Scaling the Trace Norm Distribution Partition Function*

Jason D. M. Rennie
jrennie@gmail.com

February 4, 2006

Abstract

In [1], we proposed a sampling scheme to estimate the partition function of the Trace Norm Distribution. However, for moderately large values of m (e.g. $m = 100$) the individually sampled values are too large for standard computer floating point types. We describe a simple method to scale the values to make computation easier.

1 Introduction

An important quantity in the trace norm distribution partition function is [1],

$$J = \frac{1}{m!}\int_0^\infty \cdots \int_0^\infty \prod_{i=1}^m e^{-\sigma_i/\theta}\sigma_i^{n-m}\prod_{i<j}|\sigma_i^2-\sigma_j^2|d\sigma_m \ldots d\sigma_2 d\sigma_1. \tag{1}$$

Using the $e^{-\sigma_i/\theta}\sigma_i^{n-m}$ terms to establish a Gamma distribution leaves us with a product of $|\sigma_i^2-\sigma_j^2|$ terms as the quantity over which we are taking an expectation. The drawback is that the product of $|\sigma_i^2-\sigma_j^2|$ terms can easily exceed the maximum values allowed for standard floating type values. To avoid this, we propose two different ways of scaling the difference of square terms.

The first is to scale the difference-of-square terms, $(\sigma_i\sigma_j)^r$,

$$J = \frac{1}{m!}\int_0^\infty \cdots \int_0^\infty \prod_{i=1}^m e^{-\sigma_i/\theta}\sigma_i^{n-m+r(m-1)}\prod_{i<j}\frac{|\sigma_i^2-\sigma_j^2|}{(\sigma_i\sigma_j)^r}d\sigma_m \ldots d\sigma_2 d\sigma_1, \tag{2}$$

where $r \in [0,2]$. Define $\alpha \equiv n-m+1+r(m-1)$. To prepare this for a sampling approximation, we add the necessary constants to give us a product of Gamma distributions,

$$J = \frac{\Gamma^m(\alpha)\theta^{m\alpha}}{m!}\int_0^\infty \cdots \int_0^\infty \prod_{i=1}^m g(\sigma_i|\alpha,\theta)\prod_{i<j}\frac{|\sigma_i^2-\sigma_j^2|}{(\sigma_i\sigma_j)^r}d\sigma_m \ldots d\sigma_2 d\sigma_1, \tag{3}$$

*Joint work with John Barnett and Tommi Jaakkola.

where $g(x|\alpha,\theta) = \frac{x^{\alpha-1}e^{-x/\theta}}{\Gamma(\alpha)\theta^\alpha}$ is the Gamma distribution, and $\Gamma^m(\alpha) \equiv [\Gamma(\alpha)]^m$. The sampling scheme for estimating J involves sampling $\sigma_1, \ldots, \sigma_m$ and taking a sample average of of $\prod_{i<j} \frac{|\sigma_i^2 - \sigma_j^2|}{(\sigma_i \sigma_j)^r}$ values.

The second method is to rescale our sampling distribution to unit mean. Define $\alpha \equiv n - m + 1$. Note that the Gamma distribution, $g(x|\alpha,\theta)$, has mean $\alpha\theta$. Rearranging terms, we get

$$J = \frac{\Gamma^m(\alpha)}{m!\alpha^{m\alpha}} \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^m g(\sigma_i|\alpha, 1/\alpha) e^{-\sigma_i(1/\theta - \alpha)} \prod_{i<j} |\sigma_i^2 - \sigma_j^2| d\sigma_m \ldots d\sigma_2 d\sigma_1. \tag{4}$$

$g(\sigma_i|\alpha, 1/\alpha)$ is our new sampling distribution. We average over the remaining terms inside the integral to achieve our estimate of J. One undesirable effect of this rescaling is that we introduce exponential terms outside of the sampling distribution. When $n >> m$ and $n - m >> 1/\theta$ the samples from the Gamma distribution will be too small and will yield a biased estimate.

References

[1] J. D. M. Rennie. Computing the trace norm distribution via sampling. http://people.csail.mit.edu/~jrennie/writing, January 2006.