On the Distribution of a Saddle Point Value in a Random Matrix

M. Hofri*
Department of Computer Science, WPI
100 Institute Road, Worcester MA 01609-2280
e-mail:hofri@cs.wpi.edu

September 21, 2006

1. Introduction

In [1] Hofri & Jacquet presented an analysis of algorithms to locate saddle points in a random matrix, asked by Donald E. Knuth in exercise 1.3.2-12 in *The Art of Computer Programming*.

A random matrix has (fixed) dimensions of n rows and m columns, and its elements are random variables, denoted by X, assumed drawn independently from the same given completely continuous distribution F. To remove ambiguity we sometimes write F_X.

A saddle point is defined as a matrix entry which is the minimal in its row and the maximal in its column, using sharp inequalities, which also imply that at most one could exist in a matrix.

Brief reflection suggests two results:
(1) With a continuous distribution the very *existence* of a saddle point is extremely unlikely.

(2) The distribution of the saddle point value, which we denote by R, when there is one, is very strongly "pinched" around its mean.

The purpose of this note is to flesh out those observations, especially the second.

2. Calculations of moments

The distribution of the saddle point size is immediate from it definition above, leading to the probability density function (conditional on the value being a saddle point), $r_c(x) = f(x)F(x)^{m-1}(1-F(x))^{n-1}$, where

*This work was prepared while the author was with the ALGO project at INRIA, Rocquencourt during his sabbatical from WPI. I am very grateful for the hospitality the institute and the group accorded me.

f is the probability density function of X. This is enough to provide us with

Lemma 1.: (a) The probability density function of the size of a saddle point is given by

$$r(x) = \frac{1}{P_{sp}} f(x) F(x)^{m-1} (1 - F(x))^{n-1}, \tag{1}$$

(b) The probability of an $n \times m$ matrix with independent entries drawn from a continuous distribution to have a saddle point is mnP_{sp}, where P_{sp} is the probability that a saddle point occurs at a given position, and is given by

$$P_{sp} = \frac{m+n}{mn\binom{m+n}{n}} = B(m,n), \tag{2}$$

where $B(m,n)$ is the beta function.

Proof. The integral of the unconditional density, which is given by $r_c(x)/P_{sp}$, over the entire support of $F(x)$ needs to be 1; we compute it with the change of integration variable $F(x) \to u$:

$$1 = \int r(x)dx = \frac{1}{P_{sp}} \int f(x)F(x)^{m-1}(1-F(x))^{n-1} dx = \frac{1}{P_{sp}} \int_0^1 u^{m-1}(1-u)^{n-1} du = \frac{B(m,n)}{P_{sp}}.$$

This value for P_{sp} was obtained by Knuth in [2], using purely combinatorial considerations. □

Note: if F is not continuous, then quite different, and even strange things can happen. In particular, the probability P_{sp} may depend not only on the geometry as above, but also on details of the distribution F. Here is an extreme case: X has only three possible values $a < b < c$, with probabilities $p,\ q,\ r = 1 - p - q$, respectively, then a saddle point can have only the value b, and happen with probability $P_{sp} = mnqp^{m-1}r^{n-1}$.

2.1 Using the Laplace method

The next evidence we calculate about the random variable R is its moments about the origin. Such moments require evaluating the integrals

$$E[R^i] = \frac{1}{B(m,n)} \int x^i f(x) F(x)^{m-1} (1 - F(x))^{n-1} dx \tag{3}$$

with the integration extending on the support of $F(x)$. When i is a positive integer there is no immediate integral in general. Some particular cases are of course interesting, and can be computed exactly.

Example: If F is the uniform distribution, which we may limit to (0,1), the integration is indeed immediate, and we find that $E[R] = m/(m+n)$. Then, consistent with the notion of a "pinched" distribution, the variance of R is smaller than that of X by an order of n: for $X \sim U[0,1]$ we find $V[R] = \frac{mn}{(m+n)^2(m+n+1)}$.

While Eq.(3) is not amenable usually to closed form evaluation, its appearance suggests immediately the use of the Laplace method, which we should expect to be very precise so long as we could assume the distribution F is smooth and the dimensions large. More precisely, we are looking now at

$$E[R^i] = \frac{1}{B(m,n)} \int x^i f(x)F(x)^{m-1}(1-F(x))^{n-1}\,dx \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} g_i(x)e^{-nh(x)}dx, \tag{4}$$

where

$$g_i(x) = \frac{x^i f(x)}{B(m,n)F(x)(1-F(x))}; \qquad h(x) = -\alpha \ln F(x) - \ln(1-F(x)), \tag{5}$$

and we make a natural, but in fact arbitrary choice, to let $m = \alpha n$, so that the two dimensions of the matrix grow at the same rate. The Laplace method calls for evaluating the minimum of $h(x)$. Two differentiations of $h(x)$ and solving for x in $h'(x) = 0$ reveal that the minimum position, x_0, is given by

$$h'(x) = f(x)\left(\frac{1}{1-F(x)} - \frac{\alpha}{F(x)}\right) \Longrightarrow x_0 = F^{-1}\left(\frac{\alpha}{1+\alpha}\right); \quad h''(x_0) = f(x)^2\left(\frac{1}{(1-F(x_0))^2} + \frac{\alpha}{F(x_0)^2}\right) > 0. \tag{6}$$

Then we have the known first-order Laplace method result

$$E[R^i] \sim g_i(x_0)\, e^{-nh(x_0)} \sqrt{\frac{2\pi}{n h''(x_0)}} \tag{7}$$

with

$$h(x_0) = -\alpha \ln\frac{\alpha}{1+\alpha} - \ln\frac{1}{1+\alpha} = \ln\frac{(1+\alpha)^{1+\alpha}}{\alpha^\alpha}; \quad g_i(x_0) = \frac{x_0^i(1+\alpha)^2 f(x_0)}{\alpha B(\alpha n, n)}, \quad h''(x_0) = f(x_0)^2\frac{(1+\alpha)^3}{\alpha}. \tag{8}$$

It only remains to evaluate the ratio

$$\eta \stackrel{\text{def}}{=} e^{-nh(x_0)}/B(\alpha n, n) = \left(\frac{\alpha^\alpha}{(1+\alpha)^{1+\alpha}}\right)^n \times n\binom{m+n-1}{n}. \tag{9}$$

This was done using the leading term in the Stirling approximation for the binomial coefficient, providing $\eta \approx \sqrt{n\alpha/2\pi(1+\alpha)}$. When we substitute in Eq. (7) we get the compact result that $E[R^i] \sim x_0^i$.

This is somewhat too compact: it fits some numerical evidence for $E[R]$, but for the variance it only claims that it is in $o(1)$, some lower, unspecified order of n^{-1}. That is the effect of the pinching process; it compares well with the result we obtain for the uniform distribution, but we want more specific asymptotic information about the process in a more general situation. We can show the following.

Theorem 1. Let R be the size of the saddle point value in an $n \times m$ matrix of independent entries satisfying the probability law $F(x)$ with a completely continuous distribution, then its first two moments are given to first order in n^{-1} as follows:

$$E[R] \sim x_0 - \frac{1}{n}\frac{\alpha F''(x_0)}{2(1+\alpha)^3 F'(x_0)^3}, \qquad E[R^2] \sim x_0^2 - \frac{1}{n}\frac{\alpha(x_0 F''(x_0) - F'(x_0))}{(1+\alpha)^3 F'(x_0)^3}. \tag{10}$$

Proof sketch: To obtain such estimates a more delicate calculation then the first-order Laplace method, that produces Eq. (7) is called for. The main contribution comes from refining the Laplace method, and we also added lower order terms to the ratio η.

To improve the "Laplace" estimate we make a change of integration variable, defined by $h(x) - h(x_0) \to u^2$. This keeps the convenience of working with Gaussian integrals. The main computational difficulty is then solving for x in the equation $\sqrt{h(x) - h(x_0)} = u$. This approach is essentially the same as presented in [3, §3.8], and we did it with MAPLE, which allows us to experiment with higher order terms than one can reasonably do by hand. When MAPLE is presented with such an equation, in which the left-hand side is a power expansion around x_0 (specifically, of the MAPLE type "series"), it solves for x by reversing the series and producing $x = x_0 + \sum_j a_j u^j$. The number of terms in this expansion is determined according to the desired order of the final result, since the integration converts it to an expansion in powers of $n^{-1/2}$. For the value of $r(x_0)$ we added terms to the asymptotic development of the ratio $\eta = e^{-nh(x_0)}/P_{sp}$, from Eq.(9) beyond the first one given above. This expansion underlies many of our calculations, and can be seen in the Appendix, in Eq.(20).

Since the integrand is given in terms of the unspecified distribution function $F(x)$, the expansions, and the terms a_j in solution series for x in terms of u, are all expressed via the derivatives of this function, all evaluated at the convenient point x_0.

Notes: (1) The "$-$" signs following x_0 or x_0^2 in Eq. (10) were chosen for convenience and do not mean the moments are smaller than these values. Each of the following terms there can assume any sign; the first derivative of F is a density, hence positive, but the second derivative is not so obvious. In particular, for a unimodal function, at least somewhat symmetrical distribution, we would expect $F''(x_0)$ to be positive for $\alpha < 1$ and negative for higher values, but things could be different.
(2) We calibrated our calculations by performing a similar computation with with $g_0(x)$ of Eq. (5). It should provide, naturally, 1, but as we use fairly short expansions, since we only wanted few leading coefficients, we used enough terms so that the lowest order term in our estimate of $E[R^0]$ beyond 1 it produced was n $\Theta(n^{-3})$ (which corresponds to even higher-order "error terms" in the higher order moments).

Several technical details are involved in the calculation; we posted the MAPLE program we used, with some comments. at http://www.cs.wpi.edu/∼hofri/maple1 . □

Once we have these moments we can calculate the variance, and find that as expected it is indeed of a lower order in n than $V[X]$:

Corollary 1.: *Under the conditions of Theorem 1 the variance of R is given by*

$$V[R] = \frac{1}{n}\frac{\alpha}{(1+\alpha)^3 F'(x_0)^2} + O\left(\frac{1}{n^2}\right). \tag{11}$$

The explicit coefficient of the n^{-2} term in $V[R]$ is given in the Appendix, in Eq. (18). □

This result corresponds to the intuition, that for moderate matrices the distribution of any saddle point value gets pinched around its expected value, and tends to a degenerate random variable as the matrix grows. Note moreover that the leading term in $V[R]$ does not even "acknowledge" the variance of the elements: it depends on the geometry (it is equal to $1/nh''(x_0)$, and is closely related to the curvature of the function $h(x)$ at the critical point).

3. Computing the distribution

Since the leading term in $V[R]$ is in $\theta(n^{-1})$, we should expect the random variable $\sqrt{n}R$ to have a finite variance, and the question arises: what is the shape of this scaled distribution, and in particular, what is the functional form of its tail probabilities, that give the likelihood of deviation from the mean. We define the *nearly-centered** random variable $T \stackrel{\text{def}}{=} \sqrt{n}(R-x_0)$; in this way, we let x_0 essentially stand for $E[R]$, but this choice of centering means that 0 is the mode of T, but it is its expectation only to zero-order in n^{-1}.

While $E[T]$ is in $\theta(n^{-1/2})$, its variance and standard deviation, σ_T—the natural unit for the scale of deviations which appear most informative—are in $\theta(1)$, we want to handle both small deviations, in the scale of the expectation, and more informative deviations, counted in a few σ_T units.

We can cover both needs by computing

$$t(x) \stackrel{\text{def}}{=} \Pr[T > x] = \Pr[\sqrt{n}(R-x_0) > x] = \Pr\left[R > x_0 + \frac{x}{\sqrt{n}}\right] \qquad x \in O(1). \tag{12}$$

With the density available from Eq.(1) we can write an expression for $t(x)$

$$t(x) = \frac{1}{P_{sp}} \int_{t \geq x_0 + x/\sqrt{n}} f(t)F(t)^{m-1}(1-F(t))^{n-1}\,dt. \tag{13}$$

This is even less inviting than the previous integral, because now the lower limit of the integration depends on n in a way that breaks the standard applicability of the Laplace method, except in the extreme case, when we take $x = 0$. The integral we then compute is

$$t(x) = \int_{t \geq x_0} g(t)e^{-nh(t)}\,dt, \tag{14}$$

with $h(t)$ the same as given in Eq. (5), and $g(t)$ the same as $g_0(x)$ there. We have the same integral as in Eq. (4) for $i = 0$, but extending only on the positive real line. The calculation is similar (except that we

*Normally random variables are centered at their expectation, and we would need to use $T' \stackrel{\text{def}}{=} \sqrt{n}(R-E[R])$, but the centrality of the point x_0 in our calculations as the mode of the distribution of R, the fact that all developments are in terms of the expansion of $F_X(x)$ at this point, and the nearness of x_0 to $E[R]$, make this a natural choice.

cannot use symmetry to cut on the number of contributing terms), and we find the reasonable result

$$t(0) = \Pr[T \geq 0] = \frac{1}{2} + \frac{(\alpha-1)}{3\sqrt{2n\pi\alpha(1+\alpha)}} + O\left(n^{-3/2}\right). \tag{15}$$

Two further terms are given in the appendix, in Eq. (19). Note that for a square matrix $\alpha = 1$, and x_0 is at the median of the entry distribution. Then the mean, mode and median of T are all zero, exactly.

In general we need to deal with the integral given in Eq. (13), which we transform as in the proof of Lemma 1 to the simpler looking integral

$$t(x) = \frac{1}{B(m,n)} \int_{u=v}^{1} u^{m-1} (1-u)^{n-1} du = \frac{1}{B(m,n)} \int_{u=v}^{1} g(u) e^{-nh(u)} du, \qquad v \stackrel{\text{def}}{=} F(x_0 + x/\sqrt{n}). \tag{16}$$

The simplicity consists in restricting the appearance of the arbitrary distribution $F_X(x)$ to the specification of the lower integration limit. Otherwise we continue as when refining the Laplace method above, using MAPLE at every step of the way to obtain

Theorem 2. With the conditions of Theorem 1, the limiting tails of the distribution of the size of the scaled saddle point element T are Gaussian, for $n \to \infty$:

$$\lim_{n\to\infty} t(x) = 1 - \Phi\left(\frac{x}{\sigma_T}\right) \implies T \to \mathcal{N}(0, \sigma_T^2). \tag{17}$$

In §4.4 of the Appendix we provide the ingredients of the related asymptotic expansion. Note that it is asymptotic in $n \to \infty$, while we keep x in $O(1)$.

Proof sketch: The pattern of the following steps should now be familiar, if not all the details:
(1) Define the functions $h(u) = -\alpha \ln u - \ln(1-u)$ and $g(u) = 1/[B(\alpha n, n) u(1-u)]$,
(2) Observe that $h(u)$ is minimized at $x_0 = \alpha/(1+\alpha)$, and define the new integration variable y through the relation $h(u) - h(x_0) = y^2$.
(3) Obtain the series solution of this equation, as $u = s(y)$, and use it to express $g(u)$ and du/dy, to yield an integrand in terms of y, denoted by $I(y)$. The point $y_1 = \sqrt{h(v) - h(x_0)}$ is the lower limit for integration on y; the upper limit is infinity, from $\lim_{u\to 1}(-\ln[u^{\alpha}(1-u)])$.
(4) Evaluate the integral $\int_{y \geq y_1} I(y) e^{-ny^2} dy$ to get an expression in terms of $[1 - \operatorname{erf}(\sqrt{n} y_1)]$ and terms multiplied by $e^{-ny_1^2}$. Denote the result by I1, and multiply by η from Eq. (20) for the complete value.
(5) Express the integration limit v to second order as $v = x_0 + fx/\sqrt{n} + F'' x^2/(2n)$ where the derivatives are evaluated at x_0 as well.
(6) Substitute y_1 in the results of step 4 while controlling the order of the expressions in n. Letting $n \to \infty$ produces the above result, since $\lim_{n\to\infty} \sqrt{n} y_1 = x/\sigma_T \sqrt{2}$, and $1 - \operatorname{erf}(t) = 2(1 - \Phi(t\sqrt{2}))$.

The MAPLE script for the above is in http://www.cs.wpi.edu/∼hofri/maple2 . □

Acknowledgments

Encouragement to deal with this question was provided by Boris Pittel. Bruno Salvy and Frédéric Chyzak provided guidance with MAPLE that was crucial for the calculations. Philippe Flajolet commented on earlier versions of the note.

4. Appendix

In order to simplify the presentation, the formula does not show explicitly that all functions are evaluated at $x_0 = F^{-1}(\alpha/(1+\alpha))$; f stands for F'.

4.1 Expansion of $V[R]$ to order n^{-2}

$$V[R] = \frac{1}{n}\frac{\alpha}{(1+\alpha)^3 f^2} - \alpha\frac{2f^4(1+\alpha)^2 + 2\alpha f F^{(3)} + F''(4f^2 - 4\alpha^2 f^2 - 7\alpha F'')}{2(1+\alpha)^6 f^6 n^2}. \tag{18}$$

4.2 Expansion of Eq. (15) to order $n^{-5/2}$

$$t(0) \sim \frac{1}{2} + \frac{(\alpha-1)}{3\sqrt{2\pi\alpha(1+\alpha)}\,n^{1/2}} + \frac{(\alpha-1)(\alpha^2+25\alpha+1)}{540(\alpha(1+\alpha))^{3/2}\sqrt{2\pi}n^{3/2}} - \frac{(\alpha-1)(1+\alpha+\alpha^2)(25\alpha^2+73\alpha+25)}{6048(\alpha(1+\alpha))^{5/2}\sqrt{2\pi}n^{5/2}}. \tag{19}$$

4.3 The value of the density function $r(x)$ at its mode

$$\begin{aligned} r(x_0) = f(x_0)\frac{(1+\alpha)^2}{\alpha}\eta = f(x_0)\frac{(1+\alpha)^2}{\alpha}\sqrt{\frac{n\alpha}{2\pi(1+\alpha)}}\Bigg(1 - \frac{1+A}{12An} + \frac{(1+A)^2}{288A^2n^2} \\ + \frac{417A + 139A^3 - 15A^2 + 139}{51840A^3 n^3} - \frac{(1+A)(571A^3 - 15A^2 + 1713A + 571)}{2488320A^4 n^4} \\ - \frac{(163879A^3 - 331755A^2 + 491637A + 163879)(1+A)^2}{209018880A^5 n^5} + O\left(n^{-6}\right)\Bigg) \end{aligned} \tag{20}$$

where the symbol A stands for $\alpha(1+\alpha)$.

4.4 Tails of the distribution of $T = \sqrt{n}(R - x_0)$

We give the results as produced for Theorem 2. Refer to the sketch of the proof for terminology. We found it more useful to give the 'building blocks' of the result, since putting it all together provides on paper unwieldy, and not particularly meaningful expressions.

$$y1 = \sqrt{\frac{(1+\alpha)^3}{2\alpha}} f \frac{x}{\sqrt{n}} + \frac{(1+\alpha)^{3/2}\left(3F''\alpha + 2f^2(\alpha^2-1)\right)}{6\sqrt{2\alpha^3}} \frac{x^2}{n} + \frac{(1+\alpha)^{5/2}\left(12F''\alpha(\alpha-1) + 2f^2\alpha(1+\alpha) + 7f^2(1+\alpha^3)\right)}{36\sqrt{2}\alpha^{5/2}} f \frac{x^3}{n^{3/2}} + O\left(n^{-2}\right). \tag{21}$$

Note that the first term of y_1 is $x/\sigma_T\sqrt{n}$.

$$I1 = \frac{1}{\sqrt{n}}\left\{\frac{1-\operatorname{erf}(\sqrt{n}y_1)}{\sqrt{2}}\left(\sqrt{\frac{\pi(1+\alpha)}{\alpha}} + \frac{1}{12n}\frac{\sqrt{\pi}(1+\alpha+\alpha^2)}{\sqrt{\alpha^3(1+\alpha)}}\right) + \frac{e^{-ny_1^2}}{540(1+\alpha)\alpha^2} \times \left[\frac{8}{n^{3/2}}(\alpha-1)(\alpha+2)(2\alpha+1) + \frac{1}{\sqrt{n\alpha(1+\alpha)}}\left(8y_1^2\sqrt{\alpha(1+\alpha)}(\alpha-1)(\alpha+2)(2\alpha+1) + 45\alpha\sqrt{2}y_1(1+\alpha)(1+\alpha+\alpha^2) + 180\sqrt{\alpha^3(1+\alpha)}(\alpha^2-1)\right)\right]\right\} \tag{22}$$

Finally, we note that the value of η was read from Eq.(20) and that $(1-\operatorname{erf}(t)) = 2(1-\Phi(t\sqrt{2}))$.

References

[1] M. Hofri, Ph. Jacquet: Saddle Points in Random Matrices: Analysis of Knuth Search Algorithms. *Algorithmica* **22**, #4 516–528, (1998).

[2] D. Knuth, *The Art of Computer Programming*, Addison-Wesley, volume 1: Fundamental Algorithms, 1968. 3rd. Ed. 1997.

[3] F. W. J. Olver, *Asymptotics and Special Functions*, Academic Press, New York, 1974. Reprinted with corrections, A. K. Peters, Natick Mass., 1997.

[4] B. Pittel: Personal communication, 1998, 2005.