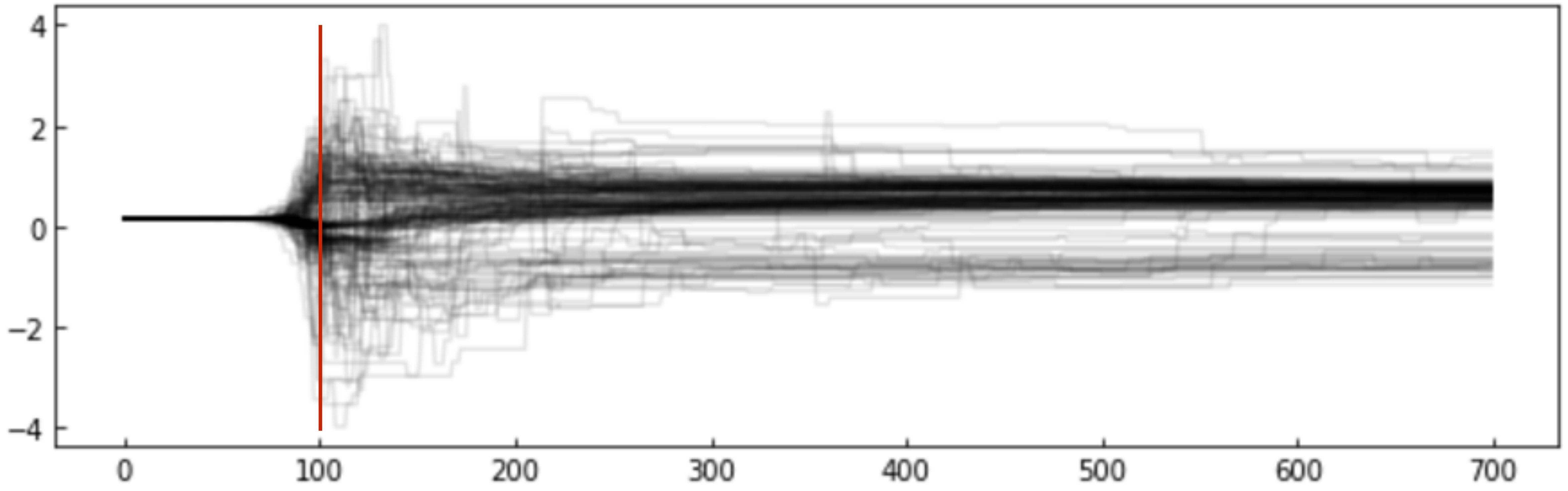

**Questions:**

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)

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

## Goal: sample the posterior distribution




**Questions:**

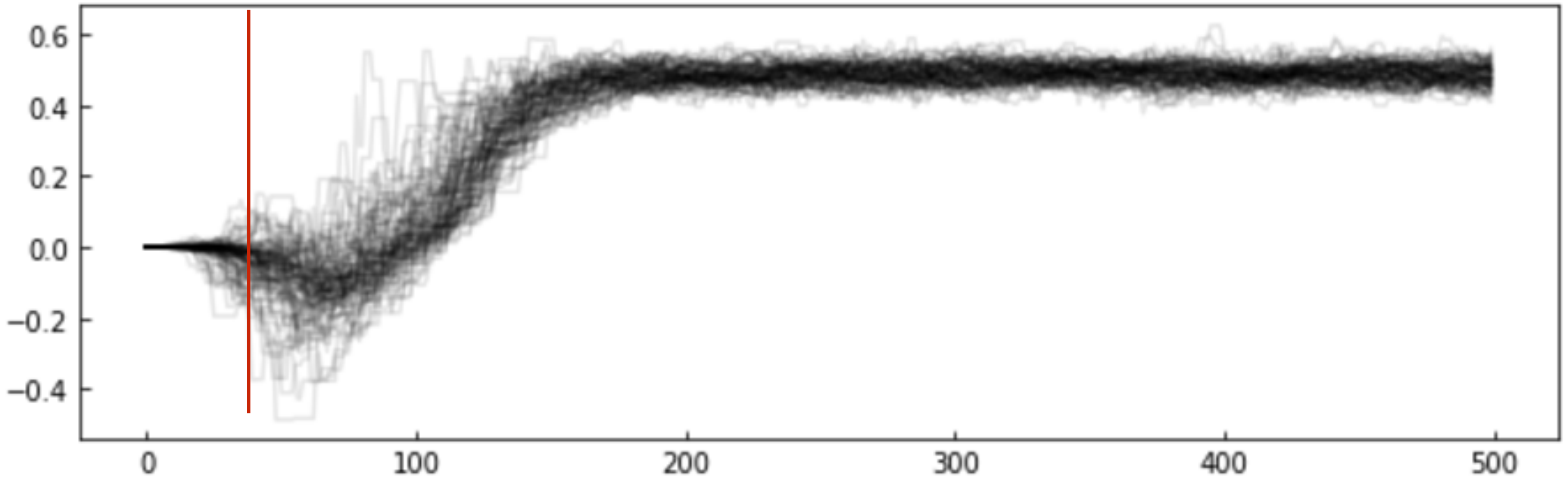
1. how do I choose the next point?

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

1. choose a starting point **current** = ***$\theta_0$= (m,b)***
2. calculate current posterior **post_orig = *P(D | $\theta$,f)***
3. WHILE convergence criterion is met:
   *//proposal*
   chose a new set of parameters **new = *$\theta_{new}$= (m,b)***

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

   IF ***post_new > post_orig:***
   current = new
   ELSE***:***
   //accept with probability ***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
4. GOTO point 2

## 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!*

## 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