

## Random Matrices and Queues in Series

Yuliy Baryshnikov

LAMA, Université de Versailles – Saint-Quentin-en-Yvelines (France)

December 11, 2000

*Summary by Marianne Durand*

### Abstract

Queues in series are defined as an infinite sequence of clients queuing in front of an infinite sequence of servers where each time a client is served by a server, it immediately enters the next queue. Simple questions about this model are very hard to solve directly. This talk describes the centralized and normalized law of the departure of the  $k$ th client from the  $n$ th server, as  $n$  tends to infinity while  $k$  remains bounded; this law is related to a sequence of largest eigenvalues of random matrices. This relation allows us to use the numerous asymptotic results known regarding the spectra of random matrices and gain useful informations about the queuing processes.

### 1. Queues and Brownian Motions

Consider an infinite series of queues corresponding to servers, and an infinity of jobs. At first all the jobs are in the first queue; then when a job leaves the server  $Q_i$ , it immediately enters the queue corresponding to the server  $Q_{i+1}$ . The question asked is: When does the  $i$ th job leave the  $j$ th server? This can be modeled by pathweights in an infinite matrix. Let  $w_{k,l}$  denote the time needed to process the  $k$ th job on the  $l$ th server. The cost of the maximal weight of a path from  $(0,0)$  to  $(i,j)$  in the matrix  $(w_{k,l})$  is noted  $c(i,j)$ . The path is made of steps of size one where only one component increases. Then one observes that  $c(i,j)$  is equal to the time when the  $i$ th job leaves the  $j$ th server. This equality illustrates the fact that server  $j$  can process job  $i$  if it has already processed job  $i-1$  and if job  $i$  has left queue  $j-1$ .

The problem of queues in series can thus be modeled by an infinite matrix, where we assume from now on that the entries are independent identically distributed random variables, with finite variance. For the main theorem and for Section 3 the distribution is assumed to be geometric with parameter  $q$ . The aim of the talk [1] is to link the queue problem to the distribution of the largest eigenvalues of random Hermitian matrix with appropriate distribution. An infinite matrix of weights is also a model for a physical problem, the interacting particle process, see [7]; there we assume that all the integers corresponds to sites that are capable of containing one particle, and that at first all the sites with negative positions are full. In this model the weight  $w_{i,j}$  is the time taken by a particle to move from  $i$  to  $i+j$ .

A preliminary remark links this queue problem to Brownian motion [3]. Given  $(B_k)_{k=1,2,\dots}$  independent standard Brownian motions, and  $D_k^{(n)} = \frac{c(k,n) - en}{\sqrt{vn}}$ , where  $e$  is the expectation of  $w_{1,1}$  and  $v$  its variance, the following theorem holds:

**Theorem 1.** *The processes  $D^{(n)} = \left(D_k^{(n)}\right)_{k=1,2,\dots}$  converge in law as  $n \rightarrow \infty$  to the stochastic process  $D = (D_k)_{k=1,2,\dots}$  where  $D_k = \sup_{0=t_0 < t_1 < \dots < t_k=1} \sum_{i=0}^{k-1} (B_i(t_{i+1}) - B_i(t_i))$ .*

This can easily be seen by modeling a path from  $(0, 0)$  to  $(k, N)$  for large  $N$  by  $k$  long vertical lines, where the path uses the  $i$ th vertical line from the time  $t_i$  to the time  $t_{i+1}$ .

We now introduce the Gaussian Unitary Ensemble (GUE) [8] as the probability distribution on the Hermitian matrices with the density  $r_{\text{GUE}}(H) = Z^{-1} e^{-\text{tr} H^2/2}$  where  $Z$  is a normalizing constant equal to  $\int e^{-\text{tr} H^2/2} dH$ . A useful property is that a Hermitian matrix  $H$  is drawn from GUE if  $\Re(h_{ij})$  and  $\Im(h_{ij})$  are i.i.d. Gaussian random variables with mean 0 and variance 1. Given a matrix  $H$ , let  $H_k = (h_{(i,j)})_{1 \leq i, j \leq k}$  be the main minor of size  $k$  of  $H$  and  $\sigma_k$  the largest eigenvalue of  $H_k$ .

**Theorem 2.** *The laws of both processes  $\sigma = \{\sigma_k\}$  for  $H$  drawn from GUE, and  $D = \{D_k\}$  coincide.*

This theorem is proven in the next sections. We first exhibit a bijection between a finite restriction of size  $M$  of the queue problem and a subspace of  $\mathbb{N}^{M(M+1)/2}$  via Young tableaux. The second part of the proof is to relate this subspace of  $\mathbb{N}^{M(M+1)/2}$  to the dominant eigenvalues of minors of the matrix  $H$ .

## 2. Combinatorics

The bijection between the matrix  $H$  and Young tableaux is a generalization of the Robinson–Schensted–Knuth correspondence (see [5]) between Young tableaux and permutations. The matrix  $W$  of size  $N \times M$  with coefficients the weights  $w_{i,j}$  can be represented as a generalized permutation  $\alpha$ ,

$$\alpha = \begin{pmatrix} i_1 & i_2 & \dots & i_k \\ j_1 & j_2 & \dots & j_k \end{pmatrix},$$

where  $i_l \in \mathbb{N}_N$ , the integers between 1 and  $N$ ,  $j_l \in \mathbb{N}_M$  and  $j_l$  represents  $\alpha(i_l)$ . The integers  $i_l$  are not necessarily distinct, this is why the permutation is said to be generalized. The number of columns of type  $\binom{i}{j}$  is equal to  $w_{i,j}$ . As the generalized permutation  $\alpha$  is written in a lexicographically sorted fashion, the bijection is quite obvious. Indeed, given a matrix, the set of columns is well defined, and sorting gives the uniqueness of the image; conversely given a generalized permutation, one simply has to count the numbers of columns of type  $\binom{i}{j}$  to reconstruct the matrix. Recall that a *Young diagram*  $\lambda$  is a decreasing sequence  $(\lambda_1, \lambda_2, \dots, \lambda_r)$  that can be represented as  $r$  rows of boxes of heights  $\lambda_i$ . A *semi-standard Young tableau* is a filling of the boxes by positive integers such that the filling is increasing rightwards in rows and strictly increasing in columns. The Young diagram underlying a Young tableau  $P$  is called the shape and is denoted by  $\text{sh}(P)$ . The Robinson–Schensted–Knuth (RSK) correspondence is a bijection between the set of generalized permutations with  $k$  columns and the set of pairs of semi-standard Young tableaux  $(P, Q)$  having the same shape  $\lambda$  such that  $\sum_i \lambda_i = k$ . With the previous bijection between matrices and generalized permutations, we thus have a bijection between matrices of size  $N \times M$  and pairs of semi-standard Young tableaux  $(P, Q)$  with shape  $\lambda$ , such that  $\sum_i \lambda_i = k$ . We denote  $(P(w), Q(w))$  Young tableaux obtained by the RSK correspondence from a matrix  $w$ .

Some properties make this correspondence really interesting. We denote by  $W_{N,M,k}$  the set of matrices of size  $N \times M$  whose coefficients add up to  $k$ , to state a result on the way the distribution of the Young tableaux behaves through this correspondence.

**Lemma 1.** *If the set  $W_{N,M,k}$  is given the uniform distribution, then the distribution of  $P(w)$  for  $w \in W_{N,M,k}$  given  $\text{sh}(P(w)) = \lambda$  is uniform on the semi-standard Young tableaux of shape  $\lambda$ .*

Moreover the shape  $\text{sh}(P)$  encodes a few characteristics of  $w$ : the length  $\lambda_1$  of the first row is the maximal weight of the monotonous paths from  $(0, 0)$  to  $(M, N)$  in the table  $w$ , because it is the length of the longest increasing subsequence of the second line of  $\alpha$ . This is a direct consequence of the RSK algorithm. It is then possible to consider the Young tableau  $P$  as an embedding of several Young tableaux  $P_1, \dots, P_M$  where  $P_M$  equals  $P$  and  $P_i$  is obtained from  $P_{i+1}$  by removing all boxes filled with  $i + 1$ . A nice consequence of the way the tableau  $P$  is built is that the sequence of the lengths of the first rows of the embedded Young tableaux coincides with the sequence of maximal paths from  $(0, 0)$  to  $(i, N)$  as  $i$  goes from 1 to  $M$ . Thus we now focus on Young tableaux instead of the weight matrix. We introduce a representation of the Young tableaux that encodes the shape of the tableaux, and also the shape of all the embedded tableaux inside. To describe a Young tableau  $P$  filled with  $\mathbb{N}_M$ , let  $x_j^i$  be the coordinate of the rightmost box filled with a number at most  $i$  in the  $j$ th row of the tableau. Equivalently, this is just the length of the  $j$ th row in the tableau  $P_{i-1}$  defined by the embedding above. The elements  $x_j^i, 1 \leq i \leq M, 1 \leq j \leq i$  can be seen as a triangular array of size  $\frac{M(M+1)}{2}$ . The image  $x$  of a tableau  $P$  by this transformation has the property that its last line is equal to  $\lambda = (\lambda_1, \dots, \lambda_M)$ , and that its first column is equal to  $c(k, N), k \in \mathbb{N}_M$ , corresponding to the length of maximal paths from  $(0, 0)$  to  $(k, N)$ . This correspondence is formalized in the following lemma.

**Lemma 2.** *Let the Gelfand–Cetlin cone  $C_{GC}$  be the set of triangular arrays  $(x_j^i)$  of size  $\frac{M(M+1)}{2}$  such that  $x_{j-1}^i \geq x_{j-1}^{i-1} \geq x_j^i$  for  $1 \leq i \leq M, 1 \leq j \leq i$ . Then the Young tableaux filled with  $\mathbb{N}_M$  are in one-to-one correspondence with the integer points in the Gelfand–Cetlin cone.*

What we have now is a mapping from  $W_{N,M,k}$ , the set of matrices of size  $N \times M$  whose coefficients add up to  $k$ , to the set of integers point in the Gelfand–Cetlin cone. This mapping has the property that if  $W_{N,M,k}$  is given the uniform distribution, then the distribution of  $x$ , given that the last line is equal to  $\lambda = (\lambda_1, \dots, \lambda_M)$ , is uniform on the integers points of the Gelfand–Cetlin cone.

### 3. Gaussian Unitary Ensemble

From now on we add the restriction that the distributions of the coefficients  $w(i, j)$  are i.i.d. geometric with parameter  $q$ , that is  $\mathbf{P}(w_{i,j} = k) = (1 - q)q^k$ . All the results of the previous section still hold in this context, as they were obtained in full generality. The aim of this section is to finish the proof of Theorem 2. For this we describe the distribution of  $x(w)$  by its distribution conditioned upon values of its last line, and the distribution of its last line. This is then linked to the distribution of the limit processes  $D_k$ .

The probability that the RSK correspondence applied to a random matrix  $w$  with i.i.d. geometric entries with parameter  $q$  yields a pair of Young tableaux of shape  $\lambda = (\lambda_1, \dots, \lambda_M)$  is

$$(1) \quad \frac{(1 - q)^{MN}}{M!} \prod_{j=0}^{M-1} \frac{1}{j!(N - M + j)!} \prod_{1 \leq i < j \leq M} (\lambda_i - \lambda_j - i + j)^2 \prod_{i=1}^M \frac{(\lambda_i + N - i)!}{(\lambda_i + M - i)!} q^k$$

where  $k = \sum \lambda_i$ . The proof of this formula can be found in [4] and is based on the fact that there are  $\prod_{1 \leq i < j \leq M} \frac{\lambda_i - \lambda_j - i + j}{j - i}$  semi-standard tableaux of shape  $\lambda$  filled with  $\mathbb{N}_N$ . The vector of centered and normalized variables  $\xi_i = \frac{\lambda_i - eN}{\sqrt{vN}}$  (with  $e$  the average and  $v$  the variance of  $w_{1,1}$ ), is noted  $\xi$ . Plugging the Stirling approximation in Equation (1) yields that for fixed  $q$  such that  $0 < q < 1$ , fixed  $M$  and  $N \rightarrow \infty$ , the distribution of  $\xi$  converges weakly to  $Z^{-1} \prod_{i < j} (\xi_i - \xi_j)^2 \prod_i e^{-\xi_i^2/2}$  ( $Z$  is a normalizing constant). This is the distribution of the vector of ordered eigenvalues of a random matrix drawn from GUE. This property leads to the following theorem:

**Theorem 3.** *The distribution of the sequence  $(D_1, \dots, D_M)$  defined in Theorem 1 is the distribution of the first column of the random triangular array  $x$  of size  $M(M+1)/2$  distributed uniformly for a fixed last line, and the distribution of its last line is the distribution of the eigenvalues of matrices drawn from GUE [6].*

The distribution of  $(D_1, \dots, D_M)$  is the same as the distribution of the first column of  $x(w)$ , up to a proper normalization. And if the distribution on  $w$  is uniform, then the distribution on  $x$ , knowing its last line, is uniform. As the probability of getting a Young tableau of shape  $\lambda$ , and thus an array  $x$  of last line  $\lambda$ , is the same as getting the vector of ordered eigenvalues of a random matrix drawn from GUE equal to  $\lambda$ , Theorem 3 is proved.

The Gelfand–Cetlin polyhedron  $GC(\lambda)$  is defined as a subset of  $C_{GC}$  such that the last line of the array is equal to  $\lambda = (\lambda_1, \dots, \lambda_M)$ . Theorem 3 means that the distribution of  $(D_1, \dots, D_M)$  is uniform on  $GC(\lambda)$ .

This allows us to state the theorem below, which is a major step in the proof of Theorem 2.

**Theorem 4.** *Let  $H = (h_{ij})$ ,  $i, j \leq M$  be a random matrix drawn from GUE with eigenvalues  $(\lambda_1, \dots, \lambda_M)$ , and*

$$x(H) = \begin{pmatrix} x_1^1 & & & & \\ \cdots & \cdots & & & \\ x_1^{M-1} & \cdots & x_{M-1}^{M-1} & & \\ x_1^M & x_2^M & \cdots & x_M^M & \end{pmatrix}$$

where  $x_j^i$  is the  $j$ th eigenvalue of the main minor of size  $i$  of  $H$ . Then the triangular array  $x(H)$  is uniformly distributed in the polyhedron  $GC(\lambda)$ .

The proof of this theorem is based on the fact that the last line of the array  $x$  corresponds to the eigenvalues of the matrix  $H$ , which is drawn from GUE, and that given this last line, the distribution is uniform. The last line of  $x$  is equal to  $\lambda$  and its first column to the vector  $(\sigma_k)$  by definition. This together with Theorems 3 and 4 proves Theorem 2.

Theorem 2 can be used to prove the conjecture of Peter Glynn and Ward Witt [3] stating that  $D_k/\sqrt{k} \rightarrow 2$ . The proof uses the already known fact that the largest eigenvalue of random Hermitian matrix drawn from GUE rescaled by  $\sqrt{k}$  converges in distribution to 2, see [2].

### Bibliography

- [1] Baryshnikov (Yu.). – GUEs and queues. *Probability Theory and Related Fields*, vol. 119, n° 2, 2001, pp. 256–274.
- [2] Geman (Stuart). – A limit theorem for the norm of random matrices. *Annals of Probability*, vol. 8, n° 2, 1980, pp. 252–261.
- [3] Glynn (Peter W.) and Whitt (Ward). – Departures from many queues in series. *Annals of Applied Probability*, vol. 1, n° 4, 1991, pp. 546–572.
- [4] Johansson (Kurt). – Shape fluctuations and random matrices. *Communications in Mathematical Physics*, vol. 209, n° 2, 2000, pp. 437–476.
- [5] Knuth (Donald E.). – Permutations, matrices, and generalized Young tableaux. *Pacific Journal of Mathematics*, vol. 34, 1970, pp. 709–727.
- [6] Kuperberg (Greg). – Random words, quantum statistics, central limits, random matrices, September 2000. To appear in *Methods Appl. Anal.*. Available from <http://front.math.ucdavis.edu/math.PR/9909104>.
- [7] Liggett (Thomas M.). – *Interacting particle systems*. – Springer-Verlag, New York, 1985, xv+488p.
- [8] Mehta (M. L.). – *Random matrices and the statistical theory of energy levels*. – Academic Press, New York, 1967, x+259p.