Estimating the Log-Partition Function of the Trace Norm Distribution*

Jason D. M. Rennie
jrennie@gmail.com

February 2, 2006

Abstract

The partition function of the Trace Norm Distribution quickly becomes too large a number for floating point standard. So, we must instead calculate the logarithm of the partition function. We discuss methods for estimating the logarithm of the partition function.

1 Introduction

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

$$J \equiv \frac{1}{m!}\int_0^\infty \cdots \int_0^\infty \prod_{i=1}^m e^{-\lambda\sigma_i}\sigma_i^{n-m}\prod_{i<j}|\sigma_i^2-\sigma_j^2|d\sigma_m \dots d\sigma_2 d\sigma_1. \quad (1)$$

We have discussed a method for estimating this quantity via sampling. In empirical tests, this sampling method seems to be highly effective for tall, skinny matrices (rows >> columns, or vice versa). But, this quantity quickly grows too large for standard computer floating point types. So, we would like to estimate $\log J$ instead. Also, if we can estimate $\log J$, we can factor a constant out of J so that the computations do not overload floating point types.

The gamma distribution pdf is $g(x|\alpha,\theta) = \frac{x^{\alpha-1}e^{-x/\theta}}{\Gamma(\alpha)\theta^\alpha}$. Let $\alpha = n-m+1$ and $\theta \equiv 1/\lambda$. We can rewrite the above integral to include g,

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

*Joint work with John Barnett and Tommi Jaakkola.

where $\Gamma^m(\alpha) \equiv [\Gamma(\alpha)]^m$. Since $\log(x)$ is a convex function,

$$\log J \geq \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^m g(\sigma_i|\alpha, \theta) \Bigg[m \log \Gamma(\alpha) + m\alpha \log \theta \\ + \sum_{i<j} \log |\sigma_i^2 - \sigma_j^2| \Bigg] - \log m!. \quad (3)$$

Thus, we can use the sampling scheme from [1] to estimate a lower bound on $\log J$. We can either use this as our estimate of $\log J$, or use this to scale values in the original sampling procedure to avoid numerical difficulties.

References

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