2.3 Exercise

1. Rejection Sampling

a) Why is sufficient knowing $p(x)$ up to a normalizing constant.

$p(x) = \frac{\tilde{p}(x)}{Z}$ $\Rightarrow$ $\boxed{\tilde{p}(x) = Z \cdot p(x) \text{ is what we know.}}$

- Since the sampling is only dependent on the proposal distribution ~~and we accept with~~

and we accept with

$$u < \frac{\tilde{p}(x^*)}{\tilde{M} q(x^*)} = \frac{Z\,p(x^*)}{\tilde{M}\; q(x^*)} = \frac{p(x^*)}{q(x^*)\, M} \quad \text{with } M = \frac{\tilde{M}}{Z}$$

there always exist another $\tilde{M} = M \cdot Z$ which if used results in the same result.

- the main problem when knowing $p(x^*)$ up to a constant is that sometimes a much bigger M is required in order to make sure the condition $\tilde{p}(x) \leq M q(x)$ is fulfilled which results in a more conservative choice of M and so higher probability of rejecting the samples

b) $$p\left(u < \frac{\tilde{p}(x)}{M q(x)}\right) = p\left(u < \frac{Z\,p(x)}{M\, q(x)}\right) = P(\text{Accept}) = ?$$

- Conditionened on x: $$P(\text{Accept} \mid X = x) = P\left(u < \frac{Z\,p(x)}{M\, q(x^*)} \,\middle|\, X = x\right) = \frac{Z\,p(x)}{M\,q(x^*)}$$

- Unconditioning and marginalizing: $$P(\text{Accept}) = P\left(u < \frac{Z\,p(x)}{M q(x)}\right) = \int P\left(u < \frac{Z p(x)}{M q(x)} \,\middle|\, X = x\right) \cdot q(x)\, dv$$

since samples are distributed according to $q(x)$.

$$= \int \frac{Z\,p(x)}{M\,\cancel{q(x)}}\, \cancel{q(x)}\, dx$$

$$\boxed{P(\text{Accept}) = \frac{Z}{M}.}$$

2) Importance Sampling

a) Show that $\mathbb{E}_{p(x)}[f(x)] \approx \frac{1}{L}\sum_{l=1}^{L} f(x^l)w(x^l)$, where $x^l \sim q(x)$

We know that $\mathbb{E}_{p(x)}[f(x)] = \int_x f(x)\,p(x)\,dx = \int_x f(x)\,w(x)\,q(x)\,dx$ with $w(x) = \frac{p(x)}{q(x)}$

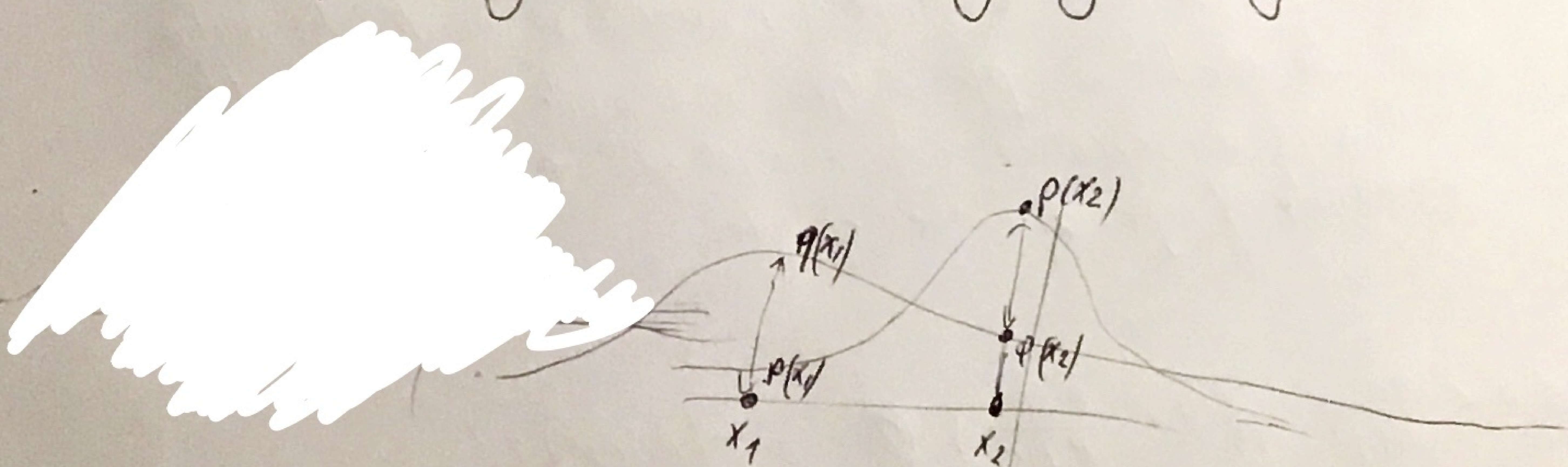
$$\approx \frac{1}{L}\sum_{l=1}^{L} f(x^l)\,w(x^l) \quad \text{where } x^l \sim q(x)$$

x^l sampled from $q(x)$

which by law of large nr converges to $\mathbb{E}_{p(x)}[f(x)]$ as $N \to \infty$

b) Importance weight are needed to set the importance of each sample relative to the target distribution we want to sample

~~So certain samples are given more importance if they come from~~



For example we often sample aroun x_1 ~~but~~ since $q(x_1)$ is high but those samples contribute not that much since the importance weight $\boxed{w(x_1) = \frac{p(x_1)}{q(x_1)} \text{ is very small.}}$

In contrast to that it is less likely to draw samples around x_2 even if they
come ~~but this is taken in consideration by~~
come from high probability region in $p(x)$.

↳ this is taken in consideration through the high importance weight $w(x_2) = \frac{p(x_2)}{q(x_2)}$

3.2.1

1) Show that the transition Kernel

$$T(x^{i+1}/x^i) = q(x^{i+1}/x^i)\,A(x^i, x^{i+1}) + \delta_{x^i}(x^{i+1})\,r(x^i)$$

satisfies the the detailed balance condition.

Solution

there are two cases for the outcome of a MH-step. ① $x^{i+1} \neq x^i$ or ② $x^{i+1} = x^i$

a) • In the first case we have $x^{i+1} \neq x^i$ which can be achieved only with an accepted MH step, i.e. $\boxed{T(x^{i+1}/x^i) = A(x^{i+1}, x^i)\,q(x^{i+1}/x^i)}$

b) In the second case $x^{i+1} = x^i$ which happens throw an accepted step that luckily lands back at the same state or rejection step:

$$T(x^i/x^i) = A(x^i, x^i)\,q(x^i/x^i) + \int q(x^*/x^i)\left(1 - A(x^i, x^*)\right) dx^*$$

since $x_{i+1} = x_i$ it is trivial that the detailed balance hold

$$T(x^{i+1}/x^i)\,p(x^i) = T(x^i/x^{i+1})\,p(x^{i+1})$$

a) For the first case we have:

$$T(x^{i+1}/x^i) = \min\left(1, \frac{p(x^{i+1})\,q(x^i/x^{i+1})}{p(x^i)\,q(x^{i+1}/x^i)}\right) \cdot q(x^{i+1}/x^i)$$

$$= \frac{1}{p(x^i)} \underbrace{\min\left(p(x^i)\,q(x^{i+1}/x^i),\, p(x^{i+1})\,q(x^i/x^{i+1})\right)}_{\text{symmetrical relative to } x^i \leftrightarrow x^{i+1}}$$

$$T(x^{i+1}/x^i) \cdot p(x^i) = \min\left(p(x^i)\,q(x^{i+1}/x^i),\, p(x^{i+1})\,q(x^i/x^{i+1})\right)$$

$$= \frac{1}{p(x^{i+1})} \min\left(p(x^{i+1}) \cdot q(x^i/x^{i+1}),\, p(x^i)\,q(x^{i+1}/x^i)\right) \cdot p(x^{i+1})$$

$$= T(x^i/x^{i+1}) \cdot p(x^{i+1})$$

2) Why use an asymmetrical proposal distribution, i.e. why MH and not just Metropolis?

~~In contrast to MH, using a symmetrical~~

Metropolis is a special case of MH limited on symmetrical proposal distributions. As such it doesn't allow usage of unsymmetric proposal distributions which lead to better performance if some known information in prior is given.

3) For a model with PGM $p(y,x) = p(x)\,p(y|x)$. What would be a reasonable proposal distribution for the Independent sampler when we want sample from posterior $p(x|y)$?

A reasonable proposal would be $p(x)$.

Exercise 9.4.7

We have:

$$\tilde{w}_t^i = \frac{P(x_{0:t}^i|y_{1:t})}{\pi(x_{0:t}^i|y_{1:t})} = \frac{P(x_{0:t-1}^i|y_{1:t-1})\,P(x_t^i|x_{0:t-1}, y_{1:t})}{\pi(x_{0:t-1}|y_{1:t-1})\,\pi(x_t^i|x_{0:t-1}, y_{1:t})}$$

$$= w_{t-1}^i \cdot \frac{P(x_t^i|x_{t-1}^i, y_t)}{\pi(x_t^i|x_{0:t-1}, y_{1:t})}$$ Markov Property

$$= w_{t-1} \cdot \frac{P(x_t^i|x_{t-1}^i)\,P(y_t|x_t^i)}{\pi(x_t^i|x_{0:t-1}, y_{1:t})} \cdot \frac{1}{P(y_t|x_{t-1})}$$ ← Independent of x_t (constant)

$$w_t^i \propto w_{t-1}^i \frac{P(x_t^i|x_{t-1}^i)\,P(y_t|x_t^i)}{\pi(x_t^i|x_{0:t-1}, y_{1:t})}$$

Since $\tilde{w}_t^i \propto w_t^i$ and $\tilde{w}_{t-1}^i \propto w_{t-1}^i$ $\Rightarrow$ $$\boxed{\tilde{w}_t^i \propto \tilde{w}_{t-1}^i \frac{P(x_t^i|x_{t-1}^i)\,P(y_t|x_t^i)}{\pi(x_t^i|x_{0:t-1}, y_{1:t})}}$$