Combinatorics, Probability and Computing (1997) 6, 179–183
Copyright © 1997 Cambridge University Press

Maximum Statistics of N Random Variables Distributed by the Negative Binomial Distribution

PETER J. GRABNER[1] and HELMUT PRODINGER[2]†

[1]Institut für Mathematik A, Steyrergasse 30, A–8010 Graz, Austria
(e-mail: grabner@weyl.math.tu_graz.ac.at)
[2]Institut für Algebra und Diskrete Mathematik,
Wiedner Hauptstrasse 8–10, A–1040 Wien, Austria
(e-mail: Helmut.Prodinger@tuwien.ac.at)

Received 9 May 1994; revised 11 October 1995

The lifetime of a player is defined to be the time where he gets his b-th hit, where a hit will occur with probability p. We consider the maximum statistics of N independent players. For $b \neq 1$ this is significantly more difficult than the known instance $b = 1$. The expected value of the maximum lifetime of N players is given by $\log_Q N + (b-1)\log_Q \log_Q N +$ smaller order terms, where $Q = 1/(1-p)$.

1. Introduction

Consider a geometrically distributed random variable X, i.e. $\mathbb{P}\{X = k\} = pq^{k-1}$ $(q = 1-p)$. If we have N independent copies $X_1, \ldots, X_N$, the *maximum* of them is well understood (cf. [13] and [9]).

Now if X is the sum of two independent geometric random variables (with the same parameter), the corresponding maximum of N such variables is surprisingly more complicated. In the present note we will work out this case and indicate how the more general case of sums of b copies can be handled. Also, we give analytic indications why the asymptotic evalution of the expected value is more complex.

In the sequel, we will use the suggestive notation of a 'player' who 'dies' at the second 'hit'. We note here that our model is quite different from the well-known 'Coupon Collector's Problem' (cf. e.g. [2, 12]), where the periodic term ((2.5) in Theorem 1 below) in the asymptotic expansion does not occur. In this context, we will concentrate on the expected value of the maximum lifetime of N independent players. To be precise, if a player gets his second hit at time k, we say that his lifetime is k. Such a waiting time distribution is generally known to be a *negative binomial* distribution (cf. [6, 11]). Now

† This research was supported by the Austrian–Hungarian cooperation 10U3

https://doi.org/10.1017/S0963548397002939 Published online by Cambridge University Press

we compute the probability that a particular player has lifetime $> k$. This is the same as to say that he is still alive at time k, which means that he obtained either zero or one hit. Hence

$$\mathbb{P}\{X > k\} = q^k + kpq^{k-1}.$$

Thus we have for the expectation of the maximum lifetime of N players

$$\mathbb{E}_N = \mathbb{E}\max(X_1, \ldots, X_N) = \sum_{k\geq 0}\Big(1 - \big(1 - (pk+q)q^{k-1}\big)^N\Big). \tag{1.1}$$

The usual procedure for the analysis of such sums would proceed as follows: Using the binomial theorem we would obtain

$$\mathbb{E}_N = \sum_{m=1}^{N}(-1)^{m-1}\binom{N}{m}\sum_{k\geq 0}\big((pk+q)q^{k-1}\big)^m. \tag{1.2}$$

Then we would have to write this sum as a contour integral (cf. [7]). But following Rice's method would require deformation of the contour of integration, and therefore analytic continuation of the function

$$f(s) = \sum_{k\geq 0}\big((pk+q)q^{k-1}\big)^s \tag{1.3}$$

beyond the line $\Re s = 0$. We tried hard to get such continuation but did not succeed. We will explain, after the analysis of $\mathbb{E}_N$, why this is so difficult.

2. Asymptotic evaluation of the sum

As indicated in the introduction, the asymptotic expansion of $\mathbb{E}_N$ cannot be obtained by standard methods. Further analysis can be done by using some approximation techniques. Our first ('exponential') approximation will give us a more suitable form of $\mathbb{E}_N$. This exponential approximation is an important tool in the study of order statistics (cf. [6]). In the following we will use the notations $\operatorname{lq} x$ to indicate $\log_{\frac{1}{q}} x$ and $Q = \frac{1}{q}$.

Let us define

$$\tilde{\mathbb{E}}_N = \sum_{k\geq 0}\Big(1 - \exp\big(-N(pk+q)q^{k-1}\big)\Big). \tag{2.1}$$

Then we have

$$\begin{aligned}\mathbb{E}_N - \tilde{\mathbb{E}}_N \;=\; &\left(\sum_{k\leq \operatorname{lq} N} + \sum_{\operatorname{lq} N < k \leq \operatorname{lq} N + 2\operatorname{lq}\operatorname{lq} N} + \sum_{k > \operatorname{lq} N + 2\operatorname{lq}\operatorname{lq} N}\right)\\ &\Big(\exp\big(-(pk+q)q^{k-1}N\big) - \big(1-(pk+q)q^{k-1}\big)^N\Big).\end{aligned} \tag{2.2}$$

We have split the sum into three parts and will give estimates for each part.

1. $k \leq \operatorname{lq} N$: in this range we have $(pk+q)q^{k-1} \geq \frac{p\operatorname{lq} N}{qN} + \frac{1}{N}$. Thus we obtain

$$\exp\big(-(pk+q)q^{k-1}N\big) \leq N^{-\frac{p}{q\log Q}} \quad \text{and} \quad \big(1-(pk+q)q^{k-1}\big)^N \leq N^{-\frac{p}{q\log Q}}$$

and consequently we have $N^{-\frac{p}{q\log Q}}\operatorname{lq} N$ as an upper bound for this first sum.

https://doi.org/10.1017/S0963548397002939 Published online by Cambridge University Press

2. $\operatorname{lq} N < k \le \operatorname{lq} N + 2 \operatorname{lq}\operatorname{lq} N$: using the Taylor expansions of $\exp(-Nu)$ and $(1-u)^N$ we obtain

$$\exp(-Nu) - (1-u)^N = \frac{Nu^2}{2}\exp(-Nu) + O(Nu^3\exp(-Nu)) + O(N^2u^4\exp(-Nu)) = O\left(\frac{\log^2 N}{N}\right)$$

for $u = (pk+q)q^{k-1}$. This yields an estimate $O\left(\frac{\log^2 N \log\log N}{N}\right)$ for the sum in this range.

3. $k > \operatorname{lq} N + 2\operatorname{lq}\operatorname{lq} N$: again, we use Taylor expansion to obtain

$$\exp(-Nu) - (1-u)^N = O\left(N(pk+q)^2 q^{2k-2}\right)$$

for $u = (pk+q)q^{k-1}$. This yields an upper bound

$$O\left(N \sum_{k > \operatorname{lq} N + 2\operatorname{lq}\operatorname{lq} N} (pk+q)^2 q^{2k-2}\right) = O\left(\frac{\log^2 N}{N}\right)$$

for the sum in this range.

Combining the three estimates yields

$$\mathbb{E}_N - \tilde{\mathbb{E}}_N = O\left(\frac{\log^2 N \log\log N}{N}\right).$$

The standard technique for deriving asymptotic estimates for $\tilde{\mathbb{E}}_N$ uses the Mellin transform, and therefore again the analytic continuation of $f(s)$. Thus we use a second approximation to $\mathbb{E}_N$ which finally yields the desired asymptotic expansion.

For this purpose we define

$$\mathbb{E}_N^* = \sum_{k \ge 0} \left(1 - \exp\left(-(p \operatorname{lq} N + q) N q^{k-1}\right)\right). \tag{2.3}$$

Now we have to give estimates for $\tilde{\mathbb{E}}_N - \mathbb{E}_N^*$. We compute

$$\tilde{\mathbb{E}}_N - \mathbb{E}_N^* = \sum_{k \le \operatorname{lq} N} \exp\left(-(pk+q)q^{k-1}N\right)\left(\exp\left(-p(\operatorname{lq} N - k)q^{k-1}N\right) - 1\right) + \sum_{k > \operatorname{lq} N} \exp\left(-(p \operatorname{lq} N + q)q^{k-1}N\right)\left(1 - \exp\left(-p(k - \operatorname{lq} N)q^{k-1}N\right)\right). \tag{2.4}$$

As above we get the estimate $O(N^{-\frac{p}{q \log Q}} \log N)$ for the first sum. For the second sum we use the estimate

$$1 - \exp\left(-p(k - \operatorname{lq} N)q^{k-1}N\right) = O\left(p(k - \operatorname{lq} N)q^{k-1}N\right).$$

We split the range of summation into two parts:

1. $\operatorname{lq} N < k \le \operatorname{lq} N + \operatorname{lq}\operatorname{lq} N + \operatorname{lq} p + 1$: let $k = \operatorname{lq} N + x$ and estimate the summand by $O(xq^x \exp(-(\frac{p}{q} \operatorname{lq} N + 1)q^x))$. This function is monotonically increasing in the interval $0 \le x \le \operatorname{lq}\operatorname{lq} N + \operatorname{lq} p + 1$, thus the summand can be estimated by $O(\frac{\operatorname{lq}\operatorname{lq} N}{\operatorname{lq} N})$ which yields a $O(\frac{(\operatorname{lq}\operatorname{lq} N)^2}{\operatorname{lq} N})$ contribution to the second sum.

https://doi.org/10.1017/S0963548397002939 Published online by Cambridge University Press

2. $k > \operatorname{lq} N + \operatorname{lq}\operatorname{lq} N + \operatorname{lq} p + 1$: in this range we have $N \operatorname{lq} N \le q^{-k}$, and thus the remaining sum can be estimated using the summation formula for the geometric series

$$O\left(\sum_{k>\operatorname{lq} N+\operatorname{lq}\operatorname{lq} N} p(k-\operatorname{lq} N)q^{k-1}N\right) = O\left(\frac{\log\log N}{\log N}\right).$$

Therefore, we arrive at the final estimate

$$\mathbb{E}_N - \mathbb{E}_N^* = O\left(\frac{(\log\log N)^2}{\log N}\right).$$

Notice now that $\mathbb{E}_N^*$ is the sum

$$S(T) = \sum_{k\ge 0}\left(1-\exp\left(-Tq^k\right)\right)$$

evaluated at $T = \frac{p}{q}N \operatorname{lq} N + N$. This sum occurred in many asymptotic investigations (e.g. [10]). By Mellin's inversion formula [1, 3–5] we have

$$S(T) = \frac{1}{2\pi i}\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty}\Gamma(s)\frac{T^{-s}ds}{Q^s-1}.$$

Hence we get the asymptotic expansion of $S(T)$ by shifting the line of integration to the right and taking residues into account

$$S(T) = \operatorname{lq} T + \frac{1}{2} + \frac{\gamma}{\log Q} + F(\operatorname{lq} T) + O\left(\frac{1}{T}\right),$$

where F is an infinitely differentiable periodic function. Thus we have proved

Theorem 1.

$$\mathbb{E}_N^* = \operatorname{lq} N + \operatorname{lq}\operatorname{lq} N + \operatorname{lq} p + \frac{3}{2} + \frac{\gamma}{\log Q} + F(\operatorname{lq} N + \operatorname{lq}\operatorname{lq} N + \operatorname{lq} p) + O\left(\frac{(\log\log N)^2}{\log N}\right) \quad (2.5)$$

where F is a periodic C^∞-function of period 1 *and mean value* 0 *whose Fourier-coefficients are given by*

$$\hat{F}(k) = -\frac{1}{\log Q}\Gamma\left(-\frac{2k\pi i}{\log Q}\right) \quad \text{for } k \in \mathbb{Z}\setminus\{0\}.$$

3. Concluding remarks

If we translate the asymptotics of $\mathbb{E}_N^*$ back to analytic properties of the Mellin-transform of $S(T \operatorname{lq} T)$, we encounter logarithmic singularities (corresponding to the log log-terms) at $s = \frac{2k\pi i}{\log Q}$ for all $k \in \mathbb{Z}$. These singularities would give us an infinity of branch-cuts which are the reason for the difficulties mentioned in the introduction. Notice that this is not a proof for the existence of an analytic continuation of the function $f(s)$ in (1.3).

The method used above to find an asymptotic expansion for the sum

$$\sum_{k\ge 0}\left(1-\left(1-(pk+q)q^{k-1}\right)^N\right) \quad (3.1)$$

https://doi.org/10.1017/S0963548397002939 Published online by Cambridge University Press

can be immediately extended to find the expansion for

$$S_g(N) = \sum_{k\geq 0} \left(1 - \left(1 - g(k)q^k\right)^N\right)$$

where $g(k) \sim \beta k^\alpha$ and $g(k) > 0$ for $k \geq 0$. The resulting asymptotic expansion is given by

$$S_g(N) = \lg N + \alpha \lg\lg N + \lg \beta + \frac{1}{2} + \frac{\gamma}{\log Q} + F(\lg N + \alpha \lg\lg N + \lg \beta) + o(1) \qquad (3.2)$$

where F is the periodic function of Theorem 1.

This general result can be used to compute the expected maximal lifetime of N players who die after b hits. It is quite plain to see that

$$\mathbb{P}\{X > k\} = q^k + kpq^{k-1} + \cdots + \binom{k}{b-1} p^{b-1} q^{k-b+1}.$$

Hence the $g(k)$ from before is

$$g(k) = 1 + k\frac{p}{q} + \cdots + \binom{k}{b-1}\left(\frac{p}{q}\right)^{k-1} \sim \frac{k^{b-1}}{(b-1)!}\left(\frac{p}{q}\right)^{k-1}$$

Therefore, this expectation is given by

$$\begin{aligned}\mathbb{E}_N^{(b)} \;=\; & \lg N + (b-1)\lg\lg N + (b-1)\lg p + (b-1) - \lg(b-1)! + \frac{1}{2} + \frac{\gamma}{\lg Q} \\ & + F\left(\lg N + (b-1)\lg\lg N + (b-1)\lg p - \lg(b-1)!\right) + o(1). \qquad (3.3)\end{aligned}$$

This result tells us that, if we allow an additional hit, the maximum of N players will roughly increase by $\lg\lg N$.

References

[1] Doetsch, G. (1950) *Handbuch der Laplace-Transformation*. Birkhäuser, Basel.

[2] Flajolet, P., Gardy, D. and Thimonier, L. (1992) Birthday paradox, coupon collectors, caching algorithms and self-organizing search. *Discr. Appl. Math.* **39** 207–229.

[3] Flajolet, P., Gourdon, X. and Dumas, P. (1995) Mellin transforms and asymtotics: Harmonic sums. *Theor. Comput. Sci.* **144** 3–58.

[4] Flajolet, P., Régnier, M. and Sedgewick, R. (1985) Some uses of the Mellin integral transform in the analysis of algorithms. In: *Combinatorics on Words*. Springer NATO ASI Series F, Vol. 12.

[5] Flajolet, P. and Sedgewick, R. (1995) Mellin transforms and asymptotics: Finite differences and Rice's integral. *Theor. Comput. Sci.* **144** 101–124.

[6] Galambos, J. (1987) *The Asymptotics of Extreme Order Statistics*. Wiley.

[7] Grabner, P. J. and Prodinger, H. (1994) The Fibonacci Killer. *The Fibonacci Quarterly* **32** 389–394.

[8] Johnson, N. L. and Kotz, S. (1977) *Urn models and Their Applications*. Wiley.

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

[10] Knuth, D. E. (1973) *The Art of Computer Programming, Vol. 3*. Addison-Wesley.

[11] Kolchin, V. F., Sevastyanov, B. A. and Chistakov, V. P. (1978) *Random Allocations*. Wiley.

[12] Nath, H. B. (1973) Waiting time in the coupon collector's problem. *Austral. J. Statist.* **15** 132–135.

[13] Szpankowski, W. and Rego, V. (1990) Yet another application of a binomial recurrence: order statistics. *Computing* **43** 401–410.

https://doi.org/10.1017/S0963548397002939 Published online by Cambridge University Press