A simple example in which importance sampling does poorly

Consider: target: $p(x) = \mathcal{N}(0, I)$

proposal: $q(x) = \mathcal{N}(0, (1+\varepsilon) I)$, $\varepsilon > 0$

$$\underbrace{\qquad\qquad\qquad\qquad}_{\text{n-dimensional Gaussians.}}$$

Note: · proposal is "good" approximation of target

↳ Gaussian ✓

↳ mean 0 ✓

↳ variance is off by a small amount, in each dimension.

· histograms of proposal & target look very similar

Compute weight and N_{eff} for this setup:

$$w(x) = \frac{p(x)}{q(x)} = \frac{(2\pi)^{-n/2} \exp(-\frac{1}{2} x^T x)}{(2\pi)^{-n/2} (1+\varepsilon)^{-n/2} \exp(-\frac{1}{2} \frac{1}{1+\varepsilon} x^T x)}$$

$$= (1+\varepsilon)^{n/2} \exp\left(-\tfrac{1}{2} x^T x \left(1 - \tfrac{1}{1+\varepsilon}\right)\right)$$

$$\boxed{w(x) = (1+\varepsilon)^{n/2} \exp\left(-\frac{1}{2} \frac{\varepsilon}{1+\varepsilon} x^T x\right)}$$

$$\boxed{w^2(x) = (1+\varepsilon)^{n} \exp\left(-\frac{\varepsilon}{1+\varepsilon} x^T x\right)}$$

$$E_q[w] = \int \frac{p(x)}{q(x)} q(x)\, dx = \int p(x)\, dx = 1$$

$$E_q[w^2] = \int (1+\varepsilon)^n \exp\left(-\frac{\varepsilon}{1+\varepsilon} x^T x\right) \cdot (2\pi)^{-n/2} (1+\varepsilon)^{-n/2} \exp\left(-\frac{1}{2} \frac{1}{1+\varepsilon} x^T x\right) dx$$

$$= (2\pi)^{-n/2} (1+\varepsilon)^{n/2} \int \exp\left(-\frac{1}{2} x^T x \left(\frac{1}{1+\varepsilon} + \frac{2\varepsilon}{1+\varepsilon}\right)\right) dx$$

$$= (2\pi)^{-n/2} (1+\varepsilon)^{n/2} \int \exp\left(-\frac{1}{2} x^T x \left(\frac{1+2\varepsilon}{1+\varepsilon}\right)\right) dx$$

$$= (2\pi)^{-n/2} (1+\varepsilon)^{n/2} \cdot (2\pi)^{n/2} \left(\frac{1+\varepsilon}{1+2\varepsilon}\right)^{n/2}$$

$$= \frac{(1+\varepsilon)^n}{(1+2\varepsilon)^{n/2}} = \underline{\left(\frac{1+\varepsilon}{\sqrt{1+2\varepsilon}}\right)^n}$$

$$\rho = \frac{E[\omega^2]}{E[\omega]^2} = \left(\frac{1+\varepsilon}{\sqrt{1+2\varepsilon}}\right)^n$$

$$N_{eff} = N_e/\rho \Rightarrow N_e = N_{eff} \cdot \rho = N_e \left(\frac{1+\varepsilon}{\sqrt{1+2\varepsilon}}\right)^n$$

<u>Result:</u>

$$\boxed{N_e \propto \left(\frac{1+\varepsilon}{\sqrt{1+2\varepsilon}}\right)^n}$$

→ Required ensemble size increases exponentially with dimension!

→ Importance sampling only useful if n is "small"

<u>But</u> $E_q[x] = E_p[x] = 0$

$var_q(x_i) = 1+\varepsilon \approx var_p(x_i) = 1$

$cov_q(x_i, x_j) = cov_p(x_i, x_j)$ if $i \neq j$

We seem to have a "good" proposal! That gets mean and variance almost right!

<u>Sidenote</u> $\frac{1+\varepsilon}{\sqrt{1+2\varepsilon}} > 1$

$\Leftrightarrow (1+\varepsilon)^2 > 1+2\varepsilon$

$1+2\varepsilon+\varepsilon^2 > 1+2\varepsilon$ ✓

What we have learned:

(1) Sampling seems difficult if dimension is large

(2) Indicators of sampling success might be misleading:

$N_{eff} \propto \exp(n)$ even though proposal is quite good.

(3) For example: our "analysis" neglects problem structure.

For diagonal covs & Gaussians, this problem is composed of n independent subproblems. This is not reflected in ρ, N_{eff}.

Current (hot) research topic

How can we describe the difficulties of sampling but taking (sparse/diagonal/banded) problem structure into account?