Normalized importance sampling.

Assume that given x^i, easy to compute $\tilde{P}(x^i) \propto P(x^i)$

$$P(x^i) = \frac{\tilde{P}(x^i)}{Z}$$ ← normalizer is hard to compute.

Applications

1) Undirected Graphical model.
2) BN and you need to sample a subset of variables conditioned on fixed values of others: eg $P(x_1, \ldots x_r \mid x_{r+1}, \ldots x_n \equiv \text{evidence})$.

$$E_p(f(x)) = \sum_{x \in \mathcal{X}} f(x)\, P(x)$$

$$= \sum_{x \in \mathcal{X}} f(x) \frac{\tilde{P}(x)}{Z} \frac{Q(x)}{Q(x)} = \frac{1}{Z} E_{Q(x)}\Big[f(x) \underbrace{\frac{\tilde{P}(x)}{Q(x)}}_{\tilde{w}(x)} \Big]$$

$$= \frac{1}{Z} E_{Q(x)}[\, f(x)\, \tilde{w}(x)]$$

Normalized importance sampling.

Given samples $S_Q^M \equiv \{x^1, x^2, \ldots x^M\}$. we approximate

$$E_P[f(x)] \approx \frac{1}{\boxed{Z}M} \sum_{i=1}^{M} f(x^i)\,\tilde{w}(x^i)$$

$$Z = \sum_{x \in X} \tilde{P}(x) = \sum_{x \in X} \tilde{P}(x) \frac{Q(x)}{Q(x)} = E_Q\left[\frac{\tilde{P}(x)}{Q(x)}\right] = E_Q[\tilde{w}(x)] \approx \frac{1}{M}\sum_{i=1}^{M} \tilde{w}(x^i)$$

$$E_P[f(x)] \approx \frac{\sum_{i=1}^{M} f(x^i)\,\tilde{w}(x^i)}{\sum_{i=1}^{M} \tilde{w}(x^i)} \leftarrow$$ estimate from normalized importance sampling.

Importance sampling: Choosing $Q(x)$

The choice of $Q(x)$ for which the expected square error of the estimate from the true expected value is minimum is when

$$Q(x) \propto |f(x)|P(x), \quad Q(x) > 0 \mathit{whenever} P(x) > 0. \tag{2}$$

$\frac{P(x)}{Q(x)}$

Normalized importance sampling is biased when M is small.

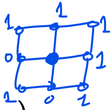
Designing a good $Q(x)$ for which the sampling is efficient is not always easy.

$P(x) = P(x_1)\,P(x_2 \mid Pa(x_2)) \ldots$

$Pa(x_i)$ are already known when x_i is sampled.

MCMC: Markov Chain Monte Carlo Sampling

When designing proposal distribution is not easy.
Easy to sample either



1. Conditional probability of one variable: $P(x_i|\mathbf{x}_{-i})$
2. Ratio of two probabilities: $P(\mathbf{x})/P(\mathbf{x}')$ — normalizer not required.

Useful when all else fails, guaranteed to converge to the optimal over infinite number of samples.

MCMC Sampling?

$Pr(\mathbf{x} = x_1 \ldots, x_n)$ where $x_i \in \{1, \ldots, m\}$ $x_i \in \mathbb{R}$
is intractable to sample from but is easy to evaluate.

- $P(x_i|\mathbf{x} - x_i) = P(x_i | \mathbf{x}_{-i})$ (Gibbs sampling) $\mathbf{x} - x_i \equiv \mathbf{x}_{-i}$
- $\frac{P(\mathbf{x})}{P(\mathbf{x}')}$ (Metropolis Hastings Sampling)

MCMC Sampling ~~Transaction~~ Transitions function

Designed by us much like the proposal distribution in Importance Sampling

$\mathcal{X}$ = Space of all x

$\sum_{x \in \mathcal{X}} T(x|x') = 1$

where

$T(x|x') \geq 0 \quad \forall\ x, x' \in \mathcal{X}$ and $|\mathcal{X}| = m^n$ (discrete case)

= infinite. (continuous case)

MCMC Sampling Algorithm

e.g. sample

1. Start with an initial sample $x^0 \equiv [\underbrace{1, 1, 1 \ldots 1}_{n \cdot \text{ones}}]$
2. For t = 1 to a large number L
   $x^{t+1} \sim T(\, x|x' = x^t \,)$
   $x^0 \rightarrow x^1 \rightarrow x^2 \rightarrow \cdots \rightarrow \boxed{x^L}$
3. Actually perform the sampling for t = L+1 to t = L+Mk
   $x^t \sim T(\, x|x^{t-1} \,)$
4. Return $x^{L+k}, x^{L+2k}, \ldots, x^{L+Mk}$ as samples

Gibbs Sampling

Gibbs Sampling is when $T(x|x')$ is defined as follows:

$T(x|x') = 0$ if x and x' differ in more than one co-ordinate.

$= \frac{1}{n}\sum_{i=1}^{n} P(x_i | x'_{-i})$ if $x = x'$

$= \frac{1}{n} \boxed{P(x_i | x'_{-i})}$ if $x \neq x'$ i.e $x_{-i} = x'_{-i}$

Example

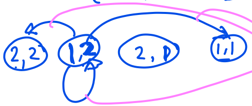
$n = 2 \quad m = 2 \quad x_i \in \{1, 2\}$.

$T(x | x' = [1, 2]) = 0$ if $x = [2, 1]$ ($\because$ we x & x' differ in 2 posi.)

$= \frac{1}{2} P(x_1 = 1 | x'_2 = 2) + \frac{1}{2} P(x_2 = 2 | x'_1 = 1)$ if $x = [1, 2]$

$\frac{1}{2} P(x_1 = 2 | x'_2 = 2)$ if $x = [2, 2]$

$\frac{1}{2} P(x_2 = 1 | x'_1 = 1)$ if $x = [1, 1]$



Example

multinomial distribution over all states.

$T(x|x^0)$

x^0 m^n

$t=0$ 1

$P_{t=0}(x) = 1$ if $x=x^0$
$= 0$ elsewhere.

$t=1$

x_0

$t=2$

$P_{t=1}(x) = T(x|x^0)$

$P_{t=2}(x) = \sum_{x'} T(x|x') T(x'|x^0)$

$= \sum_{x'} T(x|x') P_{t-1}(x')$

In general: $P_t(x) = \sum_{x'} T(x|x') P_{t-1}(x')$

$P_{t+1}(x) \approx P_t(x)$ for convergence as $t \to \infty$

$= \pi(x) \leftarrow$ Stationary Distribution

We are interested in $T(x|x')$ for which $P_{t+1}(x) = \sum_{x'} P_t(x')\, T(x|x')$

$t \to \infty$

$$\boxed{\pi(x) = \sum_{x'} \pi(x')\, T(x|x')}$$

$\pi(x) \equiv P(x) \leftarrow$ our distribution of interest.

has a unique solution for a given $T(x|x')$ and $\pi(x)$ should be reachable from any initial state x^0 via Markov walks using $T(x|x')$