Ramanujan's Q-function and Computer Science Applications

Helmut Prodinger
Technical University of Vienna

October 25, 1993

[summary by Xavier Gourdon]

Abstract

The function $Q(n) = 1 + \frac{(n-1)}{n} + \frac{(n-1)(n-2)}{n^2} + \cdots$ plays a vital role in the analysis of several algorithms and some Discrete Mathematics problems, like the classical birthday problem, hashing with linear probing, etc. We exhibit some of those connections. Also, we discuss how an original question of Ramanujan can be attacked by methods from complex analysis.

1. Ramanujan's Q-function

Let n be a positive integer. We defined the Ramanujan's Q-function of n by

$$Q(n) = 1 + \left(\frac{n-1}{n}\right) + \left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right) + \cdots = \sum_{k=1}^{n} \frac{(n)_k}{n^k}.$$

2. A first problem: Hashing with linear probing

This was Knuth's first analysis. We consider the following game: n players arrive sequentially at a random chair in $\{1, \ldots, m\}$ with $m \geq n$. If the chair is already occupied, then the corresponding player try to sit on the next chair, and continue until the chair is free. We denote by $d(m,n)$ the average distance that the n-th player has to travel.

For example, consider 6 players A, B, C, D, E, F with 6 chairs in $\{1,2,3,4,5,6\}$, and suppose

$$A \to 3, \quad B \to 1, \quad C \to 4, \quad D \to 3-4-5, \quad E \to 1-2, \quad F \to 3-4-5-6.$$

Players A, B, C travelled a distance 0, player D a distance 2, player E a distance 1 and player F a distance 3.

Our parameter of interest is

$$\delta(m,n) = \frac{d(m,1) + \cdots + d(m,n)}{n}.$$

All m^n "hash sequences" are assumed to be equally likely. An analysis leads to

$$d(m,n) = \frac{(m-n)m^{1-n}}{2} \cdot \sum_{r \geq 0} r \binom{n-1}{r} (r+1)^r (m-r-1)^{n-r-2}.$$

Thanks to Abel's binomial theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x(x-kz)^{k-1}(y+kz)^{n-k},$$

we find

$$d(m,n)=\frac{1}{2}\left(2\frac{n-1}{m}+3\frac{(n-1)(n-2)}{m^2}+\cdots\right)$$

so that

$$\delta(m,n)=\frac{1}{2}\left(\frac{n-1}{m}+\frac{(n-1)(n-2)}{m^2}+\cdots\right).$$

Asymptotic development of $\delta(m,m)$. We consider the case $m=n$, for which $\delta(m,n)=\frac{1}{2}(Q(n)-1)$. We now have to find the asymptotic expansion of $Q(n)$. We introduce

$$R(n)=1+\frac{n}{n+1}+\frac{n^2}{(n+1)(n+2)}+\cdots$$

We have the relation

$$Q(n)+R(n)=\frac{n!e^n}{n^n}\sim\sqrt{2\pi n}+\cdots$$

(the last term follows from Stirling's formula). The expression of $R(n)$ in terms of the incomplete gamma function $\gamma(a,x)=\int_0^x e^{-t}t^{a-1}\,dt$

$$R(n)=\frac{n!e^n}{n^n}\cdot\frac{\gamma(n,n)}{(n-1)!},$$

together with the asymptotic development [6, §1.2.11.3]

$$\frac{\gamma(x+1,x+y)}{\Gamma(x+1)}=\frac{1}{2}+\frac{y-2/3}{\sqrt{2\pi}}x^{-1/2}+\cdots\qquad(x\to\infty, y\text{ fixed})$$

leads to

$$R(n)=\sqrt{\frac{\pi n}{2}}+\frac{1}{3}+\frac{1}{12}\sqrt{\frac{\pi}{2n}}+\cdots\quad\text{and}\quad Q(n)=\sqrt{\frac{\pi n}{2}}-\frac{1}{3}+\frac{1}{12}\sqrt{\frac{\pi}{2n}}+\cdots$$

Another interesting proof can be found in [4, p. 5].

3. The Flajolet-Gardy-Thimonier approach

We consider the general birthday paradox. The year has m days, with probability of appearance $p_1,\ldots,p_m$. We denote by $E(B_j)$ the expected number of drawing until you first encounter j days with multiplicity at least k (the classical case is $k=2$, $j=1$, with $m=365$). The problem can be coded by an appropriate language, which translates easily in terms of exponential generating functions, and using the Laplace-Borel transform to get back to ordinary generating functions (see [3]). We have

$$E(B_j)=\sum_{q=0}^{j-1}\int_0^\infty[u^q]\prod_{i=1}^m\left(e_{k-1}(p_iz)+u(e^{p_iz}-e_{k-1}(p_iz))\right)e^{-z}\,dz,$$

where $e_{k-1}(x)=\sum_{j=0}^{k-1}x^j/j!$. Let's see special instances. When $j=1$ (first k-hit), we have

$$E(B_1)=\int_0^\infty\left(\prod_{i=1}^\infty e_{k-1}(p_iz)\right)e^{-z}\,dz,$$

and for the equiprobable case $p_i=1/m$

$$E(B_1)=\int_0^\infty\left(e_{k-1}(\frac{z}{m})\right)^m e^{-z}\,dz,$$

so that for $k = 2$

$$E(B_1) = \int_0^\infty \left(1 + \frac{z}{m}\right)^m e^{-z}\, dz = 1 + Q(m). \tag{1}$$

For $k = 1$ and $j = m$ (coupon collector's problem)

$$E(B_m) = \int_0^\infty \left(1 - \prod_{i=1}^{m}(1 - e^{-p_i z})\right) dz$$

so that in the equiprobable case

$$E(B_m) = m \left(\sum_{q=1}^{m}(-1)^{q-1}\binom{m}{q}\frac{1}{q}\right) = mH_m.$$

4. A conjecture of Ramanujan about the Q-function

Define θ by

$$\sum_{k=0}^{n-1}\frac{n^k}{k!} + \frac{n^n}{n!}\theta = \frac{e^n}{2}.$$

It follows from [1, p. 181, entry 48] that $\theta \approx \dfrac{4 + 15n}{8 + 45n}$, and Ramanujan conjectured that always $\frac{1}{3} \le \theta \le \frac{1}{2}$. This was proved independently by Watson and Szegö. A more refined conjecture (also posed by Ramanujan) was proved only recently by Flajolet, Grabner, Kirschenhofer and Prodinger [4]. We have

$$\theta = \frac{1}{3} + \frac{4}{135(n+k)} \quad \text{where always} \quad \frac{2}{21} \le k \le \frac{8}{45}.$$

All this is related to the Q-function because

$$\frac{n^n}{n!}Q(n) = \frac{n^{n-1}}{(n-1)!} + \frac{n^{n-2}}{(n-2)!} + \cdots + 1 = \frac{1}{2}e^n - \theta\frac{n^n}{n!}.$$

Since

$$Q(n) + R(n) = \frac{n!e^n}{n^n}, \quad \text{we have} \quad \theta(n) = \frac{1}{2}(R(n) - Q(n)),$$

and the Ramanujan's problem can be rephrased as

$$D(n) = R(n) - Q(n) = \frac{2}{3} + \frac{8}{135(n+k)} \quad \text{where always} \quad \frac{2}{21} \le k \le \frac{8}{45}.$$

The approaches followed by Ramanujan himself and later authors all make use of the integral (1) and proceed using the Laplace method for the asymptotic evaluation of integrals. The idea used in [4] is rather different and uses complex analysis applied to famous Knuth's "tree function" $y(z)$, defined implicitly by the equation

$$y(z) = ze^{y(z)}.$$

Thanks to Lagrange's inversion formula, we get

$$[z^n]y(z) = \frac{n^{n-1}}{n!}, \qquad Q(n)\,\frac{n^{n-1}}{n!} = [z^n]\log\frac{1}{1 - y(z)}$$

and

$$D(n)\,\frac{n^{n-1}}{n!} = [z^n]\log\frac{(1 - y(z))^2}{2(1 - y(z)e^{1-y(z)})}.$$

By singularity analysis [5] this equality can be used to get easily the first 10-terms of the asymptotic expansion of $D(n)$:

$$D(n) = \frac{2}{3} + \frac{8}{135n} + \cdots - \frac{479}{561330n^9} + O\left(\frac{1}{n^{10}}\right).$$

An estimation of the remainder was done by saddle point heuristic in the y-plane, and it is proved in [4] that this remainder is less than $10^{-7}/n^3$ for $n \geq 116$. Then one can show the desired bounds for $n \geq 116$ and check the remaining 115 cases "by hand".

5. More of Knuth's wisdom

The function $Q(n)$ shows up in unexpected places. For example, Cauchy [2, pp. 62-73] proved that

$$\frac{1}{n^n} \sum_k \binom{n}{k} k^k (n-k)^{n-k} = 1 + Q(n).$$

A Q-algebra. It is possible to develop a Q-algebra theory [7, p. 190]. For every sequence $(a_1, a_2, \ldots)$ and every positive integer n, we consider

$$Q(a_1, a_2, \ldots; n) = \sum_{k \geq 1} a_k \frac{(n)_k}{n^k}.$$

For example, $Q(1, 1, \ldots; n) = Q(n)$. We have the two identities

$$rQ(a_1, a_2, \ldots) + sQ(b_1, b_2, \ldots) = Q(ra_1 + sb_1, ra_2 + sb_2, \ldots)$$

and

$$Q(a_1, 2a_2, 3a_3, \ldots; n) = nQ(a_1, a_2 - a_1, a_3 - a_2, \ldots; n)$$

(for the latest, write $k = n - (n-k)$ — like Abel's partial summation). Therefore

- $Q(1, 2, 3, \ldots; n) = nQ(1, 0, 0, \ldots; n) = n$,
- $Q(1^2, 2^2, 3^2, \ldots; n) = nQ(1, 1, 1, \ldots; n) = nQ(n)$,
- $Q(1^3, 2^3, 3^3, \ldots; n) = 2n^2 - nQ(n)$,
- $Q(1^4, 2^4, 3^4, \ldots; n) = 3n^2Q(n) - 3n^2 + nQ(n)$.

We can always write such a formula for $Q(1^k, 2^k, 3^k, \ldots)$, with certain coefficients. They can be arranged in a triangle. The inverse triangle is also of interest, and the coefficients have combinatorial interpretations in terms of permutations of certain multisets.

Bibliography

[1] Berndt (Bruce C.). – *Ramanujan's Notebooks, Part II.* – Springer Verlag, 1989.

[2] Cauchy (A.). – *Exercices de Mathématiques.* – Paris, 1826.

[3] Flajolet (Philippe), Gardy (Danièle), and Thimonier (Loÿs). – Birthday paradox, coupon collectors, caching algorithms, and self–organizing search. *Discrete Applied Mathematics*, vol. 42, 1992.

[4] Flajolet (Philippe), Grabner (Peter), Kirschenhofer (Peter), and Prodinger (Helmut). – *On Ramanujan's Q–function.* – Research Report n° 1760, Institut National de Recherche en Informatique et en Automatique, October 1992. 13 pages. To appear in the *Journal of Computational and Applied Mathematics.*

[5] Flajolet (Philippe) and Odlyzko (Andrew M.). – Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, vol. 3, n° 2, 1990, pp. 216–240.

[6] Knuth (Donald E.). – *The Art of Computer Programming.* – Addison-Wesley, 1968, vol. 1: Fundamental Algorithms. Second edition, 1973.

[7] Knuth (Donald E.). – An analysis of optimal caching. *Journal of Algorithms*, vol. 6, 1985, pp. 181–199.