On the sum of digits of the Bell number

JAVIER CILLERUELO
Instituto de Ciencias Matemáticas (CSIC-UAM-UC3M-UCM)
and Departamento de Matemáticas
Universidad Autónoma de Madrid
28049, Madrid, España
franciscojavier.cilleruelo@uam.es

FLORIAN LUCA
Instituto de Matemáticas
Universidad Nacional Autonoma de México
C.P. 58089, Morelia, Michoacán, Mexico
fluca@matmor.unam.mx

JUANJO RUÉ
Instituto de Ciencias Matemáticas (CSIC-UAM-UC3M-UCM)
Universidad Autónoma de Madrid
28049, Madrid, Spain
juanjo.rue@icmat.es

ANA ZUMALACÁRREGUI
Instituto de Ciencias Matemáticas (CSIC-UAM-UC3M-UCM)
and Departamento de Matemáticas
Universidad Autónoma de Madrid
28049, Madrid, España
ana.zumalacarregui@uam.es

May 13, 2011

Abstract

Let $B(n)$ be the Bell number which counts the number of a partitions of a set with n elements. Let $b \geq 2$ be a fixed positive integer. In this paper, we show that for most n the sum of the digits of $B(n)$ in base b is at least $(\log n)/(60 \log b)$.

1 Introduction

Let $B(n)$ be the Bell number of n which counts the number of a partitions of a set with n elements. In this paper we study the representation of $B(n)$ in an integer base $b \geq 2$.

Denote by $s_b(m)$ the sum of the digits of the positive integer m when is written in base b. Lower bounds for $s_b(m)$ when m runs through the members of a sequence with some interesting combinatorial meaning have been investigated before. For example, it follows from a result of Stewart ([8]; see also [3] for a slightly more general result), that the inequality

$$s_b(F_n) > c_1 \frac{\log n}{\log \log n}$$

holds for all $n \geq 3$ with some positive constant $c_1 := c_1(b)$ depending on b, where F_n is the nth Fibonacci number given by $F_0 := 0,\ F_1 := 1$ and $F_{n+2} := F_{n+1} + F_n$ for all $n \geq 0$. In [4], it is shown that the inequality

$$s_b(n!) > c_2 \log n$$

holds for all $n \geq 1$, where $c_2 := c_2(b)$ is some positive constant depending on b. In [6], it was shown that if we put $b_n := \binom{2n}{n}$ and $c_n := \frac{1}{n+1}\binom{2n}{n}$ for the middle binomial coefficient and Catalan number, respectively, then both inequalities

$$s_b(b_n) \geq \varepsilon(n)\sqrt{\log n} \qquad \text{and} \qquad s_b(c_n) > \varepsilon(n)\sqrt{\log n} \tag{1}$$

hold on a set of n of asymptotic density equals to 1, where $\varepsilon(n)$ is any function tending to zero when n tends to infinity. In [7], it was shown that there is some positive constant $c_3 := c_3(b)$ depending on b such that if we put

$$A_n := \sum_{k=0}^{n} \binom{n}{k}^2 \binom{n+k}{k}^2$$

for the nth Apéry number, then the inequality

$$s_b(A_n) > c_3 \left(\frac{\log n}{\log\log n}\right)^{1/4} \tag{2}$$

holds on a set of n of asymptotic density 1. Some of the above results were superseded by the results from the recent paper [2], where it is shown that if $\mathbf{r} := (r_0, r_1, \ldots, r_m)$ is a fixed vector of nonnegative integers integers with $r_0 > 0$ and if we put

$$S(n) := \sum_{k=0}^{n} \binom{n}{k}^{r_0} \binom{n+k}{k}^{r_1} \cdots \binom{n+km}{k}^{r_m} \quad \text{for} \quad n = 0, 1, \ldots,$$

then for $\mathbf{r} \neq (1)$ there exists a positive constant $c_4 := c_4(b, \mathbf{r})$ depending on both b and $\mathbf{r}$ such that the inequality

$$s_b(S(n)) > c_4 \frac{\log n}{\log\log n} \tag{3}$$

holds for almost all n. Note that inequality (3) improves (1) for the case of the middle binomial coefficients b_n because $c_n = S(n)$ for $\mathbf{r} = (2)$, as well as inequality (2) for the case of the Apéry numbers A_n because $A_n = S(n)$ for $\mathbf{r} = (2, 2)$.

In [5], it is shown that if $p(n)$ is the partition function of n, then the inequality

$$s_b(p(n)) > \frac{\log n}{7 \log\log n}$$

holds for almost all positive integers n.

The proofs of such results use a variety of methods from number theory, such as elementary methods, sieve methods, linear forms in logarithms and the subspace theorem of Evertse–Schlickewei–Schmidt. Here, we add on the literature on the topic and prove the following theorem.

Theorem 1. *Consider an integer base $b \geq 2$. The inequality*

$$s_b(B(n)) > \frac{\log n}{60 \log b}$$

holds on a set of positive integers n of asymptotic density equals to 1.

We use the Landau symbol O and o as well as the Vinogradov symbols $\ll$, $\gg$ and $\asymp$ with their usual meanings. Recall that $A = O(B)$, $A \ll B$ and $B \gg A$ are all equivalent to the fact that the inequality $|A| \leq cB$ holds with some constant c. The constants implied by these symbols in our arguments might depend in the number b. Furthermore, $A \asymp B$ means that both $A \ll B$ and $B \ll A$ hold.

2 Preliminary results

We start with the following approximation for $B(n)$ (see formula (41) on page 562 in [1]).

Lemma 2. *Let $r := r(n)$ be defined implicitly by*

$$re^r = n + 1. \tag{4}$$

Then

$$B(n) = \frac{n! e^{e^r - 1}}{r^n \sqrt{2\pi r(r+1)e^r}} \left(1 + O\left(e^{-r/5}\right)\right). \tag{5}$$

The number $r := r(n)$ defined implicitly at (4) satisfies $r = \log n - \log\log n + o(1)$ as $n \to \infty$, therefore

$$e^{-r/5} = \left(\frac{\log n}{n}\right)^{1/5} (1 + o(1)) \quad \text{as} \quad n \to \infty. \tag{6}$$

Throughout the paper, we will assume that $r := r(z)$ is the function defined implicitly for all real numbers $z \geq 1$ as shown at (4) with n replaced by z. In particular, $r(z)$ has a derivative for real $z > 1$.

3 The proof of Theorem 1

Consider the following set positive integers:

$$\mathcal{N}_b(x) := \left\{ n \in [x/2, x) : s_b(B(n)) < \frac{\log n}{60 \log b} \right\}.$$

We need to show that $\#\mathcal{N}_b(x) = o(x)$ as $x \to \infty$, for after this the conclusion of Theorem 1 will follow by replacing x by $x/2$, then by $x/4$, and so on, and summing up the resulting estimates.

For $n \in \mathcal{N}_b(x)$, we write

$$B(n) = d_1 b^{n_1} + d_2 b^{n_2} + \cdots + d_s b^{n_s}, \tag{7}$$

where $d_1, \ldots, d_s \in \{1, \ldots, b-1\}$ and $n_1 > n_2 > \cdots > n_s$. We put $t := t(n)$ for the smallest index $i \in \{1, 2, \ldots, s-1\}$ such that $b^{n_1 - n_{i+1}} > n^{1/6}$ if it exists and set $t := s$ otherwise. From the definition of $t(n)$, we see immediately that

$$\begin{aligned} B(n) &= \left(d_1 b^{n_1} + \cdots + d_t b^{n_t}\right)\left(1 + O\left(n^{-1/6}\right)\right) \\ &:= b^{m(n)} D(n) \left(1 + O\left(n^{-1/6}\right)\right), \end{aligned} \tag{8}$$

where $m = m(n) := n_t$ and $D(n) := d_1 b^{n_1 - n_t} + d_2 b^{n_2 - n_t} + \cdots + d_t$.

Let $\mathcal{D}_b(x)$ be the subset of all possible values for $D(n)$, $n \in \mathcal{N}_b(x)$. Let us find an upper bound for the cardinality of this set. Observe first that

$$D(n) < b^{n_1 - n_t + 1} \le b^{(\log n)/(6 \log b) + 1}.$$

The positive integers $D := D(n)$ bounded by the right hand side of the above inequality quantity have at most $K := \lfloor (\log x)/(6 \log b) + 2 \rfloor$ digits in base b. As $n \in \mathcal{N}_b(x)$, the number of nonzero digits of $D(n)$ is bounded by $S := \lfloor (\log x)/(60 \log b) \rfloor$, and the number of possible values for D is at most

$$\begin{aligned} \sum_{i=0}^{S} \binom{K}{i} (b-1)^i &\le (S+1) \binom{K}{S} (b-1)^S \le (S+1) \left(\frac{(b-1)eK}{S} \right)^S \\ &\le \left(\frac{\log x}{60 \log b} + 1 \right) (10e(b-1) + o(1))^{\frac{\log x}{60 \log b}} = x^{\delta + o(1)} \end{aligned}$$

as $x \to \infty$, where

$$\delta := \frac{\log(10e(b-1))}{60 \log b}.$$

It can be checked that $\delta < 1/12$ for all integers $b \geq 2$. Thus, we get that

$$\#\mathcal{D}_b(x) = x^{\delta + o(1)} \qquad \text{as} \qquad x \to \infty \tag{9}$$

for some $\delta < 1/12$.

We now compare relations (5) (with estimate (6) for the error term), (7) and (8). Recalling that $n = O(x)$ we get that

$$\frac{n! e^{e^{r(n)}-1}}{r(n)^n \sqrt{2\pi r(n)(r(n)+1)e^{r(n)}}} = b^{m(n)} D(n) \left(1 + O\left(x^{-1/6}\right)\right).$$

Taking logarithms we get that

$$\begin{aligned} \log n! + e^{r(n)} \quad &- \quad \left(n + \frac{1}{2}\right) \log r(n) - \frac{1}{2} \log(r(n)+1) - \frac{r(n)}{2} - 1 - \frac{\log(2\pi)}{2} \\ &= \quad m(n) \log b + \log D(n) + O\left(x^{-1/6}\right). \end{aligned} \tag{10}$$

We now write

$$\mathcal{N}_b(x) = \bigcup_{D \in \mathcal{D}_b(x)} \mathcal{N}_{b,D}(x),$$

where

$$\mathcal{N}_{b,D}(x) := \{n \in \mathcal{N}_b(x) : D(n) = D\}.$$

We put $y := x^{1/12}$. Fix some $D \in \mathcal{D}_b(x)$. We take a look at the elements $n \in \mathcal{N}_{b,D}(x)$. We say that n is *separated* if the smallest element in $\mathcal{N}_{b,D}(x)$ which exceeds n, call it n', if it exists, satisfies $n' - n > y$. It is clear that there are at most $2x/y + 1 \ll x/y$ elements on $\mathcal{N}_{b,D}(x)$ which are separated. Let us now count the nonseparated $n \in \mathcal{N}_{b,D}(x)$. For such an n, there exists $n' := n + k$ with $1 \leq k \leq y$, such that $n' \in \mathcal{N}_{b,D}(x)$. Taking the difference of the relations (10) in n and $n + k$ we get

$$\sum_{i=1}^{k} \log(n+i) - (f(n+k) - f(n)) = (m(n+k)) - m(n)) \log b + O\left(x^{-1/6}\right), \tag{11}$$

where

$$f(z) := e^{r(z)} - \left(z + \frac{1}{2}\right) \log r(z) - \frac{1}{2} \log(r(z)+1) - \frac{1}{2} r(z).$$

We need to know the asymptotic behavior of the derivative of $f(z)$. Differentiating relation (4) (with n replaced by z) with respect to the variable z, we have

$$r(z)'e^{r(z)} + r(z)r'(z)e^{r(z)} = 1,$$

therefore

$$r'(z)e^{r(z)} = \frac{1}{r(z)+1} \tag{12}$$

and

$$r'(z) = \frac{r(z)}{(z+1)(r(z)+1)}. \tag{13}$$

Furthermore,

$$\begin{aligned}\frac{d}{dz}\left(\left(z+\frac{1}{2}\right)\log r(z)\right) &= \log r(z) + \frac{(2z+1)r'(z)}{2r(z)} \\ &= \log r(z) + \frac{2z+1}{2(z+1)(r(z)+1)}, \end{aligned} \tag{14}$$

$$\frac{d}{dz}\left(\frac{1}{2}\log(r(z)+1)\right) = \frac{r'(z)}{2(r(z)+1)} = \frac{r(z)}{2(z+1)(r(z)+1)^2}, \tag{15}$$

and

$$\frac{d}{dz}\left(\log r(z)\right) = \frac{r'(z)}{r(z)} = \frac{1}{(z+1)(r(z)+1)}. \tag{16}$$

Thus, using estimates (12), (13), (14) and (15), we get that

$$\begin{aligned}\frac{df(z)}{dz} &= \frac{de^{r(z)}}{dz} - \frac{d}{dz}\left(\left(z+\frac{1}{2}\right)\log r(z)\right) - \frac{1}{2}\frac{d}{dz}\left(\log(r(z)+1)\right) - \frac{1}{2}\frac{dr(z)}{dz} \\ &= \frac{1}{r(z)+1} - \log r(z) - \frac{2z+1}{2(z+1)(r(z)+1)} - \\ &\quad \frac{r(z)}{2(z+1)(r(z)+1)^2} - \frac{r(z)}{2(z+1)(r(z)+1)} \\ &= -\log r(z) + \frac{r^2(z)+r(z)-1}{2(r(z)+1)^2(z+1)} \\ &= -\log r(z) + O\left(z^{-1}\right). \end{aligned} \tag{17}$$

Furthermore, since

$$\begin{aligned}\sum_{i=1}^{k}\log(n+i) &= k\log(n+1)+\sum_{i=0}^{k-1}\log\left(1+\frac{i}{n+1}\right)\\ &= k\log(n+1)+O\left(\sum_{i=0}^{k-1}\frac{i}{n}\right)\\ &= k\log(n+1)+O\left(\frac{k^2}{x}\right),\end{aligned}$$

we get, using formula (11), the Intermediary Value Theorem (twice), and the calculation (16) and (17), together with the assumption that $k \le y = x^{1/12}$, that

$$\begin{aligned}(m(n+k)-m(n))\log b &= k\log(n+1)-(f(n+k)-f(n))+O\left(\frac{k^2}{x}+x^{-1/6}\right)\\ &= k\log(n+1)-k\left[\frac{d}{dz}f(z)\right]_{z=\zeta\in[n,n+k]}+O\left(x^{-1/6}\right)\\ &= k\log(n+1)-k\log r(\zeta)+O\left(\frac{k}{x}+x^{-1/6}\right)\\ &= k\log(n+1)-k\log r(n)+k(\log r(\zeta)-\log r(n))+O\left(x^{-1/6}\right)\\ &= k\log\left(\frac{n+1}{r(n)}\right)-k(\zeta-n)\left[\frac{d}{dz}\log r(z)\right]_{z=\nu\in[n,\zeta]}+O\left(x^{-1/6}\right)\\ &= kr(n)-\frac{k(\zeta-n)}{(\nu+1)(r(\nu)+1)}+O\left(x^{-1/6}\right)\\ &= kr(n)+O\left(\frac{k^2}{x\log x}+x^{-1/6}\right)\\ &= kr(n)+O\left(x^{-1/6}\right),\end{aligned} \tag{18}$$

where $\zeta \in [n, n+k]$ and $\nu \in [n, \zeta]$ are some points the existence of which is guaranteed by the Intermediary Value Theorem.

We first see that the above estimates imply that $m(n+k) \neq m(n)$ for large x, for if not, we would then have that

$$\log x \ll k\log x \ll kr(n) \ll x^{-1/6},$$

which is obviously impossible for $k \ge 1$ and large x.

Estimate (18) together with the fact that $m(n+k) - m(n) \in \mathbb{Z}^*$ shows that if we put $\|z\|$ for the distance from the real number z to the nearest integer, then $\|kr(n)/\log b\| = O(x^{-1/6})$. So, it remains to count such n's.

Fix $k \le y$. Since $r'(z) \sim 1/z \asymp 1/x$, for n, $n+l \in \mathcal{N}_{b,D}(x)$, we have that

$$\frac{kr\,(n+l)}{\log b} - \frac{kr\,(n)}{\log b} \asymp k\frac{l}{x\log b}.$$

From this estimate we deduce that $\lfloor kr(n)/\log b \rfloor$ takes $O(k)$ integer values. Using the same argument, we have that for each of these integer values M, the estimate $kr(n)/\log b = M + O(x^{-1/6})$ holds for $O(x^{5/6}/k)$ values of n. Thus, for each $k \le y$ the estimate $\|kr(n)/\log b\| = O(x^{-1/6})$ holds for $O(x^{5/6})$ values of n, and there are $O(yx^{5/6})$ elements of $\mathcal{N}_{b,D}(x)$ which are not separated. Hence,

$$\#\mathcal{N}_{b,D}(x) \ll yx^{5/6} + \frac{x}{y} \ll x^{11/12}. \tag{19}$$

This was for an arbitrary $D \in \mathcal{D}_b(x)$. Thus, by estimate (9),

$$\#\mathcal{N}_b(x) = \sum_{D\in\mathcal{D}_b(x)} \#\mathcal{N}_{b,D}(x) \le x^{11/12}\#\mathcal{D}_b(x) < x^{11/12+\delta+o(1)} = o(x),$$

as $x \to \infty$, which is what we wanted to prove.

Acknowledgements. F. L. thanks Professor Arnold Knopfmacher for useful suggestions. This paper was written while F. L. was in sabbatical from the Mathematical Institute UNAM from January 1 to June 30, 2011 and supported by a PASPA fellowship from DGAPA. J. R. is supported by a JAE-DOC grant from the JAE program in CSIC, Spain. A. Z. is supported by Departamento de Matemáticas of Universidad Autónoma de Madrid, Spain.

References

[1] P. Flajolet, R. Sedgewick, *Analytic Combinatorics*, Cambridge Univ. Press, 2009.

[2] A. Knopfmacher and F. Luca, "Digit sums of binomial sums", *Preprint*, 2011.

[3] F. Luca: "Distinct digits in base b expansions of linear recurrence sequences", *Quaest. Math* **23** (2000), 389–404.

[4] F. Luca: "The number of nonzero digits of $n!$", *Canadian Math. Bull.* **45** (2002), 115–118.

[5] F. Luca: "The the number of nonzero digits of the partition function", *Preprint*, 2011.

[6] F. Luca and I. E. Shparlinski, "On the g-ary expansions of middle binomial coefficients and Catalan numbers", *Rocky Mtn. J. Math.*, to appear.

[7] F. Luca and I. E. Shparlinski, "On the g-ary expansions of Apéry, Motzkin and Schröder numbers", *Annals of Combinatorics* **14** (2010), 507–524.

[8] C. L. Stewart, "On the representation of an integer in two different bases", *J. Reine Angew. Math.* **319** (1980), 63–72.