Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 65(1) 2011 15-20

Data Matrices, Entropy and Applications*

Kalyan B. Sinha
Bhatnagar Fellow of CSIR;
Dean of Facutly, J.N. Centre for Advanced Scientific Research and
Distinguished Associate, Indian Institute of Science, Bangalore

SUMMARY

Data matrices with rows which are independent samples from some population were considered by Fisher and Wishart in the 1930's. More, recently the abstract sets of such random matrices have been studied at length and their limit theorems proven. Some of these results along with the associated entropies will be discussed and their applications in modern communication theory will be mentioned.

Keywords: Random matrices, Free probability, Entropy.

1. INTRODUCTION

Statisticians often study the fluctuation theory of correlations in which they have to analyse high-dimensional data by looking at the eigenvalue (or more precisely singular values) of (random) correlation matrices. For example (taken from Diaconis (2003)), consider the 100×5 matrix $\mathbf{X}$ ($\mathbf{X}_{ij} \equiv$ score of the ith student in jth examination; $1 \leq i \leq 100$; $1 \leq j \leq 5$) and note that the norm one vector α^* which maximizes the variance of the numbers $\{\alpha.\mathbf{X}_i\}_{i=1}^{100}$ is the first principal component and the vector α^{**} maximizing the variance subject to normalisation and orthogonality to α^* is the second one and so on. The α^* is approximately the average of the five scores and α^{**} is approximately the difference between the averages of the first two and of the last three tests, and so on. What about the question of the *stability of the principal components* under random fluctuations in the data X_{ij}? This is the first question in *random matrix theory*: the fluctuations of eigenvalues or singular values of random matrices. Wishart (1928) and Fisher (1939) studied the distribution of the principal components when the entries of the $n \times p$ data matrix $\mathbf{X}$ are i.i.d. Gaussian; while more recently Marcenko and Pastur (1967) considered the empirical distribution of the whole collection of the singular values of such matrices i.e. showed that

p^{-1} {# eigenvalues $\leq nt$} $\rightarrow$ an absolutely continuous function $G(t)$; when n and $p \uparrow \infty$ such that $n/p \rightarrow \gamma > 0$. There are many other applications studied in the literature. But here I shall limit myself to a few of the earlier questions raised and refer the reader to the beautiful article/lecture by Diaconis (2003) for many more such results.

In order to focus our attention, let $\mathbf{H}$ be a $n \times n$ real symmetric matrix with independent and identical Gaussian $N(0; 1/n)$-entries. Note that we can consider the matrix family $\mathbf{H}_n$ as a "non-commutative random variable" with the expectation $\tau_n(\mathbf{H}_n) \equiv \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\,(H_{n,ii}) = \frac{1}{n}\mathbb{E}\,(\mathrm{tr}\mathbf{H}_n)$, where $\mathbb{E}$ is the classical expectation and

Dr. Rajendra Prasad Memorial Lecture delivered at the 64th Annual Conference of the Indian Society of Agricultural Statistics held at BCKV, Kalyani during 3-5 December 2010.
E-mail address : *kbs@jncasr.ac.in*

also that $\tau_n(\mathbf{H}_n^2) = 1$ (the choice of normalisation). Then one has a form of Wigner's theorem that

Theorem 1.1 $\tau_n(H_n^{2m}) \to \frac{1}{m+1}\binom{2m}{m}$, *Catalan* numbers, as $n \uparrow \infty$ which are the $2m^{\text{th}}$ moments of the semi-circle distribution density $w : w(x) = (2\pi)^{-1}(4 - x^2)^{1/2}$ for $|x| \le 2$ and $= 0$ otherwise.

In this sense, the semi-circle law plays the same role in free probability as that of the (stable) normal law of classical probability theory.

Next, we look at real symmetric matrices, dropping the Gaussian assumption. Let **H** be $n \times n$ random matrix with i.i.d. entries with finite moments of all orders, and let $\lambda_1(\mathbf{H}); \lambda_2(\mathbf{H}), ..., \lambda_n(\mathbf{H})$ be the eigenvalues in increasing order and consider the random atomic measure

$$1/n\,[\delta(\lambda_1(\mathbf{H})) + \delta(\lambda_2(\mathbf{H})) + ... + \delta(\lambda_n(\mathbf{H}))]$$

as the empirical eigenvalue distribution of **H**. Its expectation value $\mu_H = \frac{1}{n}\mathbb{E}\left(\sum_{j=1}^{n} \delta_1(\lambda_j(\mathbf{H}))\right)$ is called the mean eigenvalue distribution of **H** and it is easy to see that

$$\int x^m \mu_H(x)\, dx = \frac{1}{n}\mathbb{E}\text{tr}\,(\mathbf{H}^m) = \tau_n(\mathbf{H}^m).$$

The next result shows that the semi-circle law is independent of the details of the distribution of the entries.

Theorem 1.2 Let $\{\mathbf{H}_n\}$ be independent real symmetric matrix with finite moments such that $\mathbb{E}(H_{n,ij}) = 0$ and $\mathbb{E}(H_{n,ij}^2) = 1/n$ for $1 \le i < j \le n$. If furthermore,

$$\sup_{1 \le i < j \le n} \mathbb{E}|H_{n,ij}|^k = O(n^{-k/2}) \text{ for each } k \in \mathbb{N} \text{ as } n \to \infty,$$

then μ_H, the mean eigenvalue distribution of **H** tends to the semi-circle law.

Another way of interpreting this result is the following. $\{\mathbf{H}_n\}$ is a family of non-commutative random variable with a certain distribution family $\{\varphi_n\}$ (positive linear functionals on the *-algebra generated by **H**) and φ_n converges to the semi-circle law in weak *-topology.

Example 1.3 Again consider $n \times n$ standard symmetric Gaussian matrix $\mathbf{T}_n$ (i.e. $\mathbb{E}(T_{n,ij}) = 0$ and $\mathbb{E}(T_{n,ij}^2) = (n+i)^{-1}(1 + \delta_{ij})$ so that as before $\tau_n(\mathbf{T}_n^2) = 1$).

With respect to the Lebesgue measure $d\mathbf{T}_n = \prod_{i<j} dT_{ij}$ on $\mathbb{R}^{n(n+1)/2}$ the density $p(\mathbf{T})$ is given as $p(\mathbf{T}) = C_n \exp\left\{-\frac{n+1}{4}\text{Tr}\,\mathbf{T}^2\right\}$, where C_n is a normalising constant. It is to be noted that this measure is invariant under orthogonal transformations, i.e. $\mathbf{T} \mapsto \mathbf{O}'\mathbf{T}\mathbf{O}$ where $\mathbf{O}'\mathbf{O} = \mathbf{I}$. In fact the standard symmetric Gaussian matrices almost characterises this measure.

The measure induced by $d\mathbf{T}$ above on the ordered space of eigenvalues of **T** has the density

$$\frac{d\mathbf{T}}{\prod_{i=1}^{n} d\lambda_i} = \frac{\pi^{n(n+1)}/4}{\prod_{i=1}^{n}\Gamma(j/2)} \prod_{i<j}|\lambda_i - \lambda_j|$$

which is related to Selberg Trace formula.

Next we look at $\mathcal{U}(n)$; the compact group of $n \times n$ unitary matrices and note that there is a unique invariant Haar measure γ_n on $\mathcal{U}(n)$. Since the eigenvalues of an unitary matrix are on the unit circle $\mathcal{T}$, the joint probability of its eigenvalues will be supported on $\mathcal{T}_n$. The measure on $\mathcal{T}^n$ induced from Haar measure γ_n has the form

$$(2\pi)^{-n}(n!)^{-1}\prod_{j<k}\left|e^{i\theta_j} - e^{i\theta_k}\right|^2 \prod_{j=1}^{n} d\theta_j\,.$$

A standard unitary random matrix **U** is a non-commutative random variable with respect to the functional $\tau_n(\mathbf{U}) = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(e^{i\theta_j})$. Since the expectation $\mathbb{E}$ is with respect to the **H** invariant Haar measure γ_n, it follows that $\tau_n(\mathbf{U}^k) = 0\ \forall$ non-zero k. The above expression of the measure on $\mathcal{T}^n$ can be derived from the earlier expression for the measure induced as the ordered space of eigenvalues of a real symmetric matrix by looking at the real and imaginary parts of the unitary matrix.

2. RANDOM WALK ON FREE GROUPS AND FREE PROBABILITY

A free group $\mathbb{F}_n$ with n generators $g_1, g_2, \ldots, g_n$ is the set of all words with these n alphabets, their assigned inverses and the unit symbol e, with no 'grammar or rules' and the law of multiplication is juxtaposition. Consider a random walk on $\mathbb{F}_n$ which starts from the unit and one step is the move from the group element g to hg with probability $(2n)^{-1}$ if $h \in \{g_1, g_2, g_n, g_1^{-1}, \ldots, g_n^{-1}\}$. Then the probability of return to the unit in m steps is of the form

$$P(n, m) = \frac{1}{(2n)^m} \left\langle (Lg_1 + Lg_1^{-1} + \ldots + Lg_n + Lg_1^{-1})^m \delta_e, \delta_e \right\rangle$$

where we have observed that in $\ell_2(G)$, $g \mapsto Lg$ is a unitary representation given by

$$(Lg\xi)(g') = \xi(g^{-1}g') \text{ for } \xi \in \ell^2(G) \text{ and}$$

δ_g stands for the characteristic function of the element g. Note that $P(n, m) = 0$ if m is odd, and

$$P(n, 2m) = (2n)^{-2m} n^{-m} \left\langle \left(\sum_{j=1}^{n} X_{n,j} \right)^{2m} \delta_e, \delta_e \right\rangle \text{ where}$$

$X_{n,j} = \frac{1}{\sqrt{2}}\left(Lg_j + Lg_j^{-1}\right)$. It's asymptotic behaviour as $n \to \infty : P(n, 2m) \approx \frac{1}{(2n)^m}\left(\frac{1}{m+1}\right)\binom{2m}{m}$ may be compared with the asymptotic behaviour in Theorem 1.1 for large symmetric random matrices. Thus we have a sort of *central limit theorem* for the array $X_{n,1}, X_{n,2}, \ldots X_{n,n}$ of non-commuting random variables since $\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} X_{n,j}\right)$ converges in distribution (or in even moments) to the semi-circle law.

If we set $\varphi(X) = \langle X\delta_e, \delta_e\rangle$, then we note that $X_{n,j}$'s satisfy the following property

$$\varphi(P_1(X_{n,i(1)})\, P_2(X_{n,i(2)}) \ldots P_k(X_{n,i(k)}) = 0$$

for all polynomials $P_1, \ldots, P_k$ such that $\varphi(P_1(X_{n,i(j)})) = 0$ and with $i(1) \neq i(2) \neq \ldots \neq i(k)$. Such non-commuting random variables are called free, following Voiculescu (2000).

Now we can interpret the limit laws of symmetric random Gaussian and of unitary standard matrices in the light of the definition of "freeness" of random variables that we have just introduced. Another way of interpreting Theorem 1.1 is as follows. Let $\mathbf{X}_1(n)$, $\mathbf{X}_2(n)$ be independent random symmetric matrices with the same distribution as that of $\mathbf{H}_n$. Then it is clear by the properties of the convolution of Gaussian distributions that $\frac{1}{\sqrt{n}}\sum_{j=1}^{n} X_j(n)$ also has the same distribution. We note that

$$\tau_n([X_1(n)^k - \tau_n(X_1(n)^k)]\,[X_2(n)^\ell - \tau_n(X_2(n)^\ell)]) = 0$$

if and only if $\tau_n(X_1(n)^k X_2(n)^\ell) = \tau_n(X_1(n)^k)\tau_n(X_2(n)^\ell)$, which is equivalent to saying that

$$\frac{1}{n^2}\sum \mathbb{E}\,(A_{j(1)j(2)}A_{j(2)j(2)j(3)} \cdots A_{j(k)j(k+1)} B_{j(k+1)j(k+2)} \cdots B_{j(k+\ell)j(1)})$$
$$= \frac{1}{n^2}\sum \mathbb{E}\,(A_{j(1)j(2)}A_{j(2)j(3)} \cdots A_{j(k)j(1)})\, \mathbb{E}\,(B_{j(k+1)j(k+2)} \cdots B_{j(k+1)}B_{j(k+2)}).$$

with $\mathbf{A} = \mathbf{X}_1(n)$ and $\mathbf{B} = \mathbf{X}_2(n)$ for simplicity and with summation over all indices. The two sides are equal in many cases, in particular when k or ℓ are odd or when both are 0. However, the difference goes to 0 as $n \to \infty$ as proven by Voiculescu (2000). In other words, the property of freeness as described above, appears in the limit or equivalently random symmetric Gaussian matrices exhibit the asymptotic freeness property. Similarly one can also show that independent Haar distributed unitary matrices are also asymptotically free, thus providing a bridge between random matrix theory and free probability.

3. LARGE DEVIATIONS AND ENTROPY

First, let us consider the classical case. Let $\xi_1, \xi_2, \ldots$ be independent standard real Gaussian random variables and let G be an open set in the space $\mathcal{M}(\mathbb{R})$ of probability measures on $\mathbb{R}$ (with weak *-topology). Then Sanov's (or Varadhan's in a more general context) theorem says that if the standard Gaussian measure $\nu \notin \bar{G}$, then

$$\text{Prob}\left\{\text{the atomic measure } \frac{1}{n}\sum_{j=1}^{n} \delta_j(\xi_j) \in G\right\} \approx \exp(-nC(\nu, G)),$$

where $C(\nu, G) = \inf\{I(\mu) : \mu \in G\}$ and $I(\mu)$ is the rate function or relative entropy

$$I(m) = -\int p(x) \log p(x)\mathrm{d}x + \frac{1}{2}\int x^2\mu(\mathrm{d}x) + \frac{1}{2}\log 2\pi,$$

and $\mu(\mathrm{d}x) = p(x)\mathrm{d}x$. The first part is the entropy or the Boltzmaun-Gibbs entropy $S(\mu) = -\int p(x)\log p(x)\mathrm{d}x$. The above theorem says that the probability of a large deviation from Gaussianity decreases exponentially with n, and the coefficient multiplying n in the exponent is an infimum of an expression which is essentially the entropy.

A similar exercise can be carried out for symmetric i.i.d standard Gaussian matrices using the joint eigenvalue density of relevant random matrix $\mathbf{H}_n$,

$$\frac{1}{Z_n}\exp\left(-\frac{n+1}{4}\sum\lambda_{n,i}^2\right)\prod_{i<j}(\lambda_{n,i}-\lambda_{n,j}), \text{ where}$$

Z_n is the normalising constant. Thus we can compute for a neighbourhood G of $\mu \in \mathcal{M}(\mathbb{R})$:

Prob {the empirical eigenvalue distribution of $\mathbf{H}_n \in G$}

$$\equiv \text{Prob}\left\{\frac{1}{n}\sum_{j=1}^{n}\delta(\lambda_j(\mathbf{H}_n)) \in G\right\}$$

$$= \frac{1}{Z_n}\iint_{\tilde{G}}\exp\left(\sum_{i<1}\log|x_i - x_j| - \frac{n+1}{4}\sum_{j=1}^{n}x_j^2\right)\mathrm{d}x_1 \ldots \mathrm{d}x_n,$$

where $\tilde{G} \subseteq \mathbb{R}^n$ is defined by the requirement

$$\tilde{G} = \left\{x \in \mathbb{R}^n \,\middle|\, \frac{1}{n}\sum_{j=1}^{n}\delta(x_j) \in G\right\}.$$

Intuitively, when G tends to a point $\mu \in \mathcal{M}(\mathbb{R})$, the approximation:

$$\sum\log|x_i - x_j| - \frac{n+1}{4}\sum x_j^2$$

$$\simeq n^2\left(\frac{1}{2}\iint\log|x-y|\,\mu(\mathrm{d}x)\,\mu(\mathrm{d}y) - \frac{1}{4}\int x^2\mu(\mathrm{d}x)\right)$$

holds for $x \in \tilde{G}$ and one gets that

$$-\frac{1}{n^2}\log\text{Prob}\left\{\frac{1}{n}\sum_j\delta(\lambda_j(\mathbf{H}_n)) \in G\right\}$$

$$\approx -\frac{1}{2}\iint\log|x-y|\,\mu(\mathrm{d}x)\,\mu(\mathrm{d}y)$$

$$+\frac{1}{4}\int x^2\mu(\mathrm{d}x) + \frac{1}{n^2}\log Z_n.$$

This leads to the "large deviation" statement that

$$\text{Prob}\left\{\frac{1}{n}\sum_j\delta(\lambda_j(\mathbf{H}_n)) \in G\right\} \approx e^{-n^2C(w,G)},$$

where G is an open set in $\mathcal{M}(\mathbb{R})$ not containing the semi-circle law w and $C(w, \mathrm{G}) = \inf\{I(\mu) | \mu \in G\}$ with

$$I(\mu) = -\frac{1}{2}\iint\log|x-y|\,\mu(\mathrm{d}x)\mu(\mathrm{d}y)$$

$$+\frac{1}{4}\int x^2\mu(\mathrm{d}x) + \text{ constant.}$$

In analogy with the classical case, we can identify

$S^f(\mu) = -\frac{1}{2}\iint\log|x-y|\,\mu(\mathrm{d}x)\mu(\mathrm{d}y)$ as the free entropy.

It is important to note that the probability that the empirical eigenvalue distribution is different from the semi-circle law decreases sharply with the increase of sample size, i.e. as $\exp(-n^2C(w, G))$ in contrast to the classical case in which the probability for "large deviation" goes like $\exp(-nC(\nu, G))$. Another way of saying the same thing is that the semi-circle law is a "stronger attractor" for the empirical distribution for data matrices than the Gaussian law is for the empirical real-valued observations. This sharp decrease of 'error' with increasing 'n' opens up large number of possibilities in improvements in data-analysis.

We end this section with the following table.

	Classical	**Free**
Entropy	$-\int p(x)\log p(x)\,\mathrm{d}x$, where $p(x) = \frac{\mathrm{d}\mu}{\mathrm{d}x}(x)$	$-1/2\int\int\log\lvert x-y\rvert\,\mu(\mathrm{d}x)\mu(\mathrm{d}y)$ $= -1/2\int p(x)\mathrm{d}x\int\log\lvert x-y\rvert\,p(y)\,\mathrm{d}y$, where $p(x) = \frac{\mathrm{d}\mu}{\mathrm{d}x}(x)$.
Fisher Information	$\int\frac{[p'(x)]^2}{p(x)}\mathrm{d}x$	Constant $\int[p(x)]^3\mathrm{d}x$

4. APPLICATIONS

(i) Statistical Mechanics in Physics

Consider the constant energy surface $\{x \in \mathbb{R}^n \mid H(x) = h_0\}$ and the uniform distribution (microcanonical distribution) $u(dx)$ on it. On the other hand, Maxwell, Boltzmann and Gibbs also used the canonical measure $u_\beta(dx) \equiv Z^{-1}e^{-\beta H(x)}dx$ on $\mathbb{R}^n$ such that $\int u_\beta(dx) = 1$ and $\int H(x)u_\beta(dx) = h_0$. The principle of equivalence of ensembles of Gibbs asserts that for every continuous function f on $\mathbb{R}^n$, the ensemble averages are equal when n is large, i.e.,

$$\int f(x)u(dx) \approx \int f(x)u_\beta(dx) \text{ for } n \text{ large.}$$

For the common and the simplest case $H(x) = \sum_{j=1}^{n} x_j^2$, the microcanonical ensemble becomes the uniform distribution on the sphere and the canonical measure becomes the product Gaussian measure $u_\beta(dx) = Z^{-1}\exp(-\beta H(x))dx$.

Consider the real orthogonal group $\mathcal{O}_n$. Fill an $n \times n$ array with independent draw from standard Gaussian curve and then perform Gram-Schmidt algorithm on this array. We get a map T from $\mathbb{R}^{n^2}$ into $\mathcal{O}_n$ and under this map the product of the Gaussian measures gets mapped into the Haar measure on $\mathcal{O}_n$. If f is a continuous function of the first variable x_1 only, then one concludes that for $\mathcal{M} \in \mathcal{O}_n$, $\int f(x)u(dx) \simeq \int f(x)u_\beta(dx)$, as $n \to \infty$ using the following result.

Emile Borel proved the following theorem:

Pick M from the Haar distribution on the orthogonal group $\mathcal{O}_n$. Then

$$\text{Prob}_{\text{Haar}}\left\{\sqrt{n}M_{11} \le x\right\} \text{ converges to } \frac{1}{2\pi}\int_{-\infty}^{x^2} e^{-t^2/2}dt.$$

This result was extended by Levy and others. This says that asymptotically the entries $((\sqrt{n}\, M_{ij}))$ are Gaussian distributed, though of course that is not true for the eigenvalues. The above justifies to some extent what Physicists are doing all the time.

(ii) Communication Sciences [7]

Information theory of wireless communication channels has become increasingly important because of the necessity of the efficient use of bandwidth and power in the face of over-increasing demand of such services. Fading, wideband, multi-user and multi-antenna are some of the key characteristics of modern wireless channels.

As a simple model, consider the linear memory-less channels of the form $y = \mathbf{H}x + n$, where x is K-dim input vector, y is the N-dimensional output vector, n is the N-dim vector modeling the orthogonally symmetric Gaussian noise and $\mathbf{H}$ is the complex valued random $N \times K$ channel matrix. For example, in the single user case, K is the number of transmitting antennas while N is the same for the receiving side. On the otherhand for DS-CDMA (Direct sequence code-division multiple access) channel, K = # of users and N = the gain.

In the first, $\mathbf{H}$ is the propagation coefficient for each transmitting-receiving pair of antennas while for the second case, each entry of $\mathbf{H}$ depend on the received noise-sequence and fading coefficient, and $K = n_T J$, $N = n_R G$ with n_T and n_R the transmitting and receiving antennas and J and G the number of users and gain respectively. Of course, the simplest case is one where the entries of $\mathbf{H}$ are i.i.d. but more realistically, they are not i.i.d.

Set for a $N \times N$ positive matrix $\mathbf{A}$, the normalised distribution function as $F_A^N(x) = \frac{1}{N}\sum_{j=1}^{N} \mathcal{X}(\lambda_j(\mathbf{A}) \le x)$

so that

$$F_{HH^*}^N(x) - N\Theta(x) = KF_{H^*H}^K(x) - K\,\Theta(x),$$

where Θ is the Heavyside function. This is because the non-zero eigenvalues of $\mathbf{HH}^*$ and $\mathbf{H}^*\mathbf{H}$ are identical.

An important indicator of such channels is the mutual information conditioned on $\mathbf{H}$: $\frac{1}{N}\mathbf{I}(x;y \mid \mathbf{H}) \equiv \frac{1}{N}\log\det(\mathbf{I} + \alpha\mathbf{HH}^*)$, where

$$\alpha \equiv \text{signal-to-noise ration} = \frac{N\mathbb{E}\|x\|^2}{K\mathbb{E}\|n\|^2}.$$

One can easily see that the above $= \int_0^\infty \log(1+\alpha\lambda)\, dF^N_{HH*}(\lambda)$. Another measure is the minimum mean-square error

$$(\text{MMSE}) \equiv \frac{1}{K} \min_{A\in\mathcal{B}(C^N, C^K)} \mathbb{E}\| x - \mathbf{A}y \|^2$$
$$= \frac{1}{K}\text{tr}\{(I + \alpha\mathbf{H}^*\mathbf{H})^{-1}\}$$
$$= \frac{N}{K}\int_0^\infty (1+\alpha x)^{-1} dF^N_{HH*}(\lambda) + (1 - N/K).$$

These two measures are related by a logarithmic differentiation with respect to α. The asymptotic behaviour of any of these measures as N and K tend to ∞ keeping the aspect ratio N/K constant, is of importance. Normally, in multi-antenna systems N and K vary between 8 and 16 while in CDMA, they vary between 32 and 64. In this context, we have the following theorem of Marcenko and Pastur (1967).

Theorem 4.1 Let H be a random N × K real matrix whose entries are zero-mean i.i.d with variance N^{-1}. Then the empirical distribution of the eigen-values of H^*H converges almost surely as K, N → ∞ s.t. N/K → γ, to the Marcenko-Pastur law whose density function is given as

$$f_\gamma(x) = (1 - \gamma^{-1})^+ \delta(x) + (2\pi\gamma x)^{-1}\sqrt{(x-a)^+(b-x)^+},$$

where $y^+ = \max(0, y)$,

$$a = (1-\sqrt{\gamma})^2,\ b = (1+\sqrt{\gamma})^2.$$

This asymptotic behaviour exhibit several key engineering features:

(i) Insensitivity of the asymptotic behaviour to the details of the distribution of the random matrix entries of **H**, which implies that for a single user multi-antenna link, the asymptotic behaviour hold for any kind noise/fading satisfies and thus using binary-valued wave form for CDMA does not lead to any loss in capacity.

(ii) The eigenvalue histogram converges almost surely to a deterministic asymptotic eigenvalue distribution, displaying a kind of Ergodic behaviour.

(iii) The convergence is very fast. We have already discussed this property in detail.

REFERENCES

Diaconis, P. (2003). The 70th Willard Gibbs lectures, *Bull. Amer. Math. Soc.*, **40(2)**, 155-178.

Fisher, R.A. (1939). *Ann. Eugen.*, **9**, 238-249,

Hiai, F. and Petz, D. (2000). *The Semicircle Law, Free Random Variables and Entropy*. Math. Surveys and Monographs #77, *Amer. Math. Soc.*

Marcenko, V. and Pastur, L. (1967). *Distribution of Some Sets of Random Matrices*. Mat. Sb., **1**, 507-536.

Tulino, A.M. and Verdu, S. (2004). *Random Matrix Theory and Wireless Communications*. New Publisher-Inc., P.O. Box 1024, Hanover, MA02339, USA.

Voiculescu, D. (2000). *Lectures on Free Probability Theory*. LNM 1738, 279-349 and lectures given at Indian Statistical Institute, Delhi in (1995).

Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population, *Biometrika*, **20A**, 32-52.