Title: **On The Distribution of the Largest Principal Component**

Author(s): **Iain Johnstone**

Technical Report number (Dept. of Statistics, Stanford Univ.): **2000-27**

Date: **August 2000**

Abstract:

Let $x_{(1)}$ denote square of the largest singular value of an $n \times p$ matrix X, all of whose entries are independent standard Gaussian variates. Equivalently, $x_{(1)}$ is the largest principal component of the covariance matrix $X'X$, or the largest eigenvalue of a p variate Wishart distribution on n degrees of freedom with identity covariance.

Consider the limit of large p and n with $n/p = \gamma \geq 1$. When centered by $\mu_p = (\sqrt{n-1} + \sqrt{[b]p})^2$ and scaled by $\sigma_p = (\sqrt{n-1} + \sqrt{[b]p})(1/\sqrt{n-1} + 1/\sqrt{[b]p})^{1/3}$ the distribution of $x_{(1)}$ approaches the Tracy-Widom law of order 1, which is defined in terms of the Painlevé II differential equation, and can be numerically evaluated and tabulated in software. Simulations show the approximation to be informative for n and p as small as 5.

The limit is derived via a corresponding result for *complex* Wishart matrices using methods from random matrix theory. The result suggests that some aspects of large p multivariate distribution theory may be easier to apply in practice than their fixed p counterparts.