The Annals of Applied Probability
1996, Vol. 6, No. 2, 687–694

THE NUMBER OF WINNERS IN A DISCRETE GEOMETRICALLY DISTRIBUTED SAMPLE

BY PETER KIRSCHENHOFER AND HELMUT PRODINGER

Technical University of Vienna

In this tutorial, statistics on the number of people who tie for first place are considered. It is demonstrated that the so-called Rice's method from the calculus of finite differences is a very convenient tool both to rederive known results as well as to gain new ones with ease.

Let the random variable G be geometrically distributed. That is, $\mathbb{P}\{G = k\} = pq^{k-1}$, with $q = 1 - p$. Also, assume that n independent copies are given. Finally, let X count the number of random variables with highest value. A popular realization of this situation is to consider n "players" who independently toss coins until each of them sees the first head. In this interpretation, X is the number of players who gain their respective heads in the very last round of the game, that is, the "winners" of the game. In [1] the probability distribution of X, the expectation $\mathbb{E}X$ of X and the asymptotic behavior of $\mathbb{P}\{X = 1\}$ (probability of a single winner) for $n \to \infty$ were to be determined. In the solution [2] it was remarked that—surprisingly—this probability does not converge as $n \to \infty$ but rather has an oscillating behavior. At the same time, Eisenberg, Stengle and Strang [5] discussed this problem and related topics, exhibiting the structure of the periodic fluctuation, for which an explicit Fourier expansion was given. Also about this time, Brands, Steutel and Wilms [4] came independently to roughly the same results. A recent paper by Baryshnikov, Eisenberg and Stengle [3] deals with the existence of the limiting probability of a tie for first place.

In fact, a fluctuating behavior in asymptotic expansions is not at all uncommon. There are numerous results of that type, for example, in the analysis of divide-and-conquer recursions [7, 8, 16] or digital sums [6], that play a prominent role in the probabilistic analysis of algorithms.

Our aim in this paper is to some extent tutorial: the asymptotic technique that yields the Fourier expansions of the fluctuating functions very comfortably is called "Rice's method" (see the recent survey [9]). In order to convince the reader of the advantages of this method, we will rederive a result on the initially mentioned problem in the sequel, and afterwards present some new results concerning higher moments of distribution as well as the number of persons reaching a specified level beyond the winner(s).

Received October 1994; revised November 1995.
AMS 1991 *subject classification.* 60C05.
Key words and phrases. Geometric distribution, generating functions, Rice's method.

Let us abbreviate $Q := 1/q$ and $L := \log Q$. Also, let p_m denote the probability $p_m = \mathbb{P}\{X = m\}$, that is, the probability of having m winners (amongst n players). Then

$$p_m = p^m \binom{n}{m} \sum_{j \geq 1} q^{(j-1)m} \left(1 - q^{j-1}\right)^{n-m}. \tag{1}$$

This follows from the observation that m out of n people have a (winning) value j, while the other ones have a smaller value. Now we set $N := n - m$, expand the binomial and sum over j to get the alternative formula

$$p_m = p^m \binom{n}{m} \sum_{k=0}^{N} \binom{N}{k} (-1)^k \frac{1}{1 - Q^{-k-m}}. \tag{2}$$

The key point in analyzing this alternating sum asymptotically is the following lemma.

LEMMA 1 ([9]). *Let $f(z)$ be a function that is analytic on $[n_0, +\infty[$. Assume that $f(z)$ is meromorphic in the whole of $\mathbb{C}$ and analytic on $\Omega = \bigcup_{j=1}^{\infty} \gamma_j$, where the γ_j are concentric circles whose radius tends to ∞. Let $f(z)$ be of polynomial growth on Ω. Then, for N large enough,*

$$\sum_{k=n_0}^{N} \binom{N}{k} (-1)^k f(k) = \sum_{z} \operatorname{Res}[N; z] f(z), \tag{3}$$

where

$$[N; z] = \frac{(-1)^{N-1} N!}{z(z-1)\cdots(z-N)} = \frac{\Gamma(N+1)\Gamma(-z)}{\Gamma(N+1-z)}$$

and the sum is extended to all poles not on $[n_0, +\infty[$.

The following proposition collects the asymptotic results concerning the distribution of the number of winners among n players. We demonstrate the use of Rice's method by giving our alternative proof for the asymptotics of the probabilities, mention the (known) expectation and derive the (new) asymptotics of the variance.

PROPOSITION 1. *Let X be the random variable "number of winners among n players" as described above. Then*

$$\begin{aligned} p_m &= \mathbb{P}\{X = m\} \\ &= \frac{1}{L}\frac{p^m}{m} + \frac{p^m}{L}\delta_m(\log_Q n) + \mathcal{O}\left(\frac{1}{n}\right), \qquad m \text{ fixed}, n \to \infty, \end{aligned} \tag{4}$$

$$E_n = \mathbb{E}X = \frac{p}{q}\frac{1}{L}\left(1 + \delta_1(\log_Q n)\right) + \mathcal{O}\left(\frac{1}{n}\right) \tag{5}$$

and

$$V_n = \operatorname{Var} X$$

$$(6) \qquad = \frac{p}{q^2}\frac{1}{L} - \frac{p^2}{q^2}\frac{1}{L}\left(2\sum_{j\geq 1}\frac{Q^j}{(Q^j+1)^2} + \frac{1}{4}\right) + \tau_1(\log_Q n) + \mathcal{O}\left(\frac{1}{n}\right),$$

where $\delta_m(x)$ *and* $\tau_1(x)$ *are continuous periodic functions of period* 1, *mean* 0 *and small amplitude.*

PROOF. In order to prove (4), we apply the lemma (with $n_0 = 0$) to expression (2). In this instance $f(z) = 1/(1 - Q^{-z-m})$ has poles at $z = -m + \chi_j$, with $\chi_j = 2j\pi i/L$, and is bounded on concentric circles C_j about the origin passing through the points

$$-m \pm \frac{(2j+1)\pi i}{L}, \qquad j = 1, 2, \ldots.$$

Therefore we only have to consider the residues of

$$[N; z]\frac{1}{1 - Q^{-z-m}} \quad \text{at } z = -m + \chi_j.$$

The computation of the residues is simple;

$$\operatorname{Res}_{z=-m+\chi_j} = \frac{\Gamma(N+1)\Gamma(m-\chi_j)}{\Gamma(N+1+m-\chi_j)}\frac{1}{L},$$

whence we have [after multiplication by the factor $p^m\binom{n}{m}$ and going back to $n > m$ instead of $N \geq 1$] the formulas

$$(7) \qquad p_m = \frac{1}{L}\frac{p^m}{m!}\sum_{j\in\mathbb{Z}}\Gamma(m-\chi_j)\frac{\Gamma(n+1)}{\Gamma(n+1-\chi_j)}, \qquad n > m,$$

where—according to the previous remarks—the series stands for the Cauchy principal value.

Using Stirling's formula for the approximation of the Γ-functions, we find that

$$\frac{\Gamma(n+1)}{\Gamma(n+1-\chi_j)} = n^{\chi_j}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right), \qquad n \to \infty,$$

so that, for n getting large, the series converges to the Fourier expansion of a periodic function in $\log_Q n$. Pulling out the term with index $j = 0$ (the "mean") and denoting the remaining periodic fluctuation (of mean 0),

$$\delta_m(x) = \frac{1}{m!}\sum_{j\neq 0}\Gamma(m-\chi_j)e^{2j\pi i x},$$

we gain formula (4).

It is interesting to note that the alternating sum (2) can also be rewritten using the partial fraction decomposition of the meromorphic function $1/(1-q^{z+m})$, namely,

$$\frac{1}{1-q^{z+m}}=\frac{1}{2}+\frac{1}{L}\sum_{j\in\mathbb{Z}}\frac{1}{z+m+\chi_j}\quad\text{with } \chi_j=\frac{2j\pi i}{L}. \tag{8}$$

(Compare [11], 7.10.) Again, the sum stands for the Cauchy principal value. The usual argument to derive (8) is to compute the sum of the principal parts of the function and to show that the difference between this sum and the function—which has to be entire—is bounded, and thus a constant. Inserting (8) into (2), we find

$$p_m=p^m\binom{n}{m}\sum_{k=0}^{N}\binom{N}{k}(-1)^k\frac{1}{2}+\frac{1}{L}p^m\binom{n}{m}\sum_{j\in\mathbb{Z}}\sum_{k=0}^{N}\binom{N}{k}(-1)^k\frac{1}{k+m+\chi_j}.$$

Now, for $x\in\mathbb{C}\setminus\{-N,\dots,0\}$,

$$\sum_{k=0}^{N}\binom{N}{k}(-1)^k\frac{1}{k+x}=\frac{\Gamma(N+1)\Gamma(x)}{\Gamma(N+1+x)}=\frac{N!}{(N+x)(N+x-1)\cdots x}$$

(compare [10], (5.41)), so that (remembering $N=n-m$)

$$\begin{aligned} p_m&=\frac{p^m}{2}\delta_{n,m}+\frac{1}{L}\frac{p^m}{m!}\sum_{j\in\mathbb{Z}}\frac{n!}{(m+\chi_j)\cdots(n+\chi_j)}\\ &=\frac{p^m}{2}\delta_{n,m}+\frac{1}{L}\frac{p^m}{m!}\sum_{j\in\mathbb{Z}}\Gamma(m-\chi_j)\frac{\Gamma(n+1)}{\Gamma(n+1-\chi_j)}. \end{aligned} \tag{9}$$

In order to get (5), we observe that $E_n=Q(p_1-p\delta_{n,1})$ as was already reported in [2]. Let us now engage in the proof of (6). For this we compute the second factorial moment M_n for $n\geq 2$:

$$\begin{aligned} M_n&=\sum_{m\geq 2}m(m-1)p_m\\ &=n(n-1)\sum_{m\geq 2}\binom{n-2}{m-2}p^m\sum_{j\geq 0}q^{jm}(1-q^j)^{n-m}\\ &=\frac{n(n-1)p^2}{q^2}\sum_{j\geq 1}q^{2j}(1-q^j)^{n-2}\\ &=\frac{n(n-1)p^2}{q^2}\sum_{k=0}^{n-2}\binom{n-2}{k}(-1)^k\frac{1}{Q^{k+2}-1}\\ &=2Q^2p_2-2\frac{p^2}{q^2}\delta_{n,2}=2Q^2\left(p_2-p^2\delta_{n,2}\right). \end{aligned} \tag{10}$$

The variance is obtained in the usual way by computing $V_n = M_n + E_n - E_n^2$. Hence

$$V_n = \frac{p^2}{q^2}\frac{1}{L} + \frac{p}{q}\frac{1}{L} - \frac{p^2}{q^2}\frac{1}{L^2} + \tau(\log_Q n) + \mathscr{O}\left(\frac{1}{n}\right). \tag{11}$$

This time

$$\tau(x) = \frac{p^2}{q^2}\frac{2}{L}\delta_2(x) + \frac{p}{q}\frac{1}{L}\delta_1(x) - \frac{p^2}{q^2}\frac{1}{L^2}\delta_1^2(x)$$

has (integral) mean *different* from 0! While this quantity is quite small, it can be extracted using the methods described in [12], and we find the alternative formula (6) of the proposition. □

There is a nice way to derive the explicit forms of the expectation and the second factorial moment, using (probability) generating functions. Let the coefficient of z^k in $F_n(z)$ denote the probability that n players produce k winners. We get the following recursion:

$$F_n(z) = \sum_{k=1}^{n} \binom{n}{k} p^{n-k} q^k F_k(z) + p^n z^n, \qquad n \geq 1. \tag{12}$$

It is convenient to set $F_0(z) = 1$. This recursion is almost self-explanatory. When, at a certain level, the remaining players all fail, we label each of them by a "z" and leave the recursion (equivalently we might think of z as the probability of an event independent of the game). The expectation E_n is obtained via $F_n'(1)$; therefore

$$E_n = \sum_{k=1}^{n} \binom{n}{k} p^{n-k} q^k E_k + np^n, \qquad n \geq 1.$$

Defining the exponential generating function $E(z) = \sum_{n \geq 1} E_n z^n/n!$, we obtain

$$E(z) = e^{pz}E(qz) + pze^{pz}.$$

Using the "Poisson transform"

$$\hat{E}(z) = e^{-z}E(z) = \sum_{n \geq 1} \hat{E}_n \frac{z^n}{n!},$$

this simplifies to

$$\hat{E}(z) = \hat{E}(qz) + pze^{-qz}.$$

Equating coefficients we see that, for $n \geq 1$,

$$\hat{E}_n = \frac{np}{q}(-1)^{n-1}\frac{q^n}{1-q^n}$$

and furthermore, for $n \geq 1$,

$$E_n = \sum_{k=1}^{n} \binom{n}{k} \hat{E}_k = \sum_{k=1}^{n} \binom{n}{k} \frac{kp}{q} (-1)^{k-1} \frac{1}{Q^k - 1}$$
$$= \frac{np}{q} \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k \frac{1}{Q^{k+1} - 1},$$

which coincides with the formula in [2]. For the second factorial moment we differentiate twice and evaluate at $z = 1$. An almost identical computation gives the same expression that we obtained already.

This approach was used by Knuth in [15] under the name "binomial transform" and subsequently used by many people (see, e.g., [17]).

Finally, we want to produce some additional new results which shed some additional light on the original question about the number of winners.

Assume that the winners have reached the level j. We are interested in the number of players who reached the level $j - d$, where d is a parameter. ($d = 0$ is the case that was just considered.) Call the random variable in question X_d. Then we have the following proposition.

PROPOSITION 2. *Let X_d denote the random variable "number of players who reach level d below the winners." Then*

$$\mathbb{P}\{X_d = m_d\} = \sum \sum_{j \geq d+1} \binom{n}{m_0, \ldots, m_d, \mu} \times (pq^{j-1})^{m_0} \cdots (pq^{j-d-1})^{m_d} (1 - q^{j-d-1})^{\mu}, \tag{13}$$

where the first sum runs over all $m_0 \geq 1, m_1 \geq 0, \ldots, m_{d-1} \geq 0$ and $\mu = n - m_0 - \cdots - m_d$,

$$E_n^{(d)} = \frac{1}{L} \frac{p^2}{q^{d+1}} + \frac{1}{L} \frac{p^2}{q^{d+1}} \delta_1(\log_Q n) + \mathscr{O}\left(\frac{1}{n}\right), \qquad n \to \infty, \tag{14}$$

and

$$V_n^{(d)} = \frac{p^2}{q^{2d+2}} \frac{1}{L} (1 + q^{d+1}) - \frac{p^4}{q^{2d+2}} \frac{1}{L} \left(2 \sum_{j \geq 1} \frac{Q^j}{(Q^j + 1)^2} + \frac{1}{4}\right) + \tau_1^{(d)}(\log_Q n) + \mathscr{O}\left(\frac{1}{n}\right). \tag{15}$$

PROOF. The formula for the probabilities is self-explanatory [compare the comments on formula (1)]. The expected value is just the sum of m_d times this quantity. We get

$$E_n^{(d)} = n \sum_{j \ge d+1} pq^{j-d-1}\Big[\big(pq^{j-1} + \cdots + pq^{j-d-1} + 1 - q^{j-d-1}\big)^{n-1} - \big(pq^{j-2} + \cdots + pq^{j-d-1} + 1 - q^{j-d-1}\big)^{n-1}\Big]$$

$$= pn \sum_{j \ge d+1} q^{j-d-1}\Big[\big(1 - q^j\big)^{n-1} - \big(1 - q^{j-1}\big)^{n-1}\Big]$$

$$= \frac{p^2}{q^{d+1}} n \sum_{j \ge d} q^j\big(1 - q^j\big)^{n-1} - \frac{p}{q} n\big(1 - q^d\big)^{n-1}.$$

This formula holds only for $d \ge 1$. Since

$$E_n^{(0)} = E_n = n\frac{p}{q}\sum_{j \ge 1} q^j\big(1 - q^j\big)^{n-1},$$

we find

$$E_n^{(d)} = \frac{p}{q^d}E_n^{(0)} - \frac{p^2}{q^{d+1}} n \sum_{j=1}^{d-1} q^j\big(1 - q^j\big)^{n-1} - \frac{p}{q} n\big(1 - q^d\big)^{n-1}, \tag{16}$$

and, because the extra terms are exponentially small, we can use the asymptotic result for $E_n^{(0)}$ and have

$$E_n^{(d)} = \frac{p}{q^{d+1}}p_1 + \text{exponentially small terms in } n, \qquad n \ge 2,$$

from which (14) is immediate using (4). The second factorial moment $M_n^{(d)}$ for $d \ge 1$ is computed analogously as (with $M_n^{(0)} = M_n$)

$$M_n^{(d)} = n(n-1) \sum_{j \ge d+1} \big(pq^{j-d-1}\big)^2\Big[\big(1 - q^j\big)^{n-2} - \big(1 - q^{j-1}\big)^{n-2}\Big]$$

$$= \frac{p^3}{q^{2d+2}} n(n-1) \sum_{j \ge d} q^{2j}\big(1 - q^j\big)^{n-2} - \frac{p^2}{q^2} n(n-1)\big(1 - q^d\big)^{n-2} \tag{17}$$

$$= \frac{p}{q^{2d}} M_n^{(0)} + \text{exponentially small terms in } n, \qquad n \to \infty.$$

Thus formula (15) gives us the asymptotics for the variance. □

REFERENCES

[1] E3436 (PROBLEM) (1991). *Amer. Math. Monthly* **99** 366.

[2] E3436 (SOLUTION) (1994). *Amer. Math. Monthly* **101** 78–80.

[3] BARYSHNIKOV, Y., EISENBERG, B. and STENGLE, G. (1995). A necessary and sufficient condition for the existence of the limiting probability of a tie for first place. *Statist. Probab. Lett.* **23** 203–209.

[4] BRANDS, J. J. A. M., STEUTEL, F. W. and WILMS, R. J. G. (1994). On the number of maxima in a discrete sample. *Statist. Probab. Lett.* **20** 209–217.

[5] EISENBERG, B., STENGLE, G. and STRANG, G. (1993). The asymptotic probability of a tie for first place. *Ann. Appl. Probab.* **3** 731–745.

[6] FLAJOLET, P., GRABNER, P., KIRSCHENHOFER, P., PRODINGER, H. and TICHY, R. F. (1994). Mellin transforms and asymptotics: digital sums. *Theoret. Comput. Sci.* **123** 291–314.

[7] FLAJOLET, P. and RICHMOND, L. B. (1992). Generalized digital trees and their difference-differential equations. *Random Structures and Algorithms* **3** 305–320.

[8] FLAJOLET, P. and SEDGEWICK, R. (1986). Digital search trees revisited. *SIAM J. Comput.* **15** 748–767.

[9] FLAJOLET, P. and SEDGEWICK, R. (1995). Mellin transforms and asymptotics: finite differences and Rice's integrals. *Theoret. Comput. Sci.* **144** 101–124.

[10] GRAHAM, R., KNUTH, D. E. and PATASHNIK, O. (1989). *Concrete Mathematics.* Addison-Wesley, Reading, MA.

[11] HENRICI, P. (1977). *Applied and Computational Complex Analysis* **1**. Wiley, New York.

[12] KIRSCHENHOFER, P. and PRODINGER, H. (1991). On some applications of formulae of Ramanujan in the analysis of algorithms. *Mathematika* **38** 14–33.

[13] KIRSCHENHOFER, P. and PRODINGER, H. (1993). A result in order statistics related to probabilistic counting. *Computing* **51** 15–27.

[14] KIRSCHENHOFER, P., PRODINGER, H. and SZPANKOWSKI, W. (1989). On the variance of the external pathlength in a symmetric digital trie. *Discrete Appl. Math.* **25** 129–143.

[15] KNUTH, D. E. (1973). *The Art of Computer Programming* **3**. Addison-Wesley, Reading, MA.

[16] MAHMOUD, H. M. (1992). *Evolution of Random Search Trees.* Wiley, New York.

[17] PRODINGER, H. and SZPANKOWSKI, W. (1993). A note on binomial recurrences arising in the analysis of algorithms (letter to the editor). *Inform. Process. Lett.* **46** 309–311.

[18] SZPANKOWSKI, W. (1988). The evaluation of an alternative sum with applications to the analysis of some data structures. *Inform. Process. Lett.* **28** 13–19.

INSTITUT FÜR ALGEBRA UND DISKRETE MATHEMATIK
TECHNISCHE UNIVERSITÄT WIEN
WIEDNER HAUPTSTRASSE 8-10 / 118
A-1040 WIEN
AUSTRIA
E-MAIL: Kirsch@rsmb.tuwien.ac.at
proding@rsmb.tuwien.ac.at