Ben Sauerwine
Progress Report
April 18, 2006

Ceartainly, upon examination of the plots from previous weeks, the distribution of the generating function for the Toeplitz elements appears to be roughly a Gaussian distribution.

Here I will approach the problem from this direction.

As seen in previous chapters, I may write the diagonal entries of the matrix as

$$f_N(\lambda, D) = \sum_{-\infty}^{\infty} \binom{N}{\alpha N + xD} e^{i\lambda x} = \sum_{x=-\infty}^{\infty} e^{\ln\binom{N}{\alpha N + xD}} e^{i\lambda x} \to \binom{N}{\alpha N} + 2\sum_{x=1}^{\infty} e^{\ln\binom{N}{\alpha N + xD}} e^{i\lambda x}$$

$$\ln\binom{N}{\alpha N + xD}$$

$$\approx N \ln N - (\alpha N + xD)\ln(\alpha N + xD) - (N(1-\alpha) - xD)\ln(N(1-\alpha) - xD) - N + (\alpha N + xD) + (N(1-\alpha) - xD)$$

$$= N \ln N - (\alpha N + xD)\ln(\alpha N + xD) - (N(1-\alpha) - xD)\ln(N(1-\alpha) - xD)$$

$$\alpha \to \frac{1}{2}$$

$$= N \ln N - \alpha N\left(1 + \frac{xD}{\alpha N}\right)\left[\ln\left(1 + \frac{xD}{\alpha N}\right) + \ln \alpha N\right] - \alpha N\left(1 - \frac{xD}{\alpha N}\right)\left[\ln\left(1 - \frac{xD}{\alpha N}\right) + \ln \alpha N\right]$$

$$= N \ln N - \alpha N\left[\left(1 + \frac{xD}{\alpha N}\right)\ln\left(1 + \frac{xD}{\alpha N}\right) + \left(1 + \frac{xD}{\alpha N}\right)\ln \alpha N + \left(1 - \frac{xD}{\alpha N}\right)\ln\left(1 - \frac{xD}{\alpha N}\right) + \left(1 - \frac{xD}{\alpha N}\right)\ln \alpha N\right]$$

$$= N \ln N - 2\alpha N \ln \alpha N - \alpha N\left[\left(1 + \frac{xD}{\alpha N}\right)\ln\left(1 + \frac{xD}{\alpha N}\right) + \left(1 - \frac{xD}{\alpha N}\right)\ln\left(1 - \frac{xD}{\alpha N}\right)\right]$$

Now if I expand in small $\frac{xD}{\alpha N}$ where most of the function's mass should lie, I have:

```
In[139]:=
        Simplify[Series[(1 + x) Log[1 + x] + (1 - x) Log[1 - x], {x, 0, 6}]]

Out[139]= x^2 + x^4/6 + x^6/15 + O[x]^7
```

$$\approx N \ln N - 2\alpha N \ln \alpha N - \alpha N\left[\left(\frac{xD}{\alpha N}\right)^2\right]$$

Returning to the generating function, then, I have

$$f_N(\lambda, D) = \sum_{-\infty}^{\infty} \binom{N}{\alpha N + xD} e^{i\lambda x} \approx (\alpha N)^{-2\alpha N} N^N \sum_{x=-\infty}^{\infty} e^{-\alpha N\left(\frac{xD}{\alpha N}\right)^2} e^{i\lambda x} \qquad k = \frac{D^2}{\alpha N}$$

$$= (\alpha N)^{-2\alpha N} N^N e^{\frac{-\lambda^2}{4k}} \sum_{x=-\infty}^{\infty} e^{-k\left(x - \frac{i\lambda}{2k}\right)^2}$$

which is a very nice Gaussian form. Now if only I could perform the sum.
At the very least, I should be able to verify the validity of this approximation versus the known good approximation. (Look at the first section of GVLongterm15) It looks like while the actual distribution initially looks Gaussian, this function misses the height dramatically. The cause seems to be due to when we exponentiate the approximation, we have missed the indicated square-root like factor in N.

$$\ln N! = N\ln N - N + \ln\left(\sqrt{2\pi N}\right)$$

$$n\begin{pmatrix} N \\ \alpha N + xD \end{pmatrix}$$

$$\alpha \to \frac{1}{2}$$

$$\approx N\ln N - 2\alpha N\ln\alpha N - \alpha N\left[\left(1+\frac{xD}{\alpha N}\right)\ln\left(1+\frac{xD}{\alpha N}\right)+\left(1-\frac{xD}{\alpha N}\right)\ln\left(1-\frac{xD}{\alpha N}\right)\right]+E(N,D,\alpha,x)$$

$$E(N,D,\alpha,x)=\frac{1}{2}\ln 2\pi N-\frac{1}{2}\ln(2\pi(\alpha N+xD))-\frac{1}{2}\ln(2\pi(\alpha N-xD))$$

$$=\frac{1}{2}\ln 2\pi N-\frac{1}{2}\ln\left(2\pi\left(1+\frac{xD}{\alpha N}\right)\right)-\frac{1}{2}\ln\left(2\pi\left(1-\frac{xD}{\alpha N}\right)\right)$$

$$\approx\frac{1}{2}\ln 2\pi N-\ln\alpha N+\frac{1}{2}\left(\frac{xD}{\alpha N}\right)^2$$

This is all very small with respect to a large N, and so it's surprising that it makes such a difference. However, it seems probable that only the terms $E(N,D,\alpha,x)\approx\frac{1}{2}\ln 2\pi N-\ln\alpha N$ make any substantial difference at all.

$$f_N(\lambda,D)\to\frac{2\sqrt{2\pi N}}{N}(\alpha N)^{-2\alpha N}N^N e^{\frac{-\lambda^2}{4k}}\sum_{x=-\infty}^{\infty}e^{-k\left(x-\frac{i\lambda}{2k}\right)^2}\qquad k=\frac{D^2}{\alpha N}$$

This, it turns out, still doesn't work particularly well: it misses by some factor.

Suppose that I did want to try taking the integral of the logarithm of this (I can't find a proper generating function for this, but it converges quickly.) I wonder if the results would agree closely with the expected value in the long-term since I take the logarithm anyhow. Looking at the results in Mathematica, it does indeed appear that one took the second factor to account for area-wise entropy, this expansion would work. The problem now is performing this sum.