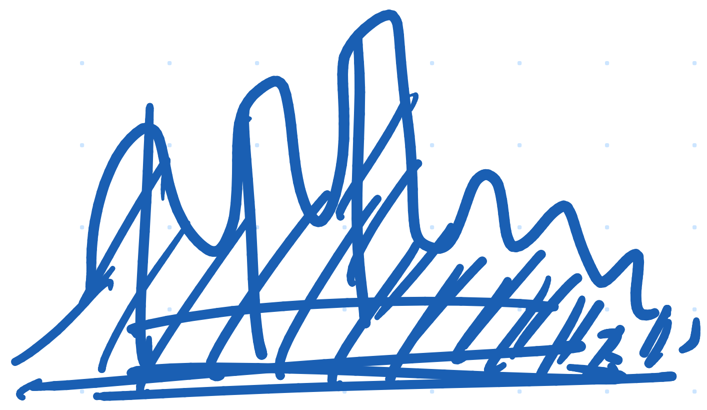
Sampling Techniques.

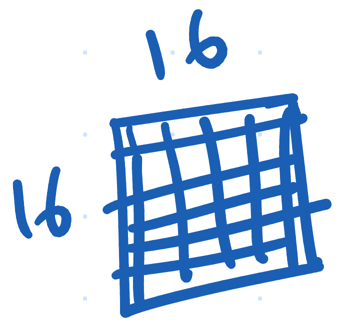
01/02/2025

Latent

x $\quad \mathcal{E}_\theta \quad [\;] \quad D_\theta \quad$ $\tilde{x} \simeq x$. MSE.

$x \sim p_{data}$. $\quad p(z|x)$ $\quad z \quad$ $p(x|z)$.





$x \in \mathbb{R}^{n \times n}$

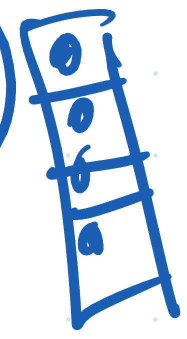
VAE.

Simple dist

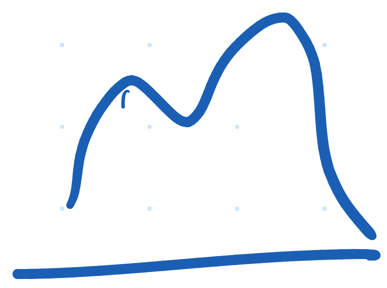
$\rightarrow z^{(i)}$. $\quad D_\theta$

$\tilde{x} \simeq x$.

$\mathcal{N}(0, I)$.

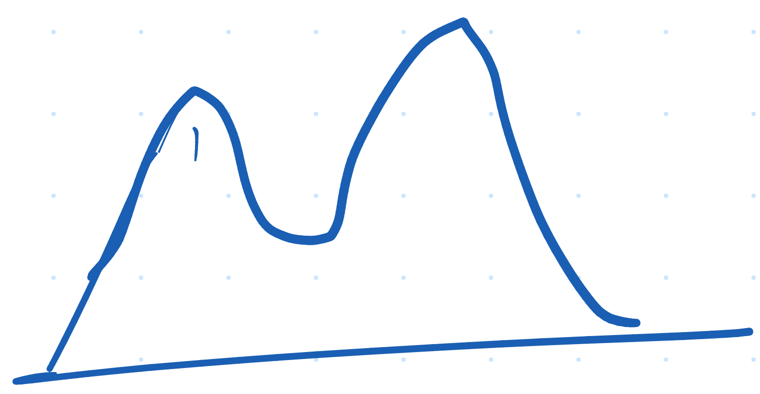


$4 \longrightarrow n \times n$



$n \times n >> 4$

$$P(x) = 3e^{-\frac{x^2}{2}} + e^{-\frac{(x-4)^2}{2}}$$



$f(x) = x$, $\qquad g(x) = \sin x$.

$E[x] = ?$ $\qquad x\, p(x)\, dx$

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} = \sqrt{2\pi}$$

Z = Normalizing constant.

$$Z = \int \underline{P(x)dx} = \int_{-\infty}^{\infty} \underbrace{\left(3e^{-\frac{x^2}{2}} + e^{-\frac{(x-4)^2}{2}}\right)}dx$$

$$= 3\underbrace{\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}}\, dx + \int_{-\infty}^{\infty} e^{-\frac{(x-4)^2}{2}} dx$$

$\underline{(-\infty, \infty)}$

$$3\sqrt{2\pi} \quad + \quad \sqrt{2\pi}$$

$$E[x]. \quad = 4\sqrt{2\pi} \approx 10.026.$$

~~Law of unconscious statistician~~

$$E[\underbrace{\sin(x)}_{g(x)}] = \frac{1}{Z}\int_{-\infty}^{\infty} \underbrace{\sin(x)}_{g(x).} P(x)\,dx.$$

$$Z = \int_{-\infty}^{\infty} P(x)dx \approx 10.026.$$

$$\int_{-\infty}^{\infty} \sin x \left(3e^{-\frac{x^2}{2}} + e^{-\frac{(x-4)^2}{2}}\right) dx$$

$$\downarrow$$

$$\underbrace{3\int_{-\infty}^{\infty} \sin(x)\, e^{-\frac{x^2}{2}}\, dx}. + \int_{-\infty}^{\infty} \sin(x)\, e^{-\frac{(x-4)^2}{2}}\, dx$$

Imaginary part of the Fourier transform of $e^{-\frac{x^2}{2}}$ evaluated at $k=1$

$$\mathcal{F}\left\{e^{-\frac{x^2}{2}}\right\}(k) = \sqrt{2\pi}\, e^{-\frac{k^2}{2}}$$

$$\int_{-\infty}^{\infty} \sin(x)\, e^{-\frac{x^2}{2}}\, dx = \operatorname{Im}\left(\sqrt{2\pi}\, e^{-\frac{1^2}{2}}\right) = 0.$$

odd function   even function.

$$\int_{-\infty}^{\infty} \sin(x) \quad e^{-\frac{(x-4)^2}{2}} \, dx$$

$y = x - 4.$ $\qquad\qquad x = y + 4.$

$$\int_{-\infty}^{\infty} \sin(y+4) \; e^{-\frac{y^2}{2}} dy$$

$$\downarrow$$

$$\sin(A+B) = \sin A \cos B + \cos A \sin B.$$

$$\sin(y+4) = \sin(y)\cos(4) + \cos(y)\sin(4)$$

$$\cos(4) \int_{-\infty}^{\infty} \sin(y) \; e^{-\frac{y^2}{2}} dy \; + \sin(4) \int_{-\infty}^{\infty} \cos(y) \, e^{-\frac{y^2}{2}} dy$$

0

Real part of the Fourier transform

$e^{-\frac{y^2}{2}}$ evaluated at $k=1$

$$\int_{-\infty}^{\infty} \cos(y) \cdot e^{-\frac{y^2}{2}} dy = \mathrm{Re}\left(\sqrt{2\pi} \; e^{-\frac{1^2}{2}}\right) = \sqrt{2\pi}\, e^{-1/2}$$

$$\int_{-\infty}^{\infty} \sin(x)\, e^{-\frac{(x-4)^2}{2}}\, dx \doteq \sin(4)\,\sqrt{2\pi}\, e^{-1/2}$$

$$E[\sin(x)] = \frac{1}{Z}\,\sin(4)\,\sqrt{2\pi}\, e^{-1/2}$$

$\frac{1}{10.02}$ $\quad 0.75 \quad 2.50 \quad 0.60$

$\Rightarrow -1.15$

$(0 - 1)$

Coin toss problem.

$p(H) = 0.6$

$p(T) = 0.4$

$p(H) + p(T) = 1.$

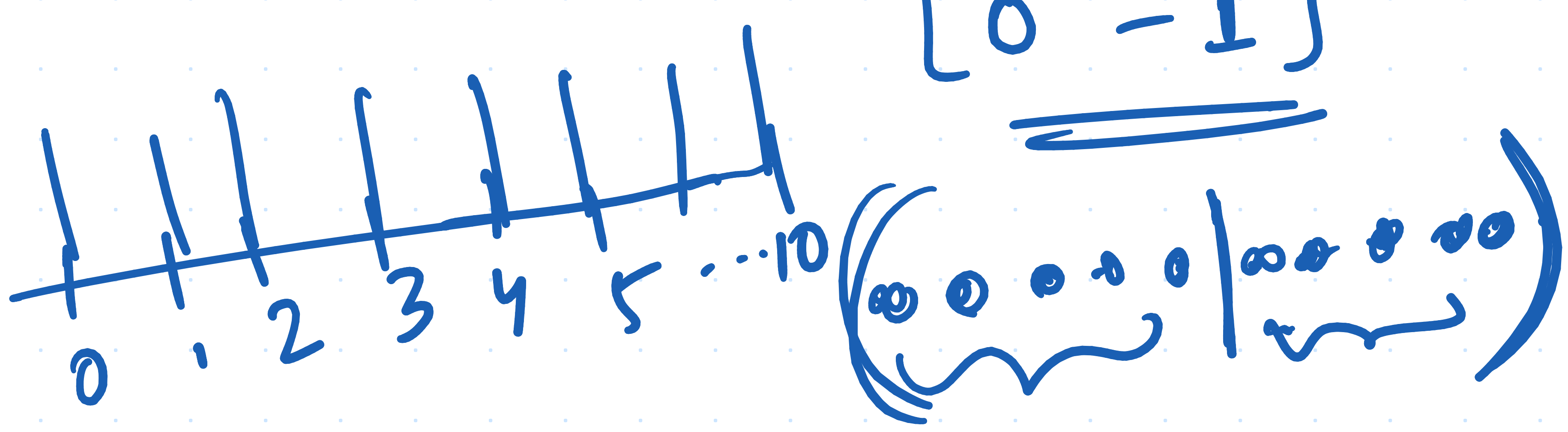
Biased coin.

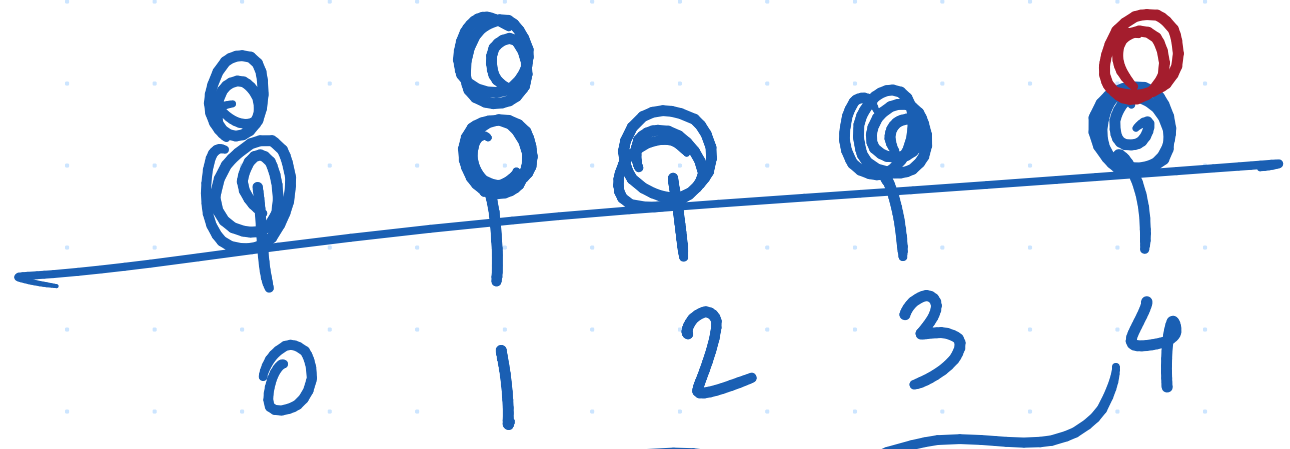
programatically simulate.

$[0 - 1]$

$(0 - 1)$

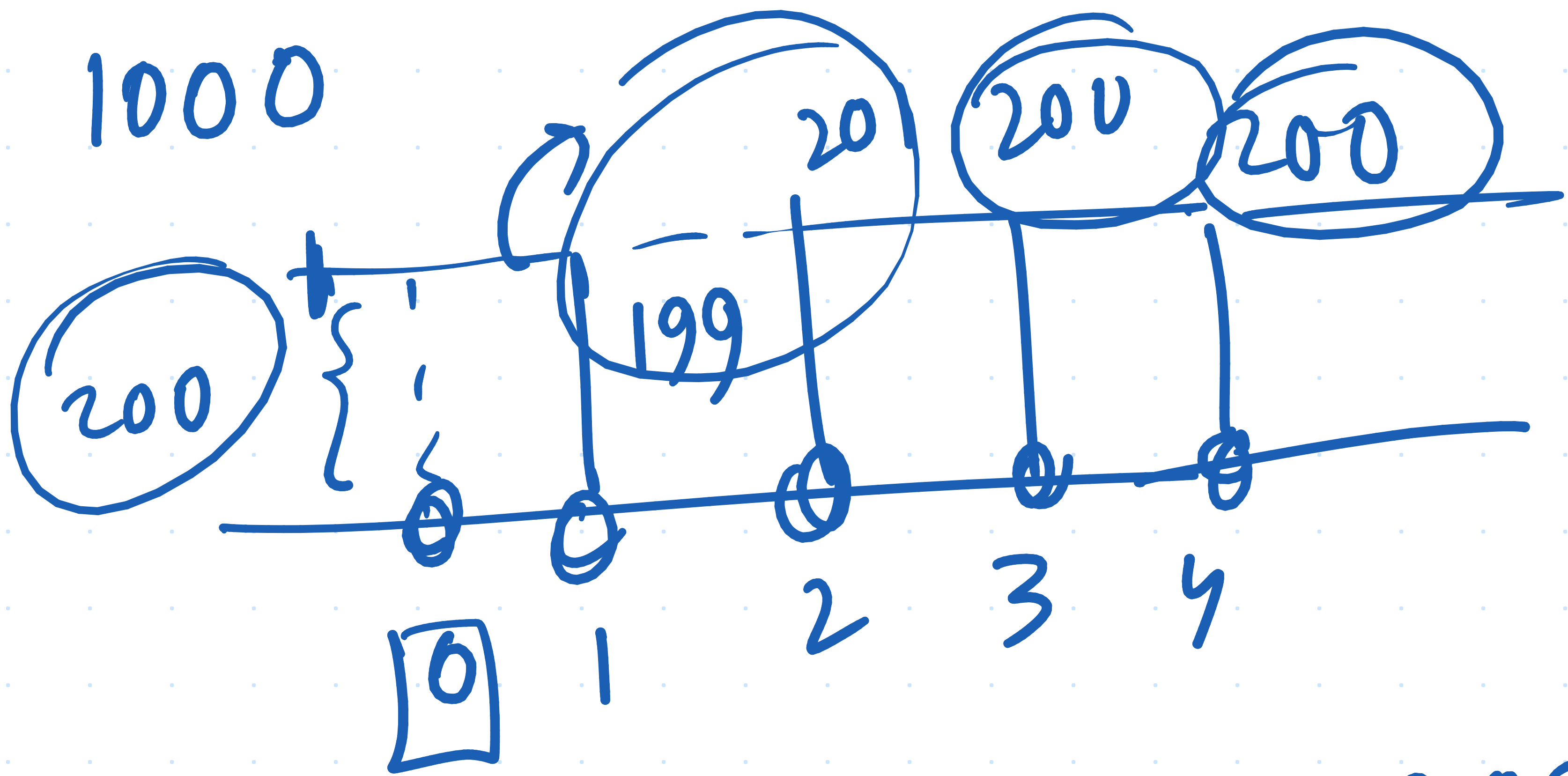
$(0 - 10)$



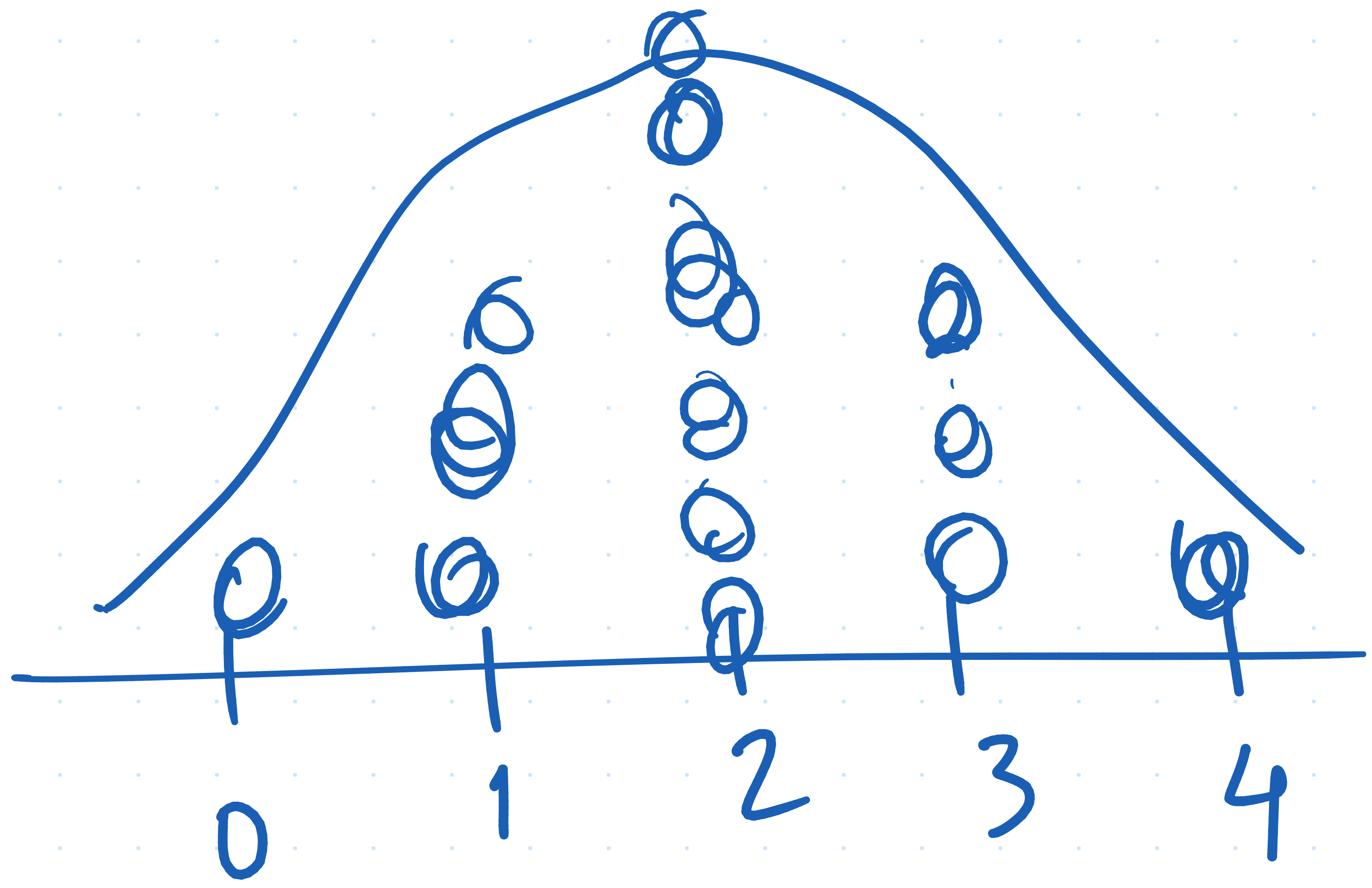


Uniform Random

$(0,5) \approx 1000$



$100{,}000 / 5 = 20000$



mean = 2.

Rejection Sampling :-

For a $cQ(x) > P(x) \; \forall x \in \mathbb{R}$, the rejection sampling :-

- Sample $x_i \sim Q(x)$
- Sample $k_i \sim \text{Uniform } [0, cQ(x_i)]$
- Accept x_i if $k_i < P(x_i)$

Accepted x_i is automatically sampled from normalized $P(x)$.

$P(x) \rightarrow$ difficult to sample.

$Q(x) \rightarrow$ simpler proposal distribution.

Reject samples from $Q(x)$ that are unlikely under $P(x)$ and keep those that are likely under $P(x)$

Probability of accepting a sample x_i is proportional to $\frac{P(x_i)}{c\,Q(x_i)}$ → this ensures that the accepted samples are distributed according to $P(x)$.

$P = 0.6$, head
$P = 0.4$, tail.

} brand.

coin.

$(- - - - -) \rightarrow 10.$

$(- - - - -) \rightarrow 1000$

$\boxed{(- - - - -) \rightarrow 10{,}000}$

Importance Sampling

Technique for estimating the expectation of a function $f(x)$ under a target dist $P(x)$, using samples drawn from a proposal dist $Q(x)$

Sampling from $P(x)$ is difficult while sampling from $Q(x)$ is easy

$$\mathbb{E}_{x \sim p}[f(x)] = \int f(x)\, \underbrace{p(x)}\, dx$$

using proposal dist $Q(x)$

difficult to sample from.

$$\mathbb{E}_{x \sim p}[f(x)] = \int f(x) \frac{P(x)}{Q(x)} Q(x)\, dx$$

$$= \mathbb{E}_{x \sim Q}\left[f(x) \frac{P(x)}{Q(x)} \right]$$

We can estimate the expectations by sampling from $Q(x)$ and re-weighting the samples using the ratio $\frac{P(x)}{Q(x)}$.

Normalized Case :-

$$E_{x \sim p}[f(x)] \approx \frac{1}{n} \sum_{i=1}^{n} f(x_i) \frac{P(x_i)}{Q(x_i)}.$$

$$x_i \sim Q(x).$$

Un-normalized Case :-

$P(x) \rightarrow$ unnormalized $\qquad P(x) = \frac{1}{Z} \tilde{P}(x).$

$\tilde{P}(x) \rightarrow$ unnormalized density and $Z \rightarrow$ normalizing constant.

$$\mathbb{E}_{x \sim p}[f(x)] \approx \frac{\sum_{i=1}^{n} f(x_i) \frac{\tilde{P}(x_i)}{Q(x_i)}}{\sum_{i=1}^{n} \frac{\tilde{P}(x_i)}{Q(x_i)}}$$

$\rightarrow Z$ - normalizing constant.

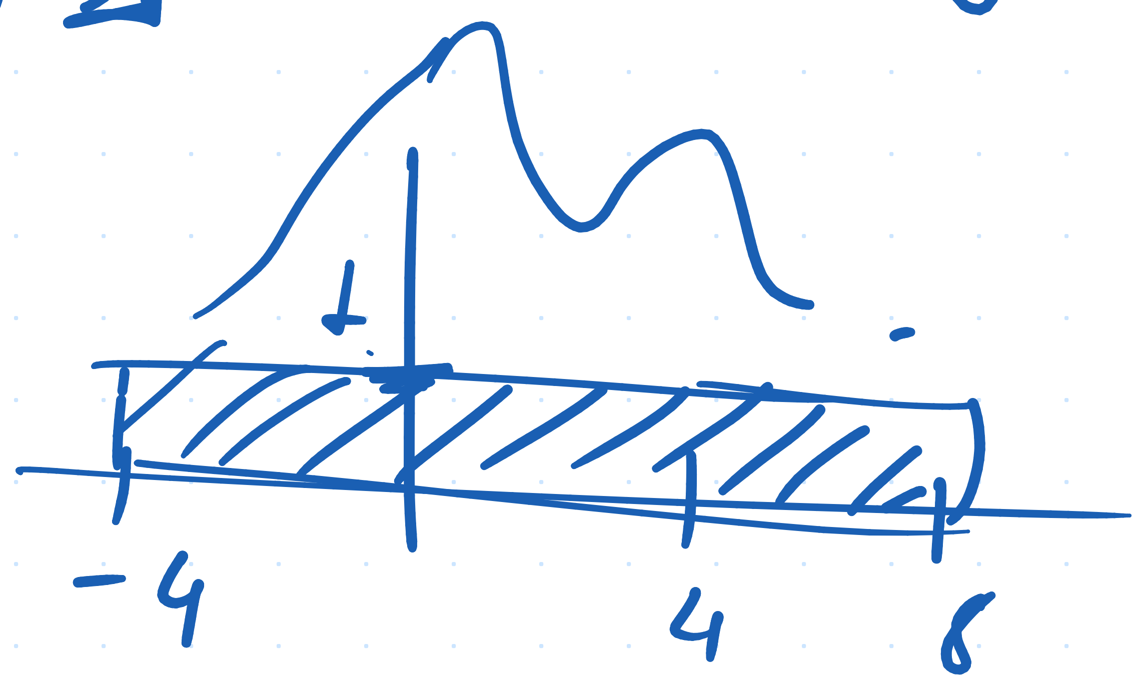
$$P(x) = 3e^{-x^2/2} + e^{-\frac{(x-4)^2}{2}}$$

$\downarrow$

Sampling is not easy

$Q(x) =$ Uniform dist. over $[-4, 8]$ $\rightarrow$ sampling is easy.

$$Q(x) = \frac{1}{12}, \quad x \in [-4, 8]$$



$x_1, x_2, \ldots, x_n$ samples drawn from Uniform distribution & compute its weight,

$$w_i = \frac{P(x_i)}{Q(x_i)}$$

Normalized :-

$$\mathbb{E}_{x\sim p}[f(x)] \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i) w_i$$

Un-normalized :-

$$\mathbb{E}_{x\sim p}[f(x)] \approx \frac{\sum_{i=1}^{n} f(x_i) w_i}{\sum_{i=1}^{n} w_i}.$$

Gibbs Sampling

It is a Markov chain Monte Carlo (MCMC) method to generate samples from a multivariate prob. dist. $P(X)$, where $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

useful, when sampling directly from $P(x_1, x_2)$ is difficult but sampling from $P(x_1 | x_2)$ and $P(x_2 | x_1)$ is feasible

Algorithm :-

→ start with an initial value $X^0 = \begin{bmatrix} x_1^0 \\ x_2^0 \end{bmatrix}$

→ For each iteration t :-

conditional dist.

- Sample x_1^{t+1} from $\underline{P(x_1 | x_2^t)}$
- Sample x_2^{t+1} from $P(x_2 | x_1^{t+1})$
- Update $X^{t+1} = \begin{bmatrix} x_1^{t+1} \\ x_2^{t+1} \end{bmatrix}$

Remove the first few samples as burn-out.

Gibbs Sampling constructs a Markov chain where each step updates one variable at a time, conditioned on current & other variables.

$P(x_1, x_2)$ → bivariate normal dist.

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \text{Covariance Matrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}$$

Conditional Dist :-

$P(x_1|x_2)$ → normal dist. with mean

$$\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$$

variance ⇒ $\sigma_1^2(1-\rho^2)$

$P(x_2|x_1)$ → normal dist. with mean,

$$\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$$

variance ⇒ $\sigma_2^2(1-\rho^2)$

Gibbs Sampling requires the ability to sample from the conditional distributions $P(x_1|x_2)$ & $P(x_2|x_1)$

$$P(x) = P(x_1, x_2) = \frac{1}{\sqrt{|2\pi\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 1 & b \\ b & 1 \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Conditional probabilities are given by:-

$$P(x_1|x_2) = N(bx_2, 1-b^2)$$

$$P(x_2|x_1) = N(bx_1, 1-b^2)$$

General formula for conditional distribution of Bivariate Normal.

$$\boxed{p(x_1|x_2) = N\left(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho(x_2-\mu_2), \sigma_1^2(1-\rho^2)\right)}$$

μ_1, μ_2 = means of x_1 & x_2

σ_1, σ_2 = std. of x_1 & x_2

ρ = correlation coefficient.

$$\mu_1 = \mu_2 = 0 \quad , \sigma_1 = \sigma_2 = 1, \quad \rho = b = 0.8$$

$$p(x_1 | x_2) = N(0.8 x_2 , 1 - 0.64) = N(0.8 x_2 \, 0.36)$$

For this particular example the Gibbs Sampling algorithm :-

Initialize

$$X^0 = \begin{bmatrix} x_1^0 \\ x_2^0 \end{bmatrix}$$

Iterate

These are the Normal dist, hence sampling is very easy

Sample $x_1^{t+1} \sim p(x_1 | x_2^t) = N(0.8 x_2^t, 0.36)$

Sample x_2^{t+1} from $p(x_2 | x_1^{t+1}) = N(0.8 x_1^{t+1}, 0.36)$

Update $X^{t+1} = \begin{bmatrix} x_1^{t+1} \\ x_2^{t+1} \end{bmatrix}$