Notes for Standard Normal Example

March 4, 2024

1 Intro

This document shows the theoretical steps that we took to implement stochastic gradiant optimized adaptive importance sampling (OAIS) to generate sample from standard normal distribution. The calculation is baes on section 3.2 of the paper ”Convergence rates for optimised adaptive importance samplers by Akyildiz and Miguez, 2021”. The contents of this document are acting as supporting docuemnt for the Julia script stochgradOAIS.jl which implements the algorithm.

2 Normal Location Model

In this example, we work on implementing OAIS to simulate random samples from a standard normal distribution. Our target distribution (to sample from) is standard normal distribution $X \sim N(0, 1)$ where density is:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}x^2}$$

Our proposal distribution to aid in sampling is $N(\mu, 1)$ where μ is the unknown mean with density:

$$g(x, \mu) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(x-\mu)^2}$$

We start by trying to write the $g(x, \mu)$ density in the form of exponential family distributions:

$$h(x)e^{(\theta T(x) - A(\theta))}$$

where θ is the unknown parameter (in our example μ). The calculation follows:

$$
\begin{aligned}
g(x,\mu) &= \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(x^2+\mu^2-2\mu x)} \\
&= \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}x^2-\frac{1}{2}\mu^2+\mu x} \\
&= \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}x^2} e^{\mu x-\frac{1}{2}\mu^2}
\end{aligned}
$$

Therefore we can say that in our example $\theta T(X) = \mu x$, $A(\theta) = \frac{1}{2}\mu^2$, and $h(x) = e^{\frac{-1}{2}x^2}$. Now we compute formula 2.8 from the paper as follows:

$$
\begin{aligned}
\nabla\rho(\theta) &= E_{q_\theta}\left[(\nabla A(\theta) - T(X))\frac{f^2(x)}{g^2(x)}\right] \\
&= E_{q_\theta}\left[(\mu - x)\frac{f^2(x)}{g^2(x)}\right]
\end{aligned}
$$

An unbiased estimator for $\nabla\rho(\theta)$ based on a sample of size N is:

$$
\sim \frac{1}{N}\sum_{i=1}^{N}(\mu - x_i)\frac{f^2(x_i)}{g^2(x_i)}
$$

2.1 Objective Function

I would like to compare the objective function used by Akyildiz and Míguez (2021), ρ, with the $\hat{k}$ diagnostic we are working with. To this end, it will be helpful to have an analytic expression for $\rho = \mathbb{G}w^2$, and $w = f/g$ is the importance weight.

First, note that

$$
\begin{aligned}
w &:= \frac{f}{g_\mu} && (1) \\
&= \exp(\mu^2 - 2x\mu) && (2)
\end{aligned}
$$

Next, since $\rho = \mathbb{G}w^2$, we need only integrate $g \cdot w^2$. The result ends up being $\exp(\mu^2)$. As a brief sanity check, we observe that this expression for ρ is minimized at $\mu = 0$, corresponding to $\mathbb{G} = \mathbb{F}$, with a value of 1. This value of ρ implies that the $\mathbb{G}$-variance of w is 0, which is exactly what we would expect when the proposal equals the target (specifically, these weights equal 1). This all sounds plausible to me. Furthermore, differentiating our expression for ρ matches the gradient computed elsewhere using identities for exponential family distributions. I'm happy with this expression now.

3 Normal Dispersion Model

What about the family $\mathcal{G} := \{\phi(0, \sigma^2) : \sigma \in \mathbb{R}\}$? I.e. A Normal scale family with known mean (for convenience, set $\mu = 0$).

4 Gamma Distribution

Target distribution is Gamma(α, β) with density:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad x > 0$$

and proposal distribution is Exp(λ) with density:

$$g(x) = \lambda e^{-\lambda x} \qquad x > 0$$

We start by writing the proposal distribution in the form of exponential family distirbutions, i.e:

$$g(x) = h(x) e^{\theta T(X) - A(\theta)}$$

We can write the proposal distribution as follows:

$$g(x) = 1 \times e^{ln(\lambda) - \lambda x}$$

This concludes that:

$$h(x) = 1 \quad T(X) = -X \quad A(\lambda) = -ln(\lambda)$$

The gradient of effective sample size is:

$$\begin{aligned}
\nabla \rho(\lambda) &= E_g\left[(\nabla A(\lambda) - T(X)) \frac{f^2(x)}{g^2(x)}\right] \\
&= E_g\left[(X - \frac{1}{\lambda}) \frac{f^2(x)}{g^2(x)}\right] \\
&= \int_0^\infty (X - \frac{1}{\lambda}) \frac{f^2(x)}{g^2(x)} g(x) dx \\
&= \int_0^\infty (X - \frac{1}{\lambda}) \frac{f^2(x)}{g(x)} dx
\end{aligned}$$

$$\begin{aligned}
\frac{f^2(x)}{g(x)} &= \frac{\left(\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\right)^2}{\lambda e^{-\lambda x}} = \frac{\frac{\beta^{2\alpha}}{\Gamma^2(\alpha)} x^{2\alpha-2} e^{-2\beta x}}{\lambda e^{-\lambda x}} \\
&= \frac{\beta^{2\alpha}}{\Gamma^2(\alpha)} x^{(2\alpha-1)-1} e^{-2\beta x} \lambda^{-1} e^{\lambda x} \\
&= \frac{\beta^{2\alpha}}{\Gamma^2(\alpha)} \frac{1}{\lambda} x^{(2\alpha-1)-1} e^{-(2\beta-\lambda)x} \\
&= A(\alpha, \beta, \lambda) x^{(2\alpha-1)-1} e^{-(2\beta-\lambda)x}
\end{aligned}$$

where $A(\alpha, \beta, \lambda) = \frac{\beta^{2\alpha}}{\lambda \Gamma^2(\alpha)}$. Continue the calculation of $\nabla\rho(\lambda)$ as follows:

$$\begin{aligned}
\nabla\rho(\lambda) &= \int_0^\infty (X - \frac{1}{\lambda}) A(\alpha, \beta, \lambda) x^{(2\alpha-1)-1} e^{-(2\beta-\lambda)x} dx \\
&= A(\alpha, \beta, \lambda) \left[\int_0^\infty x^{(2\alpha-1)} e^{-(2\beta-\lambda)x} dx - \frac{1}{\lambda} \int_0^\infty x^{(2\alpha-1)} e^{-(2\beta-\lambda)x} dx \right] \\
&= A(\alpha, \beta, \lambda) \left[\frac{\Gamma(2\alpha)}{(2\beta-\lambda)^{2\alpha}} - \frac{1}{\lambda} \frac{\Gamma(2\alpha-1)}{(2\beta-\lambda)^{2\alpha-1}} \right] \\
&= \frac{\beta^{2\alpha}}{\lambda\Gamma^2(\alpha)} \left[\frac{\lambda\Gamma(2\alpha) - (2\beta-\lambda)\Gamma(2\alpha-1)}{\lambda(2\beta-\lambda)^{2\alpha}} \right] \\
&= \left(\frac{\beta}{2\beta-\lambda} \right)^{2\alpha} \left[\frac{\Gamma(2\alpha)}{\lambda\Gamma^2(\alpha)} - \frac{(2\beta-\lambda)}{\lambda^2} \frac{\Gamma(2\alpha-1)}{\Gamma^2(\alpha)} \right] \\
&= \left(\frac{\beta}{2\beta-\lambda} \right)^{2\alpha} \frac{1}{\lambda\Gamma^2(\alpha)} \left[\Gamma(2\alpha) - \frac{(2\beta-\lambda)\Gamma(2\alpha-1)}{\lambda} \right]
\end{aligned}$$

Provided $2\beta - \lambda \neq 0$, we can solve $\nabla\rho(\lambda) = 0$ to find λ_{opt} as:

$$\begin{aligned}
&\left(\frac{\beta}{2\beta-\lambda} \right)^{2\alpha} \frac{1}{\lambda\Gamma^2(\alpha)} \left[\Gamma(2\alpha) - \frac{(2\beta-\lambda)\Gamma(2\alpha-1)}{\lambda} \right] = 0 \\
&\Gamma(2\alpha) - \frac{(2\beta-\lambda)\Gamma(2\alpha-1)}{\lambda} = 0 \\
&\Gamma(2\alpha) = \frac{(2\beta-\lambda)\Gamma(2\alpha-1)}{\lambda} \\
&\frac{\Gamma(2\alpha)}{\Gamma(2\alpha-1)} = \frac{2\beta}{\lambda} - 1 \\
&\frac{\Gamma(2\alpha)}{\Gamma(2\alpha-1)} + 1 = \frac{2\beta}{\lambda} \\
&\frac{\Gamma(2\alpha-1) + \Gamma(2\alpha)}{\Gamma(2\alpha-1)} = \frac{2\beta}{\lambda} \\
&\Rightarrow \lambda_{opt} = (2\beta) \frac{\Gamma(2\alpha-1)}{\Gamma(2\alpha-1) + \Gamma(2\alpha)}
\end{aligned}$$

and since $\Gamma(2\alpha) = (2\alpha-1)\Gamma(2\alpha-1)$ we can write:

$$\begin{aligned}
\lambda_{opt} &= (2\beta) \frac{\Gamma(2\alpha-1)}{\Gamma(2\alpha-1) + (2\alpha-1)\Gamma(2\alpha-1)} \\
&= (2\beta) \frac{1}{1 + 2\alpha - 1} = \frac{2\beta}{2\alpha} = \frac{\beta}{\alpha}
\end{aligned}$$

Therefore $\lambda_{opt} = \frac{\beta}{\alpha}$

References

Akyildiz, O. D. and Míguez, J. (2021). Convergence rates for optimized adaptive importance samplers. *Statistics and Computing*, 31(2).