APPROXIMATE COUNTING VIA EULER TRANSFORM

Helmut Prodinger

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

Approximate Counting might be described as follows. There are the states $1, 2, \dots$, and in one step one may either advance from state i to state $i+1$ with probability q^i, or stay in the state i with probability $1-q^i$. The interesting parameter is the state that one reaches after n random steps, starting in state 1.

The original analysis was performed by Flajolet [2] and consists of an enumerative (or algebraic) part and an asymptotic (or analytic) part. The latter was done by the Mellin transform. In [5], some additional manipulations allowed to rewrite the sought quantities in such a way that an alternative asymptotic technique (Rice's method) could be used. See also the related papers [6,7].

In this short note we would like to emphasize how some elements of "q–analysis" (basic hypergeometric functions) allow some shortcuts in the enumerative part. Let $H_l(x)$ be the generating function, where the coefficient of x^n is the probability that n random steps have led to state l. We find a rather explicit form for the super generating function $\sum_{l\ge1} H_l(x)y^l$ which is interesting in itself and might lead to additional insight. We only use it here to find representations for the expection and the second factorial moment as alternating sums, with binomial coefficients and some simple quantities, which is essential for the use of Rice's method. One could avoid to use such explicit formulæ and, using the (new) generating functions, derive the asymptotics directly by the ingenious method in [4]. However, here, we only deal with the enumerative part.

We need a few concepts from q–analysis which are taken from [1].
q–**Pochhammer symbol**:

$$(a)_n := (1-a)(1-aq)\dots(1-aq^{n-1}), \qquad (a)_0 = 1, \qquad (a)_\infty = \lim_{n\to\infty}(a)_n$$

Cauchy's formula:

$$\sum_{n\ge0}\frac{(a)_n t^n}{(q)_n} = \frac{(at)_\infty}{(t)_\infty}.$$

Heine's transformation:

$$\sum_{n\ge0}\frac{(a)_n(b)_n t^n}{(q)_n(c)_n} = \frac{(b)_\infty(at)_\infty}{(c)_\infty(t)_\infty}\sum_{n\ge0}\frac{(c/b)_n(t)_n b^n}{(q)_n(at)_n}.$$

I thank Christian Krattenthaler for his help with q–hypergeometric functions

Euler's transformation: If

$$f(x)=\sum_{n\ge0}a_nx^n,$$

then

$$\frac{1}{1-x}f\Big(\frac{x}{x-1}\Big)=\sum_{n\ge0}\bigg(\sum_{k=0}^{n}\binom{n}{k}(-1)^ka_k\bigg)x^n.$$

This is very easily computed directly and was used with great success in [4] and [3].

The generating function $H_l(x)$ was computed in [2]. Using a decomposition of a path from 1 to l in stages, it is not hard to see that

$$H_l(x)=\frac{x^{l-1}q^{\binom{l}{2}}}{\prod\limits_{i=1}^{l}\big(1-x(1-q^i)\big)}=\frac{\frac{1}{x}\big(\frac{x}{1-x}\big)^lq^{\binom{l}{2}}}{\big(\frac{xq}{x-1}\big)_l}.$$

The coefficient of x^n in $\sum_{l\ge1}H_l(x)$ must be 1 for all n, since each path of length n must simply lead somewhere. Let us see how we find the formula

$$\sum_{l\ge1}H_l(x)=\frac{1}{1-x}$$

by some properties of "q-analysis". It is equivalent to show

$$\sum_{l\ge1}\frac{(-z)^lq^{\binom{l}{2}}}{(qz)_l}=-z,$$

with $z=\frac{x}{x-1}$. Let us show that

$$\sum_{l\ge0}\frac{(-z)^lq^{\binom{l}{2}}}{(qz)_l}=1-z.$$

First, write

$$\begin{aligned}(-1)^lq^{\binom{l}{2}}&=(0-1)(0-q)\dots(0-q^{l-1})\\&=\lim_{\varepsilon\to0}\varepsilon^l(1/\varepsilon)_l.\end{aligned}$$

Then we can use the transformation of Heine and compute

$$\sum_{l\ge0}\frac{(1/\varepsilon)_l(\varepsilon z)^l}{(qz)_l}\cdot\frac{(q)_l}{(q)_l}=\frac{(q)_\infty(z)_\infty}{(qz)_\infty(\varepsilon z)_\infty}\sum_{n\ge0}\frac{(z)_n(\varepsilon z)_nq^n}{(q)_n(z)_n}.$$

Now we can perform the limit $\varepsilon\to\infty$ on the right hand side without problems and obtain

$$\frac{(q)_\infty(z)_\infty}{(qz)_\infty}\sum_{n\ge0}\frac{q^n}{(q)_n}.$$

We use the trivial fact $(z)_\infty = (1-z)(qz)_\infty$ for the first factor and Cauchy's identity (the special case which is attributed to Euler), which evaluates the sum to $1/(q)_\infty$, which gives us the desired $1-z$. Great.

Now let us attack the generating function of the expectations,

$$\sum_{l\ge 1} l\, H_l(x).$$

For that, we consider the super generating function

$$H(x,y) = \sum_{l\ge 0} H_l(x) y^l,$$

differentiate it w.r.t. y and evaluate at $y=1$. We obtain

$$H(x,y) = \frac{1}{x}\sum_{l\ge 0} \frac{(-yz)^l q^{\binom{l}{2}}}{(qz)_l}, \qquad z = \frac{x}{x-1}.$$

Using Heine's transformation as before, we get

$$\sum_{l\ge 0} \frac{(-yz)^l q^{\binom{l}{2}}}{(qz)_l} = \frac{(q)_\infty (yz)_\infty}{(qz)_\infty} \sum_{n\ge 0} \frac{(z)_n q^n}{(q)_n (yz)_n}.$$

For $y=1$ we could evaluate the sum immediately. Now we transform the sum again by "Heine", with $a=0$, $b=z$, $c=yz$ and $t=q$ and find (for the sum only)

$$\frac{(z)_\infty}{(yz)_\infty (q)_\infty} \sum_{n\ge 0} \frac{(y)_n (q)_n z^n}{(q)_n},$$

which gives

$$H(x,y) = \frac{1}{x}(1-z)\sum_{n\ge 0} (y)_n z^n, \qquad z = \frac{x}{x-1}.$$

Now observe that

$$\frac{d}{dy}(y)_n\Big|_{y=1} = -(q)_{n-1}, \qquad n \ge 1,$$

so that the generating function of the expectations $E(x)$ is

$$E(x) = -\frac{1}{x}(1-z)\sum_{n\ge 0} (q)_n z^{n+1} = (1-z)^2 \sum_{n\ge 0} (q)_n z^n.$$

According to Euler's transformation we have to extract the coefficient of z^n in

$$\frac{1}{1-z}\cdot (1-z)^2 \sum_{n\ge 0} (q)_n z^n = (1-z)\sum_{n\ge 0} (q)_n z^n,$$

which is

$$(q)_n - (q)_{n-1} = -q^n (q)_{n-1}$$

for $n \geq 1$ and 1 for $n = 0$. Hence the expected value E_n is

$$E_n = 1 - \sum_{k=1}^{n} \binom{n}{k} (-1)^k q^k (q)_{k-1},$$

a formula already reported in [5].

To obtain the generating function $E_2(x)$ for the second factorial moments we have to differentiate twice w.r.t. y and evaluate at $y = 1$. Observe that for $n \geq 2$

$$\frac{d^2}{dy^2}(y)_n\Big|_{y=1} = 2(q)_{n-1} \sum_{k=1}^{n-1} \frac{q^k}{1-q^k}.$$

Let us abbreviate

$$T_n = \sum_{k=1}^{n} \frac{q^k}{1-q^k}.$$

Then

$$E_2(x) = \frac{1}{x}(1-z) \cdot 2 \sum_{n \geq 1} (q)_n T_n z^{n+1}.$$

As before, we have to extract the coefficient of z^n in

$$\frac{1}{1-z} E_2(x) = -2(1-z) \sum_{n \geq 1} (q)_n T_n z^n,$$

which is

$$-2\Big((q)_n T_n - (q)_{n-1} T_{n-1}\Big) = 2q^n (q)_{n-1} \Big(T_{n-1} - 1\Big).$$

Hence the second factorial moment $E_n^{(2)}$ is given by

$$E_n^{(2)} = \sum_{k=1}^{n} \binom{n}{k} (-1)^k \cdot 2q^k (q)_{k-1} \Big(T_{k-1} - 1\Big).$$

This is equivalent to the formula given in [5].

As stated before, asymptotics follow, as demonstrated in [5], by Rice's method. However, since we have the explicit forms of the generating functions of the moments, an approach like the one in [4] would give the asymptotics even a little bit more directly. We do not work it out, because the computations are somehow similar to the ones which occur with Rice's method.

References

1. G. E. Andrews, *The Theory of Partitions*, Addison Wesley, 1976.
2. P. Flajolet, *Approximate Counting: A Detailed Analysis*, BIT **25** (1985), 113–134.
3. P. Flajolet, G. Labelle, L. Laforest, and B. Salvy, *Hypergeometrics and the Cost Structure of Quadtrees*, Random Structures and Algorithms (1995) (to appear).
4. P. Flajolet and B. Richmond, *Generalized Digital Trees and their Difference–Differential Equations*, Random Structures and Algorithms **5** (1992), 305–320.
5. P. Kirschenhofer and H. Prodinger, *Approximate Counting: An Alternative Approach*, RAIRO Informatique Théorique et Applications **25** (1991), 43–48.
6. P.Kirschenhofer and H. Prodinger, *A coin tossing algorithm for counting large numbers of events*, Mathematica Slovaca **42** (1992), 531–545.
7. H. Prodinger, *Hypothetic Analyses: Approximate Counting in the Style of Knuth, Path Length in the Style of Flajolet*, Theoretical Computer Science **100** (1992), 243–251.

TU Vienna
Wiedner Hauptstrasse 8–10
A-1040 Vienna
Austria
E-mail address: proding@email.tuwien.ac.at