Monte Carlo methods : introduction

○ Typical situation : uncertain variables , random systems , many variables (high dim.) , num. integration (later)

$$\Rightarrow \underline{pdfs}$$

○ Pdfs way too complicated for analytical approach.

○ MC approach : work with <u>samples from pdfs</u>, <u>not pdfs directly</u>

pdfs $\longrightarrow$ histograms

expectation values $\longrightarrow$ sample averages

○ Basic challenges :

1) How to generate random numbers using a deterministic system ?   (later)

2) How to generate samples from arbitrary pdf ?

3) Efficiency , in particular for pdfs of many variables (high-dim.)

Markov Chain Monte Carlo (MCMC)

Need to look at :
- Markov chains
- MCMC algorithm example
- Metropolis rule

Monte Carlo approach

- Variable $x \sim p(x)$ $\left[\text{In proj. 4: } \vec{s} \sim p(\vec{s}) = \text{Boltzmann distr.} \right]$

① Draw sample x from $p(x)$ $\left[\text{This is where we will use MCMC} \right]$

② Compute other quantities that dep. on x:

$$f(x), \quad g(x), \quad h(f(x), g(x)), \quad \dots$$

③ Store results you are interested in and repeat from ①

- End up with a set of samples $\{x_1, x_2, \dots\}$, and corresponding samples $\{f_1, f_2, \dots\}$, $\{g_1, g_2, \dots\}$, $\{h_1, h_2, \dots\}$

Typical sample file

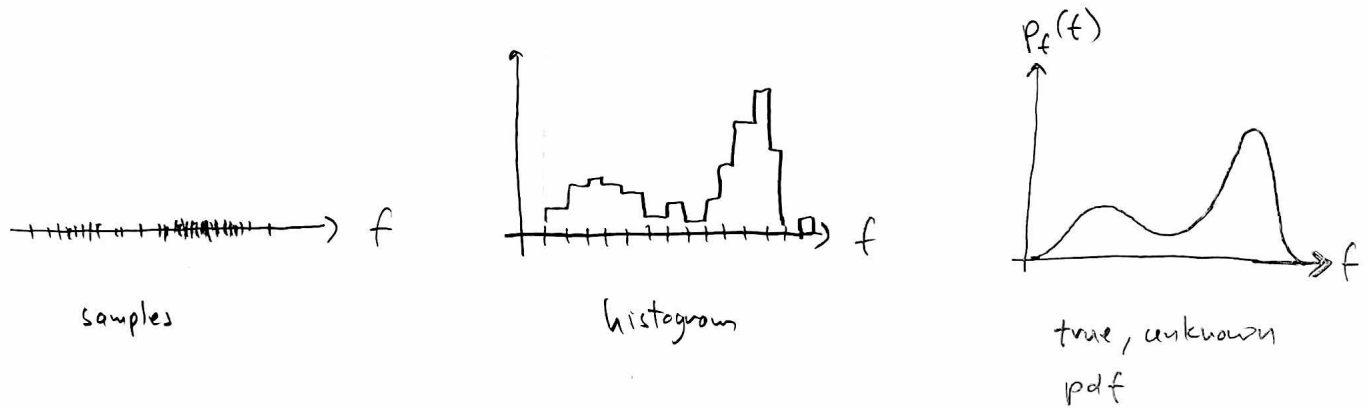| x	$f(x)$	$g(x)$	$h(f, g)$
x_1	f_1	g_1	h_1
x_2	f_2	g_2	h_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$

- Note: Can't always store all information ...

  (Might need billions of samples...)

o Typical ways to use samples :

    o Create histograms to approximate unknown pdfs,

      e.g. histogram of f_i samples to approx. $p_f(f)$

samples   histogram   true, unknown
                                                              pdf

    ○ Approximate expectation values using <u>sample means</u>

$$E[f] \approx \bar{f} = \frac{1}{N} \sum_{i=1}^{N} f_i$$

N : number of samples

true, unknown
expectation value

Note :

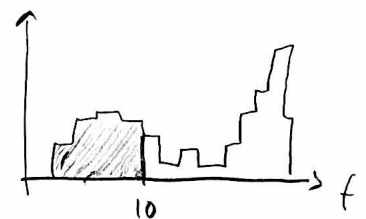$E[\bar{f}] = E[f]$ ($\bar{f}$ unbiased estimator)

$Var[\bar{f}] = \frac{Var[f]}{N}$

    o Compute some integral / probability , e.g.

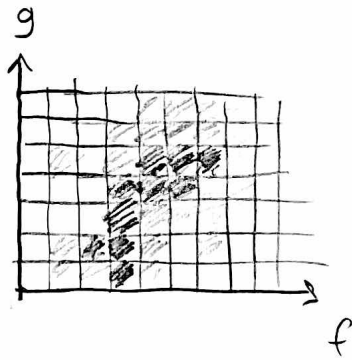$$Prob(f < 10) = \int_{-\infty}^{10} p_f(f) df \approx \frac{N_{f<10}}{N}$$
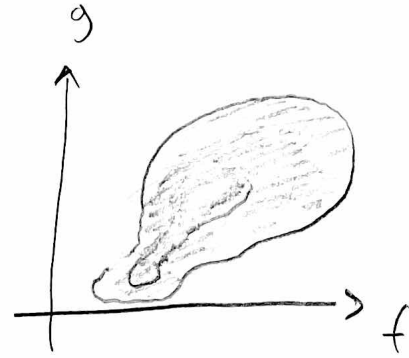
true, unknown
result

    o Estimate position of maxima of $p_f(f)$

○ Approximate multivariate pdfs with multivariate histograms

g

f

histogram of

(f_i, g_i) samples

(not Ising model state ☺)

g

f

true, unknown

joint pdf $P_{f,g}(f, g)$

○ Approximate conditional pdfs

® Note: marginalization ⟷ just don't bin your samples
over variables   in those variable

$$P_f(f) = \iiint P_{x,f,g,h}(x, f, g, h) \, dx \, dg \, dh$$ ⟿ shape of 1D
histogram of the
f samples

(ignoring the x, g, h
samples)

⓪ Subject samples to detector simulation,
data analysis filters, etc ⟹ resulting histogram can be
<u>realistic</u> prediction of experiment
outcome.

(Impossible pdf shape to find
analytically...)