The predictive power of noisy round-robin tournaments

Dmitry Ryvkin[1] and Andreas Ortmann[2]

CERGE-EI
Charles University and Academy of Sciences of the Czech Republic
Politických vězňů 7, 11121 Prague, Czech Republic

December 2003

Abstract

The round-robin tournament format for N players is a scheme that matches players with one another in all possible $N(N-1)/2$ pairwise comparisons. A noisy round-robin tournament adds to the particular matching scheme performance fluctuations of the participating players, or noise. With noise, upsets become possible and hence the possibility that the ex ante best player is not the winner ex post. Thus, noise reduces the predictive power of a tournament. In this article we study theoretically (analytically and by way of computational simulations) the predictive power of a noisy round-robin tournament for three prominent distributions of players' abilities, as a function of the level of noise, and the number of players. At first sight, some of our results (e.g., non-monotonicity as a function of the number of players N and the noise level σ, which can make some ranges of N non-optimal) are quite counterintuitive but should be of help to a tournament designer who tries to maximize, or maybe minimize, the probability of the best player winning.

Keywords: round-robin tournaments, noise, power distributions, design economics
JEL Classification: C73, C90, D21

[1]dmitry.ryvkin@cerge-ei.cz
[2]andreas.ortmann@cerge-ei.cz or aortmann@yahoo.com

1 Introduction

Agents (individuals or teams) are usually rewarded based on their performance. Often it is the relative performance that matters. As a means of assessing the relative performance of agents, principals extensively use tournaments.

A *tournament* is a procedure that ranks a set of agents. Such a ranking can be done in various ways. Prominent examples are "contests", round-robin tournaments, and elimination tournaments. Contests are essentially one-shot all-pay auctions whose properties have been widely discussed in the literature [e.g., theoretically by Lazear and Rosen (1981), Green and Stokey (1983), Taylor (1995), Hvide (2002); empirically by Knoeber and Thurman (1994); experimentally by Schotter and Weigelt (1992), Gneezy *et al* (2001); see also reviews by Lazear (1999) and Prendergast (1999)]. They can be thought of as a complete set of pairwise comparisons where every agent performs, however, only once. In contrast, round-robin tournaments and elimination tournaments are polar cases of schemes that compare agents pairwise and sequentially in various degrees of completeness. Any of these schemes allows agents to perform repeatedly, typically against a stream of ever changing other agents, or competitors. A round-robin tournament is a complete sequential pairwise matching scheme.

Sport provides us with a simple and useful language to describe tournaments. Competitors are typically called *players*, and pairwise comparisons are called *matches*. In every match, there is a *winner* and a *loser*, or there is a tie. Below we will use this terminology.

In the presence of environmental randomness ("noise") which affects the performance of players either positively or negatively, a tournament can be thought of as a probabilistic device whose output - the ranking - is a statistic of sorts of the "true" ordering of the set of players. Such an ordering identifies who is the best player ex ante.

In the present paper, we analyze theoretically (analytically and by way of computational simulations), the properties of round-robin tournaments as a function of noise level, number of players, and distribution of players' abilities. In later papers, we will analyze similarly elimination tournaments, and compare the properties of these two polar matching schemes, as well as variants thereof.

The investigation of round-robin tournaments is an important step towards solving the problem of optimal design [see Moldovanu and Sela (2002) for a theoretical study of "contest architecture"; see more generally Roth (2002), for a discussion of how simulations with artificial agents and experiments with human subjects serve to extend simple theoretical models, which become too complex in engineering-like situations].

In economics, round-robin tournaments have been discussed in the context of public choice models such as voting schemes and decision rules in committees [see, e.g., Levin and Nalebuff (1995), Ben-Yashir and Nitzan (1997), Esteban and Ray (2001)].

In mathematics, round-robin tournaments have been studied as complete directed graphs [see Harary and Moser (1966) for a review; Moon and Pullman (1970) for a discussion of tournament matrices]. Importantly, Rubinstein (1980) shows that the ranking that assigns 1 point to the winner and 0 to the loser of a match, and then sums up every player's points across all matches he or she played, is a very "good" ranking scheme in the sense that it satisfies certain natural axioms. Rubinstein (1980) also shows that it is *the only* such scheme. Below we will make use of this result.

Our paper contributes to both literatures by analyzing theoretically the probabilistic properties of a round-robin tournament and by studying these properties across what we consider the most prominent distributions of player abilities.

The present paper is organized as follows: In Section 2 we present the general setup. In Section 3 we discuss a model underlying the key ingredient of our model, winning probabilities. In Section 4 we calculate analytically the predictive power. In Section 5 we discuss the individual distribution of scores which we use in the approximate calculation of the predictive power for large N for which the theory developed in Section 4 becomes computationally too time consuming. In Section 6 we present and discuss our results. We conclude in Section 7.

2 General setup

Let $\mathcal{P} = \{1, \dots, N\}$ be a set of N players. A *round-robin tournament* on the set $\mathcal{P}$ consists of all $M = N(N-1)/2$ possible pairwise matches (i, j), $1 \le i < j \le N$, of the players.

Every match (i, j) has one of the two allowed outcomes: either $i \to j$ (player i defeated player j) or $j \to i$ (player j defeated player i) [3]. We introduce a variable p_{ij} such that

$$p_{ij} = 1 \quad \text{if} \quad i \to j, \quad p_{ij} = 0 \quad \text{otherwise} \tag{1}$$

Since the ordering of matches is not important (all matches are assumed to be statistically independent) we will, for convenience, adopt the following (lexicographic) ordering $m = 1, \dots, M$:

$$\begin{array}{cccccccccc} (i,j): & (1,2) & \dots & (1,N) & (2,3) & \dots & (2,N) & \dots & (N-1,N) \\ \downarrow & \downarrow & & \downarrow & \downarrow & & \downarrow & & \downarrow \\ m(i,j): & 1 & \dots & N-1 & N & \dots & 2N-2 & \dots & M, \end{array}$$

which can be described by

$$m(i,j) = N(i-1) - \frac{i(i+1)}{2} + j. \tag{2}$$

The outcome of the tournament can then be represented as an M-bit binary number $\mathbf{b} = \langle b_1 \dots b_M \rangle$, where $b_{m(i,j)} = p_{ij}$. There are 2^M possible tournament outcomes, ranging from $\langle 00 \dots 0 \rangle$ to $\langle 11 \dots 1 \rangle$.

The result of the tournament is an N-dimensional vector of scores $\mathbf{s} = (s_1, \dots, s_N)$ where every player's score is the number of wins she has, i.e. players start with zero scores and get 1 point for each win. Since every player plays $N-1$ matches and the total number of matches is M, for any score vector $\mathbf{s}$

$$\textstyle\sum_{i=1}^{N} s_i = M, \quad 0 \le s_i \le N-1. \tag{3}$$

If the outcome of the tournament $\mathbf{b}$ is known, then the score vector can be calculated directly, using the function $\mathbf{S} : \{0,1\}^M \to \{0, \dots, N-1\}^N$ from the set of all M-bit binary strings (outcomes) into the set of possible score vectors:

$$S_i(\mathbf{b}) = \textstyle\sum_{j=1}^{i-1} (1 - b_{m(j,i)}) + \sum_{j=i+1}^{N} b_{m(i,j)}, \quad i = 1, \dots, N. \tag{4}$$

The winners of the tournament are the players with maximal scores. There may be more than one such player, and then an additional rule (perhaps another tournament) has to be applied if one

[3] Note that we ignore the possibility of ties, which admittedly is, compared to elimination tournaments, a somewhat realistic occurence in round-robin schemes. We did so because the literature [e.g., Rubinstein (1980)] does not provide us with a consistent point counting scheme. Introducing ties would have required us to build a different model. We believe that such a model would not materially affect the results that we are interested in: the effect of noise and distribution of players' abilities.

needs to determine the best among them. In the present paper, we do not adopt any additional rules and consider all the players with a maximal score as winners.

Assume now that the outcomes of the matches are random, i.e. in a match (i, j) the result is $i \to j$ with some probability w_{ij} and $j \to i$ with probability $w_{ji} = 1 - w_{ij}$. Then every p_{ij} becomes a Bernoulli random variable with the probability of success w_{ij}, and the tournament outcome $\mathbf{b}$ becomes a multivariate Bernoulli vector with independent components. The probability of every outcome is

$$P(\mathbf{b}) = \prod_{i=1}^{N-1} \prod_{j=i+1}^{N} w_{ij}^{b_{m(i,j)}} (1 - w_{ij})^{1-b_{m(i,j)}}. \tag{5}$$

Also, the score vector $\mathbf{s}$ becomes a random vector satisfying conditions (3). The distribution of $\mathbf{s}$ is non-trivial since its components are not independent. The probability density function (pdf) of $\mathbf{s}$ can be written as

$$\pi(\mathbf{s}) = \sum_{\mathbf{b}\in\{0,1\}^M} P(\mathbf{b}) \boldsymbol{\delta}[\mathbf{s} - \mathbf{S}(\mathbf{b})]. \tag{6}$$

Here the $\boldsymbol{\delta}$-function is M-dimensional; summation goes over all M-bit binary numbers; $\mathbf{S}(\mathbf{b})$ is determined by Eq. (4); $P(\mathbf{b})$ is given by Eq. (5).

We are interested in the probability ρ_1 for a specific player (player 1, for concreteness) to be among the winners of the tournament, i.e. to have a maximal score. We will calculate this probability both analytically and using computational simulations. Along the way, some other important quantities will be calculated.

An additional remark on the winning probabilities w_{ij} should be made. They can be given exogenously, through past statistics, rating data, or another tournament. Alternatively, we can calculate the winning probabilities using a simple model in which it is assumed that every player has a *power level* (or ability) which can be represented by a single real number. The power level is assumed to be distributed in the population of players with some known pdf $f(\cdot)$. The randomness is introduced through the assumption that in every match the power levels of players are distorted by additive noise, whose pdf $g(\cdot)$ is known. It is then possible to calculate the average winning probabilities w_{ij}, where the indices indicate the *ordering* of players by their power level.

For actual calculations, we have chosen three prominent distributions of players' abilities, or power levels: uniform, normal, and Pareto distributions. The uniform and normal distributions are useful, and frequently used, benchmarks and need no further justification as such. Empirical evidence [e.g., Reed (2001); see also Hertwig et al. (1999), or Harrison (2004)] suggests in addition that the Pareto distribution is a widespread and pervasive phenomenon. In the next section we present the general model that allows to calculate w_{ij}.

3 Winning probabilities

Let $f(x)$ be the pdf of the power levels in the population of players. Suppose N players are drawn from that population and ordered by their power levels $x_1, ..., x_N$ so that $x_1 \geq x_2 \geq ... \geq x_N$.

Consider an arbitrary match (i, j). The performance levels of the players in this match will be random numbers $Y_i = x_i + \epsilon_i$ and $Y_j = x_j + \epsilon_j$, where ϵ's represent the noise, which is i.i.d. across players and across matches with pdf $g(\epsilon)$. Since (x_i, x_j) are fixed numbers at this point, the Y's will be distributed with pdf's $h_i(y|x_i) = g(y - x_i)$ and $h_j(y|x_j) = g(y - x_j)$. Therefore, the probability for player i to be the winner in this match is

$$\tilde{w}(x_i, x_j) \equiv \Pr\{Y_i - Y_j \geq 0 | x_i, x_j\} = \int_0^\infty dy \int_{-\infty}^\infty dy' g(y + y' - x_i) g(y' - x_j). \tag{7}$$

Particularly, if $g(\epsilon)$ is normal with zero mean and variance σ^2, one obtains

$$\tilde{w}(x_i, x_j) = \Phi\left(\frac{x_i - x_j}{\sigma\sqrt{2}}\right), \tag{8}$$

where $\Phi(\cdot)$ is the cumulative standard normal density.

Now we can average over all possible realizations of $x_1, ..., x_N$ such that $x_1 \geq x_2 \geq ... \geq x_N$, to get the average probability for player ranked i to win player ranked j:

$$w_{ij} = N! \int_{-\infty}^{\infty} dx_1 f(x_1) \int_{-\infty}^{x_1} dx_2 f(x_2) ... \int_{-\infty}^{x_{N-1}} dx_N f(x_N) \tilde{w}(x_i, x_j). \tag{9}$$

Here $(N!)^{-1} = \Pr\{x_1 \geq x_2 \geq ... \geq x_N\} = \int_{-\infty}^{\infty} dx_1 f(x_1) \int_{-\infty}^{x_1} dx_2 f(x_2) ... \int_{-\infty}^{x_{N-1}} dx_N f(x_N)$ is the renormalization denominator, which arises because we fixed a specific permutation of x's.

The N-dimensional integral in Eq. (9) can be reduced (see Appendix A) to give

$$\begin{aligned} w_{ij} &= \frac{N!}{(i-1)!(j-i-1)!(N-j)!} \int_{-\infty}^{\infty} dx_i f(x_i) \int_{-\infty}^{x_i} dx_j f(x_j) \tilde{w}(x_i, x_j) \\ &\times [1 - F(x_i)]^{i-1} [F(x_i) - F(x_j)]^{j-i-1} [F(x_j)]^{N-j}. \end{aligned} \tag{10}$$

4 The predictive power

Our main objective is to calculate the probability ρ_1 for a specific player (player 1) to be among the winners of the tournament, i.e. to have a maximal score. For a score vector $\mathbf{s}$ this can be expressed as

$$\rho_1 = \Pr\{(s_1 \geq s_2) \& \ldots \& (s_1 \geq s_N)\}. \tag{11}$$

By introducing variables $q_1 = s_1, q_2 = s_1 - s_2, ..., q_N = s_1 - s_N$, we need to require that they all be non-negative, i.e. $\rho_1 = \Pr\{q_1 \geq 0, ..., q_N \geq 0\}$. The transformation $\mathbf{s} \leftrightarrow \mathbf{q}$ has a unitary Jacobian, with the inverse transfromation being $s_1 = q_1, s_2 = q_1 - q_2, ..., s_N = q_1 - q_N$. Therefore the joint pdf of $\mathbf{q} = (q_1, ..., q_N)$ is just $\pi(q_1, q_1 - q_2, ..., q_1 - q_N)$, and the probability that all q_i are non-negative is

$$\rho_1 = \int_0^{\infty} dq_1 ... \int_0^{\infty} dq_N \pi(q_1, q_1 - q_2, ..., q_1 - q_N), \tag{12}$$

where function π is given by Eq. (6).

Then we write

$$\begin{aligned} \rho_1 &= \sum_{\mathbf{b} \in \{0,1\}^M} P(\mathbf{b}) \int_0^{\infty} dq_1 ... \int_0^{\infty} dq_N \delta[q_1 - S_1(\mathbf{b})] \delta[q_1 - q_2 - S_2(\mathbf{b})] \ldots \delta[q_1 - q_N - S_N(\mathbf{b})] \\ &= \sum_{\mathbf{b} \in \{0,1\}^M} P(\mathbf{b}) H[S_1(\mathbf{b}) - S_2(\mathbf{b})] \ldots H[S_1(\mathbf{b}) - S_N(\mathbf{b})]. \end{aligned} \tag{13}$$

Here $H(z)$ is the step function defined as 1 for $z \geq 0$ and 0 for $z < 0$. The result is very intuitive: we sum over all mutually exclusive tournament outcomes and add up the probabilities of those of them for which the player 1's score is maximal.

In Appendix C we describe computational simulations of the round-robin tournament and demonstrate that they agree with the analytical result (13) for moderate N. For large N, Eq. (13) becomes inapplicable practically, since its computational time grows as $2^{N(N-1)/2}$. In the following chapter we develop an asymptotic theory that allows to approximately calculate ρ_1 much faster for large N. We then compare the exact and asymptotic analytical results with the results of direct simulations.

In the limit of $\sigma \to \infty$ (for a fixed number of players N) all the winning probabilities $w_{ij} \to \frac{1}{2}$ independent of the ability distribution $f(\cdot)$. The predictive power therefore has a limiting behavior $\rho_1 \to \rho_1^\infty(N)$, where

$$\rho_1^\infty(N) = \frac{1}{2^M} \sum_{\mathbf{b}\in\{0,1\}^M} H[S_1(\mathbf{b}) - S_2(\mathbf{b})] \dots H[S_1(\mathbf{b}) - S_N(\mathbf{b})]. \tag{14}$$

For example, $\rho_1^\infty(2) = \rho_1^\infty(3) = \frac{1}{2}$, $\rho_1^\infty(4) = \frac{13}{32}$.

5 Individual distribution of scores

The joint distribution of scores is given by Eq. (6). That equation implies a summation over all 2^M binary strings of length M, whose number grows as $\sim \exp(N^2/2)$ for large N. This circumstance makes Eq. (6) practically inapplicable for $N \sim 10$ and larger[4]. In this Section we develop an asymptotic theory that is based on the fact that the correlation among scores s_i gets weaker as N increases. Indeed, for arbitrary pair of scores s_i,s_j

$$|\mathrm{Cov}(s_i, s_j)| = (w_{ij} - w_{ij}^2) \ll \mathrm{Var}(s_i) = \sum\nolimits_{k\neq i}(w_{ik} - w_{ik}^2), \tag{15}$$

so that the correlation coefficient between them drops as $\sim 1/N$. Of course, the fact that the pairwise correlation is weak does not imply that the correlation among all scores is weak. Conditions (3) ensure that the latter correlation stays finite for arbitrary N. However, we can hope (and the calculations support this hope) that since the quantity of our interest, ρ_1, only refers to one player, it is influenced less and less by scores of players that are ranked low.

Thus, in our approximation, the scores of individual players are independent. This situation can be thought of as a modified round-robin setting, in which every match is conducted twice, but every player has a chance to score in only one of the two matches. For example, imagine a round-robin soccer tournament, in which every team plays one home game with every other team, but only the host team can score (i.e., if the host team wins it gets 1 point, but if it loses both teams get 0).

We will now calculate the individual distribution of score for player i. This calculation is exact regardless the assumption about score independence made above. We will use the independence assumption later when the *joint* distribution of scores will be approximated by a product of individual score distributions.

Throughout the tournament, every player i plays $N-1$ matches with players $\mathcal{P} \setminus i$. Every match (i,j) is a Bernoulli trial with the probability of success w_{ij}. In case of success, player i gets 1 point. In a match (i,j), the number of points player i gets can be represented by a discrete random variable p_{ij} [cf. Eq. (1)] with pdf $\phi_{ij}(p) = w_{ij}\delta(p-1) + (1-w_{ij})\delta(p)$.

The total score player i will get is $s_i = \sum_{j\neq i} p_{ij}$. Let $\mathcal{B}_N^{-i}$ denote a set of $(N-1)$-bit binary vectors $\mathbf{b} = (b_1, ..., b_{i-1}, b_{i+1}, ..., b_N)$, whose every component is 0 or 1. The pdf of the total score for player i can then be written as

$$\pi_i(s) = \sum_{\mathbf{b}\in\mathcal{B}_N^{-i}} \prod\nolimits_{j\neq i} [b_j w_{ij} + (1-b_j)(1-w_{ij})]\, \delta(s - \sum\nolimits_{j\neq i} b_j). \tag{16}$$

[4]In fact, we only used it for N up to 8.

It is clear that actually the pdf for s_i has the form

$$\pi_i(s) = \sum_{k=0}^{N-1} A_k^i \delta(s-k), \tag{17}$$

where A_k^i is the probability for player i to get k points. Eq. (16) allows to calculate those probabilities.

Theorem 1. Let $\mathbf{w}^{-i} = (w_{i1}, ..., w_{i,i-1}, w_{i,i+1}, ..., w_{iN})$ be a $(N-1)$-dimensional vector of winning probabilities for player i. Then the probability for player i to get k points in a round-robin tournament of N players is

$$A_k^i = \sum_{l=k}^{N-1} (-1)^{l-k} C_l^k Q_l(\mathbf{w}^{-i}). \tag{18}$$

The proof of the theorem, as well as the definition of polynomials Q_l, is given in the Appendix B.

Using the probabilities A_k^i and the independence approximation discussed above, one can write the joint distribution of scores approximately as

$$\tilde{\pi}(\boldsymbol{s}) = \prod_{i=1}^{N} \pi_i(s_i), \tag{19}$$

where $\pi_i(\cdot)$ is given by Eq. (17). Then using the same approach as in Secion 4 we obtain the approximate expression for the predictive power,

$$\rho_1 \approx \sum_{k_1,...,k_N=0}^{N-1} A_{k_1}^1 ... A_{k_N}^N H(k_1 - k_2) ... H(k_1 - k_N), \tag{20}$$

which has a computational time much smaller than Eq. (13).

6 Results and discussion

We first present, and later discuss, the results for the uniform distribution, then those for the Pareto distribution, and then those for the normal distribution because the Pareto distribution, at least for our purposes here, can be thought of as the upper half of a normal distribution (albeit a slowly falling one)

Figures 1 through 6 below illustrate the predictive power, ρ_1, as a function of noise level, number of players, and distribution of players' abilities. Specifically, we analyzed by way of computational simulations uniform, normal, and Pareto distributions of players' abilities, with the variance of all normalized to 1. Figures 1-3 illustrate predictive power as a function of noise level, for selected numbers of players (N = 2, 4, 8, 16, 32 or 64, 256) and for the three distributions. Figures 4-6 illustrate predictive power as a function of number of players, for selected noise levels, and for the three distributions. The computation of the Figures is detailed in Appendix C. [We restrict our computational explorations to those noise and number parameters that can be observed in "real life" (e.g., we are not aware of elimination tournaments that involve more than 256 individuals or teams).]

The following discussion of the key results draws on Figures 1-6:

1. For uniform distributions of abilities, predictive power is a monotonically decreasing function of noise and number of players. See Figures 1 and 4. The intuition for this result is straightforward: As the performance fluctuations of equally distanced (in terms of ability) players increase, the probability for upsets increases uniformly for all players (except, of course, the top-ranked player).

This result also implies, however, that the probability of beating a lower-ranked player decreases (except, of course, the lowest-ranked player), possibly neutralizing the net effect on the expected score. This neutralization will be complete only for the median player, and will happen partially for other "interior" players (i.e., players other than the top-ranked and lowest-ranked players). In fact, interior players to the right of the median player will experience a negative effect because relative to the median player there is more mass of the distribution to their left (and hence more chances of being upset, and less chances of upsetting). Likewise, interior players to the left of the median player will experience a positive effect because relative to the median player there is less mass of the distribution to their left (and hence less chances of being upset, and more chances of upsetting).

2. The previous result illustrates a fundamental principle: Noise is a redistributor of scores in that it gives from the "haves" to the "havenots". Obviously this distributional process is affected by the shape of the distribution of players' abilities.

3. Another observation which holds for all distributions is that for any given N the predictive power converges to a constant which is a function of N only (and not the distributional specification). In effect, this result is a straightforward application of combinatorics, as we demonstrated in Section 4: Note that for both $N = 2$ and $N = 3$, the predictive power converges to $1/2$ as the noise goes to infinity. For $N = 4$, the predictive power converges to $13/32$. This result sheds light on the asymptotic properties of predictive power when noise goes to infinity. However, for noise levels that are of relevance for practical purposes, these results are of lesser importance. The basic intuition is worth mentioning though: As noise increases, ability becomes less and less important in determing the outcome of a match, while chance becomes more and more important.

4. Our simulation suggests that the speed of convergence decreases in N.

5. For uniform distributions of abilities, competition at the top is higher than for normal and Pareto distributions (as it will be for all distributions with falling upper tail) and therefore the predictive power is lower. This is true for all noise levels, and increasingly true for increasing number of competitors. See Figure 1, as well as Figures 2-6. The intuition is clear: Interior players are symmetrically affected in that players indexed i and $N - i + 1$ have the same net effect in expected scores (albeit with opposite signs).

6. For Pareto distributions, players are differentially affected by increasing noise. Specifically, the probabilities of upsets will be lower for the top players for the simple reason that lower-ranked players are bunched more tightly and hence have more chances to score upsets. Of course, again, this good news is counterbalanced by the bad news of a decreased probability of beating a lower-ranked player, possibly neutralizing the net effect on the expected score. However, now the net effect of noise on the expected score is determined by the number of players in a more complicated (but ultimately still intuitive) fashion. Specifically, we get qualitatively different behavior for small N (such as 2, or 4) and large N (such as 256), with yet different behavior for intermediate N. In fact, our computations suggest that the qualitatively different behaviors for small N and large N are homotopic, with a critical $N = N_c$ that defines the switching point between monotonic and non-monotonic behavior. See Figure 2.

7. So, what intuition then drives the curious behavior of the predictive power for large N (such as 256, and already to some extent 32) in Figure 2?

As N increases the long tail of the Pareto distribution implies that relative to the low-ability players the top-ranked player moves away from the second-ranked player ever more. Therefore, for sufficiently small noise, the predictive power drops less for larger N. This trend is countered by a countervailing trend as noise increases. Let us distinguish the cases of small N and large

N. Obviously, as noise increases, so does the probability of upsets. However, this increase in probability is decreasing as we move up the ranking since the distance between players increases, on average. When N is small the increase in the probability of upsets is the only effect. When N is large then the top-ranked players score increases relatively to the lower-ranked players' because the lower-ranked a player is, the more ferocious competition that player faces. This result leads to the surprising and initially unintuitive upward swing in the predictive power as a function of noise for large N. In fact, the lowest-ranked player face the most ferocious competition because in that player's neighborhood of the ability distribution players are bunched the tightest.

8. For normal distributions, the intuition stemming from the Pareto distribution is a useful point of departure. Note that, ignoring for the time being the difference in tails, the Pareto distribution is in a sense the upper half of the normal distribution. Hence, we should expect qualitatively somewhat similar behavior in Figure 3. And indeed, for low noise (roughly up to $\sigma = 1$) we see very similar behavior. (Of course, the quantitative behavior differs somehow.) Now look at the median player in the normal distribution. As the performance fluctuations of every player increase, the probability for upsets increases (but not uniformly) for all players (except, of course, the top-ranked player). But note that this result also implies that the probability of beating a lower-ranked player decreases (except, of course, the lowest-ranked player), possibly neutralizing the net effect on the expected score. This neutralization will be complete only for the median player, and will happen partially for other "interior" players (i.e., players other than the top-ranked and lowest-ranked players). This, of course, is an argument similar to the one we made above for the case of the uniform distribution. Qualitatively, the net effect of this balancing act is negative for players to the right of the median player, and positive for players to the left of the median players, for the same rationale laid out for the uniform distribution above. The difference in behavior reflected in Figures 2 and 3 stems from the fundamental difference in the gestalt of the tails of the distribution.

7 Conclusion

In the present paper, we analyzed analytically and by way of computational simulations, the properties of round-robin tournaments as a function of noise level, number of players, and distribution of players' abilities.

A planner who has decided to conduct a round-robin tournament might benefit from our insights in a number of ways: If he or she knows, or at least has some inkling, about the distribution of abilities, and if the distribution happens to be one of the three that we studied, then the planner can estimate from the probabilities of upsets the noise level σ for our model. In fact, the planner could infer the properties of a distribution from the probabilities of upsets which, in principle, are oberservable. We note that these assumptions are not quite as far fetched as they seem. Certainly most professional sports have ranking schemes. In fact, some of these schemes are based on the statistics of upsets in various ways [e.g., in chess, table tennis, soccer, and American football)].

If the planner knows distribution and noise, then he can identify the optimal number of players. This decision, obviously, requires the definition of an objective function such as to maximize, or minimize, the probability of the best team wining. But once that decision has been made, our results suggest interesting choices. For the maximization case, as regards the Pareto distribution at $\sigma = 3$ and moderate N, it makes sense to decrease the number of participants. For the minimization case, with the same distribution but $\sigma = 1$, the number of players again should be decreased. The planner also wants to keep in mind that for large N ($\sim$32 or more) and non-uniform distributions, the predictive power depends non-monotonically on σ.

In later papers, we will analyze similarly elimination tournaments, and compare the properties of these two polar matching schemes, as well as variants thereof. We will also analyze the robustness of the present results to different (asymmetric, heteroscedastic) specifications of the noise. We conjecture that such modification will have quantitative effects but will not change the results reported here in any fundamental manner.

Appendix

A Winning probabilities

In this section of the Appendix we show how the expression (10) is obtained. The variables in the ordered integral in Eq. (9) are assumed to change in the following ranges:

$x_1 \in (-\infty, \infty)$, $x_2 \in (-\infty, x_1)$, ..., $x_N \in (-\infty, x_{N-1})$.

However, they can be rearranged so that

$x_i \in (-\infty, \infty)$, $x_j \in (-\infty, x_i)$,
$x_1 \in (x_i, \infty)$, $x_2 \in (x_i, x_1)$, ..., $x_{i-1} \in (x_i, x_{i-2})$,
$x_{i+1} \in (x_j, x_i)$, $x_{i+2} \in (x_j, x_{i+1})$, ..., $x_{j-2} \in (x_j, x_{j-1})$,
$x_{j+1} \in (-\infty, x_j)$, $x_{j+2} \in (-\infty, x_{x+1})$, ..., $x_N \in (-\infty, x_{N-1})$.

This reordering corresponds to the following integral:

$$\begin{aligned} w_{ij} &= N! \int_{-\infty}^{\infty} dx_i f(x_i) \int_{-\infty}^{x_i} dx_j f(x_j) \tilde{w}(x_i, x_j) \\ &\times \int_{x_i}^{\infty} dx_1 f(x_1) \int_{x_i}^{x_1} dx_2 f(x_2) \dots \int_{x_i}^{x_{i-2}} dx_{i-1} f(x_{i-1}) \\ &\times \int_{x_j}^{x_i} dx_{i+1} f(x_{i+1}) \int_{x_j}^{x_{i+1}} dx_{i+2} f(x_{i+2}) \dots \int_{x_j}^{x_{j-2}} dx_{j-1} f(x_{j-1}) \\ &\times \int_{-\infty}^{x_j} dx_{j+1} f(x_{j+1}) \int_{-\infty}^{x_{j+1}} dx_{j+2} f(x_{j+2}) \dots \int_{-\infty}^{x_{N-1}} dx_N f(x_N). \end{aligned}$$

In the last three lines, the integrals are easily calculated, and we obtain Eq. (10).

B Individual distribution of scores

In this section of the Appendix we obtain Eq. (18) for the probability of player i to get k points. Let $\mathbf{u} = (u_1, ..., u_n)$ be a vector of n real numbers, and

$$\begin{aligned} &Q_0^n(\mathbf{u}) = 1, \\ &Q_1^n(\mathbf{u}) = u_1 + ... + u_n, \\ &Q_2^n(\mathbf{u}) = u_1 u_2 + u_1 u_3 + ... + u_1 u_n + u_2 u_3 + ... + u_2 u_n + ... + u_{n-1} u_n, \\ &... \\ &Q_k^n(\mathbf{u}) = \sum_{j_1 < ... < j_k} u_{j_1} u_{j_2} ... u_{j_k}, \\ &... \\ &Q_n^n(\mathbf{u}) = u_1 u_2 ... u_n. \end{aligned} \tag{21}$$

Then we can formulate the following theorem.

Theorem 1. Let $\mathbf{w}^{-i} = (w_{i1}, ..., w_{i,i-1}, w_{i,i+1}, ..., w_{iN}) \in [0,1]^{N-1}$ be a vector of winning probabilities for player of *ex ante* rank i. Then the probability for player i to get j points in a round-robin tournament of N players is

$$A^i_j(N) = \sum_{k=j}^{N-1} (-1)^{k-j} C^j_k Q^{N-1}_k(\mathbf{w}^{-i}). \tag{22}$$

Proof. The only place i enters the right-hand side of Eq. (22) is through $Q^{N-1}_k(\mathbf{w}^{-i})$, but the polynomials Q^n_k are completely symmetric with respect to permutations of their arguments, therefore we only need to prove the theorem for any particular i, for example $i = 1$. We prove by induction over number of players N. The induction base, $N = 2$, is obvious [it can be directly calculated using Eq. (16)]. Now assume that Eq. (22) holds for N players. Suppose we add one more player, $N + 1$ (again, using the symmetricity of Q^n_k, we can always assume that the added player occupies the last ranking position), with the winning probability for player 1 over her being $w_{1,N+1}$. Then $A^1_j(N + 1)$ can be expressed as follows:

$$A^1_j(N + 1) = A^1_j(N)(1 - w_{1,N+1}) + A^1_{j-1}(N)w_{1,N+1}. \tag{23}$$

Indeed, there are the only two mutually exclusive ways to get j points: to get j points playing with the former $N - 1$ players and to lose to the new player (first term), and to get $j - 1$ points and win the new player (second term), respectively.

Note that the polynomials Q^n_k have the following property:

$$Q^{n+1}_k(u_1, ..., u_n, u_{n+1}) = Q^n_k(u_1, ..., u_n) + u_{n+1}Q^n_{k-1}(u_1, ..., u_n), \tag{24}$$

if we set $Q^n_{n+1} = Q^n_{-1} = 0$.

From Eq. (22) (that holds for N by the induction assumption) we have

$A^1_{j-1}(N) - A^1_j(N) = \sum_{k=j-1}^{N-1}(-1)^{k-j+1}C^{j-1}_k Q^{N-1}_k(\mathbf{w}^{-1}) - \sum_{k=j}^{N-1}(-1)^{k-j}C^j_k Q^{N-1}_k(\mathbf{w}^{-1}) =$
$= \sum_{k=j}^{N-1}(-1)^{k-j}[-C^{j-1}_k - C^j_k]Q^{N-1}_k(\mathbf{w}^{-1}) + Q^{N-1}_{j-1}(\mathbf{w}^{-1}) =$
$= \sum_{k=j}^{N-1}(-1)^{k-j+1}C^j_{k+1}Q^{N-1}_k(\mathbf{w}^{-1}) + Q^{N-1}_{j-1}(\mathbf{w}^{-1}) =$
$= \sum_{k=j-1}^{N-1}(-1)^{k-j+1}C^j_{k+1}Q^{N-1}_k(\mathbf{w}^{-1})$.

Therefore, using Eqs. (22), (23) and (24), one obtains

$A^1_j(N + 1) = A^1_j(N) + w_{1,N+1}[A^1_{j-1}(N) - A^1_j(N)] =$
$= \sum_{k=j}^{N-1}(-1)^{k-j}C^j_k Q^{N-1}_k(\mathbf{w}^{-1}) + w_{1,N+1}\sum_{k=j-1}^{N-1}(-1)^{k-j+1}C^j_{k+1}Q^{N-1}_k(\mathbf{w}^{-1}) =$
$= \sum_{k=j}^{N}(-1)^{k-j}C^j_k Q^N_k(\mathbf{w}^{-1}, w_{1,N+1})$,

which completes the proof of the induction iteration. *Q.E.D.*

C Simulations

In this section of the Appendix we describe how the simulations were done. The whole procedure consisted of two stages: (i) simulating the winning probabilities w_{ij}; (ii) simulating the predictive power ρ_1.

Winning probabilities. The algorithm for this simulation was the following:

1) Populate the matrix Z_{ij} with zeros;
2) Independently draw N numbers $x_1, \ldots, x_N$ from the distribution $f(\cdot)$ and order them so that $x_1 > x_2 > \ldots > x_N$;
3) Populate the matrix p_{ij} for all $1 \leq i < j \leq N$ according to the following rule: take x_i and x_j such that $i < j$, draw independently two noise terms ϵ_i and ϵ_j, let $y_i = x_i + \epsilon_i$ and $y_j = x_j + \epsilon_j$, and set $p_{ij} = 1$ if $y_i > y_j$ and $p_{ij} = 0$ if $y_i < y_j$;
4) Add the matrix p_{ij} from step 3 to Z_{ij};
5) Go to step 2.
The whole procedure is to be repeated a large number of times, T, and the average over realizations matrix p_{ij} will be the matrix of winning probabilies:

$$w_{ij} = \langle p_{ij} \rangle = \frac{Z_{ij}}{T}. \tag{25}$$

Predictive power. The algorithm for this simulation was the following:
1) Set a counter $c = 0$;
2) Populate the matrix p_{ij} for all $1 \leq i < j \leq N$ according to the following rule: take $i < j$, draw a uniform random number r from interval $(0, 1)$; set $p_{ij} = 1$ if $r > w_{ij}$ and $p_{ij} = 0$ if $r < w_{ij}$;
3) Calculate the scores of players as $s_i = \sum_{j \neq i} p_{ij}$;
4) If player 1 has a maximal score, i.e. $s_1 \geq s_i$ for all $1 < i \leq N$, then increment c by 1;
5) Go to step 2.
The whole procedure is to be repeated a large number of times, T, and then the share of times when the counter was incremented gives the predictive power:

$$\rho_1 = \frac{c}{T}. \tag{26}$$

The winning probabilities and the predictive power (the latter only for moderate N) were also calculated analytically using Eqs. (10) and (13). The results perfectly agree with those of the simulations.

References

Ben-Yashar R., Nitzan S. The optimal decision rule for fixed-size committees in dichotomous choice situations: the general result. *International Economic Review*, **38**(1), Feb. 1997, 175-186.

Esteban J., Ray D. Social desicion rules are not immune to conflict. *Economics of Governance*, 2001, 2:59-67.

Gneezy U., Niederle M., Rustichini A. Performance in competitive environments: gender differences. *Quarterly Journal of Economics*, Aug. 2003, 1049-1074

Green J., Stokey N. A comparison of tournaments and contracts. *The Journal of Political Economy*, **91**(3), Jun. 1983, 349-364.

Harary E., Moser L. The theory of round robin tournaments. *American Mathematical Monthly*,

73(3), March 1966, 231-246.

Harrison W. Field experiments and control. *Mimeo*, 2004.

Hertwig R., Hoffrage U., Martignon L. Quick estimation: letting the environment do the work, in Gigerenzer G. et al, *Simple heuristics that make us smart*, NY Oxford, Oxford University Press, 1999, pp. 209-234

Hvide H. Tournament rewards and risk taking. *Journal of Labor Economics*, **20**(4), 2002, 877-898.

Knoeber C., Thurman W. Testing the theory of tournaments: An empirical analysis of broiler production. *Journal of Labor Economics*, **10**(4), Oct. 1994, 357-379.

Lazear E. Personnel economics: past lessons and future directions. Presidential address to the society of labor economists, San Francisco, May 1, 1998. *Journal of Labor Economics*, **17**(2), Apr. 1999, iv+199-236.

Lazear E., Rosen S. Rank-order tournaments as optimal labor contracts. *The Journal of Political Economy*, **89**(5), Oct. 1981, 841-864.

Levin J., Nalebuff B. An introduction to vote-counting schemes. *The Journal of Economic Perspectives*, **9**(1), 1995, 3-26.

Moldovanu B., Sela A. Contest Architecture. Discussion paper, University of Mannheim, 2002.

Moon J., Pullman N. On generalized tournament matrices. *SIAM Review*, **12**(3), Jul. 1970, 384-399.

Prendergast C. The provision of incentives in firms. *Journal of Economic Literature*, **37**(1), Mar. 1999, 7-63.

Reed W. The Pareto, Zipf, and other power laws. *Economics Letters*, **74**(1), May 2001, 15-19.

Roth A. The economist as engineer: game theory, experimentation, and computation as tools for design economics. *Econometrica*, **70**(4), July 2002, 1341-1378.

Rubinstein A. Ranking the participants in a tournament. *SIAM Journal of Applied Mathematics*, **38**(1), Feb. 1980, 108-111.

Schotter A., Weigelt K. Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results *The Quarterly Journal of Economics*, **107**(2), May 1992, 511-539.

Taylor C. Digging for golden Carrots: an analysis of research tournaments. *The American Economic Review*, **85**(4), Sep. 1995, 872-890.

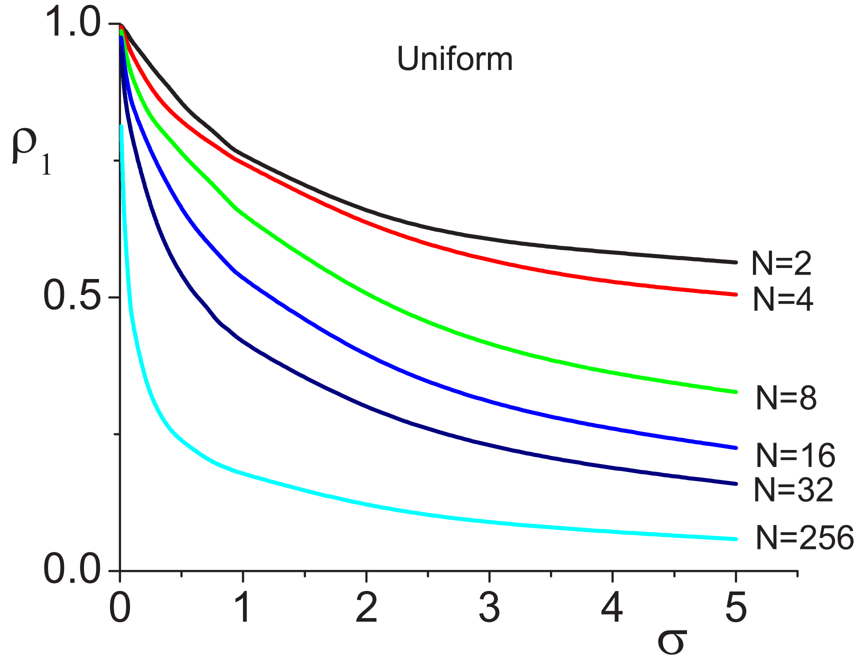

Figure 1: The predictive power ρ_1 as a function of noise intensity σ for various N, for a uniform distribution of players' abilities.

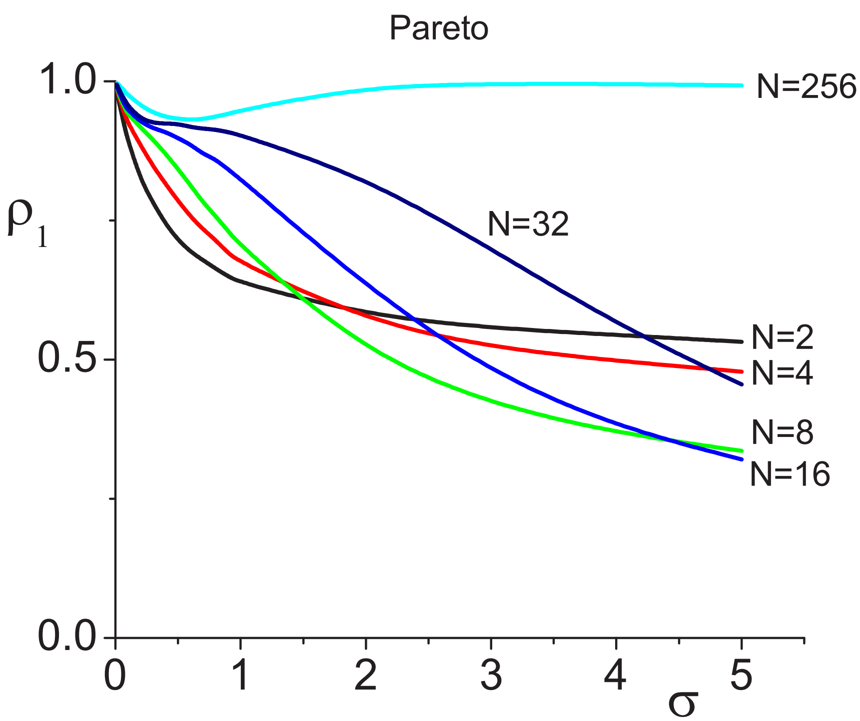

Figure 2: The predictive power ρ_1 as a function of noise intensity σ for various N, for a Pareto distribution of players' abilities.

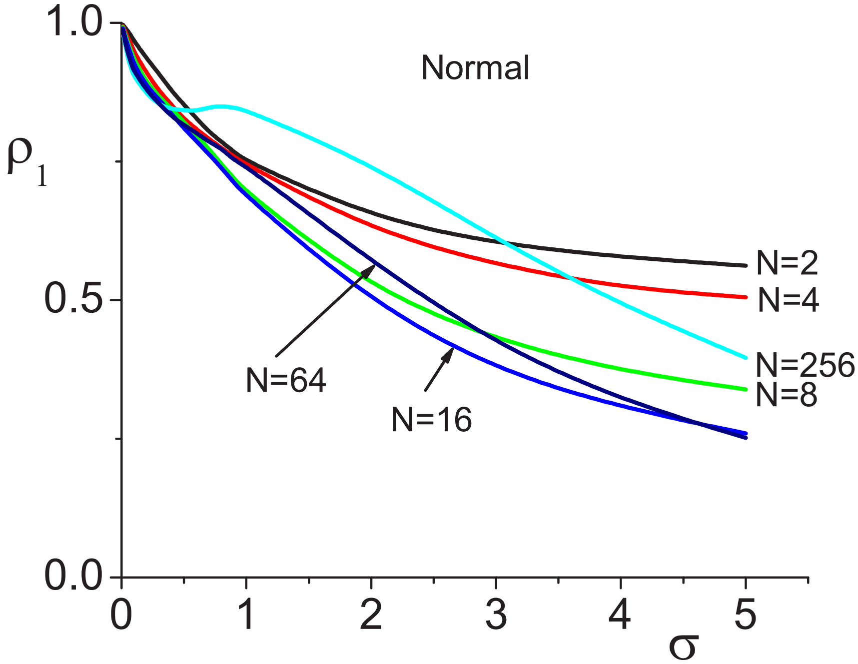

Figure 3: The predictive power ρ_1 as a function of noise intensity σ for various N, for a normal distribution of players' abilities.

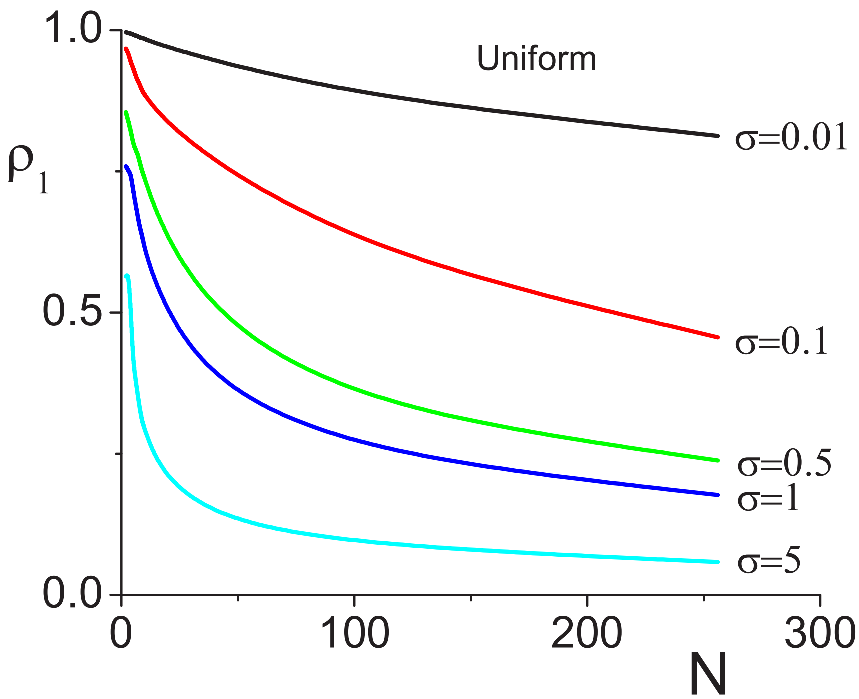

Figure 4: The predictive power ρ_1 as a function of number of players N for various σ, for a uniform distribution of players' ability.

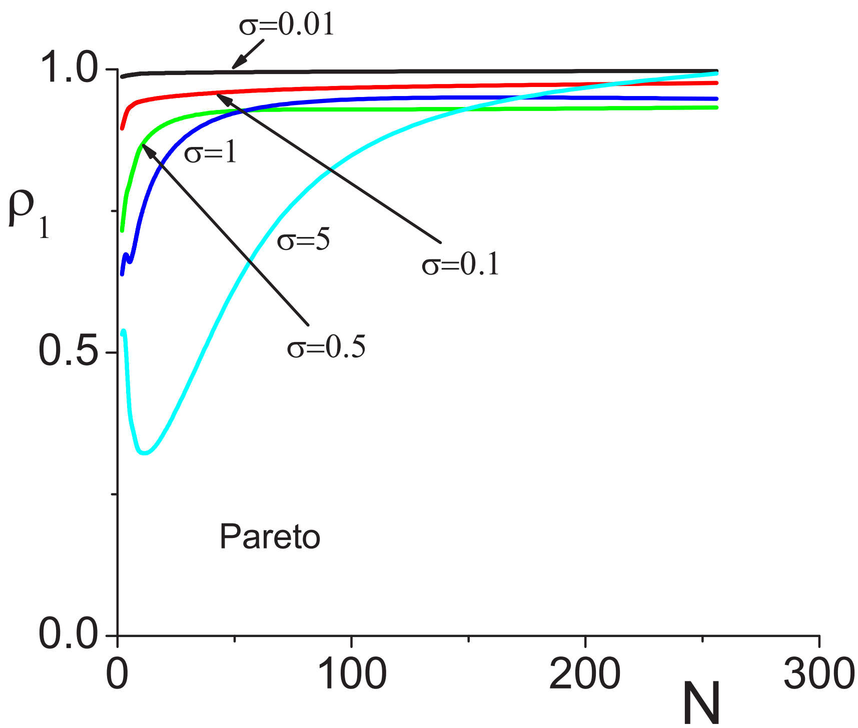

Figure 5: The predictive power ρ_1 as a function of number of players N for various σ, for a Pareto distribution of players' ability.

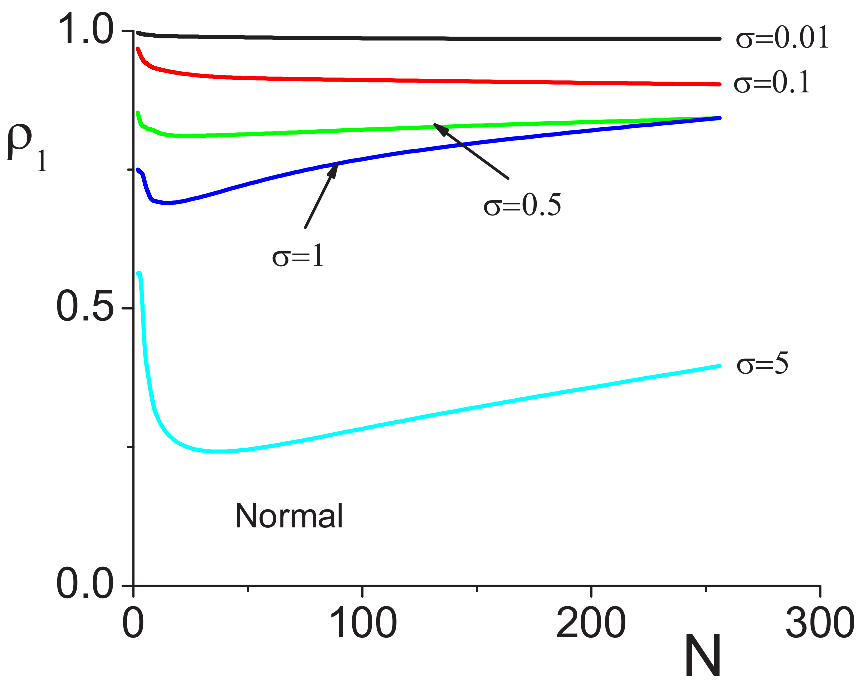

Figure 6: The predictive power ρ_1 as a function of number of players N for various σ, for a normal distribution of players' ability.