Computing the Trace Norm Distribution via Sampling*

Jason D. M. Rennie
jrennie@gmail.com

January 30, 2006†

Abstract

The Trace Norm Distribution is a distribution over matrices such that the negative log-likelihood is proportional to the trace norm of the matrix. The partition function of this distribution is intractible to compute exactly for large matrices, so we must use approximation techniques. We discuss a sampling scheme to approximate the integral.

1 Introduction

The trace norm distribution of a matrix, $X \in \mathbb{R}^{n \times m}$ (we assume $n > m$) is defined as

$$P_\lambda(X) = \frac{1}{Z_\lambda} \exp(-\lambda \|X\|_{\mathrm{tr}}), \tag{1}$$

where $\|X\|_{\mathrm{tr}}$ is the trace norm of X and Z_λ is the partition function (aka normalization constant). A change of variables to a singular value decomposition (SVD) factorization gives us

$$Z_\lambda = \frac{1}{2^m} \int \prod_{i=1}^{m} e^{-\lambda \sigma_i} \sigma_i^{n-m} \prod_{i<j} (\sigma_i^2 - \sigma_j^2) d\Sigma^\wedge (H^T dU)^\wedge (V^T dV)^\wedge, \tag{2}$$

where $H \in \mathbb{R}^{n \times n}$ is an orthogonal matrix with first m columns identical to U [1]. As discussed in [2], the singular value integral is intractible for large matrices. Hence, we must turn to approximation techniques. In the next section, we discuss a sampling scheme.

---

*Joint work with John Barnett and Tommi Jaakkola.

†Updated February 1, 2006.

2 Approximation via Sampling

We are concerned with calculating the singular value integral from above. Let

$$J \equiv \int_0^\infty \int_0^{\sigma_2} \cdots \int_0^{\sigma_{m-1}} \prod_{i=1}^m e^{-\lambda\sigma_i} \sigma_i^{n-m} \prod_{i<j} (\sigma_i^2 - \sigma_j^2) d\sigma_m \dots d\sigma_2 d\sigma_1. \quad (3)$$

Since the set of matrices with duplicate singular values make up a zero-measure set,

$$J = \frac{1}{m!} \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^m e^{-\lambda\sigma_i} \sigma_i^{n-m} \prod_{i<j} |\sigma_i^2 - \sigma_j^2| d\sigma_m \dots d\sigma_2 d\sigma_1. \quad (4)$$

This form suggests a sampling scheme:

- for $i = 1$ to k:
  - Sample $\{\sigma_1, \dots, \sigma_m\}$ iid from exponential(λ)
  - Let $z_i = \prod_{i<j} |\sigma_i^2 - \sigma_j^2| \prod_{i=1}^m \sigma_i^{n-m}$.

Then, $J \approx \frac{1}{km!} \sum_{i=1}^k z_i$, with $\lim_{k\to\infty} \frac{1}{km!} \sum_{i=1}^k z_i = J$.

We can also incorporate the product of singular value terms into the sampling distribution. The gamma distribution pdf is $g(x|a,\theta) = \frac{x^{a-1}e^{-x/\theta}}{\Gamma(a)\theta^a}$; we can write our integral in terms of a product of gamma pdf's. Let $\alpha \equiv n - m + 1$ and $\theta \equiv 1/\lambda$. Then,

$$J = \frac{\Gamma^m(\alpha)\theta^{m\alpha}}{m!} \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^m g(\sigma_i|\alpha,\theta) \prod_{i<j} |\sigma_i^2 - \sigma_j^2| d\sigma_m \dots d\sigma_2 d\sigma_1, \quad (5)$$

where $\Gamma^m(\alpha) \equiv [\Gamma(\alpha)]^m$. The updated sampling scheme is similar to above except that $z_i = \prod_{i<j} |\sigma_i^2 - \sigma_j^2|$ and $\{\sigma_1, \dots, \sigma_m\}$ are sampled iid from $g(\sigma_i|\alpha,\theta)$.

References

[1] A. Edelman. Volumes and integration. http://web.mit.edu/18.325/www/handouts.html, March 2005. 18.325 Class Notes: Finite Random Matrix Theory, Handout #4.

[2] J. D. M. Rennie. Approximating the trace norm distribution partition function. http://people.csail.mit.edu/~jrennie/writing, January 2006.