Acceptance probability.

$$\alpha = \frac{f(x_{t+1})}{f(x_t)} \cdot \boxed{\frac{q(x_t \mid x_{t+1})}{q(x_{t+1} \mid x_t)}}$$

Hastings's Ratio.

Constant.

$$I = \int_a^b f(x)\, dx \qquad\qquad \int_a^b p(x) = 1.$$

$$= \int_a^b \frac{f(x)}{p(x)}\, p(x)\, dx \qquad\qquad p(x) \int_a^b dx = 1.$$

$$p(x) = \frac{1}{(b-a)}.$$

$$= (b-a) \underbrace{\int_a^b f(x)\, p(x)\, dx}_{\langle f \rangle}$$

$$\bar{f} = \frac{1}{N} \sum_{i=1}^{N} f(x_i).$$

$$= (b-a) \langle f \rangle$$

$$\lim_{N \to \infty} \hat{I} = I.$$

$$\hat{I} = (b-a)\, \bar{f} = \frac{(b-a)}{N} \sum_{i=1}^{N} f(x_i).$$

$$\mathrm{Var}(\hat{I}) = \mathrm{Var}((b-a)\, \bar{f}) = (b-a)^2 \cdot \mathrm{Var}(\bar{f}).$$

$$= (b-a)^2\, \sigma_N^2$$

where the std

$$\text{std} = \frac{1}{N-1} \sum_{i=1}^{N} \left(f(x_i) - \bar{f}\right)^2$$

$$\delta \hat{I} = \sqrt{\mathrm{var}\, \hat{I}}$$

$$= \frac{(b-a)\, \sigma_N}{\sqrt{N}}.$$

and

However

Therefore.

Monte Carlo Integration

importance sampling.
5, 53, 55, 181, 184.

Importance Sampling

Back. $I = \int_a^b f(x)\, dx.$

PDF.

$I = (b-a) \langle f(x) \rangle$

$$I \approx \frac{(b-a)}{N} \sum_{i=1}^{N} f(x_i).$$

$\langle f(x) \rangle$ unweighted average.

$$\langle f(x) \rangle = \frac{1}{N} \sum_{i=1}^{N} f(x_i).$$

$(b-a)$ length of interval.

Importance sampling is when a function is weighted. achieve better estimate with fewer points.

$$I = \int_0^1 dx\, w(x) \frac{f(x)}{w(x)}$$

(3.1.5)

$$I = \int_0^1 du \frac{f(x(u))}{w(x(u))}$$

(3.1.6).

$$I \approx \frac{1}{L} \sum_{i=1}^{L} \frac{f(x(u_i))}{w(x(u_i))}$$

$w(x)$ prob density.

$u(0) = 0, u(1) = 1$

$$x = g(y)$$

$$\frac{dx}{dy} = g'(y).$$

$$I = \int_a^b f(x)\, dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(y)) \frac{dx}{dy}\, dy$$

$$(*) \quad = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(y))\, g'(x)\, dy.$$

$$y \in [g^{-1}(a), g^{-1}(b)].$$

$$I \approx \frac{g^{-1}(b) - g^{-1}(a)}{N} \sum_{i=1}^{m} f(g(y_i))\, g'(y_i)$$

$$I = \int_a^b f(x)\, dx \approx \frac{1}{N} \sum_{i=}^{N} \frac{f(x_i)}{p(x_i)}.$$

$p(x)$ prob distribution.

$$I = \int_a^b f(x)\,dx \qquad \langle f(x)\rangle = \frac{1}{N}\sum_{i=1}^{N} f(x_i).$$

$$I = (b-a)\,\langle f(x)\rangle$$

$$I \approx \frac{(b-a)}{N}\sum_{i=1}^{N} f(x_i).$$

OG MC.

$x = g(y)$

$\frac{dx}{dy} = g'(y).$

$$\int_a^b f(x)\,dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(y))\,\frac{dx}{dy}\,dy.$$

$$= \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(y))\,g'(y)\,dy.$$

$$I \approx \frac{g^{-1}(b) - g^{-1}(a)}{N}\sum_{i=1}^{N} f(g(y_i))\,g'(y_i)$$

$$E_p[f(x)] = \int f(x)\,p(x)\,dx.$$

$$I = \int_a^b f(x)\,dx = \int_a^b \frac{f(x)}{p(x)}\,p(x)\,dx.$$

$$I = E_p\left[\frac{f(x)}{p(x)}\right].$$

$$I \approx \frac{1}{N}\sum_i \frac{f(x_i)}{p(x_i)}.$$

or instead of uniform prob dist chose x with any prob dist $p(x)$.

$\therefore\ I \approx \frac{1}{N}\sum_{i=1}^{N} \frac{f(x_i)}{p(x_i)}$ $\quad p(x)$ prob dist $\int p(x)\,dx = 1.$

$w(x)$ not normalised. $w(x)$ weighting func.

$$p(x) = \frac{w(x)}{\int w(x)\,dx.} \qquad I \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{w(x_i)}\int w(x_i)\,dx.$$

normalising $w(x)$

rewrite using prob density $u(x)$

$$I = \int \frac{f(x)}{w(x)}\,w(x)\,dx.$$

$w(x)$ derivation of $u(x)$, $u(0)=0$, $u(1)=1$ implies $w(x)$ normalised.

$$I = \int \frac{f(x(u))}{w(x(u))}\,du$$

generate L random values of u uniformly distributed $\therefore$

$$I \approx \frac{1}{L}\sum_{i=1} \frac{f(x(u_i))}{w(x(u_i))}.$$

Weighted average of function of random Variable

$E_{p_\theta}[h(x)] = \int_{\mathbb{R}} h(x)\, p_\theta(x)\, dx.$ $\quad E_{p_\theta}[h(x)] = \sum_{i=1}^{\infty} h(x_i) p_\theta(x_i).$ $\quad \frac{1}{N}\sum_{i=1}^{N} h(x_i) \approx E_{p_\theta}[h(x)].$

1) Can't sample from p_θ
2) Inefficient to sample from p_θ • variance reduction • Rare event probabilities.
3) p_θ is not normalized $p_\theta(x) = \frac{\hat{p}_\theta(x)}{Z_p}$ ← Z_p can't compute.

$E_{p_\theta}[h(x)] = \int_{\mathbb{R}} h(x) \frac{q_\phi(x)}{q_\phi(x)} p_\theta(x)\, dx.$

↑ probabilistic one

$E_{q_\phi}[h'(x)] = \int_{\mathbb{R}} h(x) \frac{p_\theta(x)}{q_\phi(x)} q_\phi(x)\, dx$ $\quad h'(x) = h(x)\frac{p_\theta(x)}{q_\phi(x)}.$

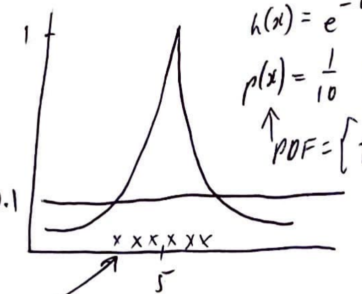
Sampling (prob density distribution func) is now q_ϕ.

Swapped places.

Likelihood Ratio (Importance weight).

$E_{q_\phi}\left[h(x)\frac{p_\theta(x)}{q_\phi(x)}\right] \approx \frac{1}{N}\sum_{i=1}^{N} h(x_i)\frac{p_\theta(x_i)}{q_\phi(x_i)}.$

How to choose q_ϕ?



$h(x) = e^{-2|x-5|}$

$p(x) = \frac{1}{10}$ $x \sim U(0,10)$.

↑ PDF $= \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise.} \end{cases}$

$p(x) = \frac{1}{10} = 0.1$

"Region of Importance"

$\int_0^{10} e^{-2|x-5|}\, dx. = 1 - \frac{1}{e^{10}} \approx 0.99995.$

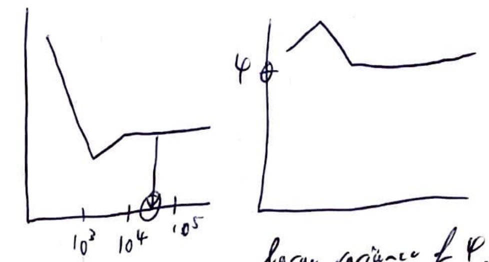
$= 10 \int_0^{10} e^{-2|x-5|} \frac{1}{10}\, dx.$

$= 10 \int_0^{10} h(x) p(x)\, dx.$

$\approx \frac{10}{N}\sum_{i=1}^{N} h(x_i)$ $x_i \sim U(0,10)$. Standard MC.

Law of Large Numbers. |Wiki

The average of the results obtained from a large no. of trials should be close to the expected value and will tend to become closer to the expected value as more trials are performed!



large variance of Ψ.

$q(x)$ should be chosen st. it helps in sampling from a region of importance.

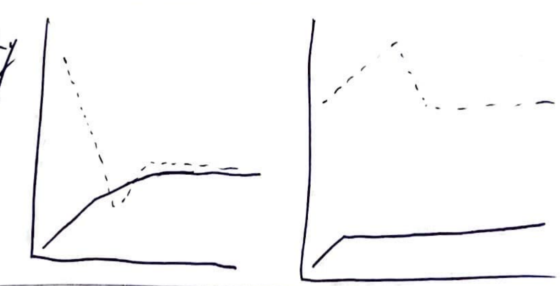
$q(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-5)^2}{2}}.$ $x \sim N(5,1)$

$\therefore \int_0^{10} e^{-2|x-5|}\, dx$

$= 10 \int_0^{10} h(x) p(x)\, dx.$

$= 10 \int_0^{10} h(x) \frac{p(x)}{q(x)} q(x)\, d(x).$

$\approx 10 \sum_{i=1}^{N} h(x_i) \frac{p(x_i)}{q(x_i)}$



Emphasize that the region of importance depends on the problem, for this problem it was in the middle of the function $h(x)$. However, for another problem the region of importance could have been somewhere else eg LHS, RHS.

Importance Sampling in Statistical Physics.

$e^{-\beta E_i}$, Boltzmann factor. States are weighted by a prob that depends on state energy E_i

inverse temp $\beta = \frac{1}{kT}$.

normalised prob func for each energy state.

$$P_B(E_i) = \frac{e^{-\beta E_i}}{Z}.$$

where $Z = \sum_i e^{-\beta E_i}$ partition function.

[want to calculate expectation value for a quantity for a system that is in thermal equilibrium at temp T.].

Expectation value of a physical quantity X that depends on the state is

$$\langle X \rangle = \sum_i P_B(E_i) X_i.$$

MCMC

Markov Chain MC.

Boltzmann factor $e^{-\beta E_i}$ state energy E_i

inverse temperature $\beta = \frac{1}{kT}$.

normalised prob func. for each energy state.

$$P_\beta(E_i) = \frac{e^{-\beta E_i}}{Z}$$

$Z \equiv \sum_i e^{-\beta E_i}$ partition function

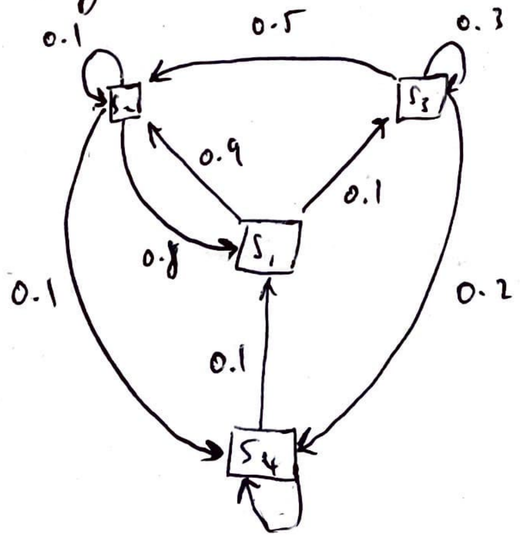
Could weight by $P_\beta(E_i)$, but this involves knowing the partition function. A work around is to generate a series of energy values with distribution consistent with a Boltzmann distribution without knowing or computing the partition function. This motivates our next topic, Markov chains.

Motivation + Conclusion:
- Multidimensional
- Boltzmann
- Quantum.

output1, output2, ... = Function(input1, input2, ...)

importance sampling → occurrence of one sample has no impact on the chances of occurrence of another sample.

- Independent Sampling.



$$\pi = \begin{bmatrix} P(X_0 = s_1) \\ P(X_0 = s_2) \\ P(X_0 = s_3) \\ P(X_0 = s_4) \\ \vdots \\ P(X_0 = s_N) \end{bmatrix} = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \vdots \\ \pi_N \end{bmatrix}$$

$$P(X_0 = i) = \pi_i$$

to X_{t+1}.

$$\begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \end{array} \begin{array}{c} \begin{array}{cccc} s_1 & s_2 & s_3 & s_4 \end{array} \\ \begin{pmatrix} 0 & 0.9 & 0.1 & 0 \\ 0.8 & 0.1 & 0 & \boxed{0.1} \\ 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0 & 0.9 \end{pmatrix} \end{array}$$

X_t from

$$P(X_{t+1} = s_2 \mid X_t = s_3) = 0.5$$

$$\square \; P(X_{t+1} = s_4 \mid X_t = s_2) = 0.1$$

$$P(X_{t+1} = \underset{\uparrow \text{ column}}{j} \mid X_t = \underset{\uparrow \text{ row}}{i}) = P_{ij}$$

Transition probabilities form a 2 dimensional square matrix

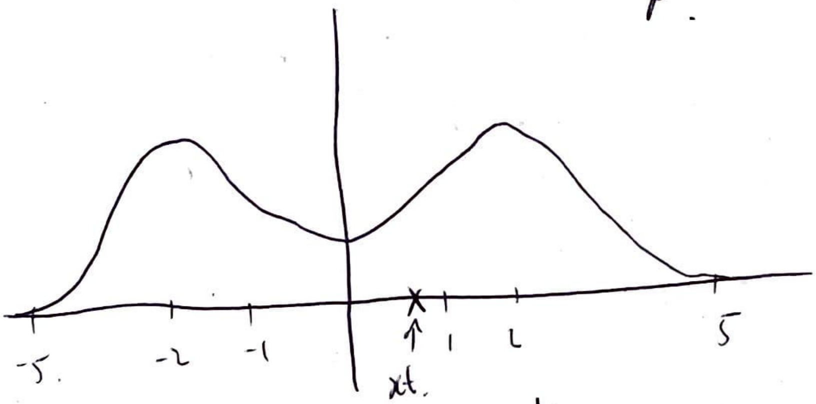
start single position $\pi_0 = (0\,0\,0\,1\,0\,...)$

After one step, new distribution of probabilities becomes new vector of probabilities $\pi_1 = \pi_0 P$

$$P(X_t = j \mid X_0 = i) = P(X_{n+t} = j \mid X_n = i) = (P^t)_{ij} \quad \text{for any } n.$$

$$\pi_1 = \pi_0 P \qquad \underset{\pi}{(\cdots)} \underset{P}{\begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}}$$

lecture notes for stats 325
Fewster.



posterior distribution.

$f(x)$

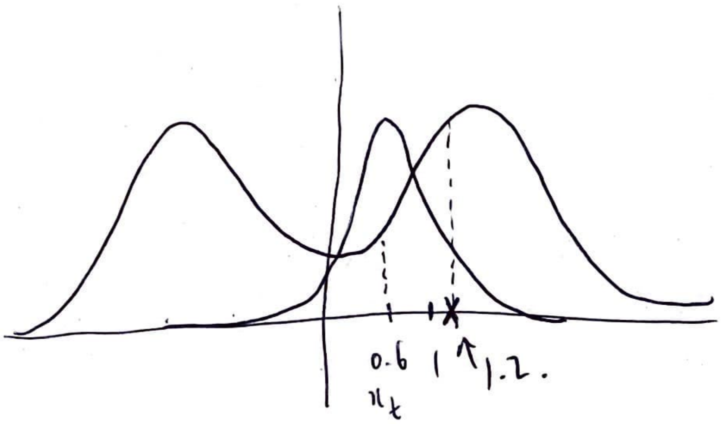
$X \sim 0.5\, N(-2,1) + 0.5\, N(2,1)$

$x_t = 0.6$.

add some random perturbance. $x_{t+1} = x_t + N(0, \boxed{\sigma^2})$.

step size.

$x_{t+1} \sim N(x_t, \sigma^2)$.



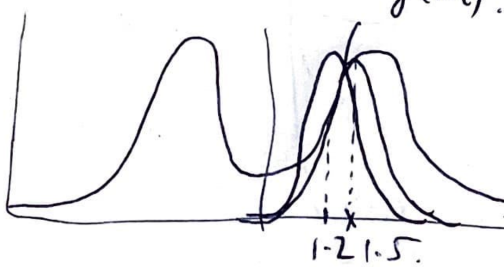
$q(x_{t+1} | x_t)$.

$x_{t+1} \sim N(\overset{0.6}{x_t}, 0.4)$. standard deviation

$$\alpha = \frac{f(x_{t+1})}{f(x_t)}.$$

$u \sim U(0,1)$.

if $\alpha > u$
move to proposed sample
else
stay at current sample



Proposal function doing job of transition matrix

Acceptance Criterion (Move).

MCMC methods $\rightarrow$ transition Kernel.
continuous.

storage = [0.6].
if we move
storage = [0.6, 1.2].
else
storage = [0.6, 0.6].

$\left.\begin{array}{l}\text{- step size} \\ \text{- how many samples}\end{array}\right\}$ serious challenges to generate stationary Markov Chain.

No guarantee you would have approximated the target distribution.