Applications of free probability and random matrix theory

Øyvind Ryan

January 2008

Some important concepts from classical probability

- Random variables are functions (i.e. they commute w.r.t. multiplication) with a given p.d.f. (denoted f)
- Expectation (denoted E) is integration
- Independence
- Additive convolution ($*$) and the logarithm of the Fourier transform
- Multiplicative convolution
- Central limit law, with special role of the Gaussian law
- Poisson distribution P_c: The limit of $\left(\left(1-\frac{c}{n}\right)\delta(0)+\frac{c}{n}\delta(1)\right)^{*n}$ as $n \rightarrow \infty$.
- Divisibility: For a given a, find i.i.d. $b_1, ..., b_n$ such that $f_a = f_{b_1+\cdots+b_n}$.

- Can we find a more general theory, where the random variables are matrices (or more generally, operators), with their eigenvalue distribution (or spectrum) taking the role as the p.d.f.?
- What are the analogues to the above mentioned concepts for this theory? In particular, what plays the role as central limit?
- What are the applications of such a theory?

Free probability

Free probability was developed as a probability theory for random variables which do not commute, like matrices

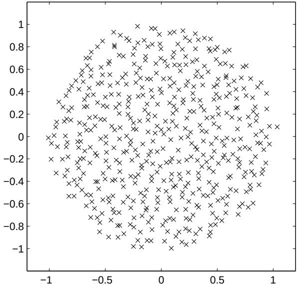
- The random variables are elements in a unital $*$-algebra (denoted A), typically $B(\mathcal{H})$, or $M_n(\mathbb{C})$.
- Expectation (denoted ϕ) is a normalized linear functional on A. The pair (A, ϕ) is called a *noncommutative probability space*.
- For matrices, ϕ will be the normalized trace tr_n, defined by

$$tr_n(a) = \frac{1}{n} \sum_{i=1}^{n} a_{ii}.$$

  For random matrices, we set $\phi(a) = \tau_n(a) = E(tr_n(a))$ is defined by.
- Useful way to think about free probability: Two independent random matrices, where the eigenvectors for one of them point in each direction with equal probability, then the two matrices are "free" (to be defined).

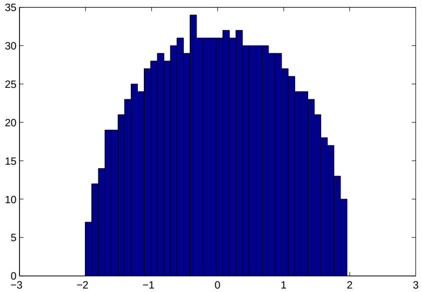
The full circle law

Let $\mathbf{X}_n = \frac{1}{\sqrt{n}}\mathbf{Y}_n$ where $\mathbf{Y}_n$ is $n \times n$ and has i.i.d. complex standard Gaussian entries. This is a "central limit candidate". What happens when n is large? The eigenvalues converge to what is called the *full circle law*. Here for $n = 500$.




```
plot(eig( (1/sqrt(1000)) * (randn(500,500) +
j*randn(500,500)) ),'kx')
```

The semicircle law




```
A = (1/sqrt(2000)) * (randn(1000,1000) +
j*randn(1000,1000));
A = (sqrt(2)/2)*(A+A');
hist(eig(A),40)
```

The Marčhenko Pastur law

What happens with the eigenvalues of $\frac{1}{N}\mathbf{X}_n\mathbf{X}_n^H$ when $\mathbf{X}_n$ is an $n \times N$ random matrix with standard complex Gaussian entries?

- One can show that

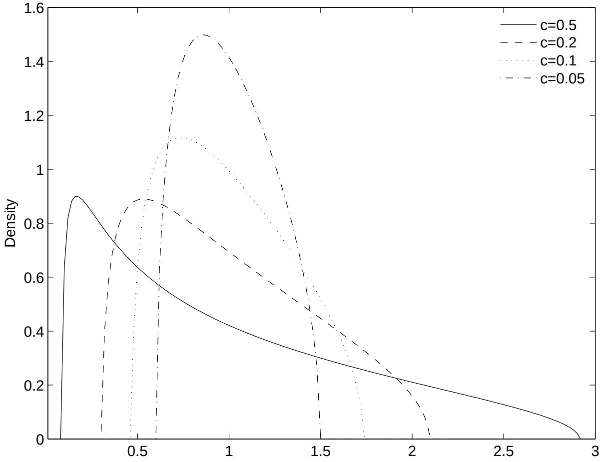
$$\tau_n\left(\left(\frac{1}{N}\mathbf{X}_n\mathbf{X}_n^H\right)^p\right) = \sum_{\hat{\pi}\in NC_{2p}} 1 + \sum_k \frac{a_k}{N^{2k}}.$$

  Convergence is "almost sure", i.e. we have very accurate eigenvalue prediction when the matrices are large.
- When $\frac{n}{N} \rightarrow c$, the eigenvalue distribution converges to the Marčhenko Pastur law with parameter c, denoted $\mu_{\frac{n}{N}}$, with density

$$f^{\mu_c}(x) = (1 - \frac{1}{c})^+\delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi cx}, \qquad (1)$$

  where $(z)^+ = \max(0, z)$, $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$.

Four different Marčhenko Pastur laws $\mu_{\frac{n}{N}}$ are drawn.



Motivation for free probability

One can show that for the Gaussian random matrices we considered, the limits

$$\phi\left(X^{i_1}Y^{j_1}\cdots X^{i_l}Y^{j_l}\right) = \lim_{n\to\infty} tr_n\left(\mathbf{X}_n^{i_1}\mathbf{Y}_n^{j_1}\cdots\mathbf{X}_n^{i_l}\mathbf{Y}_n^{j_l}\right)$$

exist. If we linearly extend the linear functional ϕ to all polynomials in A and B, the following can be shown:

Theorem

If P_i, Q_i are polynomials in X and Y respectively, with $1 \leq i \leq l$, and $\phi(P_i(X)) = 0$, $\phi(Q_i(Y)) = 0$ for all i, then

$$\phi\left(P_1(X)Q_1(Y)\cdots P_l(X)Q_l(Y)\right) = 0.$$

This motivates the definition of freeness, which is the analogy to independence.

Definition of freeness

Definition

A family of unital $*$-subalgebras $(A_i)_{i\in I}$ is called a free family if

$$\left\{ \begin{array}{c} a_j \in A_{i_j} \\ i_1 \neq i_2, i_2 \neq i_3, \cdots, i_{n-1} \neq i_n \\ \phi(a_1) = \phi(a_2) = \cdots = \phi(a_n) = 0 \end{array} \right\} \Rightarrow \phi(a_1 \cdots a_n) = 0. \quad (2)$$

A family of random variables a_i is called a free family if the algebras they generate form a free family.

The free central limit theorem

Theorem

If

- $a_1, ..., a_n$ *are free and self-adjoint,*
- $\phi(a_i) = 0$,
- $\phi(a_i^2 = 1$,
- $\sup_i |\phi(a_i^k)| < \infty$ *for all* k,

then the sequence $(a_1 + \cdots + a_n)/\sqrt{n}$ *converges in distribution to the semicircle law.*

In free probability, the semicircle law thus has the role of the Gaussian law. it's density is density $\frac{1}{2\pi}\sqrt{4 - x^2}$

Similarities between classical and free probability

1. Additive convolution $\boxplus$: The p.d.f. of the sum of free random variables. The role of the logarithm of the Fourier transform is now taken by the R-transform, which satisfies $R_{\mu_a \boxplus \mu_b}(z) = R_{\mu_a}(z) + R_{\mu_b}(z)$.
2. The S-transform: Transform on probability distributions which satisfies $S_{\mu_a \boxtimes \mu_b}(z) = S_{\mu_a}(z) S_{\mu_b}(z)$
3. Poisson distributions have their analogy in the free Poisson distributions: These are given by the Marčenko Pastur laws μ_c with parameter c, which also can be written as the limit of $\left(\left(1 - \frac{c}{n}\right)\delta(0) + \frac{c}{n}\delta(1)\right)^{\boxplus n}$ as $n \to \infty$
4. Infinite divisibility: There exists an analogy to the Lévy-Hinčin formula for infinite divisibility in classical probability.

Main usage of free probability in my papers

- Let $\mathbf{X}$ and $\mathbf{Y}$ be random matrices. How can we make a good prediction of the eigenvalue distribution of $\mathbf{X}$ when one has the eigenvalue distribution of $\mathbf{XY}$ and $\mathbf{Y}$? Simplest case is when one assumes that $\mathbf{Y}$ is Gaussian (Noise). What about the eigenvectors?
- Assume that we have the eigenvalue distribution of $\frac{1}{N}(\mathbf{R}+\mathbf{X})(\mathbf{R}+\mathbf{X})^H$, where $\mathbf{R}$ and $\mathbf{X}$ are independent $n \times N$ random matrices, with $\mathbf{X}$ Gaussian. If the columns of $\mathbf{R}$ are realizations of some random vector $\mathbf{r}$, what is the covariance matrix $E(\mathbf{r}_i\mathbf{r}_j^*)$?
- Multiplicative free convolution with the Marčhenko Pastur law is particularly useful. This has an efficient implementation.

Application of free probability

The capacity per receiving antenna of a channel with $n \times m$ channel matrix $\mathbf{H}$ and signal to noise ratio $\rho = \frac{1}{\sigma^2}$ is given by

$$C = \frac{1}{n} \log_2 \det \left(\mathbf{I}_n + \frac{1}{m\sigma^2} \mathbf{H}\mathbf{H}^H \right) = \frac{1}{n} \sum_{l=1}^{n} \log_2 (1 + \frac{1}{\sigma^2} \lambda_l) \qquad (3)$$

where λ_l are the eigenvalues of $\frac{1}{m}\mathbf{H}\mathbf{H}^H$. We would like to estimate C.

To estimate C, we will use free probability tools to estimate the eigenvalues of $\frac{1}{m}\mathbf{H}\mathbf{H}^H$ based on some observations $\hat{\mathbf{H}}_i$

Observation model

The following is a much used observation model:

$$\hat{\mathbf{H}}_i = \mathbf{H} + \sigma \mathbf{X}_i \tag{4}$$

where

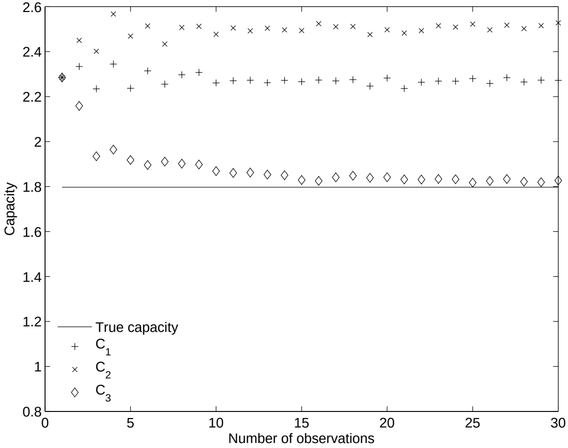
- The matrices are $n \times m$ (n is the number of receiving antennas, m is the number of transmitting antennas)
- $\hat{\mathbf{H}}_i$ is the measured MIMO matrix,
- $\mathbf{X}_i$ is the noise matrix with i.i.d standard complex Gaussian entries.

Existing ways to estimate the channel capacity

Several channel capacity estimators have been used in the literature:

$$
\begin{aligned}
C_1 &= \frac{1}{nL}\sum_{i=1}^{L} \log_2 \det\left(\mathbf{I}_n + \frac{1}{m\sigma^2}\hat{\mathbf{H}}_i\hat{\mathbf{H}}_i^H\right) \\
C_2 &= \frac{1}{n}\log_2 \det\left(\mathbf{I}_n + \frac{1}{L\sigma^2 m}\sum_{i=1}^{L}\hat{\mathbf{H}}_i\hat{\mathbf{H}}_i^H\right) \qquad (5)\\
C_3 &= \frac{1}{n}\log_2 \det\left(\mathbf{I}_n + \frac{1}{\sigma^2 m}\left(\frac{1}{L}\sum_{i=1}^{L}\hat{\mathbf{H}}_i\right)\left(\frac{1}{L}\sum_{i=1}^{L}\hat{\mathbf{H}}_i\right)^H\right)
\end{aligned}
$$

Why not try to formulate an estimator based on free probability instead?



Comparison of the classical capacity estimators for various number of observations. $\sigma^2 = 0.1$, $n = 10$ receive antennas, $m = 10$ transmit antennas. The rank of $\mathbf{H}$ was 3.

Main free probability result we will use

Define

$$\Gamma_n = \frac{1}{N}\mathbf{R}_n\mathbf{R}_n^H$$
$$\mathbf{W}_n = \frac{1}{N}(\mathbf{R}_n + \sigma\mathbf{X}_n)(\mathbf{R}_n + \sigma\mathbf{X}_n)^H,$$

where $\mathbf{R}_n$ and $\mathbf{X}_n$ are independent $n \times N$ random matrices, $\mathbf{X}_n$ is complex, standard, Gaussian.

Theorem

If e.e.d.$(\Gamma_n) \rightarrow \nu_\Gamma$, then e.e.d.$(\mathbf{W}_n) \rightarrow \nu_W$ where ν_W is uniquely identified by

$$\nu_W \boxslash \mu_c = (\nu_\Gamma \boxslash \mu_c) \boxplus \delta_{\sigma^2}$$

($\boxslash$ = "the opposite of $\boxtimes$").

Realization of the theorem for the problem at hand

Form the compound observation matrix

$$
\begin{aligned}
\hat{\mathbf{H}}_{1\ldots L} &= \mathbf{H}_{1\ldots L} + \frac{\sigma}{\sqrt{L}}\mathbf{X}_{1\ldots L}, \text{ where} \\
\hat{\mathbf{H}}_{1\ldots L} &= \frac{1}{\sqrt{L}}\left[\hat{\mathbf{H}}_1, \hat{\mathbf{H}}_2, \ldots, \hat{\mathbf{H}}_L\right], \\
\mathbf{H}_{1\ldots L} &= \frac{1}{\sqrt{L}}\left[\mathbf{H}, \mathbf{H}, \ldots, \mathbf{H}\right], \\
\mathbf{X}_{1\ldots L} &= \left[\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_L\right].
\end{aligned}
$$

For the problem at hand, the theorem takes the form

$$
\nu_{\frac{1}{m}\hat{\mathbf{H}}_{1\ldots L}\hat{\mathbf{H}}_{1\ldots L}^H} \boxslash \mu_{\frac{n}{mL}} \approx \left(\nu_{\frac{1}{m}\mathbf{H}_{1\ldots L}\mathbf{H}_{1\ldots L}^H} \boxslash \mu_{\frac{n}{mL}}\right) \boxplus \delta_{\sigma^2} \qquad (6)
$$

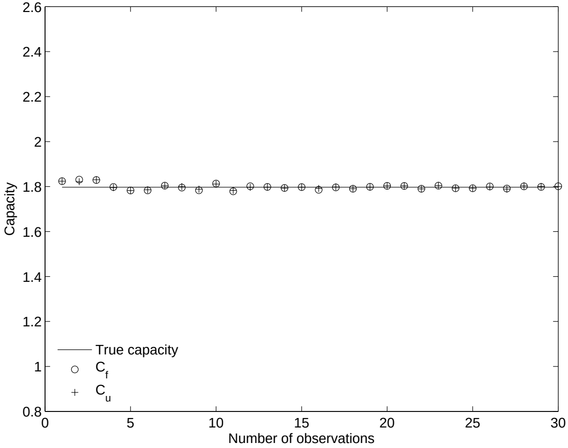
Since $\frac{1}{m}\mathbf{H}_{1\ldots L}\mathbf{H}_{1\ldots L}^H = \frac{1}{m}\mathbf{H}\mathbf{H}^H$, we can now estimate the moments of $\frac{1}{m}\mathbf{H}\mathbf{H}^H$ from the moments of the observation matrix $\frac{1}{m}\hat{\mathbf{H}}_{1\ldots L}\hat{\mathbf{H}}_{1\ldots L}^H$, and thereby estimate the eigenvalues, and hence the channel capacity.

Free probability based estimator for the moments of the channel matrix

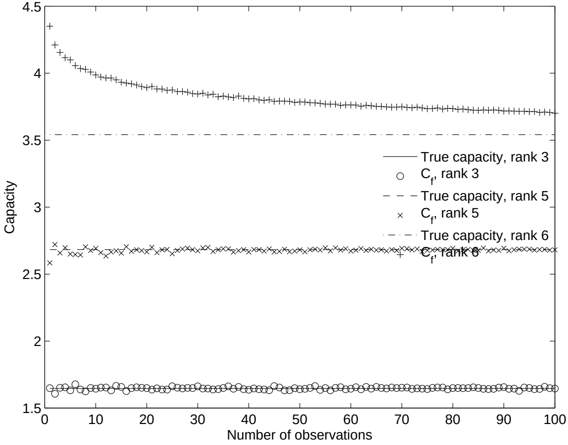
Can also be written in the following way for the first four moments:

$$\begin{aligned}
\hat{h}_1 &= h_1 + \sigma^2 \\
\hat{h}_2 &= h_2 + 2\sigma^2(1+c)h_1 + \sigma^4(1+c) \\
\hat{h}_3 &= h_3 + 3\sigma^2(1+c)h_2 + 3\sigma^2 c h_1^2 \\
&\quad +3\sigma^4\left(c^2+3c+1\right)h_1 \\
&\quad +\sigma^6\left(c^2+3c+1\right) \\
\hat{h}_4 &= h_4 + 4\sigma^2(1+c)h_3 + 8\sigma^2 c h_2 h_1 \\
&\quad +\sigma^4(6c^2+16c+6)h_2 \\
&\quad +14\sigma^4 c(1+c)h_1^2 \\
&\quad +4\sigma^6(c^3+6c^2+6c+1)h_1 \\
&\quad +\sigma^8\left(c^3+6c^2+6c+1\right),
\end{aligned} \tag{7}$$

where $\hat{h}_i$ are the moments of the observation matrix $\frac{1}{m}\hat{\mathbf{H}}_{1\ldots L}\hat{\mathbf{H}}_{1\ldots L}^H$, h_i are the moments of $\frac{1}{m}\mathbf{H}\mathbf{H}^H$.



Comparison of C_f and C_u for various number of observations. $\sigma^2 = 0.1$, $n = 10$ receive antennas, $m = 10$ transmit antennas. The rank of $\mathbf{H}$ was 3.



C_f for various number of observations. No phase off-set/phase drift. $\sigma^2 = 0.1$, $n = 10$ receive antennas, $m = 10$ transmit antennas. The rank of $\mathbf{H}$ was 3, 5 and 6.

List of papers

- Free Deconvolution for Signal Processing Applications. Submitted to IEEE Trans. Inform. Theory. arxiv.org/cs.IT/0701025.
- Multiplicative free Convolution and Information-Plus-Noise Type Matrices. arxiv.org/math.PR/0702342.
- Channel Capacity Estimation using Free Probability Theory. Submitted to IEEE. Trans. Signal Process. arxiv.org/abs/0707.3095.
- Random Vandermonde Matrices-Part I: Fundamental results. Work in progress.
- Random Vandermonde Matrices-Part II: Applications to wireless applications. Work in progress.
- Vandermonde Frequency Division Multiplexing: An Approach for Cognitive Radio. Work in progess.

This talk is available at
http://heim.ifi.uio.no/∼oyvindry/talks.shtml.
My publications are listed at
http://heim.ifi.uio.no/∼oyvindry/publications.shtml