Combinatorial identities associated with CFTP

by

Gareth O. Roberts* and Jeffrey S. Rosenthal**

Abstract. We explore a method of obtaining combinatorial identities by analysing partially-completed runs of the Coupling from the Past (CFTP) algorithm. In particular, using CFTP for simple symmetric random walk (ssrw) with holding boundaries, we derive an identity involving linear combinations of $C_{ab}(s)$ for different a and b, where $C_{ab}(s)$ is the probability that unconstrained ssrw run from 0 for time n has maximum value a, and minimum value b, and ends up at s at time n.

1. Introduction.

This paper shall prove the following combinatorial identity:

Theorem 1. *For any fixed integers N, n, and i with $0 < n \le N$ and $0 \le i \le N$, we have*

$$\frac{1}{N+1} \;=\; \sum_{k=0}^{N}\sum_{\ell=0}^{N}\sum_{m=0}^{N-\ell}\left[\frac{\ell}{N+1}\,\mathbf{1}_{j=i} + \frac{m}{N+1}\,\mathbf{1}_{k=i} + \frac{1}{N+1}\,\mathbf{1}_{k\le i\le j}\right] C_{\ell,-m}(k-m)\,, \quad (1)$$

where $j \equiv j(k,\ell,m) \equiv k + N - \min(N, \ell+m)$, and where $C_{ab}(s)$ is the probability that unconstrained ssrw run from 0 for time n has maximum value a, and minimum value b, and ends up at s at time n.

The identity (1) can probably be proved directly, though it appears to be non-trivial to do so. Furthermore, we acknowledge that applications of this identity may be limited. However, more important than the actual form of the identity is the manner in which it is derived. Indeed, (1) emerges as a consequence of the correctness of the Coupling from the

* Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF, England. Internet: `g.o.robert@lancaster.ac.uk`.

** Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Internet: `jeff@math.toronto.edu`. Supported in part by NSERC of Canada.

Past (CFTP) algorithm for a very simple Markov chain – the case of constrained simple symmetric random walk on $\{0, 1, \ldots, N\}$.

Since CFTP can be applied to essentially any ergodic Markov chain, it appears that similar (though more complicated) methods could be used to derive other combinatorial identities, using other Markov chains in other settings (e.g. for the Ising model, as in Propp and Wilson, 1996).

We explain the CFTP algorithm in Section 2. In Section 3, we explain how CFTP gives rise to combinatorial identities. We derive the identity (1) in Sections 4 and 5. Further remarks about the quantities $C_{ab}(j)$, and how to compute them, are presented in Section 6.

2. Coupling from the Past (CFTP).

Markov chain Monte Carlo (MCMC) algorithms are an extremely popular tool in statistics to approximately sample from a probability distribution $\pi(\cdot)$, by designing a Markov chain $P(x, \cdot)$ such that π is stationary for P (see e.g. Smith and Roberts, 1993; Tierney, 1994; Gilks, Richardson, and Spiegelhalter, 1996).

More recently, Propp and Wilson (1996) have developed an algorithm called Coupling from the Past (CFTP), which makes use of the Markov chain P in a novel way to sample from π *exactly*. This has led to an explosion of research in this area; see e.g. Thönnes (2000) and the multitude of papers described in Wilson (1998).

To define CFTP, let us assume that we have an ergodic Markov chain $\{X_n\}_{n \in \mathbf{Z}}$ with transition kernel $P(x, \cdot)$ on a state space $\mathcal{X}$, and a probability measure π on $\mathcal{X}$, such that π is stationary for P (i.e. $(\pi P)(dy) \equiv \int_{\mathcal{X}} \pi(dx)\, P(x, dy) = \pi(dy)$). Let us further assume that we have defined the Markov chain as a *stochastic recursive sequence*, so there is a function $\phi : \mathcal{X} \times \mathbf{R} \to \mathcal{X}$ and an i.i.d. sequence of random variables $\{U_n\}_{n \in \mathbf{Z}}$, such that we always have $X_{n+1} = \phi(X_n, U_n)$. (It is not strictly necessary to use stochastic recursive sequences to define CFTP, see e.g. Murdoch and Rosenthal, 2000. On the other hand, it is easiest, and most common, to define CFTP in this way. Furthermore, the use of stochastic recursive sequences involves essentially no loss of generality, cf. Borovkov and Foss, 1992.)

CFTP involves considering *negative* times n, rather than positive times. Specifically, let

$$\phi^{(n)}(x; u_{-n}, \ldots, u_{-1}) \;=\; \phi(\phi(\phi(\ldots\phi(x,\, u_{-n}),\, u_{-n+1}),\, u_{-n+2}),\, \ldots),\, u_{-1})\,. \qquad (2)$$

Then CFTP proceeds by considering various increasing choices of $T > 0$, in the search for a value $T > 0$ such that

$$\phi^{(T)}(x; U_{-T}, \ldots, U_{-1}) \text{ does not depend on } x \in \mathcal{X}\,, \qquad (3)$$

i.e. such that the chain has *coalesced* in the time interval from time $-T$ to time 0.

Once such a T has been found, the resulting value

$$W \equiv \phi^{(T)}(x; U_{-T}, \ldots, U_{-1})$$

(which does not depend on x) is the output of the algorithm. Note in particular that, because of the backward composition implicit in (2), $W = \phi^{(n)}(y; U_{-n}, \ldots, U_{-1})$ for any $n \geq T$ and any $y \in \mathcal{X}$. In particular, letting $n \to \infty$, it follows by ergodicity that $W \sim \pi(\cdot)$. (See Propp and Wilson, 1996, for a formal proof of this.)

Note that the values $\{U_n\}$ should be thought of as being fixed in advance, even though of course they are only computed as needed. In particular, crucially, all previously-used values of $\{U_n\}$ must be used again, unchanged, as T is increased.

In the special case in which ϕ is *monotone*, meaning that there is an ordering $\preceq$ on $\mathcal{X}$ such that $\phi(x,u) \preceq \phi(y,u)$ whenever $x \preceq y$ (which implies that P is stochastically monotone), and there are maximal and minimal elements $x_{\max}, x_{\min} \in \mathcal{X}$, then to check (3) it suffices to check that

$$\phi^{(T)}(x_{\min}; U_{-T}, \ldots, U_{-1}) \;=\; \phi^{(T)}(x_{\max}; U_{-T}, \ldots, U_{-1})\,. \qquad (4)$$

3. Identities arising from CFTP.

As mentioned above, CFTP does a search for a value of T satisfying (3), by checking increasing values of T (often simply *doubling* the value of T each time) until (3) is satisfied.

Suppose in particular that some fixed value of T has been tried, so that the values $U_{-T}, \ldots, U_{-1}$ have been generated, and the paths $\{\phi^{(n)}(x,; U_{-n}, \ldots, U_{-1})\}_{1 \leq n \leq T}$ have been computed (and perhaps even displayed), though coalescence may have occured by then. *Conditional* on the values $U_{-T}, \ldots, U_{-1}$, the output value W will have some conditional distribution. Since overall $W \sim \pi(\cdot)$, we must have from the Law of Total Probability that

$$\int \ldots \int \mathbf{P}(U_{-1} \in du_{-1}, \ldots, U_{-T} \in du_{-T})\, \mathbf{P}(W \in A \,|\, U_{-1} = u_{-1}, \ldots, U_{-T} = u_{-T}) \;=\; \pi(A)\,.$$

Now, $\mathbf{P}(W \in A \,|\, U_{-1} = u_{-1}, \ldots, U_{-T} = u_{-T})$ could be computed as

$$\mathbf{P}(W \in A \,|\, U_{-1} = u_{-1}, \ldots, U_{-T} = u_{-T}) \;=\; \mathbf{P}(\phi^{(T)}(X_{-T}; U_{-1}, \ldots, U_{-T}) \in A)\,, \qquad (5)$$

except that X_{-T} is unknown. On the other hand, given the values of $\{U_n\}$, suppose we choose $S > T$ large enough that $\phi^{(S-T)}(x; U_{-S}, \ldots, U_{-S+T-1})$ does not depend on $x \in \mathcal{X}$, i.e. that coalescence occurs during the time interval from time $-S$ to time $-S+T$. Let $Z = \phi^{(S-T)}(x; U_{-S}, \ldots, U_{-S+T-1})$ (which does not depend on x). Then, by the validity of CFTP, we see that $Z \sim \pi(\cdot)$. Also, by construction,

$$\begin{aligned}\phi^{(S)}(x;\, u_{-S}, \ldots, u_{-1}) \;&=\; \phi^{(T)}(\phi^{(S-T)}(x;\, u_{-S}, \ldots, u_{-T-1});\, u_{-T}, \ldots, u_{-1}) \\ &=\; \phi^{(T)}(Z;\, u_{-T}, \ldots, u_{-1})\,.\end{aligned}$$

This suggests that in (5), we can assume that $X_{-T} = Z \sim \pi(\cdot)$. The following result confirms this.

Theorem 2. *Consider any CFTP algorithm as described above. Then for any* $T \in \mathbf{N}$,

$$\int \ldots \int \mathbf{P}(U_{-1} \in du_{-1}, \ldots, U_{-T} \in du_{-T})\, \pi\{x \in \mathcal{X};\; \phi^{(T)}(x; u_{-1}, \ldots, u_{-T}) \in A\} \;=\; \pi(A)\,. \qquad (6)$$

Proof. This follows immediately from the stationarity of π, since all says is that if we start a Markov chain in stationarity at time $-T$, then it will still be in stationarity at time 0. ∎

Remark. The proof of Theorem 2 shows that the result does not *really* depend on CFTP at all, just on the stationarity of $\pi(\cdot)$ for the Markov chain. However, it was CFTP that led us to consider these identities, and provides the proper context for them.

In principle, (6) gives a separate identity for every single Markov chain. The difficulty lies in interpreting (6) in a way that is meaningful and insightful. To do this, it is necessary to consider particular Markov chains.

4. The case of ssrw on $\{0, \ldots, N\}$.

Consider simple symmetric random walk (ssrw) on $\mathcal{X} = \{0, \ldots, N\}$, with holding boundaries. That is, let $\{U_n\}_{n \in \mathbf{Z}}$ be i.i.d., with $\mathbf{P}(U_n = +1) = \mathbf{P}(U_n = -1) = \frac{1}{2}$, and define a Markov chain by

$$X_{n+1} = \max\left[0,\ \min[N,\ X_n + U_n]\right]. \tag{7}$$

As a stochastic recursive sequence, we can write this as $X_{n+1} = \phi(X_n, U_n)$, where

$$\phi(x, u) \;=\; \max\left[0,\ \min[N,\ x + u]\right].$$

This Markov chain has as its stationarity distribution the uniform distribution on $\mathcal{X}$, so that $\pi(x) = 1/(N+1)$ for $x \in \mathcal{X}$. (Indeed, the chain is reversible with respect to $\pi(\cdot)$, and in fact it may be regarded as a Metropolis algorithm for $\pi(\cdot)$ with proposal given by unconstrained ssrw.) Hence, we can run CFTP for this Markov chain, to output a value $W \sim \pi(\cdot)$. See Rosenthal (1998) for an interactive display of CFTP for this Markov chain, which illustrates the status of the algorithm after a particular choice of T has been tried but coalescence has not been achieved.

Note that this chain is indeed monotone, with maximal and minimal values $x_{\max} = N$ and $x_{\min} = 0$. Hence, to check for coalescence we need only check (4). Thus, we consider just the "bottom process" starting at 0, and the "top process" starting at N.

We wish to specialise the general identity (6) to this particular case. To do so, we proceed as follows.

We first define a "hold" for the top or bottom process to be a time such that the chain remains in the same state. That is, for fixed N, we let

$$H_n^{\mathrm{top}} = \#\{m;\ 1 \le m \le n;\ \phi^{(n-m)}(N; U_{-n}, \ldots, U_{-n+m}) = \phi^{(n-m+1)}(N; U_{-n}, \ldots, U_{-n+m-1})\},$$

and

$$H_n^{\mathrm{bot}} = \#\{m;\ 1 \le m \le n;\ \phi^{(n-m)}(0; U_{-n}, \ldots, U_{-n+m}) = \phi^{(n-m+1)}(0; U_{-n}, \ldots, U_{-n+m-1})\}.$$

Let $R_{N,n,\ell,m,k}$ be the probability that the Markov chain given by (7), when run for a time n, has all of the following properties: (a) the bottom process ends up at $k \in \mathcal{X}$; (b) the top process has ℓ holds; and (c) the bottom process has m holds. It necessarily follows from this that (d) the top process ends up at $j = j(k, \ell, m) = k + N - \min(N,\ \ell + m)$.

In symbols, if we fix N throughout, then we can write this as

$$R_{N,n,\ell,m,k} = \mathbf{P}(\phi^{(n)}(0, U_{-n}, \ldots, U_{-1}) = k,\ H_n^{top} = \ell,\ H_n^{bot} = m).$$

We have the following.

Lemma 3. *Consider a partial run of CFTP for ssrw, from time $-n$ to time 0, as above. Let $j \equiv j(k, \ell, m) \equiv k + N - \min(N,\ \ell + m)$. Then if $\ell + m < N$ (so that $j < k$), then*

$$P(W = k \,|\, \phi^{(n)}(0, U_{-n}, \ldots, U_{-1}) = k,\ H_n^{top} = \ell,\ H_n^{bot} = m) = \frac{\ell + 1}{N + 1},$$

$$P(W = j \,|\, \phi^{(n)}(0, U_{-n}, \ldots, U_{-1}) = k,\ H_n^{top} = \ell,\ H_n^{bot} = m) = \frac{m + 1}{N + 1},$$

and for $j < i < k$,

$$P(W = i \,|\, \phi^{(n)}(0, U_{-n}, \ldots, U_{-1}) = k,\ H_n^{top} = \ell,\ H_n^{bot} = m) = \frac{1}{N + 1}.$$

If instead $\ell + m = N$, so that $j = k$, then

$$P(W = k \,|\, \phi^{(n)}(0, U_{-n}, \ldots, U_{-1}) = k,\ H_n^{top} = \ell,\ H_n^{bot} = m) = 1.$$

Proof. As in the proof of Theorem 2, we may assume (either by extending the CFTP construction back far enough, or by simply using stationarity of π) that the run $\{X_t\}$ is already in stationarity at time $-n$.

Now, from time $-n$ to 0, assuming the processes have not yet coalesced, then every time the top process holds, the process $\{X_t\}$ moves one unit closer to the top process. Similarly, every time the bottom process holds, the process $\{X_t\}$ moves one unit closer to the bottom process. It follows that $X_0 = j$ if and only if $X_{-n} \geq N - \ell$, and $X_0 = k$ if and only if $X_{-n} \leq m$. Otherwise, the values of X_{-n} strictly between m and $N - \ell$ are in a one-one correspondence with the values of X_0 strictly between k and j.

Hence, if $\ell + m < N$ so that $j < k$, then

$$P(W = k \,|\, \phi^{(n)}(0, U_{-n}, \ldots, U_{-1}) = k,\ H_n^{top} = \ell,\ H_n^{bot} = m) \;=\; P(X_{-n} \leq m)\,,$$

and

$$P(W = j \,|\, \phi^{(n)}(0, U_{-n}, \ldots, U_{-1}) = k,\ H_n^{top} = \ell,\ H_n^{bot} = m) \;=\; P(X_{-n} \geq N - \ell)\,,$$

while for $j < i < k$,

$$P(W = i \,|\, \phi^{(n)}(0, U_{-n}, \ldots, U_{-1}) = k,\ H_n^{top} = \ell,\ H_n^{bot} = m) \;=\; P(X_{-n} = m - k + i)\,.$$

Since $X_{-n} \sim \pi(\cdot)$, the result for the case $j < k$ follows.

On the other hand, if $\ell + m = N$ so that $j = k$, then

$$P(W = k \,|\, \phi^{(n)}(0, U_{-n}, \ldots, U_{-1}) = k,\ H_n^{top} = \ell,\ H_n^{bot} = m) \;=\; P(X_{-n} \in \mathcal{X}) \;=\; 1\,,$$

thus establishing the result in the case $\ell + m = N$ as well. ∎

Using this lemma and the definition of $R_{N,n,\ell,m,k}$, we see that Theorem 2 can now be written as follows. (For the case $\ell + m = N$, we use the observation that if $i = j = k$, then $\frac{\ell}{N+1}\,\mathbf{1}_{j=i} + \frac{m}{N+1}\,\mathbf{1}_{k=i} + \frac{1}{N+1}\,\mathbf{1}_{k \leq i \leq j} = 1$.)

Proposition 4. *For fixed integers N, n, and i, with $0 < n \le N$ and $0 \le i \le N$, letting $j = j(k, \ell, m) = k + N - \min(N,\ \ell + m)$, we have the identity*

$$\frac{1}{N+1} \;=\; \sum_{k=0}^{N} \sum_{\substack{\ell, m \ge 0 \\ m+\ell \le n}} \left[\frac{\ell}{N+1}\, \mathbf{1}_{j=i} + \frac{m}{N+1}\, \mathbf{1}_{k=i} + \frac{1}{N+1}\, \mathbf{1}_{k \le i \le j} \right] R_{N,n,\ell,m,k}\,.$$

5. Interpreting the quantities $R_{N,n,\ell,m,k}$.

To clarify the result of Proposition 4, we wish to better interpret the quantities $R_{N,n,\ell,m,k}$. Fix $N, n \in \mathbf{N}$ throughout. We first note the following.

Proposition 5. *If $\ell + m \le N$, then $R_{N,n,\ell,m,k}$ is equal to the probability that unconstrained ssrw starting at 0 and run for time n, has maximum value ℓ, minimum value $-m$, and ends up at $k - m$. In symbols,*

$$R_{N,n,\ell,m,k} \;=\; C_{\ell,-m}(k-m)\,,$$

where

$$C_{ab}(s) \;=\; \mathbf{P}\left[\max_{0 \le t \le n} Z_t = a,\ \min_{0 \le t \le n} Z_t = b,\ Z_n = s \right]\,,$$

where $\{Z_t\}$ is unconstrained ssrw started at 0 and run for time n.

Proof. Consider running CFTP as described above, using random variables $U_{-n}, \ldots, U_{-1}$. Let $Z_0 = 0$, and $Z_t = U_{-n} + \ldots + U_{-n+t-1}$ for $0 < t \le n$. Then $\{Z_t\}$ is indeed equal to an unconstrained ssrw started at 0 and run for time n. However, it has now been coupled with the CFTP algorithm under consideration, and we shall make use of this fact.

Now, suppose that for the CFTP algorithm, the event whose probability is $R_{N,n,\ell,m,k}$ has indeed occurred. That is, suppose that the bottom process of CFTP ends up at k, the top process has ℓ holds, and the bottom process has m holds. We wish to see what effect these suppositions have on the unconstrained process $\{Z_t\}$.

Let $-s$ be the time just after the last hold of the bottom process of the CFTP algorithm. Then the corresponding value of the $\{Z_t\}$ process, namely Z_{n-s}, must be equal to $-m$. Furthermore, we will then have $Z_t \ge -m$ for all $t \ge n - s$. It follows that we will have $\min_{0 \le t \le n} Z_t = -m$. Similarly, we will have $\max_{0 \le t \le n} Z_t = -m$.

Finally, the value of Z_n must be equal to the value of the bottom process in the CFTP algorithm, minus the number of times the bottom process held at 0 (so that it didn't decrease, even though $\{Z_t\}$ did). Hence, we must have $Z_n = k - m$.

Conversely, it is easily checked that if $\min_{0 \le t \le n} Z_t = -m$, and $\max_{0 \le t \le n} Z_t = -m$, and $Z_n = k - m$, then it must be that the bottom process of CFTP ends up at k, the top process has ℓ holds, and the bottom process has m holds.

We conclude that

$$R_{N,n,\ell,m,k} \;=\; \mathbf{P}\big(\min_{0 \le t \le n} Z_t = -m\,,\; \max_{0 \le t \le n} Z_t = -m\,,\; Z_n = k - m\big)\,.$$

This gives the result. ∎

Combining Proposition 5 with Proposition 4, we immediately obtain Theorem 1.

Remark. If $\ell + m > N$, then it is possible that the *bottom* process will actually reach the top value of N, and then hold at the *top* value. This implies that the final value of the corresponding unconstrained ssrw will no longer be just a function of k, ℓ, m, which causes additional complications. It is possible to overcome these complications, however to do so is somewhat messy and not particularly useful. Hence, in the above proposition, we restrict to the case $\ell + m \le N$. In Theorem 1, we restrict to the case $n \le N$, precisely so that these additional complications do not arise.

6. Computing the quantities $C_{ab}(s)$.

In order to verify the result of Theorem 1 numerically (say), it is necessary to know how to compute the quantities $C_{ab}(s)$. The reflection principle for ssrw is of some help here. However, since the events corresponding to $C_{ab}(s)$ specify both a maximum value *and* a minimum value, it seems that the best we can do is compute $C_{ab}(s)$ in terms of a sum of $O(n)$ terms. We present that here.

To begin, let

$$Q_n(x) \;=\; \begin{cases} 2^{-n}\dbinom{n}{\frac{n+x}{2}}\,, & -n \le x \le n, \quad x+n \text{ even} \\ 0\,, & \text{otherwise} \end{cases}$$

be the probability that *unconstrained* ssrw on the set of all integers, starting from 0, is at the point x at time n.

Now, for any integers $x_1, x_2, \ldots, x_r, j$, let $A_{x_1 x_2 \ldots x_r}(j)$ be the probability that unconstrained ssrw, starting from 0 and run for time n, hits in turn the values $x_1, x_2, \ldots, x_r$, and then ends up at j at time n. Also, say a sequence $x_1, x_2, \ldots, x_r$ is *sign alternating* if $(x_q - x_{q-1})(x_{q+1} - x_q) \leq 0$ for $2 \leq q \leq r-1$.

Lemma 6. *If the sequence $x_1, x_2, \ldots, x_r, j$ is sign-alternating, then*

$$A_{x_1 x_2 \ldots x_r}(j) \;=\; Q_n\left(|x_1| + |x_2 - x_1| + \ldots + |x_r - x_{r-1}| + |j - x_r|\right) .$$

Proof. Just use the reflection principle for ssrw, r times (cf. pp. 72, 96 of Feller, 1968). ∎

Now let $B_{ab}(j)$ be the probability that unconstrained ssrw hits run from time 0 to n, hits both a and b, and furthermore ends up at j at time n.

Lemma 7. *We always have $B_{00}(j) = 0$. Otherwise, for $(a,b) \neq (0,0)$ and $ab \leq 0$,*

$$B_{ab}(j) \;=\; [A_{ab}(j) - A_{aba}(j) + A_{abab}(j) - A_{ababa}(j) + \ldots]$$

$$+ [A_{ba}(j) - A_{bab}(j) + A_{baba}(j) - A_{babab}(j) + \ldots] .$$

(Since $A_{x_1 x_2 \ldots x_r}(j) = 0$ whenever $|x_1| + |x_2 - x_1| + \ldots + |x_r - x_{r-1}| > n$, the above sums each terminate after at most $n/|b-a|$ terms.)

Proof. By the inclusion-exclusion principle, the probability that we *first* hit b, and *then* hit a, finally ending up at j, is equal to

$$A_{ab}(j) - A_{bab}(j) + A_{abab}(j) - A_{babab}(j) + \ldots .$$

Similarly, the probability that we *first* hit a, and *then* hit b, finally ending up at j, is equal to the same sum but with a and b reversed. Adding these two sums, we obtain the formula for $B_{ab}(j)$ as claimed. ∎

Remark. Note that we cannot compute $B_{ab}(j)$ in simple closed form, as

$$[A_{ab}(j) - A_{bab}(j)] + [A_{ba}(j) - A_{aba}(j)] \ .$$

The problem with this is that paths of the form *abab* would end up getting added and subtracted equally often, and therefore counted a total of zero times, rather than one time as desired.

Now recall that $C_{ab}(j)$ is the probability that unconstrained ssrw has maximum value a, and minimum value b, and ends up at j at time n.

Lemma 8. *For $a \ge 0 \ge b$,*

$$C_{ab}(j) \ = \ B_{ab}(j) - B_{a+1,b}(j) - B_{a,b-1}(j) + B_{a+1,b-1}(j) \,.$$

Proof. $B_{ab}(j)$ is the probability that the max is *at least* a and the min is *at most* b (and we end up at j). The formula then follows from the basic probability result that $\mathbf{P}(A_1 \cap A_2^C \cap A_3^C) \ = \ \mathbf{P}(A_1) - \mathbf{P}(A_1 \cap A_2) - \mathbf{P}(A_1 \cap A_3) + \mathbf{P}(A_1 \cap A_2 \cap A_3)$, where $A_1 = \{\max \ge a\,,\ \min \le b\}$, $A_2 = \{\max \ge a+1\}$, $A_3 = \{\min \le b-1\}$. ∎

The above three lemmas together provide a fully computable formula for $R_{N,n,\ell,m,j}$, whenever $\ell + m \le N$. This allows us to numerically verify the identity (1), for any appropriate choices of N, n, and i. Such verification has been done in the C program `combin.c`, freely available at `http://markov.utstat.toronto.edu/jeff/comp/combin.c`

The three lemmas also allow the possibility of analytically expanding the identity of Theorem 1 further, in an effort to simplify or better understand it. However, it is not clear that such an expanded version provides any helpful new insights, so we do not pursue it here.

REFERENCES

A.A. Borovkov and S.G. Foss (1992), Stochastically recursive sequences and their generalizations. Siberian. Adv. Math. **2**, 16–81.

S.P. Brooks (1998), Markov Chain Monte Carlo Method and its Application. The Statistician **47**, 69–100.

W. Feller (1968), An Introduction to Probability Theory and its Applications, Volume I (3rd ed.). Wiley, New York.

W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, eds. (1996), Markov chain Monte Carlo in practice. Chapman and Hall, London.

D.J. Murdoch and P.J. Green (1998), Exact sampling from a continuous state space. Scandinavian Journal of Statistics **25(3)**, 483–502.

D.J. Murdoch and J.S. Rosenthal (2000), "Efficient use of exact samples". Statistics and Computing **10**, 237–243.

J.G. Propp and D.B. Wilson (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Structures and Algorithms **9**, 223–252.

J.S. Rosenthal (1998). CFTP Java Applet. Available at

`http://markov.utstat.toronto.edu/jeff/java/cftp.html`

A.F.M. Smith and G.O. Roberts (1993), Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). J. Roy. Stat. Soc. Ser. B **55**, 3-24.

E. Thönnes (2000), A primer on perfect simulation. In *Statistical Physics and Spatial Statistics* (K.R. Mecke and D. Stoyan, eds.), Springer Lecture Notes in Physics, volume **554**, 349–378.

L. Tierney (1994), Markov chains for exploring posterior distributions (with discussion). Ann. Stat. **22**, 1701-1762.

D.B. Wilson (1998). Annotated bibliography of perfectly random sampling with Markov chains. Available at `http://dimacs.rutgers.edu/~dbwilson/exact/`