Monte Carlo methods: introduction

Named after Monte Carlo in Monaco.
Famous Monte Carlo Casino
Stanislaw Ulam, John von Neumann

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

  $\Rightarrow$ pdfs

- Pdfs way too complicated for analytical approach.
- MC approach: work with samples from pdfs, not pdfs directly

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.)

$\boxed{\text{Markov Chain Monte Carlo (MCMC)}}$

Need to look at:

- Markov chains
- MCMC algorithm example
  - Metropolis rule

Monte Carlo approach

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

(1) Draw sample x from $p(x)$ [This is where we will use MCMC]

(2) Compute other quantities that dep. on x:

$$f(x), \; g(x), \; h(f(x), g(x)), \ldots$$

(3) Store results you are interested in and repeat from (1)

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

Typical sample file

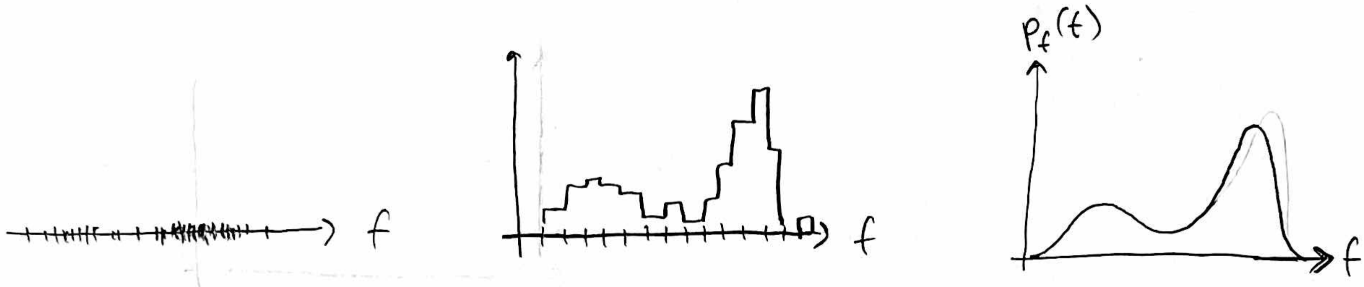
x	$f(x)$	$g(x)$	$h(f,g)$
x_1	f_1	g_1	h_1
x_2	f_2	g_2	h_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$

- Note: Can't always store all information ...
  (Might need billions of samples ...)

Typical ways to use samples:

- 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>

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

true, unknown expectation value

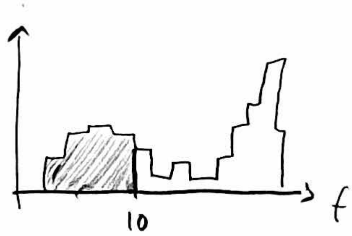
N : number of samples

> Note:
> $$E[\bar{f}] = E[f] \quad (\bar{f} \text{ unbiased estimator})$$
> $$Var[\bar{f}] = \frac{Var[f]}{N}$$

- Compute some integral / probability, e.g.

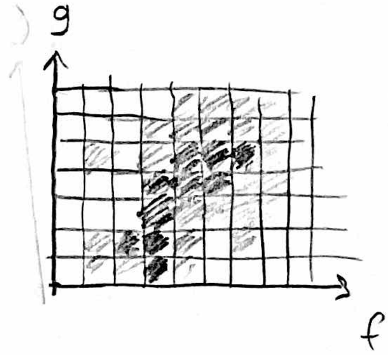
$$Prob(f < 10) = \int_{-\infty}^{10} p_f(f)\, df \approx \frac{N_{f<10}}{N}$$

true, unknown result



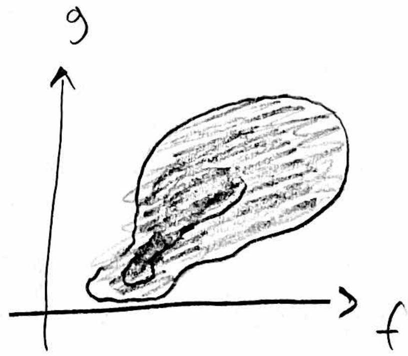
- Estimate position of maxima of $p_f(f)$

- Approximate multivariate pdfs with multivariate histograms



histogram of (f_i, g_i) samples

(not Ising model state ☺)



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

- Approximate conditional pdfs

- Note: marginalization over variables $\longleftrightarrow$ just don't bin your samples in those variable

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

(ignoring the x, g, h samples)

- Subject samples to detector simulation, data analysis filters, etc $\Rightarrow$ resulting histogram can be realistic prediction of experiment outcome.

(Impossible pdf shape to find analytically...)