MAXIMUM STATISTICS OF N RANDOM VARIABLES DISTRIBUTED BY THE NEGATIVE BINOMIAL DISTRIBUTION

Peter J. Grabner and Helmut Prodinger

Department of Mathematics
Technical University of Graz, Austria
and
Department of Algebra and Discrete Mathematics
Technical University of Vienna, Austria

Abstract. The lifetime of a player is defined to be the time where he gets his bth 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 average value of this parameter 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 $X_1, \dots, X_N$, the *maximum* of them is well understood, compare [7] and [5].

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 expectated value is more complex.

In the sequel, we will use the suggestive notation of a "player" who "dies" at the second "hit". 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. [4]). Now 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

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

1991 *Mathematics Subject Classification.* 05A15.

This research was supported by the Austrian-Hungarian cooperation 10U3

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

$$\mathbb{E}_N = \mathbb{E}\max(X_1,\dots,X_N) = \sum_{k\geq 0}\Bigl(1-\bigl(1-(pk+q)q^{k-1}\bigr)^N\Bigr). \tag{1.1}$$

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

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

Then we would have to write this sum as a contour integral (cf. [3]). But following Rice's method would require deformation of the contour of integration and therefore analytic continuation of the function

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

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 we have 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 $\mathbb{E}_N$. 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

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

Then we have

$$\mathbb{E}_N - \tilde{\mathbb{E}}_N = \left(\sum_{k\leq \operatorname{lq} N} + \sum_{\operatorname{lq} N<k\leq \operatorname{lq} N+2\operatorname{lq}\operatorname{lq} N} + \sum_{k>\operatorname{lq} N+2\operatorname{lq}\operatorname{lq} N}\right) \Bigl(\exp\bigl(-(pk+q)q^{k-1}N\bigr)-\bigl(1-(pk+q)q^{k-1}\bigr)^N\Bigr). \tag{2.2}$$

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

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

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

(2) $\operatorname{lq} N < k \leq \operatorname{lq} N + 2 \operatorname{lq} \operatorname{lq} N$: 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 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

$$\mathbb{E}^*_N = \sum_{k\geq 0}\Big(1 - \exp\big(-(p \operatorname{lq} N + q)Nq^{k-1}\big)\Big). \tag{2.3}$$

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

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

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

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

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

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

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

evaluated at $T = \frac{p}{q} N\,\mathrm{lq}\, N + N$. This sum occurred in many asymptotic investigations, e.g. [6]. By Mellin's inversion formula [1,2] 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

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

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

Theorem 1.
(2.5)
$$\mathbb{E}_N^* = \mathrm{lq}\, N + \mathrm{lq}\,\mathrm{lq}\, N + \mathrm{lq}\, p + \frac{3}{2} + \frac{\gamma}{\log Q} + F\left(\mathrm{lq}\, N + \mathrm{lq}\,\mathrm{lq}\, N + \mathrm{lq}\, p\right) + O\left(\frac{(\log\log N)^2}{\log N}\right)$$

where F is a C^∞ periodic 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

$$\sum_{k\geq 0}\Big(1-\big(1-(pk+q)q^{k-1}\big)^N\Big) \tag{3.1}$$

can be immediately extended to find the expansion for

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

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

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

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

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

Therefore this expectation is given by

$$\begin{aligned}\mathbb{E}_N^{(b)} &= \operatorname{lq} N + (b-1)\operatorname{lq}\operatorname{lq} N + (b-1)\operatorname{lq} p + (b-1) - \operatorname{lq}(b-1)! + \frac{1}{2} + \frac{\gamma}{\operatorname{lq} Q}\\ &\quad + F\big(\operatorname{lq} N + (b-1)\operatorname{lq} N + (b-1)\operatorname{lq} p - \operatorname{lq}(b-1)!\big) + o(1).\end{aligned} \tag{3.3}$$

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

References

1. G.Doetsch, *Handbuch der Laplace-Transformation*, Birkhäuser, Basel, 1950.
2. P.Flajolet, M.Régnier, and R.Sedgewick, *Some Uses of the Mellin Integral Transform in the Analysis of Algorithms*, Combinatorics on Words, Springer NATO ASI Series F, Volume 12, Berlin, 1985.
3. P.Grabner and H.Prodinger, *The Fibonacci Killer*, The Fibonacci Quarterly (1994) (to appear).
4. N.L.Johnson and S.Kotz, *Urn models and Their Applications*, John Wiley and Sons, New York, 1977.
5. P.Kirschenhofer and H.Prodinger, *A Result in Order Statistics Related to Probabilistic Counting*, Computing **51** (1993), 15–27.
6. D.E.Knuth, *The Art of Computer Programming, Vol.3*, Addison–Wesley, Reading, MA, 1973.
7. W.Szpankowski and V.Rego, *Yet another application of a binomial recurrence: Order Statistics*, Computing **43** (1990), 401–410.

P. Grabner
Institut für Mathematik A
Steyrergasse 30
A–8010 Graz, Austria

H. Prodinger
Institut für Algebra
und Diskrete Mathematik
Wiedner Hauptstrasse 8–10
A–1040 Wien, Austria

E-mail address: grabner@ftug.dnet.tu–graz.ac.at, proding@email.tuwien.ac.at