A probabilistic model for the degree of the cancellation polynomial in Gosper's Algorithm

Robin Pemantle [1]

ABSTRACT: Gosper's algorithm is a cornerstone of automated summation of hypergeometric series. Milenkovic and Compton in 2002 gave an analysis of the run time of Gosper's algorithm applied to a random input. The main part of this was an asymptotic analysis of the random degree of the cancellation polynomial $c(k)$ under various stipulated laws for the input. Their methods use probabilistic transform techniques. These results may also be obtained via probabilistic methods. Limit laws of the type proved by Milenkovic and Compton will be shown to follow from a general functional central limit theorem. Advantages to this approach include applicability to a more general class of input distributions and transparancy, in the sense that a non-quantitative limit theorem may be proved with no computation. A disadvantage to probabilistic methods is that one may typically derive only the leading term behavior. This is due to the inherent limit on precision in the Central Limit Theorem.

Keywords: Urn model, central limit, functional CLT, Brownian motion, Brownian bridge, conditioned IID

[1]Dept. of Math, University of Pennsylvania, 209 South 33rd Street, Philadelphia, PA 19104, pemantle@math.upenn.edu; research supported in part by NSF grant DMS-0103635.

1 Introduction

Great strides have been made recently in automatic summation of series, particularly hypergeometric series. A source for this is [PWZ96], which includes a historical development of the problem as well as a fine exposition of the recent and seminal work of the three authors. A cornerstone of the automation of hypergeometric summation is Gosper's algorithm. In [MC02a], it is pointed out that "despite the fact that Gosper's algorithm is one of the most important achievements in computer algebra, to date there are no results concerning the average running time of the algorithm." In that same work, Milenkovic and Compton undertake an analysis of the run time under various stipulated probabilistic models for the inputs. A notable achievement of [MC02a] is the harnessing and further development of the relatively new apparatus of de-Poissonization (as set out in [JS99] and [Mil02]), to analyze a procedure of significant interest and utility.

To describe the results of [MC02a], let f and g be polynomials, let $r_k := f(k)/g(k)$, and let $t_n := \prod_{k=1}^{n-1} r_k$. The sequence $\{t_n\}$ and the partial sums of this series are known as *hypergeometric*, and r_k is called the *hypergeometric ratio*. The purpose of Gosper's algorithm is to find a closed form expression for the partial sum $S_N := \sum_{n=1}^{N} t_n$. Its input is often specified as the rational function r_k in factored form. Roots of f and g differing by integers play a crucial role in the algorithm. Milenkovic and Compton observe that not much generality is lost in assuming the roots have been classified according to their remainders modulo 1, and that the problem has been restricted to one of these moduli classes. In other words, they assume that f and g have integer roots. They go on to stipulate joint probability distributions for f and g, which have as parameters an *a priori* bound on the location of the roots. Specifically, they assume that

$$f(k) = \prod_{j=1}^{m} (k-j)^{A_j} , \qquad g(k) = \prod_{j=1}^{m} (k-j)^{B_j} , \tag{1.1}$$

so all roots of f and g lie in $[m] := \{1, \dots, m\}$.

The "uniform **R** model," considered by Milenkovic and Compton, may be described as follows. For each of f and g, a sequence of n IID uniform picks is made from $[m]$. This gives the multisets of roots for f and g. In other words, for $j \in [m]$, the random variables A_j and B_j count how many of n IID picks from $[m]$ are equal to j. Milenkovic and Compton point out that this is an urn model of Maxwell-Boltzman type. A key to their analysis is the representation of the variables $\{A_j\}$ as a set of m IID picks from a Poisson distribution (of any mean), conditioned to sum to n. A second distribution of roots they analyze is the "multi-set **R** model," in which the multisets of roots (equivalently the sequences $(A_j)_{j=1}^m$ and $(B_j)_{j=1}^m$) are chosen uniformly from all multisets (equivalently all sequences of m nonnegative integers summing to n). This is a Bose-Einstein urn model, and is equivalent to conditioning two IID sequences of geometric random variables (with any mean) both to sum to n. They also discuss two models, the "Uniform **T** model" and the "Multiset **T** model," in which the roots of the numerator and denominator of the partial product terms t_k are directly modeled by the two respective urn models; these models are not addressed in this paper.

Milenkovic and Compton give a partial average case analysis, meaning that they focus on a few quantities which are highly determinative of the run time and give average case analyses of these. The most important such quantity is the degree of the *cancellation polynomial*, $c(k)$. This is defined as the minimal polynomial c for which we may write

$$\frac{f(k)}{g(k)} = \frac{a(k)}{b(k)} \frac{c(k+1)}{c(k)} \tag{1.2}$$

and also satisfying

$$\mathrm{GCD}(a(x), b(x-h)) = 1 \text{ for all nonnegative integers } h\,. \tag{1.3}$$

The determination of this polynomial is Step 2 in the version of Gosper's algorithm described in [Wis03], which is distilled from [PWZ96]. Milenkovic and Compton obtain the following results (they use N in place of the m in the present paper). The draft of their manuscript cited here is a very preliminary version which the authors have kindly provided. Consequently, only results independently proved in the present paper are quoted here, though in fact the manuscript [MC02a] obtains explicit constants for the asymptotic expectations.

Theorem 1 (Milenkovic and Compton (2002) Theorems 20 and 21) *In the uniform* $\mathbf{R}$ *model, if* $n/m \to \lambda$*, the expected degree of* $c(k)$ *is asymptotic to*

$$C_1(\lambda)m^{3/2}\,.$$

In the multiset $\mathbf{R}$ *model, when* λ *is sufficiently large, the expected degree of* c *is asymptotically*

$$C_2(\lambda)m^{3/2}\,.$$

The method of [MC02a] is to compute transforms (generating functions) for the unconditioned distributions, in which the variables A_j and B_j are independent Poissons or geometrics, and then de-Poissonize, according to machinery they developed in [MC02b].

Analytic de-poissonization may be technically somewhat involved; see for example [JS99]. The present paper also relies on the representation of the stipulated distributions as IID conditioned on a fixed sum. After that, however, the method herein is purely probabilistic, relying on limit theory for the random walk whose increments are $A_j - B_j$. Theorem 3 below, whose proof is a direct application of random walk limit theory, shows that as $n, m \to \infty$ with $n/m = \lambda + o(\lambda^{-1/2})$, the expected degree of c is

$$(c + o(1))m^{3/2} \tag{1.4}$$

where c is a certain expectation taken with respect to the Brownian bridge. For the uniform and multiset $\mathbf{R}$ models, the constant c is calculated respectively as

$$c_{\text{unif}} = \frac{\pi\sqrt{2\lambda}}{16}; \tag{1.5}$$

$$c_{\text{multi}} = \frac{\pi\sqrt{2\lambda(\lambda+1)}}{16}. \tag{1.6}$$

When comparing methods, it should be noted that the authors of [MC02a] are aware of the random walk representation. In fact this representation plays a role in their derivation of some transforms associated with the run time. If one wants results more precise than just the leading asymptotic behavior, it has been known since the work of Feller, Spitzer, Chung, and others that transforms, or equivalently, generating functions, offer the only route to complete asymptotics. But it is also worth knowing how to continue in the probabilistic vein, via scaling limits of random walk paths. This way, as is shown below, one may often obtain qualitative results with soft arguments and quantitative results with only a modest amount of computation.

2 Definitions and results

Let F be a distribution on the nonnegative integers with mean λ and variance $V < \infty$. We assume throughout that the GCD of the support of F is 1, as is true for Poisson distributions and geometric distributions of any mean. Let $\mathbf{P}_{F,m}$ be the probability measure on $\mathcal{S} := \mathbf{Z}^{2m}$ making the coordinates IID with common distribution F. Denote the first m coordinates by $A_1, \ldots, A_m$ and the last m coordinates by $B_1, \ldots, B_m$. Let $\mathbf{Q}_{F,m;n}$ be the result of conditioning $\mathbf{P}_{F,m}$ so that

$$\sum_{j=1}^{m} A_j = \sum_{j=1}^{m} B_j = n \, .$$

Associated to each $\omega \in \mathcal{S}$ are the polynomials f and g defined by (1.1), and the associated hypergeometric series with ratio $r_k = f(k)/g(k)$. Define a function $\beta : \mathcal{S} \to \mathbf{Z}^+$ by letting $\beta(\omega)$ be the minimal degree of a polynomial c satisfying (1.2) for the polynomials f and g. The next result, proved at the end of the section, provides an alternative expression for β, which is essentially the random walk representation in [MC02a, Theorem 1].

Define $X_j := A_j - B_j$ and $S_k := \sum_{j=1}^{k} X_j$, with the convention that $S_0 := 0$. Define $M_k := \min\{S_j : 0 \leq j \leq k\}$ and define $\tilde{M}_k := \min\{S_j : k \leq j \leq m\}$. Let $\tau := \min\{k : S_j \geq S_k \; \forall j > k\}$ denote the time of the first minimum of the process $\{S_j : 0 \leq j \leq m\}$. For each j, define $Y_j := S_j - M_j$ and $\tilde{Y}_j := S_j - \tilde{M}_j$. Define $I_j := Y_j \mathbf{1}_{j \leq \tau} + \tilde{Y}_j \mathbf{1}_{j > \tau}$. At the end of this section we will prove:

Lemma 2

$$\beta = \sum_{j=1}^{m-1} I_j = \sum_{j=1}^{\tau} Y_j + \sum_{j=\tau+1}^{m-1} \tilde{Y}_j \, .$$

Under the measures $\mathbf{P}_{F,m}$ and $\mathbf{Q}_{F,m;n}$, the quantities $A_j, B_j, X_j, Y_j, \beta$ and so forth become random variables. In order to state the main results, some definitions are required that mirror the definitions of these quantities but on the space of continuous limits. These definitions are quite familiar to probabilists, but likely not for applied mathematicians and computer scientists. Of central importance is the notion of a functional central limit theorem, that is, convergence of a sequence of partial sums, when rescaled in both space and time, to a Brownian path. This relies in turn on the apparatus of stochastic processes defined on filtrations on the space of CADLAG paths, where CADLAG stands for "continue a droite, limite a gauche", that is, the space of right-continuous paths with left limits. For an introduction to these notions, which develops only as much terminology as is needed for a basic functional central limit theorem, see [Dur96, Chapter 7].

Let Ω be the space of CADLAG paths on $[0,1]$, with filtration $\{\mathcal{F}_t\}$. For $\omega \in \Omega$, define the minimum process M^* by $M^*(\omega)(t) = \inf\{\omega(s) : s \leq t\}$. Define $\tilde{M}^*$ to be the right to left minimum process $\tilde{M}^*(\omega)(t) = \inf\{\omega(s) : s \geq t\}$. Define the process Y to be $\omega - M^*$ and $\tilde{Y}$ to be $\omega - \tilde{M}^*$. Define $\tau = \inf\{t : \omega(s) \geq \omega(t) \, \forall s > t\}$ to be the time of the first minimum and let I be the process $Y\mathbf{1}_{t\leq\tau} + \tilde{Y}\mathbf{1}_{t>\tau}$. Finally, we define $\beta^*(\omega)$ to be the quantity $\int_0^1 I(\omega)(t)\, dt$.

Let $\mathbf{P}_{2V}$ be the law of a centered continuous Gaussian process with covariances

$$\mathbf{E}\omega(s)\omega(t) = 2V(s \wedge t)\,.$$

In other words, $\mathbf{P}_{2V}$ is a Brownian motion with amplitude $\sqrt{2V}$. Let $\mathbf{Q}_{2V}$ be the law of a centered continuous Gaussian process with covariances

$$\mathbf{E}\omega(s)\omega(t) = 2V((s \wedge t) - st)\,. \tag{2.7}$$

In other words, $\mathbf{Q}_{2V}$ is a Brownian bridge of amplitude $\sqrt{2V}$. Recall that the mean and variance of F are denoted by λ and V. The first main result of the paper, proved in the next section, is:

Theorem 3 *Let $m \to \infty$ with $n = \lambda m + o(m^{1/2})$. Then*

$$\int \beta \, d\mathbf{P}_{F,m} = (K_1 + o(1))m^{3/2} \tag{2.8}$$

and

$$\int \beta \, d\mathbf{Q}_{F,m;n} = (K_2 + o(1))m^{3/2} \tag{2.9}$$

with

$$K_1 := \int \beta^* \, d\mathbf{P}_{2V}\,; \tag{2.10}$$

$$K_2 := \int \beta^* \, d\mathbf{Q}_{2V}\,. \tag{2.11}$$

The integrals (2.10) and (2.11) may be evaluated, leading to quantitative versions:

Theorem 4 *Consider the measure $\mathbf{P}_{F,m}$, under which $\{A_j, B_j : 1 \leq j \leq m\}$ are IID and have common variance V. Then as $m \to \infty$,*

$$\mathbf{E}\beta = (1 + o(1))\frac{2}{3}\sqrt{\frac{V}{\pi}}m^{3/2}\,.$$

Theorem 5 *Suppose that $N(m)$ is an integer satisfying $N(m)/m = \lambda + o(m^{-1/2})$ as $m \to \infty$. Consider the measure $\mathbf{Q}_{F,m;n}$, under which $\{A_j, B_j : 1 \le j \le m\}$ have the distribution of IID picks from F conditioned on $\sum_{j=1}^m A_j = \sum_{j=1}^m B_j = N$. Then as $m \to \infty$,*

$$\mathbf{E}\beta = (1 + o(1)) \frac{\pi\sqrt{2}}{16} \sqrt{V} m^{3/2} \, .$$

Remarks:

(1) Theorem 3 emerges without much difficulty from the convergence of the random walks to Brownian paths. Thus not only does $m^{-3/2}$ times the $\mathbf{P}_{F,m}$ expectation of β converge to $\mathbf{E}_{2V}\beta^*$, but the $\mathbf{P}_{F,m}$ distribution of $m^{-3/2}\beta$ converges to the $\mathbf{P}_{2V}$ distribution of β^*.

(2) Although it is relatively easy and is superseded by the two quantitative results, Theorem 3 is worth stating separately for the following reason: the computations in Theorems 4 and 5 are a little tricky, and it is instructive to see that the form of the result does not depend on calculations which are not transparent.

(3) The asymptotics obtained by [MC02a] for both the uniform and multiset models pertain to Theorem 5. Theorem 4 corresponds to two models

discussed in [MC02a] but not quoted here, where the roots of f and g are assumed to be unconditioned picks from the Poisson (respectively geometric) distributions, hence not necessarily equinumerous.

The rest of the organization of this paper is as follows. The functional central limit arguments are spelled out in Section 3. The values of the constants K_1 and K_2 are then computed in Section 4. In the remainder of this section we prove Lemma 2. We begin with an intermediate representation.

For each j between 1 and m, place A_j red balls and B_j blue balls in an urn marked j; this will be called "position j" or "time j". An *admissible* matching of the balls is a set of pairs of balls such that

(i) each pair contains one red ball and one blue ball; say that these two balls are *matched*, and any ball not in the union of pairs is called *unmatched*;

(ii) the pairs are pairwise disjoint;

(iii) if a red ball in position i is matched with a blue ball in position j, then $i \leq j$;

(iv) if there is an unmatched red ball in position i and an unmatched blue ball in position j then $i > j$ (that is, among unmatched balls, all red balls sit to the right of all blue balls).

For each admissible matching ξ, define the weight $w(\xi)$ to be the sum over all pairs in the matching of $j - i$, where j is the position of the blue ball and i is the position of the red ball.

Proposition 6

$$\beta = \min\{w(\xi) : \xi \textit{ is an admissible matching}\}.$$

PROOF: For every matched pair in positions i and j, we have

$$\frac{x-i}{x-j} = \prod_{s=i}^{j-1} \frac{x-s}{x-(s+1)} = \frac{e(k+1)}{e(k)}$$

where $e(x) = \prod_{s=i}^{j-1}(x-s)$. Thus any admissible matching ξ of weight w yields a solution $H(\xi)$ to (1.2) where c has degree w. Conversely, let c solve (1.2) and have degree w. If $w = 0$ then the empty matching is admissible. Assume now, for induction, that $w > 0$ and that for any f', g' represented by red and blue balls in urns, and any solution c' to (1.2) of degree less than w, there is an admissible ξ for that urn problem with $c' = H(\xi)$. For some j which is a root of c, let c' denote c with the linear factor $(x - j)$ removed. By induction, there is a matching ξ' of weight $w-1$, with $H(\xi') = c'$, admissible for the urn problem gotten from the original one by adding a blue ball at position j and a red ball at $j + 1$. Both of these new balls must be matched in ξ', say in pairs of positions (s, j) and $(j + 1, t)$, whence replacing these with the pair (s,t) produces a ξ admissible for the original problem with $H(\xi) = c$.

$\square$

Proof of Lemma 2: For any admissible matching ξ, define $d_j(\xi)$ to be the number of red balls in positions $1, \ldots, j$ matched with blue balls in positions $j+1, \ldots, m$. Elementarily, summing by parts,

$$w(\xi) = \sum_{\substack{\text{pairs of} \\ \text{positions } (i,j)}} \sum_{j \le t < i} 1 = \sum_t d_t(\xi)\,. \tag{2.12}$$

Fix any $j \le \tau$ and let $j_* := \min\{i : 0 \le i \le j \text{ and } S_i = M_j\}$. Then in positions $j_* + 1, \ldots, j$ there are a total of Y_j more red balls than blue balls. In positions $j+1, \ldots, \tau$, there is an excess of $S_j - S_\tau \ge S_j - M_j = Y_j$ blue balls over red balls. In an admissible matching, either every red ball at a position at most j is matched or every blue ball at a position at least $j+1$ is matched. It follows that either at least Y_j red balls from positions $j_* + 1, \ldots, j$ are matched to blue balls in positions beyond j, or at least Y_j blue balls from positions $j+1, \ldots, \tau$ are matched with red ball in positions up to j. In either case, $d_j(\xi) \ge Y_j$.

Similarly, fix any $j > \tau$ and let $j^* \ge j$ satisfying $S_{j^*} = \tilde{M}_j$. There is an excess of $\tilde{Y}_{j-1}$ blue balls in positions $j, \ldots, j^* - 1$. There is an excess of red balls in positions $\tau+1, \ldots, j-1$ of $S_{j-1} - S_\tau \ge \tilde{Y}_{j-1}$. Reasoning as before, one sees that $d_{j-1}(\xi) \ge \tilde{Y}_{j-1}$.

Summing over j now gives

$$\sum_{j=1}^{m-1} d_j(m) \geq \sum_{j=1}^{\tau} Y_j + \sum_{j=\tau}^{m-1} \tilde{Y}_j \,. \tag{2.13}$$

Note that $Y_\tau = \tilde{Y}_\tau = 0$, so this is equal to $\sum_{j=1}^{m-1} I_j$. Minimizing over ξ then gives half of the conclusion of the theorem: $\beta \geq \sum_{j=1}^{m-1} I_j$.

To prove the other half, we produce an admissible matching ξ with $d_j(\xi) = I_j$ for all $1 \leq j \leq m-1$. In particular, $d_\tau = 0$, so no ball in a position at most j is matched with a ball in a position beyond j, and ξ may be decomposed into a matching on balls in urns $1, \ldots, j$ and another on balls in urns $j+1, \ldots, m$. We construct these separately. An algorithm for constructing the matching on urns $1, \ldots, j$ is as follows.

> Initialize $j := 1$. Initialize a LIFO stack. Pull red balls out of urn j and place them on the stack until there are no more red balls in urn j. Pull blue balls out of urn j: while the stack is non-empty, match each blue ball with the top element of the stack; once the stack is empty, label each new blue ball "unmatched" and discard it. When urn j is empty, increment j and execute the loop until finished with the step $j = \tau$.

It is easy to see that all red balls in positions up to τ will be matched, since every time a red ball goes on the stack, there are more blues and reds to follow by time τ and no blue will be discarded until that red ball is matched. Inductively, it is easy to check that:

- the stack size after step j is Y_j;
- the change in stack size from time $j-1$ to j is $\max\{X_j, -Y_j\}$;
- the total number of balls discarded through time j is $-M_j$.

From this one sees that $d_j(\xi)$ is equal to the stack size after time j, and is therefore equal to Y_j. A stack algorithm dual to this works in the case $j > \tau$, working backward from time m to τ, stacking blue balls and matching or discarding reds. It constructs the other half of ξ so that $d_j(\xi) \geq Y_j$ for all $j > \tau$. This completes the proof of Lemma 2. $\square$

3 Proof of Theorem 3

Let $\mathcal{T}$ be the topology on Ω generated by the sup norm $|\omega| := \sup_{0 \leq t \leq 1} |\omega(t)|$. The following lemma is necessary only because τ is not a continuous function.

Lemma 7 *β^* is a continuous function on Ω with respect to $\mathcal{T}$.*

Proof: Suppose $\sup_s |\omega_1(s) - \omega_2(s))| \leq \epsilon$. Fix any t. It is clear that $|M(t, \omega_1) - M(t, \omega_2)| \leq \epsilon$ and likewise for $\tilde{M}$, hence these are continuous. If $\tau(\omega_1), \tau(\omega_2) \geq t$, it follows immediately also that $|I(t,\omega_1) - I(t,\omega_2)| \leq 2\epsilon$. Likewise, if $\tau(\omega_1), \tau(\omega_2) > t$, it follows that $|I(t,\omega_1) - I(t,\omega_2)| \leq 2\epsilon$.

Suppose now that $\tau_1 := \tau(\omega_1) < t \leq \tau_2 := \tau(\omega_2)$. Then

$$\tilde{M}(t,\omega_1) \geq \omega_1(\tau_1) \geq \omega_2(\tau_1) - \epsilon \geq M(t,\omega_2) - \epsilon$$

since τ_1 is one of the times over which the inf defining $M(t,\omega_2)$ is taken. Similarly,

$$\tilde{M}(t,\omega_2) \geq \omega_2(\tau_2) \geq \omega_1(\tau_2) - \epsilon \geq M(t,\omega_1) - \epsilon\,.$$

It follows that

$$\left|\tilde{M}(t,\omega_1) - M(t,\omega_2)\right| \leq \epsilon\,. \tag{3.14}$$

Together with $|\omega_1(t) - \omega_2(t)| \leq \epsilon$, this shows that $|I(t,\omega_1) - I(t,\omega_2)| \leq 2\epsilon$. A similar argument shows this in the case that $\tau_2 < t \leq \tau_1$. This establishes continuity of I, with continuity of D following by integration. $\square$

Proof of (2.8): Recall the definition of the partial sums S_j on $\mathcal{S}$ and define a map $\kappa : \mathcal{S} \to \Omega$ by

$$\kappa(\omega)(t) = m^{-1/2} S_{\lfloor mt \rfloor}\,.$$

The following relations are evident:

$$\begin{aligned} \tau^* \circ \kappa &= m^{-1}\tau\,; \\ M^*(\kappa(\omega)(t)) &= m^{-1/2} M(\omega(\lfloor mt \rfloor))\,; \end{aligned}$$

$$
\begin{aligned}
\tilde{M}^*(\kappa(\omega)(t)) &= m^{-1/2}\tilde{M}(\omega(\lfloor mt \rfloor))\,; \\
Y^*(\kappa(\omega)(t)) &= m^{-1/2}Y(\omega(\lfloor mt \rfloor))\,; \\
\tilde{Y}^*(\kappa(\omega)(t)) &= m^{-1/2}\tilde{Y}(\omega(\lfloor mt \rfloor))\,; \\
I^*(\kappa(\omega)(t)) &= m^{-1/2}I(\omega(\lfloor mt \rfloor))\,; \\
\beta^* \circ \kappa &= m^{-3/2}\beta\,.
\end{aligned}
$$

Let $\mathbf{P}_{(m)}$ denote the image under κ of $\mathbf{P}_{F,m}$. The functional central limit theorem says that the laws under $\mathbf{P}_{(m)}$ of κ converge weakly as $m \to \infty$ to the measure $\mathbf{P}_{2V}$. See [Bil86, Theorem 37.8] for a proof when $p \geq 4$ or [Dur96, Theorem 6.3 of Chapter 7] for a general proof using Skorohod embedding. This and Lemma 7 would complete the proof of (2.8) if β^* were bounded. Since β^* is not bounded, we may define for each $L > 0$,

$$I_L := \operatorname{sgn}(I^*)\left(|I^*| \wedge L\right) .$$

We may then conclude that the expectation with respect to $\mathbf{P}_{(m)}$ of the bounded continuous function $\beta_{(L)} := \int I_{(L)}$ converges as $m \to \infty$ to its expectation with respect to $\mathbf{P}_{2V}$.

Lemma 8 *Let* $\mathsf{max} := \sup_{0 \leq t \leq 1} \omega(t)$. *Then*

$$\lim_{L \to \infty} \sup_m \mathbf{E}_{(m)} \mathsf{max} \mathbf{1}_{\mathsf{max} > L} = 0\,.$$

Assuming this for the moment, we observe that $\mathbf{P}_{(m)}$ is symmetric so the same holds with inf in place of sup, and thus

$$\lim_{L\to\infty}\sup_m \mathbf{E}_{(m)}\mathsf{span}\,\mathbf{1}_{\mathsf{span}>L} = 0 \tag{3.15}$$

where $\mathsf{span} = \sup_t \omega(t) - \inf_t \omega(t)$. Since

$$|\beta^*(\omega) - \beta_{(L)}(\omega)| \leq \mathsf{span} := \sup_t \omega(t) - \inf_t \omega(t)$$

and $\beta^* = \beta_{(L)}$ on the event $\{\mathsf{span} \leq L\}$, the inequality (2.8) follows from (3.15) and the convergence of $\mathbf{E}_{(m)}\beta_{P(L)}$ to $\mathbf{E}_{2V}\beta_{(L)}$.

Proof of Lemma 8: By the L^2 maximum inequality ([Dur96, Theorem 4.4.3]),

$$\begin{aligned}
\mathbf{E}_{(m)}\mathsf{max}^2 &\leq 4\mathbf{E}\omega(1)^2 \\
&= 4\mathbf{E}_{F,m}(m^{-1/2}S_m)^2 \\
&= 4V\,, \qquad (3.16)
\end{aligned}$$

and hence

$$\mathbf{E}_{(m)}\mathsf{max}\mathbf{1}_{\mathsf{max}>L} \leq L^{-1}\mathbf{E}_{(m)}\mathsf{max}^2 \leq \frac{4V}{L} \tag{3.17}$$

for all m, proving the lemma. □

Proof of (2.9): Recall that $\mathbf{Q}_{2V}$ denotes the law on Ω of a Brownian bridge of amplitude $\sqrt{2V}$. The proof proceeds analogously to the proof of (2.8). In place of the standard functional central limit theorem is a well known result that may be found, among other places, in [Pit02, equation (6) of Section 0.4] (refer to [DK63] for the proof).

Lemma 9 (Conditional Functional CLT) *Let $n(m) \to \infty$ as $m \to \infty$ with $n(m)/m = \lambda + o(m^{-1/2})$. Recall that $\mathbf{Q}_{F,m;n}$ is the measure on $(\mathbf{Z}^+)^{2m}$ whose coordinates have the distribution of IID draws from F conditioned on $\sum_{j=1}^m A_j = \sum_{j=1}^m B_j = n$ and recall the aperiodicity assumption on F. Let $\mathbf{Q}_{(m)}$ denote the image under κ of $\mathbf{Q}_{F,m;n}$. Then the $\mathbf{Q}_{(m)}$ law of $\{S_j : 1 \leq j \leq m\}$ converges weakly to $\mathbf{Q}_{2V}$.* $\square$

All that remains is to show the analogue of Lemma 8 with $\mathbf{Q}_{(m)}$ in place of $\mathbf{P}_{(m)}$. As in (3.17), this will follow once we have shown a uniform bound on the $\mathbf{Q}_{(m)}$ second moment of max^2 analogous to (3.16). This in turn follows immediately from the inequality

$$\mathbf{Q}_{F,m;n}(\mathsf{max} > L) \leq C\mathbf{P}_{F,m}(\mathsf{max} > L) \tag{3.18}$$

where hereafter C may change from equation to equation but will depend only on F. To prove (3.18), let G be the event that $\sum_{j=1}^m A_j = \sum_{j=1}^m B_j = n$. By the local central limit theorem, $\mathbf{P}_{F,m}(G)$ is asymptotic to C/n, hence, using time-reversal symmetry of the path $\{S_j\}$ under $\mathbf{P}_{F,m}$,

$$\begin{aligned}
\mathbf{Q}_{F,m;n}(\mathsf{max} > L) &= \frac{\mathbf{P}_{F,m}(\mathsf{max} > L; G)}{\mathbf{P}_{F,m}(G)} \\
&\leq C\,n\mathbf{P}_{F,m}(\mathsf{max} > L; G)
\end{aligned}$$

$$= \quad 2C\,n \sum_{j>L} \sum_{t \le m/2} \mathbf{P}_{F,m}(\mathsf{max} = j = S_t; G)\,. \tag{3.19}$$

It then suffices to show

$$\mathbf{P}_{F,m}(\mathsf{max} = j = S_t; G) \le Cn^{-1}\mathbf{P}_{F,m}(\mathsf{max} = j = S_t) \tag{3.20}$$

since then resumming (3.19) proves (3.18).

Let $l = \lfloor 3m/4 \rfloor$, let $\mathcal{F}_l$ be the σ-field generated by $\{A_i, B_i : i \le l\}$ and let H be the event that $S_t = \max\{S_i : i \le l\}$. The local central limit theorem [Dur96, (5.2) in Chapter 2] gives $\mathbf{P}_{F,m}(G \mid \mathcal{F}_l) \le Cn^{-1}$, whence

$$\begin{aligned}
\mathbf{P}_{F,m}(\mathsf{max} = j = S_t; G) &\le \mathbf{P}_{F,m}(S_t = j; H \cap G) \\
&\le \mathbf{P}_{F,m}(S_t = j; H)\mathbf{P}_{F,m}(G \mid \mathcal{F}_l) \\
&\le Cn^{-1}\mathbf{P}_{F,m}(S_t = j; H)\,.
\end{aligned}$$

Conditioning again on $\mathcal{F}_l$, we see that $\mathbf{P}_{F,m}(S_t = j; H) \le Cn^{-1}\mathbf{P}_{F,m}(S_t = j = \mathsf{max})$, which establishes (3.20), hence (3.18) and the theorem. □

4 Evaluation of the constants

Let

$$D_1(\omega) := \int_0^\tau (\omega(t) - M(t))\,dt\,.$$

By symmetry, $\int D_1 \, d\mathbf{P}_{2V} = (1/2) \int \beta^* \, d\mathbf{P}_{2V}$ and similarly, $\int D_1 \, d\mathbf{Q}_{2V} = (1/2) \int \beta^* \, d\mathbf{Q}_{2V}$. Thus it suffices to compute expectations of D_1.

Proof of Theorem 4: For a process with law $\mathbf{P}_{2V}$, the process $Y(t) = \omega(t) - M(t)$ is well known to have the same distribution as the law under $\mathbf{P}_{2V}$ of the reflected Brownian motion $\{|\omega(t)| : 0 \leq t \leq 1\}$ (see, e.g., [Kal02, Prop. 13.13]). Clearly, the map $\omega \mapsto \omega - M$ has the property that t is a left-to-right minimum for ω if and only if t is a zero of $\omega - M$. The last left-to-right minimum is the global minimum, which occurs at the last zero of $\omega - M$. Because the process $\omega - M$ is distributed as $|\omega|$, the location of the last zero of $\omega - M$ is distributed as the last zero of ω. This has an arc-sine density $\pi^{-1}(x(1-x))^{-1/2} \, dx$ [Dur96, Example 4.4]. The following lemma writing Brownian motion as a mixture of bridges up to the last zero follows directly from the strong Markov property, scaling, and the fact that the bridge is a Brownian motion conditioned to return to zero:

Lemma 10 *Let $L = L(\omega) = \sup\{t \leq 1 : \omega(t) = 0\}$ be the last zero of Brownian motion and let $g : \Omega \to \mathbf{R}^+$ depend only on $\omega|_{[0,L(\omega)]}$. Then*

$$\int_\Omega g \, d\mathbf{P}_{2V} = \int_0^1 \frac{dt}{\pi\sqrt{t(1-t)}} \int_\Omega g \, d\mathbf{Q}_{2V}^{(t)}(\omega)$$

where the inner integral is on $\mathcal{F}_t$ and $\mathbf{Q}_{2V}^{(t)}$ is the law on $\mathcal{F}_t$ of a bridge of amplitude $\sqrt{2V}$ on $[0,t]$, that is, a centered Gaussian process with covariance

$$\mathbf{E}\omega(u)\omega(v) = u \wedge v - \frac{uv}{t} \, .$$

$\square$

Using this, c may be evaluated as follows. By definition, by [Kal02, Prop. 13.13], and lastly by Lemma 10 applied to the integral up to L of $|\omega|$, we have

$$\begin{aligned}
\int_\Omega D_1 \, d\mathbf{P}_{2V}(\omega) &= \int_\Omega \int_0^\tau (\omega(s) - M(\omega)(s)) \, ds \, d\mathbf{P}_{2V} \\
&= \int_\Omega \int_0^{L(\omega)} |\omega(s)| \, ds \, d\mathbf{P}_{2V} \\
&= \int_0^1 \frac{dt}{\pi\sqrt{t(1-t)}} \int_\Omega d\mathbf{Q}_{2V}^{(t)} \int_0^t |\omega(s)| \, ds \, .
\end{aligned}$$

From the covariance structure of $\mathbf{Q}_{2V}^{(t)}$, we see this law makes $\omega(s)$ a centered Gaussian with variance $s(1 - s/t)$. The expected absolute value of a $N(0, a)$ random variable is $\sqrt{2a/\pi}$. Switching the order of the two inner integrals, we may then write

$$\begin{aligned}
\int_\Omega D_1 \, d\mathbf{P}_{2V}(\omega) &= \int_0^1 \frac{dt}{\pi\sqrt{t(1-t)}} \int_0^t ds \int_\Omega |\omega(s)| \, d\mathbf{Q}_{2V}^{(t)}(\omega) \\
&= \int_0^1 \frac{dt}{\pi\sqrt{t(1-t)}} \int_0^t \frac{ds}{\pi\sqrt{t(1-t)}} \sqrt{\frac{4V}{\pi}} \sqrt{s\left(1 - \frac{s}{t}\right)} \, .
\end{aligned}$$

The evaluation is now straightforward integration. Substitute $s = tu$ and $ds = t \, du$ to get

$$\begin{aligned}
\int_\Omega D_1 \, d\mathbf{P}_{2V}(\omega) &= \sqrt{\frac{4V}{\pi}} \int_0^1 \frac{dt}{\pi\sqrt{1-t}} \int_0^1 \sqrt{u(1-u)} \, du \\
&= \sqrt{\frac{4V}{\pi}} \int_0^1 \frac{dt}{\pi\sqrt{1-t}} \frac{\pi}{8} \\
&= \frac{1}{4} \sqrt{\frac{V}{\pi}} \int_0^1 \frac{t}{\sqrt{1-t}} \, dt
\end{aligned}$$

$$= \frac{1}{3}\sqrt{\frac{V}{\pi}}\,.$$

Doubling yields $\int \beta^* \, d\mathbf{P}_{2V}$ and finishes the proof of Theorem 4, that is, $K_1 = (2/3)\sqrt{V/\pi} \approx 0.376\sqrt{V}$. □

Proof of Theorem 5: If the distribution of $\omega(t) - m(t)$ for a Brownian bridge were explicitly known in a usable form, the computation of the $\mathbf{Q}$-expectation would be analogous to the $\mathbf{P}$-expectation of D. In the absence of such a representation, the second computation ignores the representation of the law of $\omega - M$ as that of $|\omega|$ and proceeds as follows.

The counterpart to Lemma 10 is the following joint density for the pair $(\omega(t), M(\omega)(t))$ under $\mathbf{Q}_1$ conditioned on $\tau > t$.

Lemma 11 *For fixed $t \in (0, 1)$, define the positive function f_H on the set $R := \{(x, y) : y \geq 0\,,\, x \geq -y\}$ as follows:*

$$f_H(x, y) := \frac{2}{\pi t^3(1-t)^3}(x + 2y)e^{-(x+2y)^2/(2t(1-t))}\,.$$

Then f_H is a conditional density for $(\omega(t), -M(\omega)(t))$ under $\mathbf{Q}_1$ conditioned on $\tau > t$.

Proof: Begin with the computation of a density for (ω, M) under the standard Brownian measure $\mathbf{P}$. By the reflection principle, assuming $(x, y) \in R$ and using $\mathbf{P}^a$

to denote standard Brownian motion starting at a, one has

$$\begin{aligned} \mathbf{P}^0(\omega(t) \in [x, x+dx], M(t) \leq -y) &= \mathbf{P}^{-2y}(\omega(t) \in [x, x+dx]) \\ &= \sqrt{\frac{1}{2\pi t}} e^{-(x+2y)^2/(2t)}\, dx\,. \end{aligned}$$

Differentiating with respect to y yields

$$\mathbf{P}^0(\omega(t) \in [x, x+dx], M(t) \in [-y, -y+dy]) = \frac{4(x+2y)}{2t}\sqrt{\frac{1}{2\pi t}} e^{-(x+2y)^2/(2t)}\, dx\, dy \tag{4.21}$$

for $(x, y) \in R$.

Next, compute the joint density

$$\mathbf{P}(\omega(t) \in [x, x+dx], M(t) \in [y, y+dy], \omega(1) \in [0, dz], \tau > t)\,.$$

To do this, according to the Markov property, one must multiply (4.21) by

$$\mathbf{P}^x(\omega(1-t) \in [0, dz], \min_{0 \leq s \leq 1-t} \omega(s) \leq -y)\,.$$

By the reflection principle, this last factor is equal to

$$\mathbf{P}^{-2y-x}(\omega(1-t) \in [0, dz]) = \sqrt{\frac{1}{2\pi t}} e^{-(x+2y)^2/(2(1-t))}\, dz\,,$$

and multiplying and simplifying $1/t + 1/(1-t)$ to $1/(t(1-t))$ in the exponent gives a joint density of

$$\frac{x+2y}{\pi\sqrt{t^3(1-t)}} e^{-(x+2y)^2/(2t(1-t))}\, dz\, dx\, dy\,. \tag{4.22}$$

A change of variables simplifies the computation a little. Let $u = x + 2y$ and $v = (2x - y)/5$, so that $du\, dv = dx\, dy$ and $x = (u + 10v)/5, y = (2u - 5v)/5$. The region R is transformed into the region $R' := \{u \geq 0, (-3/5)u \leq v \leq (2/5)u$. Now rewrite the density (4.22) as

$$\frac{u}{\pi\sqrt{t^3(1-t)}} e^{-u^2/(2t(1-t))}\, dz\, du\, dv\,. \tag{4.23}$$

The conditional density of (u, v) given $\omega(1) \in [0, dz]$ and $\tau < t$, is given by normalizing this. One must divide by the integral of (4.23) over R', computed by a simple linear change of variables $u = r\sqrt{t(1-t)}$ in the third line:

$$\begin{aligned}
&\int_0^\infty \int_{-(3/5)u}^{(2/5)u} \frac{u}{\pi\sqrt{t^3(1-t)}} e^{-u^2/(2t(1-t))}\, dv\, du\, dz \\
= &\int_0^\infty \frac{u^2}{\pi\sqrt{t^3(1-t)}} e^{-u^2/(2t(1-t))}\, du\, dz \\
= &\int_0^\infty t(1-t) \frac{r^2}{\pi\sqrt{t^3(1-t)}} e^{-r^2/2} \sqrt{t(1-t)}\, dr\, dz \\
= &\frac{1-t}{\sqrt{2\pi}}\, dz
\end{aligned}$$

using the fact that $\int_0^\infty r^2 e^{-r^2/2}\, dr = \sqrt{\pi/2}$. Dividing,

$$f_H(x(u, v), y(u, v)) = \sqrt{\frac{2}{\pi t^3 (1-t)^3}}\, u e^{-u^2/(2t(1-t))}$$

and plugging in $u = x + 2y$ proves the lemma. $\square$

Proof of Theorem 5 continued: Let G denote the CDF for the time at which a Brownian bridge on $[0, 1]$ reaches its minimum. Set the amplitude $2V = 1$ for convenience, and note that $\omega(t)-M(t) = x-(-y) = (3u+5v)/5$. Then by Lemma 11, we have

$$
\begin{aligned}
\int_\Omega D_1 \, d\mathbf{Q}_1 &= \int_\Omega \int_0^1 (\omega(t) - M(t))\, \mathbf{1}_{\tau > t}\, dt\, d\mathbf{Q}_1(\omega) \\
&= \int_0^1 dt\, (1 - G(t)) \int_\Omega (\omega(t) - M(\omega)(t))\ d(\mathbf{Q}_1 \,|\, \tau > t)(\omega) \\
&= \int_0^1 dt\, (1 - G(t)) \int_{R'} \frac{3u + 5v}{5} f_H(x(u,v), y(u,v))\, du\, dv\,. \quad (4.24)
\end{aligned}
$$

The integral over R' may be computed by substituting $u = r\sqrt{t(1-t)}$ as before to get

$$
\begin{aligned}
\int_{R'} \frac{3u+5v}{5} f_H(x,y) dv\, du &= \sqrt{\frac{2}{\pi t^3(1-t)^3}} \int_0^\infty du\, u e^{-u^2/(2t(1-t))} \int_{(-3/5)u}^{(2/5)u} \frac{3u+5v}{5}\, dv \\
&= \sqrt{\frac{2}{\pi t^3(1-t)^3}} \int_0^\infty u e^{-u^2/(2t(1-t))} (\frac{3}{5}u^2 - \frac{1}{10}u^2)\, du \\
&= \sqrt{\frac{1}{2\pi t^3(1-t)^3}} \int_0^\infty u^3 e^{-u^2/(2t(1-t))}\, du \\
&= \sqrt{\frac{1}{2\pi}} \sqrt{t(1-t)} \int_0^\infty e^{-r^2/2}\, dr \\
&= \frac{1}{2}\sqrt{t(1-t)}\,.
\end{aligned}
$$

Finally, one may evaluate the integral in (4.24) without knowing the exact distribution function G. By symmetry, we know $G(t) + G(1-t) = 1$. The integral over R' is symmetric under $t \mapsto 1 - t$. Thus

$$
\int_\Omega D_1 \, d\mathbf{Q}_1 = \int_0^1 (1 - G(t)) \frac{1}{2} \sqrt{t(1-t)}\, dt
$$

$$
\begin{aligned}
&= \int_0^1 \frac{1}{4}\sqrt{t(1-t)}\,dt \\
&= \frac{\pi}{32}\,.
\end{aligned}
$$

Now doubling to get D and multiplying by the amplitude of $\sqrt{2V}$ leads to

$$\int_\Omega D\,d\mathbf{Q}_{2V} = \frac{\pi\sqrt{2V}}{16}\,,$$

which establishes the value of K_2 and finishes the proof of Theorem 5.

References

[Bil86] Billingsley, P. (1986). Probability and measure. John Wiley & Sons: New York.

[Dur96] Durrett, R. (1996). *Probability, theory and examples.* Second Edition. Duxbury Press: Belmont, CA.

[DK63] Dwass, M. and Karlin, S. (1963). Conditioned limit theorems. *Ann. Math. Stat.* **34** 1147–1167.

[JS99] Jacquet, P. and Szpankowski, W. (1999). Analytical de-poissonization and its applications. *Theoret. Comp. Sci.* **201**, 1–62.

[Kal02] Kallenberg, O. (2002). *Foundations of modern probability, second edition.* Springer: Berlin.

[Mil02] Milenkovic, O. (2002). *Probabilistic transforms in the analysis of algorithms.* Ph.D. Thesis, University of Michigan.

[MC02a] Milenkovic, O. and Compton, K. (2002). Probabilistic analysis of Gosper's algorithm. *Preprint.*

[MC02b] Milenkovic, O. and Compton, K. (2002). Probabilistic transforms for combinatorial urn models. *Preprint.*

[PWZ96] Petkovsek, M., Wilf, H. and Zeilberger, D. (1996). *$A = B$.* A. K. Peters: Wellesley, MA.

[Pit02] Pitman, J. (2002). *Combinatorial stochastic processes.* Manuscript dated August 30, 2002, available at http://www.stat.berkeley.edu/users/pitman/621.ps .

[Wis03] Weisstein, E. *Mathworld.* http://mathworld.wolfram.com/GospersAlgorithm.html