Probability in the Engineering and Informational Sciences (2022), **36**:4 1098–1102
doi:10.1017/S0269964821000267

CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

On the distribution of winners' scores in a round-robin tournament

Yaakov Malinovsky

Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD 21250, USA.
E-mail: yaakovm@umbc.edu.

Keywords: Extremes, Negative correlation

MSC2020: 62G32, 05C20

Abstract

In a classical chess round-robin tournament, each of n players wins, draws, or loses a game against each of the other $n-1$ players. A win rewards a player with 1 points, a draw with 1/2 point, and a loss with 0 points. We are interested in the distribution of the scores associated with ranks of n players after $\binom{n}{2}$ games, that is, the distribution of the maximal score, second maximum, and so on. The exact distribution for a general n seems impossible to obtain; we obtain a limit distribution.

1. Introduction

In a classical chess round-robin tournament, each of n players wins, draws, or loses a game against each of the other $n-1$ players. A win rewards a player with 1 points, a draw with 1/2 point, and a loss with 0 points. Denoting by X_{ij} the score of the player i after the game with the player j, $j \neq i$, in this article, we consider the following model:

Model M:

For $i \neq j$, $X_{ij} + X_{ji} = 1$, $X_{ij} \in \{0, 1/2, 1\}$; we assume that all players are equally strong, that is, $P(X_{ij} = 1) = P(X_{ji} = 1)$, and that the probability of a draw is the same for all games, denoted by $p = P(X_{ij} = 1/2)$. We also assume that all $\binom{n}{2}$ pairs of scores $(X_{12}, X_{21}), \ldots, (X_{1n}, X_{n1}), \ldots, (X_{n-1,n}, X_{n,n-1})$ are independent.

Let $s_i = \sum_{j=1, j\neq i}^{n} X_{ij}$ be a score of the player i $(i = 1, \ldots, n)$ after playing with $n-1$ opponents. We use a standard notation and denote by $s_{(1)} \leq s_{(2)} \leq \cdots \leq s_{(n)}$ the order statistics of the random variables $s_1, s_2, \ldots, s_n$, and further denote normalized scores (zero expectation and unit variance) by $s_1^*, s_2^*, \ldots, s_n^*$ with the corresponding order statistics $s_{(1)}^* \leq s_{(2)}^* \leq \cdots \leq s_{(n)}^*$.

For the case where there are no draws, that is, $X_{ij} \in \{0, 1\}, X_{ij} + X_{ji} = 1, p_{ij} = P(X_{ij} = 1) = \frac{1}{2}$, Huber [7] proved that

$$s_{(n)}^* - \sqrt{2\log(n-1)} \longrightarrow 0$$

in probability as $n \longrightarrow \infty$ (see also [12]), where *log(x) is the logarithm of x, to base e*. The main step in his proof was establishing the following inequality (Lemma 1 in [7]):

$$P(s_1 < k_1, \ldots, s_m < k_m) \leq P(s_1 < k_1) \cdots P(s_m < k_m)$$

for any probability matrix (p_{ij}) and any numbers $(k_1, \ldots, k_m)$, $m \leq n$.

Malinovsky and Moon [11] extended Huber's lemma for the case with draws and a general p_{ij}, where $X_{ij} \in \{0, 1/2, 1\}, X_{ij} + X_{ji} = 1, p_{ij} = P(X_{ij} = 1)$ (Model H1), and for the case of a symmetric distribution where $X_{ij} \in \{1, 2, \ldots, m\}$, $X_{ij} + X_{ji} = m$, $p_u = P(X_{ij} = u) = p_{m-u}$, $u = 0, 1, \ldots, m$ (Model H2). As a byproduct, in Models H1 and H2, similar convergence in the probability of the normalized maximal score holds. Ross [13] found some bounds on the distribution of the number of wins of the winning team for a related model.

In this work, we are interested in the marginal distribution of the scores associated with the ranks of n players after $\binom{n}{2}$ games under Model M, where rank 1 is the winner's rank, rank 2 is the second best, and so on. This means that we are interested in finding the marginal distribution of $s_{(i)}$. The exact distribution for a general n seems impossible to obtain; we obtain a limit distribution, and demonstrate it with the three best scores in Model M.

2. Asymptotic distribution

Under Model M, we have the following properties of the scores $s_1, s_2 \ldots, s_n$ that satisfy $s_1 + s_2 + \cdots + s_n = n(n-1)/2$:

(a) $E_n = E(s_1) = (n-1)/2, \sigma_n = \sigma(s_1) = \sqrt{(n-1)(1-p)/4}$,
(b) $\rho_n = \mathrm{corr}(s_1, s_2) = -1/(n-1)$,
(c) The random variables $s_1, s_2, \ldots, s_n$ are exchangeable.

From the multivariate central limit theorem (e.g., see [10]), it follows that when $n \to \infty$, the joint distribution of $s_1, s_2, \ldots, s_n$ is multivariate normal. Therefore, the problem of finding the limit distribution of $s_{(i)}$ reduces to the problem of the distribution of the ith largest term in the multivariate normal vector.

The investigation of the distribution of maximum and extreme elements of i.i.d. random variables has a long history (see, e.g., the books by [3,4,6]). In the case of dependent random variables, the fundamental work of Berman [1] will lead to a limit distribution of $s_{(n)}$, and using the results presented in Leadbetter *et al.* [9], we can obtain a limit distribution of $s_{(i)}$. Due to the properties of a multivariate normal distribution and its particular correlation structure under Model M, the limit distribution of the maximal score is identical to the limit distribution of the corresponding independent random variables. The intuition for this "surprising" phenomenon was initially explained by Markov and Bernstein, as is described in Leadbetter [8].

2.1. Maximal score

First, we present the result from Berman [1] and therefore have to introduce the definition of a stationary sequence of random variables (see, e.g., [9]).

Definition 2.1. *A sequence of random variables is called a stationary sequence if the distributions of the vectors $(X_{j_1}, \ldots, X_{j_n})$ and $(X_{j_1+s}, \ldots, X_{j_n+s})$ are identical for any choice of n, $j_1, \ldots, j_n$, and s.*

Theorem 2.1 (Berman[1]). *Let $X_0, X_1, X_2, \ldots$ be a stationary Gaussian sequence with $E(X_1) = 0$, $E(X_1^2) = 1$, $E(X_0 X_j) = r_j$ for $j \geq 1$ and let sequences $\{a_n\}$ and $\{b_n\}$ be defined as*

$$a_n = (2\log n)^{-1/2}, \quad b_n = (2\log n)^{1/2} - \tfrac{1}{2}(2\log n)^{-1/2}(\log\log n + \log 4\pi). \tag{1}$$

If $\lim_{n\to\infty} r_n \log n = 0$ or $\sum_{n=1}^{\infty} r_n^2 < \infty$, then

$$\lim_{n\to\infty} P(X_{(n)} \leq a_n t + b_n) = e^{e^{-t}} \equiv G(t) \quad (\text{"Gumbel"}),$$

for all t.

https://doi.org/10.1017/S0269964821000267 Published online by Cambridge University Press

We obtain:

Result 2.1.

$$\lim_{n\to\infty} P(s^*_{(n)} \le a_n + b_n t) = G(t).$$

Proof. The random variables $s_1, s_2, \ldots, s_n$ are exchangeable random variables and therefore stationary. In our case, $\mathrm{corr}(s_1, s_2) = -1/(n-1)$ and $\lim_{n\to\infty}(1/(n-1))\log(n) = 0$, and therefore, Berman's Theorem 2.1 holds for Model M. Combining the multivariate central limit theorem with Theorem 2.1, we obtain the limiting distribution of $s^*_{(n)}$. □

Corollary 2.1.

$$E(s_{(n)}) \sim \frac{n-1}{2} + \sqrt{\frac{(n-1)\log(n)(1-p)}{2}} + \sqrt{\frac{(n-1)(1-p)}{2\log(n)}}\left\{\frac{\gamma}{2} - \frac{1}{4}(\log\log(n) + \log(4\pi))\right\} \equiv \hat{E}_{(n)}$$

$$\sigma(s_{(n)}) \sim \frac{\pi}{4\sqrt{3}}\sqrt{\frac{(n-1)(1-p)}{2\log(n)}} \equiv \hat{\sigma}_{(n)},$$

where $\gamma = 0.5772156649\ldots$ *is the Euler constant and* $a_n \sim b_n$ *means* $\lim_{n\to\infty} a_n/b_n = 1$.

Proof. The moments under the distribution function G can be obtained based on the following consideration. If $Y_1, \ldots, Y_n$ are independent $\exp(1)$ random variables, then straightforward calculation shows (see, e.g., [5]):

$$\lim_{n\to\infty} P(Y_{(n)} - \ln(n) \le t) = G(t), \tag{2}$$

and for $r = 1, 2, \ldots, n$,

$$(n+1-r)(Y_{(r)} - Y_{(r-1)})$$

are independent exponential random variables with rate parameter 1, where $Y_{(0)}$ is defined as zero. Since

$$Y_{(k)} = Y_{(1)} + (Y_{(2)} - Y_{(1)}) + \cdots + (Y_{(k)} - Y_{(k-1)}), \tag{3}$$

we obtain that

$$E(Y_{(n)}) = \sum_{j=1}^{n} \frac{1}{j}, \quad \mathrm{Var}(Y_{(n)}) = \sum_{j=1}^{n} \frac{1}{j^2}.$$

From $\lim_{n\to\infty}\{\sum_{j=1}^{n} 1/j - \log(n)\} = \gamma$, $\lim_{n\to\infty} \sum_{j=1}^{n} 1/j^2 = \pi^2/6$ (see, e.g., [2]), and (2), we obtain the expectation and variance under the distribution function G as $E_G = \gamma$, $\mathrm{Var}_G = \pi^2/6$. Combining this with Result 2.1, we have

$$E(s^*_{(n)}) \sim \gamma b_n + a_n, \sigma(s^*_{(n)}) \sim \sqrt{\frac{\pi^2}{6}}\, b_n. \tag{4}$$

Then, upon substituting $s^*_{(n)} = (s_{(n)} - E_n)/\sigma_n$, Corollary 2.1 follows. □

In Table 1, we compare $E(s_{(n)})$ with $\hat{E}_{(n)}$ and $\sigma(s_{(n)})$ with $\hat{\sigma}_{(n)}$ in this manner: We fix $p = 2/3$ and for $n = 10$, 20, 50, 100, 1,000, and 10,000 we evaluate $E(s_{(n)})$ and $\sigma(s_{(n)})$ using Monte-Carlo (MC) simulation. Values of $\hat{E}_{(n)}$ and $\hat{\sigma}_{(n)}$ obtained based on Corollary 2.1.

https://doi.org/10.1017/S0269964821000267 Published online by Cambridge University Press

Table 1. *The number of Monte-Carlo repetitions is 100,000 for* $n = 10, 20, 50, 100; 10,000$ *for* $n =$ 1,000*; and* 500 *for* $n = 10{,}000$.

n	$E(s_{(n)})$	$\hat{E}_{(n)}$	$\lvert\hat{E}_{(n)}/E(s_{(n)}) - 1\rvert * 100\%$	$\sigma(s_{(n)})$	$\hat{\sigma}_{(n)}$	$\lvert\hat{\sigma}_{(n)}/\sigma(s_{(n)}) - 1\rvert * 100\%$
10	5.833	5.912	1.360	0.469	0.518	10.454
20	11.89	11.944	0.456	0.627	0.659	5.189
50	29.08	29.162	0.283	0.912	0.927	1.563
100	56.73	56.843	0.199	1.219	1.214	0.426
1,000	529.12	529.352	0.044	3.259	3.148	3.529
10,000	5110.23	5111.295	0.0212	8.949	8.626	3.742

Table 2. *The number of Monte-Carlo repetitions is 100,000 for* $n = 10, 20, 50, 100; 10,000$ *for* $n =$ 1,000*; and* 500 *for* $n = 10{,}000$*;* $r_j = \lvert\widehat{E}_{(j)}/E(s_{(j)}) - 1\rvert * 100\%$*,* $j = n-1, n-2$.

n	$E(s_{(n-1)})(\sigma(s_{(n-1)}))$	$\widehat{E}_{(n-1)}(\widehat{\sigma}_{(n-1)})$	$E(s_{(n-2)})(\sigma(s_{(n-2)}))$	$\widehat{E}_{(n-2)}(\widehat{\sigma}_{(n-2)})$	$r_{(n-1)}$	$r_{(n-2)}$
10	5.400 (0.338)	5.509 (0.324)	5.093 (0.273)	5.307 (0.254)	2.009	4.195
20	11.305 (0.446)	11.43 (0.413)	10.95 (0.374)	11.173 (0.323)	1.106	2.037
50	28.277 (0.649)	28.44 (0.580)	27.816 (0.541)	28.079 (0.454)	0.576	0.946
100	55.695 (0.858)	55.896 (0.760)	55.113 (0.712	55.423 (0.595)	0.361	0.563
1,000	526.48 (2.154)	526.9 (1.971)	525.05 (1.764)	525.67 (1.543)	0.080	0.118
10,000	5103.2 (5.866)	5104.6 (5.401)	5099.5 (4.672)	5101.2 (4.227)	0.027	0.033

2.2. *Second and third largest scores*

Result 2.2. *For* $j = 1, \ldots, n-1$,

$$\lim_{n\to\infty} P(s^*_{(n-j)} < a_n + b_n t) = G(t)(1 + e^{-t} + \cdots + e^{-jt}/j!),$$

where a_n *and* b_n *are defined in (1).*

Proof. Follows from combining the multivariate central limit theorem with Theorems 4.5.2. and 5.3.4. in Leadbetter *et al.* [9]. □

Corollary 2.2.

$$E(s^*_{(n-1)}) \sim \gamma b_n + a_n - b_n, \quad \sigma(s^*_{(n-1)}) \sim \sqrt{\left(\frac{\pi^2}{6} - 1\right)}\, b_n, \tag{5}$$

$$E(s^*_{(n-2)}) \sim \gamma b_n + a_n - 3/2 b_n, \quad \sigma(s^*_{(n-2)}) \sim \sqrt{\left(\frac{\pi^2}{6} - 1.25\right)}\, b_n \tag{6}$$

Proof. Combing (2) with Theorem 2.2.2 in Leadbetter *et al.* [9], we obtain the following result: if $Y_1, \ldots, Y_n$ are independent exp(1) random variables, then for $j = 1, \ldots, n-1$

$$\lim_{n\to\infty} P(Y_{(n-j)} - \log(n) \le t) = G(t)(1 + e^{-t} + \cdots + e^{-jt}/j!). \tag{7}$$

The rest of the proof is similar to the proof of Corollary 2.1. □

https://doi.org/10.1017/S0269964821000267 Published online by Cambridge University Press

Substituting $s^*_{(j)} = (s_{(j)} - E_n)/\sigma_n$ for $j = n-1, n-2$, we obtain the values $E(s_{(j)}), \sigma(s_{(j)}), \widehat{E}_{(j)}, \widehat{\sigma}_{(j)}$, which are similar to the corresponding values obtained in Corollary 2.1 for the case $j = n$.

In the case where $p = 2/3$, we provide (in Table 2) numerical comparisons for the second and third largest scores in a similar manner as was done in Table 1.

Acknowledgments. The author thank Abram Kagan for describing a score issue in chess round-robin tournaments with draws. The author also thank the Editor and anonymous referee for the comments.

Funding

This research was supported by grant no. 2020063 from the United States – Israel Binational Science Foundation (BSF), Jerusalem, Israel.

Conflict of interest. The authors declare no conflict of interest.

References

[1] Berman, S.M. (1964). Limiting distribution of the maximum of a diffusion process. *Annals of Mathematical Statistics* 35: 319–329.
[2] Courant, R. & Robbins, H. (1996). *What is mathematics? An elementary approach to ideas and methods*, 2nd ed., revised by Ian Stewart. New York: Oxford University Press.
[3] Cramér, H. (1946). *Mathematical methods of statistics*. Princeton: Princeton University Press.
[4] Galambos, J. (1987). *The asymptotic theory of extreme order statistics*. 2nd ed. Malabar, FL: Krieger.
[5] Grimmett, G.R. & Stirzaker, D.R. (2020). *Probability and random processes*, 4th ed. Oxford: Oxford University Press.
[6] Gumbel, E.J. (1958). *Statistics of extemes*. New York: Columbia University Press.
[7] Huber, P.J. (1963). A remark on a paper of Trawinski and David entitled: Selection of the best treatment in a paired comparison experiment. *Annals of Mathematical Statistics* 34: 92–94.
[8] Leadbetter, M.R. (2017). Extremes under dependence–historical development and parallels with central limit theory. In *Extreme events in finance*, Wiley Handb. Finance Eng. Econom. Edited by Francois Longin. Hoboken: Wiley, pp. 11–23.
[9] Leadbetter, M.R., Lindgren, G., & Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. New York-Berlin: Springer-Verlag.
[10] Lehmann, E.L. (1998). *Elements of large-sample theory*. New York: Springer.
[11] Malinovsky, Y. & Moon, J.W. (2021). On the negative dependence inequalities and maximal score in round-robin tournament. https://arxiv.org/abs/2104.01450.
[12] Moon, J.W. (2013). Topics on tournaments. [Publicaly available on website of Project Gutenberg https://www.gutenberg.org/ebooks/42833].
[13] Ross, S.M. (2021). Team's seasonal win probabilities. *Probability in the Engineering and Informational Sciences*. doi:10.1017/S026996482100019X. In press.

Cite this article: Malinovsky Y (2022). On the distribution of winners' scores in a round-robin tournament. *Probability in the Engineering and Informational Sciences* **36**, 1098–1102. https://doi.org/10.1017/S0269964821000267