Generating random elements in finite groups

John D. Dixon
School of Mathematics and Statistics
Carleton University
Ottawa, Ontario K2G 0E2, Canada
jdixon@math.carleton.ca

Submitted: Aug 8, 2006; Accepted: Jul 9, 2008; Published: Jul 21, 2008
Mathematics Subject Classification: 20P05, 20D60, 20C05, 20-04, 68W20

Abstract

Let G be a finite group of order g. A probability distribution Z on G is called ε-uniform if $|Z(x) - 1/g| \leq \varepsilon/g$ for each $x \in G$. If $x_1, x_2, \ldots, x_m$ is a list of elements of G, then the *random cube* $Z_m := Cube(x_1, \ldots, x_m)$ is the probability distribution where $Z_m(y)$ is proportional to the number of ways in which y can be written as a product $x_1^{\varepsilon_1} x_2^{\varepsilon_2} \cdots x_m^{\varepsilon_m}$ with each $\varepsilon_i = 0$ or 1. Let $x_1, \ldots, x_d$ be a list of generators for G and consider a sequence of cubes $W_k := Cube(x_k^{-1}, \ldots, x_1^{-1}, x_1, \ldots, x_k)$ where, for $k > d$, x_k is chosen at random from W_{k-1}. Then we prove that for each $\delta > 0$ there is a constant $K_\delta > 0$ independent of G such that, with probability at least $1 - \delta$, the distribution W_m is 1/4-uniform when $m \geq d + K_\delta \lg |G|$. This justifies a proposed algorithm of Gene Cooperman for constructing random generators for groups. We also consider modifications of this algorithm which may be more suitable in practice.

1 Introduction

In 2002 Gene Cooperman posted a manuscript "Towards a practical, theoretically sound algorithm for random generation in finite groups" on arXiv:math [4]. He proposed a new algorithm for generating (almost) random elements of a finite group G in which the cost to set up the generator is proportional to $\lg^2 |G|$ (where lg denotes the logarithm to base 2), and the average cost to produce each of the successive random elements from the generator is proportional to $\lg |G|$. The best theoretically justified generator previously known is due to Babai [2] and has a cost proportional to $\lg^5 |G|$. Another widely studied algorithm is the product replacement algorithm [3] (see also [9]). Although Pak (see [12]) has shown that the product replacement algorithm produces almost random elements in time polynomial in $\lg |G|$, there still exists a wide gap between the theoretical performance of this algorithm and what the original proposers hoped for (see [11]). (Igor Pak has informed me that he has now been able to show that the time complexity to construct the product replacement generator is $O(\lg^5 |G|))$.

Unfortunately, [4] is flawed. It has never been published, and it is not clear to me how it can be repaired in its original form. However, in the present paper I shall present a simplified variant of the proposed algorithm of Cooperman (see Theorem 1). Using a different approach (generating functions), but similar underlying ideas, I give a short proof that this variant algorithm is valid and has the asymptotic behaviour predicted by Cooperman. (Igor Pak has informed me that he has proved a similar result using a different approach. His proof is so far unpublished.)

Throughout this paper, G will denote a finite group of order g. We consider probability distributions on G. The uniform distribution U has the property that $U(x) = 1/g$ for all $x \in G$, and a distribution Z on G is said to be *ε-uniform* for $0 \leq \varepsilon < 1$ if $(1-\varepsilon)/g \leq Z(x) \leq (1+\varepsilon)/g$ for all x. For any list $x_1, x_2, \ldots, x_m$ of elements of G, the *random cube* $Cube(x_1, x_2, \ldots, x_m)$ of *length* m is the probability distribution on G induced by the mapping $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m) \mapsto x_1^{\varepsilon_1} x_2^{\varepsilon_2} \cdots x_m^{\varepsilon_m}$ from the the uniform distribution on the vertex set $\{0,1\}^m$ of the hypercube. It takes an average of $(m-1)/2$ group operations (multiplications) to construct an element of the cube. The concept of a random cube goes back to [7].

Theorem 1 (Cooperman) *Let $x_{1,} x_{2,\ldots,} x_d$ be a set of generators for G. Consider the random cubes*

$$Z_m := Cube(x_1, x_2, \ldots, x_m)$$

where for each $m > d$ we choose $x_m := y_m^{-1} z_m$ where y_m, z_m are random elements from Z_{m-1}.

Then for each $\delta > 0$ there exist a constant $K > 0$ (depending on δ but independent of d or G) such that, with probability at least $1 - \delta$,

$$Cube(x_m^{-1}, x_{m-1}^{-1}, \ldots, x_1^{-1}, x_1, x_2, \ldots, x_m)$$

is $1/4$-uniform for all $m \geq d + K \lg |G|$.

Remark 2 *A more precise statement appears in Section 4. If $m = d + \lceil K \lg |G| \rceil$, then the construction of the cube requires only $O((d + \lg |G|) \lg |G|)$ basic group operations (multiplication or inversion).*

In order to discuss these and related questions, we need some further measures of "almost" uniform. The deviation of Z from the uniform distribution in the *variational norm* is defined in [6, page 21] by

$$\|P - U\|_{var} := \frac{1}{2} \sum_{x \in G} |P(x) - U(x)| = \max_{A \subseteq G} |P(A) - U(A)|.$$

Clearly $\|P - U\|_{var} \leq \frac{1}{2}\varepsilon$ whenever P is ε-uniform, but the condition $\|P - U\|_{var} \leq \frac{1}{2}\varepsilon$ is a great deal weaker than being ε-uniform. We shall discuss this at greater length in

Section 5. As well as the variational norm we shall use the Euclidean norm whose square is given by

$$\|P - U\|^2 := \sum_{x \in G} (P(x) - U(x))^2$$

The value of the constant K in Theorem 1 which we obtain in Section 4 and the fact that the number of group operations to construct the random element generator is proportional to $\lg^2 |G|$ still means that a direct implementation of an algorithm based on Theorem 1 may be impractical. In Section 5 we examine some numerical examples, possible ways in which the process may be speeded up, and how shorter random element generators might be constructed. Some of these results reflect the following theorem which shows how a faster generator can be constructed if we have available a distribution which is close to uniform in the variational norm.

Theorem 3 *Let U be the uniform distribution on G and suppose that W is a distribution such that $\|W - U\|_{var} \leq \varepsilon$ for some ε with $0 \leq \varepsilon < 1$. Let $x_1, x_2, \ldots, x_m$ be random elements of G chosen independently according to the distribution W. If $Z_m := Cube(x_1, x_2, \ldots, x_m)$, and E denotes the expected value, then*

$$E(\|Z_m - U\|^2) < \left(\frac{1+\varepsilon}{2}\right)^m \text{ for all } m \geq 1. \tag{1}$$

Hence, if $\beta := 1/\lg(2/(1+\varepsilon))$, then:
- *(a) $E(\|Z_m - U\|^2_{var}) < 2^{-h}$ when $m \geq \beta(\lg|G| + h - 2)$;*
- *(b) $\Pr(\|Z_m - U\|_{var} > 2^{-k}) < 2^{-h}$ when $m \geq \beta(\lg|G| + h + 2k - 2)$;*
- *(c) with probability at least $1-2^{-h}$, Z_m is 2^{-k}-uniform when $m \geq \beta\left(2\lg|G| + h + 2k\right)$.*

Remark 4 *Part (c) was proved in [7] in the case where $W = U$, that is, when $\varepsilon = 0$ and $\beta = 1$. (Their theorem is stated for abelian groups but the proof is easily adapted to the general case.) It is shown in [2] that a result analogous to [7] holds if W is ε-*uniform *(a much stronger assumption than we have here).*

2 Some known results

Lemma 5 (Random subproducts) *[5, Prop. 2.1] If $x_1, x_2, \ldots, x_m$ generate G, and H is a proper subgroup of G then, with probability $\geq \frac{1}{2}$, a random element of G chosen using the distribution $Cube(x_1, x_2, \ldots, x_m)$ does not lie in H.*

Lemma 6 *Let λ, p and b be positive real numbers. Suppose that $Y_1, Y_2, \ldots$ are independent nonnegative random variables such that $\Pr(Y_k \geq 1/\lambda) \geq p$ for each k, and define the random variable M to be the least integer m such that $Y_1 + Y_2 + \cdots + Y_m \geq b$. Then*

$$\Pr(M > n) < \exp\left(-\frac{2(np - b\lambda)^2}{n}\right).$$

Proof. Chernoff's inequality shows that if X has the binomial distribution $B(n,p)$ then for all $a > 0$ we have $\Pr(X - np < -a) < \exp(-2a^2/n)$ (see, for example, Theorem A.1.4 in [1], and replace p by $1-p$ and X by $n - X$). Now define

$$X_k := \begin{cases} 1 \text{ if } Y_k \geq 1/\lambda \\ 0 \text{ otherwise} \end{cases} .$$

Thus, if X has the binomial distribution $B(n,p)$, then

$$\Pr(X < np - a) \geq \Pr(X_1 + \cdots + X_n < np - a) \geq \Pr(Y_1 + \cdots + Y_n < (np-a)/\lambda)$$

and so Chernoff's inequality shows that

$$\Pr(M > n) = \Pr(Y_1 + \cdots + Y_n < b) < \exp\left(-\frac{2(np - b\lambda)^2}{n}\right)$$

as required. ■

3 Generating functions

The use of group representations to analyze probability distributions on finite groups is widely used, particularly since the publication of the influential book [6]. What appears to be less common is a direct use of properties of the group algebra which on one hand reflect independence properties of probability distributions in a natural way and on the other hand enable manipulation of these distributions as linear transformations on a normed space.

We fix the group G. Let Z be a probability distribution on G. We identify Z with the element $\sum_{x \in G} \zeta_x x$ in the group ring $\mathbb{R}[G]$ where $\zeta_x = Z(x)$. Note that ZW (product in the group ring) is the convolution of distributions Z and W. This means that ZW is the distribution of the product of two independent random variables from Z and W, respectively (in general, when G is nonabelian, $ZW \neq WZ$). In particular, putting $g := |G|$, the uniform distribution is $U := (1/g)\sum_{x \in G} x$. We write $\operatorname{supp}(Z) := \{x \in G \mid \zeta_x \neq 0\}$ for the support of Z.

For each $x \in G$, $(1+x)/2$ is the distribution of a random variable which takes two values, 1 and x, with equal probability. Hence $Cube(x_1, x_2, \ldots, x_m)$ has distribution $Z_m := 2^{-m} \prod_{i=1}^m (1 + x_i)$.

There is a natural involution $*$ on $\mathbb{R}[G]$ given by $\sum_{x \in G} \zeta_x x \mapsto \sum_{x \in G} \zeta_x x^{-1}$, and a corresponding inner product on $\mathbb{R}[G]$ given by $\langle X, Y \rangle := tr(X^* Y)$ $(= \langle Y, X \rangle)$ where the *trace* $tr(\sum_{x \in G} \zeta_x x) := \zeta_1$. A simple calculation shows that this inner product is just the dot product of the vectors of coefficients with respect to the obvious basis. In particular, if $Z = \sum_{x \in G} \zeta_x x$, then the square of the Euclidean norm $\|Z\|^2 := \langle Z, Z \rangle = \sum_{x \in G} \zeta_x^2$. In general it is *not* true that $\|XY\| \leq \|X\| \, \|Y\|$, but $\|Xx\| = \|X\|$ for all $x \in G$.

The Euclidean norm is generally easier to work with than the variational norm, although the latter has a more natural interpretation for probability distributions. By the Cauchy-Schwarz inequality

$$4 \|Z - U\|_{var}^2 \leq g \|Z - U\|^2 . \tag{2}$$

On the other hand, if Z is any probability distribution, then $ZU = UZ = U$, and so

$$\|Z-U\|^2 = \|Z\|^2 + \|U\|^2 - 2tr(Z^*U) = \|Z\|^2 - 1/g. \tag{3}$$

In particular $1/g \leq \|Z\|^2 \leq 1$.

Let Z be a distribution and consider the distribution $Z^*Z = \sum_{t\in G} \omega_t t$, say. Note that Z^*Z is symmetric with respect to $*$ and that $\omega_x = \langle Z, Zx\rangle$. In particular, $\omega_x \leq \omega_1 = \|Z\|^2$ for all x by the Cauchy-Schwarz inequality.

Lemma 7 *For all* $x, y \in G$

$$\sqrt{\omega_1 - \omega_{xy}} \leq \sqrt{\omega_1 - \omega_x} + \sqrt{\omega_1 - \omega_y}$$

Proof. $\|Z(1-x)\|^2 = \|Z\|^2 + \|Zx\|^2 - 2\langle Z, Zx\rangle = 2(\omega_1 - \omega_x)$. On the other hand, the triangle inequality shows

$$\begin{aligned}\|Z(1-xy)\| &= \|Z(1-y) + Z(1-x)y\| \\ &\leq \|Z(1-y)\| + \|Z(1-x)y\| = \|Z(1-y)\| + \|Z(1-x)\|\end{aligned}$$

so the stated inequality follows. ■

The next lemma is the central core of our proof of Theorem 1. Our object in that proof will be to show that by successively extending a cube Z we shall (with high probability) push $\|Z\|^2$ down towards $1/g$. Then (3) shows that the series of cubes will have distributions converging to uniform. The following lemma proves that at each step we can expect the square norm of the cube to be reduced at least by a constant factor $(1-\frac{1}{2}\delta)$ unless the distribution of Z^*Z is already close to uniform.

Lemma 8 *Suppose that* $Z := Cube(x_1, x_2, \ldots, x_m)$ *and that* $x_1, x_2, \ldots, x_m$ *generate* G. *Set* $Z^*Z = \sum_{t\in G} \omega_t t$. *Then* $\|Z(1+x)/2\|^2 = \frac{1}{2}(\omega_1 + \omega_x) \leq \|Z\|^2$ *for all* $x \in G$. *Moreover, for each* δ *with* $0 < \delta < \frac{1}{12}$, *either*

(a) $(1-4\delta)\frac{1}{g} \leq \omega_t \leq \frac{1}{1-4\delta}\frac{1}{g}$ *for all* $t \in G$, *or*

(b) the probability that

$$\|Z(1+x)/2\|^2 < (1-\frac{1}{2}\delta)\|Z\|^2 \tag{4}$$

holds for $x \in G$ *(under the distribution* Z^*Z*) is at least* $(1-12\delta)/(2-13\delta)$.

Remark 9 *Taking* $\delta = 0.05$ *in (b) we find that the norm is reduced by* 2.5% *with probability nearly* 0.3. *Note that* $Z^*Z = Cube(x_m^{-1}, x_{m-1}^{-1}, \ldots, x_1^{-1}, x_1, x_2, \ldots, x_m)$.

Proof. We have $\|Z(1+x)/2\|^2 = \frac{1}{4}\left\{\|Z\|^2 + \|Zx\|^2 + 2\langle Z, Zx\rangle\right\} = \frac{1}{2}(\omega_1 + \omega_x)$.In particular, $\|Z(1+x)/2\|^2 \leq \omega_1 = \|Z\|^2$ and inequality (4) holds if and only if $\omega_x < (1-\delta)\omega_1$.

Set $C := \{t \in G \mid \omega_t \geq (1-\delta)\omega_1\}$. We have $1 \in C$ and $C = C^{-1}$ since Z^*Z is symmetric under *. The probability that $x \in C$ under the distribution Z^*Z is $\alpha := \sum_{t\in C} \omega_t$.

Now $\omega_1 - \omega_x \leq \delta\omega_1$ for all $x \in C$, so Lemma 7 shows that for all $x, t \in C$ we have

$$\sqrt{\omega_1 - \omega_{xt}} \leq \sqrt{\omega_1 - \omega_t} + \sqrt{\delta\omega_1}$$

which shows that

$$\omega_1 - \omega_{xt} \leq \omega_1 - \omega_t + 2\sqrt{\omega_1 - \omega_t}\sqrt{\delta\omega_1} + \delta\omega_1 \leq \omega_1 - \omega_t + 3\delta\omega_1.$$

Thus

$$\omega_{xt} \geq \omega_t - 3\delta\omega_1 \geq \omega_t(1 - \frac{3\delta}{1-\delta}) \text{ for all } x, t \in C.$$

Again Lemma 7 shows that

$$\sqrt{\omega_1 - \omega_y} \leq 2\sqrt{\delta\omega_1} \text{ for all } y \in C^2 \tag{5}$$

and so a similar argument shows that

$$\omega_{yt} \geq \omega_t(1 - \frac{8\delta}{1-\delta}) \text{ for all } t \in C \text{ and } y \in C^2.$$

Therefore for all $x \in C$ and $y \in C^2$

$$\sum_{t\in C} \omega_{xt} + \sum_{t\in C} \omega_{yt} \geq \beta := (2 - \frac{11\delta}{1-\delta}) \sum_{t\in C} \omega_t = \alpha\frac{2-13\delta}{1-\delta}.$$

First suppose that $\beta > 1$. Then, since $\sum_{z\in G} \omega_z = 1$, there exist $s, t \in C$ such that $xs = yt$ and this implies that $x^{-1}y = st^{-1} \in C^2$. Since this holds for all $x \in C = C^{-1}$ and $y \in C^2$, we conclude that $C^2C^2 = C(CC^2) \subseteq CC^2 \subseteq C^2$, and so the nonempty set C^2 is a subgroup of G. If C^2 were a proper subgroup of G, then Lemma 5 would show that an element x chosen using the cube distribution Z^*Z is not in C^2 with probability at least $\frac{1}{2}$. Since $1 \in C$, this shows that $\Pr(x \notin C) \geq \frac{1}{2}$, contrary to the fact that $\alpha > \beta/2$. Thus the subgroup C^2 equals G. But now equation (5) shows that

$$\omega_1 \geq \omega_x \geq (1 - 4\delta)\omega_1$$

for all $x \in G$. Since $g\omega_1 \geq \sum_{x\in G} \omega_x = 1$, this shows that $1 \geq (1-4\delta)g\omega_1 \geq 1 - 4\delta$. Thus $1/(1-4\delta) \geq g\omega_1 \geq g\omega_x \geq 1 - 4\delta$ and (a) holds in this case.

On the other hand, suppose that $\beta \leq 1$. Then the probability that $\omega_x < (1-\delta)\omega_1$ (that is, $x \notin C$) is

$$1 - \alpha = 1 - \frac{\beta(1-\delta)}{2-13\delta} \geq \frac{1-12\delta}{2-13\delta}.$$

By the observation at the beginning of this proof, alternative (b) holds in this case. ∎

4 Proof of Theorem 1

We shall prove the theorem in the following form. Note that, for all positive K and p, a unique positive solution of the equation $\varepsilon^2 = K(p-\varepsilon)$ exists and lies in the interval $(Kp/(K+p), p)$.

Theorem 10 *Let $x_1, x_2, \dots, x_d$ be a set of generators of a finite group G of order g. Consider the random cubes*

$$Z_m := Cube(x_1, x_2, \dots, x_m)$$

where for each $m > d$ we choose $x_m := y_m^{-1} z_m$ where y_m, z_m are random elements from Z_{m-1}.

Now, for each $\eta > 0$ define ε as the positive solution of $\varepsilon^2 = (0.3-\varepsilon)\lg(1/\eta)/(56\lg g)$, and note that $\varepsilon \to 0$ as $g \to \infty$. Then, with probability at least $1-\eta$, $Z_m^* Z_m$ is $1/4$-uniform for all $m \geq d + \lceil 28 \lg g/(0.3-\varepsilon)\rceil$.

Proof. We can assume that the generators $x_1, x_2, \dots, x_d$ are all nontrivial. Consider the random variable $\phi_m := \lg(1/\|Z_m\|^2)$. Since $\|Z_1\|^2 = \frac{1}{2}$, it follows from Lemma 8 (with close-to-optimal $\delta = 0.049$) that $1 = \phi_1 \leq \phi_2 \leq \cdots$ and that, for $m \geq d$, there is a probability > 0.3 that $\phi_{m+1} - \phi_m \geq \lg(1/0.9755) > 1/28$ unless the coefficients of $Z_m^* Z_m$ all lie between $0.804/g$ and $1/(0.804g)$. In the latter case $Z_m^* Z_m$ is a $1/4$-uniform distribution.

The minimum value for the square norm of a distribution is $\|U\|^2 = 1/g$, and so each $\phi_m \leq \lg g$. Define the random variable M to be the least value of n for which $Z_{n+d}^* Z_{n+d}$ is a $1/4$-uniform distribution. Then Lemma 6 (with $\lambda = 28$, $p = 0.3$ and $b = \lg g$) shows that $\Pr(M > n) < \eta$ whenever

$$\exp\left(-\frac{2(0.3n - 28\lg g)^2}{n}\right) < \eta.$$

Putting $\varepsilon := (0.3 - 28\lg g)/n$, we require that $2\varepsilon^2 n > \lg(1/\eta)$, and the given estimate is now easily verified. ■

5 Faster random element generators

The results proved in the previous section are undoubtedly weaker than what is really true. To compare them with some numerical examples, GAP [8] was used to compute $2^{2m} Z_m^* Z_m$ ($m = 1, 2, \dots$) in the group ring $\mathbb{Z}[G]$ for various groups G until $Z_m^* Z_m$ was $1/4$-uniform. This experiment was repeated 20 times for each group and a record kept of the number r of random steps required in each case (so the resulting cube had length $d + r$ where d was the number of generators). The results are summarized in the table below.

Group G	**d**	$\lvert G\rvert$	$\lg\lvert G\rvert$	**r**
S_5	2	120	6.9	8–16
Cyclic group C_{128}	1	128	7.0	13–39
$17:8$	2	136	7.1	8–20
$\mathrm{PSL}(2,7)$	2	168	7.4	9–16
Dihedral group D_{256}	2	256	8.0	18-32
$(A_4\times A_4):2$	2	288	8.2	8–18
$(2^4:5).4$	2	320	8.3	9–15
$AGL(1,16):2$	2	480	8.9	10–15
$2^4.(S_4\times S_4)$	3	576	9.2	8–13
$ASL(3,2)$	2	1344	10.4	10–17
$P\Gamma L(2,9)$	2	1440	10.5	12–17
$ASL(2,4):2$	2	1920	10.9	10–15

For comparison, if we calculate $m-d$ from Theorem 10 at the 90% confidence level ($\eta=0.1$), the bounds we obtain for r range from 790 (for $|G|=120$) up to 1190 (for $|G|=1920$) which are several orders of magnitude larger than the experimental results. Although the groups considered in the table are necessarily small (limited by the time and space required for the computations), the values for r suggest that the best value for the constant K in Theorem 1 is much smaller than that given by Theorem 10. Note that the experimental values obtained for r are largest for C_{128} and D_{256}, both of which contain an element of order 128.

Remark 11 *It should be noted that for permutation groups there are direct ways to compute (pseudo-)random elements via a stabilizer series and such series can be computed for quite large groups. The practical problem of generating random elements by other means is of interest only for groups of much larger size (see the end of this section).*

Also in practice we would use a different approach to generate random elements when the group is abelian. If $x_1,x_2,\ldots,x_d$ *generate an abelian group* G *of order* g *and* $2^m\geq g$, *then define* $Z_i := Cube(1,x_i,x_i^2,\ldots,x_i^{2^{m-1}})$ *for each* i. *Write* $2^m=gq+r$ *for integers* q,r *with* $0\leq r<g$. *We define the partial ordering* $\succcurlyeq$ *on* $\mathbb{R}[G]$ *by:* $X\succcurlyeq Y$ *if all coefficients of* $X-Y$ *are nonnegative. Now it is simple to verify that*

$$(1+(g-r)/2^m)U_i \succcurlyeq Z_i = 2^{-m}\sum_{j=0}^{2^m-1} x_i^j \succcurlyeq (1-r/2^m)U_i \text{ where } U_i := (1/g)\sum_{j=0}^{g-1} x_i^j.$$

Since $U_1U_2\cdots U_d=U$ *(the uniform distribution on* G*),* $Z:=Z_1Z_2\cdots Z_d$ *lies between*

$$(1+(g-r)/2^m)^dU \text{ and } (1-r/2^m)^dU.$$

Thus Z *is a random cube of length* md *which is* ε*-uniform on* G *where*

$$\varepsilon=\max\left((1+(g-r)/2^m)^d-1,1-(1-r/2^m)^d\right).$$

For an alternative approach see [10].

An examination of Lemma 8 shows that we should be able to do considerably better if we choose x using a different distribution. The $(m+1)$st generator of the cube in Cooperman's algorithm is chosen using the distribution $Z_m^* Z_m$ which gives a value of ω_x with probability ω_x. This is biased towards relatively large value of ω_x and hence towards large values of $\|Z_{m+1}\|^2$. We do better if we can choose x so as to obtain smaller values of ω_x. Theorem 3 examines what happens if we choose x using a distribution close to uniform on G. Leading up to the proof of that theorem, Lemma 13 lists a number of related results, part (c) being the primary result needed to prove the theorem. We begin by proving a simple property of the variational norm (valid even if G is not a group).

Lemma 12 *Let W be a probability distribution on G, and ϕ be any real valued function on G. Denote the maximum and minimum values of ϕ by $\phi_{\max}$ and $\phi_{\min}$, respectively, and put $\bar{\phi} := \left(\sum_{t \in G} \phi(t)\right)/g$. If $\|W-U\|_{var} \le \varepsilon$, then the expected value of $\phi - \bar{\phi}$ under the distribution W satisfies*

$$\left|E(\phi-\bar{\phi})\right| \le \varepsilon(\phi_{\max}-\phi_{\min}).$$

Proof. (Compare with Exercise 2 in [6, page 21].) Set $W = \sum_{t\in G} \lambda_t t$, say. Enumerate the elements $x_1, x_2, \ldots, x_g$ of G so that $\phi_{\max} = \phi(x_1) \ge \phi(x_2) \ge \cdots \ge \phi(x_g) = \phi_{\min}$ and define $\Lambda_i := \sum_{j=1}^{i} \left(\lambda_{x_j} - 1/g\right)$ for each i. Then

$$\begin{aligned} E(\phi-\bar{\phi}) &= \sum_{i=1}^{g} (\lambda_{x_i} - 1/g)\phi(x_i) = \sum_{i=1}^{g} (\Lambda_i - \Lambda_{i-1})\phi(x_i) \\ &= \sum_{i=1}^{g-1} \Lambda_i \left(\phi(x_i) - \phi(x_{i+1})\right) + \Lambda_g \phi(x_g). \end{aligned}$$

The hypothesis on W shows that $|\Lambda_i| \le \varepsilon$ for all i, and $\Lambda_g = 0$. Since $\phi(x_i) \ge \phi(x_{i+1})$ for all i, we conclude that

$$\left|E(\phi-\bar{\phi})\right| \le \sum_{i=1}^{g-1} \varepsilon\left(\phi(x_i) - \phi(x_{i+1})\right) = \varepsilon(\phi(x_1) - \phi(x_g))$$

as claimed. ■

Lemma 13 *Let Z and W be probability distributions on G. Then*

(a) If $s := |\mathrm{Supp}(Z)|$ and $\|W-U\|_{var} \le \varepsilon$, then for x chosen from the distribution W

$$E(|\mathrm{Supp}\left(Z(1+x)/2\right)|) \text{ lies in the range } s\left(2 - s/g \pm \varepsilon\right).$$

(b) Suppose that $2^m \le g$. If $Z := Cube(x_1, x_2, \ldots, x_m)$ and $s := |\mathrm{Supp}(Z)|$, then $\|Z-U\|_{var} = 1 - s/g$. Moreover, if $x_1, x_2, \ldots, x_m$ are independent and uniformly distributed, then

$$E(\|Z-U\|_{var}) \le (1-1/g)^{2^m} \le \exp(-2^m/g).$$

(c) If $\|W-U\|_{var} \leq \varepsilon$ and x is chosen from the distribution W, then

$$E(\|Z(1+x)/2\|^2 - 1/g) \leq \tfrac{1}{2}(1+\varepsilon)(\|Z\|^2 - 1/g).$$

Hence if $Z = Cube(x_1, x_2, \ldots, x_m)$ where $x_1, x_2, \ldots, x_m$ are independent and from the distribution W, then

$$E(\|Z-U\|^2) < \left(\frac{1+\varepsilon}{2}\right)^m.$$

(Note that the inequalities in (c) are for the Euclidean norm).

Proof. (a) Set $W = \sum_{t \in G} \lambda_t t$ and $\mathcal{S} := \mathrm{Supp}(Z)$. For each $u \in \mathcal{S}$ define $F(u) := \{x \in G \mid u \in \mathcal{S}x \cap \mathcal{S}\}$. Then each $F(u)$ has size $|\mathcal{S}|$ and so

$$\sum_{x \in G} |\mathcal{S}x \cap \mathcal{S}| = \sum_{u \in S} |F(u)| = |\mathcal{S}|^2.$$

Now $|\mathrm{Supp}(Z(1+x)/2)| = |\mathcal{S} \cup \mathcal{S}x| = 2|\mathcal{S}| - |\mathcal{S}x \cap \mathcal{S}|$, and so

$$\begin{aligned} E(|\mathrm{Supp}(Z(1+x)/2)|) &= \sum_{t \in G} \lambda_t (2|\mathcal{S}| - |\mathcal{S}t \cap \mathcal{S}|) \\ &= 2|\mathcal{S}| - \frac{1}{g}|\mathcal{S}|^2 - \sum_{t \in G} (\lambda_t - 1/g)|\mathcal{S}t \cap \mathcal{S}|. \end{aligned} \tag{6}$$

Applying Lemma 12 we conclude that the absolute value of $E(|\mathrm{Supp}(Z(1+x)/2)|) - 2|\mathcal{S}| + \frac{1}{g}|\mathcal{S}|^2$ is at most $\varepsilon(|\mathcal{S}| - 0) = \varepsilon|\mathcal{S}|$ as claimed.

(b) Write $Z = \sum_{t \in G} \zeta_t t$. Since $Z = 2^{-m}\prod_{i=1}^m (1+x_i)$, we have $\zeta_t \geq 2^{-m} \geq 1/g$ for each $t \in \mathrm{Supp}(Z)$ and so

$$\begin{aligned} \|Z-U\|_{var} &= \frac{1}{2}\sum_{t \in G} |\zeta_t - 1/g| \\ &= \frac{1}{2}\left\{\sum_{t \in G}(\zeta_t - 1/g) + 2\sum_{t \notin \mathrm{Supp}(Z)} 1/g\right\} = (g-s)/g. \end{aligned}$$

This proves the first part. Now let S_k be the support of $Z_k := Cube(x_1, x_2, \ldots, x_k)$ with $S_0 = \{1\}$, and put $s_k := |S_k|$ for each k. Then (6) with $\lambda_t = 1/g$ shows that

$$E(s_{k+1}) = 2E(s_k) - \tfrac{1}{g}E(s_k^2) \leq 2E(s_k) - \tfrac{1}{g}E(s_k)^2 \text{ for } k = 0, 1, \ldots, m-1$$

because $E(X^2) \geq E(X)^2$ for every real valued random variable X. Hence $E(1 - s_{k+1}/g) \leq (E(1 - s_k/g))^2$. Now induction on m gives

$$E(\|Z_m - U\|_{var}) = E(1 - s_m/g) \leq (1 - 1/g)^{2^m}$$

whenever $2^m \leq g$.

(c) Write $Z^*Z = \sum_{t\in G}\omega_t t$ and $W = \sum_{t\in G}\lambda_t t$. From Lemma 8 we know that $\|Z(1+x)/2\|^2 = \frac{1}{2}(\omega_1 + \omega_x)$ and $\|Z\|^2 = \omega_1$. By hypothesis $E(\omega_x) = \sum_{t\in G}\lambda_t\omega_t$. Since $\omega_1 \geq \omega_x \geq 0$ for all x, Lemma 12 shows that $|E(\omega_x - 1/g)| \leq \varepsilon\omega_1$. Thus $E(\|Z(1+x)/2\|^2 - 1/g) = \frac{1}{2}(\omega_1 + E(\omega_x)) - 1/g \leq \frac{1}{2}(1+\varepsilon)(\omega_1 - 1/g)$ as required.

Since $\|Z_m - U\|^2 = \|Z_m\|^2 - 1/g$, the final inequality in (c) follows from a simple induction. ■

Proof of Theorem 3. The initial inequality has been proved in Lemma 13. It remains to prove the consequences (a)-(c).

(a) Equations (1) and (2) show

$$E(\|Z_m - U\|^2) < \frac{g}{4}\left(\frac{1+\varepsilon}{2}\right)^m \leq 2^{-h}$$

when $m \geq \beta(\lg g + h - 2)$.

(b) If we replace h by $h + 2k$ in (a) and apply the Markov inequality we obtain

$$\Pr(\|Z_m - U\| > 2^{-k}) = \Pr(\|Z_m - U\|^2 > 2^{-2k}) < 2^{-h}$$

when $m \geq \beta(\lg g + h + 2k - 2)$.

(c) Clearly $\|Z_m - U\|^2 \leq 2^{-2k}/g^2$ implies that Z_m is 2^{-k}-uniform. On the other hand, (1) and Markov's inequality show that

$$\Pr\left(\|Z_m - U\|^2 > 2^{-2k}/g^2\right) < \left(\frac{1+\varepsilon}{2}\right)^m g^2 2^{2k} < 2^{-h}$$

when $m \geq \beta(2\lg g + h + 2k)$. ■

Theorem 3 says roughly that if we have a source of approximately random elements then we can construct a cube which is not too long and produces (with high probability) elements which are more closely random. It might also be interpreted as saying that it is not much harder to construct an ε-uniform random generator than to construct a random distribution Z satisfying $\|Z - U\|_{var} \leq \varepsilon$ which is a little surprising since the latter seems much cruder than the former.

Lemma 13 (b) suggested the following procedure which we carried out in GAP. Given generators $x_1, x_2, \ldots, x_d$ for a group G and an integer $l \geq d$, two random cubes X_{l+d} and Y_{l+d} of lengths $l + d$ were constructed as follows. Let $X_d := Cube(x_1, x_2, \ldots, x_d)$ and $Y_d := Cube(y_1, y_2, \ldots, y_d)$ where $y_1 := x_1$, $y_2 := x_2x_1$, $\ldots$, $y_d := x_dx_{d-1}\cdots x_1$. Then, for $k = d+1, \ldots, l$, $X_k := Cube(x_1, \ldots, x_k)$ where x_k is a random element from Y_{k-1}, and $Y_k := Cube(y_1, \ldots, y_k)$ where y_k is a random element from X_k (we also added a technical condition to ensure that for each cube the generators were distinct and nontrivial). Finally, $Z_m := Cube(x_1, \ldots, x_{l+d}, y_{d+1}, \ldots, y_{l+d})$ is a cube of length $m := 2l + d$. The idea behind this ad hoc construction is that the distributions of X_k and Y_k should be approximately independent and so the arguments used in the proof of Lemma 13 (b) may possibly apply. The final cube Z_m was then used to generate a list of 2000 random elements of G which were classified according to the conjugacy classes into which they fell. Then, if G had

$k := k(G)$ conjugacy classes of sizes $h_1, h_2, \ldots, h_k$ and the number of random elements which lay in the ith class was f_i, we computed

$$var_m := \frac{1}{2}\sum_{i=1}^{k}\left|\frac{h_i}{g} - \frac{f_i}{\sum f_j}\right|$$

as an approximation to $\|Z_m - U\|_{var}$. The table below compares values of var_m for different lengths m of cubes for some groups which are available in the permutation group library of GAP. Since var_m is computed from a statistical sample of size 2000, there is always some part of this variation which is due simply to this sampling. We therefore also calculated as a bench mark a value of var_∞ in which the frequencies f_i arise from sampling the various classes in exact proportion to their sizes. It is not easy to interpret these figures (all the groups have relatively few conjugacy classes), but for the groups listed it appears that random samples of size 2000 from cubes of length 25 and random samples from the uniform distribution are essentially indistinguishable in terms of how they are distributed over the conjugacy classes of G.

Group G	$\lvert G\rvert$	**d**	$k(G)$	var_{10}	var_{15}	var_{25}	var_∞
Cyclic group	512	1	512	0.93	0.59	0.20	0.19
$7^3 \cdot 2016$	6.9×10^5	2	74	0.12	0.07	0.06	0.06
HS	4.4×10^7	2	24	0.12	0.03	0.03	0.04
M_{24}	2.4×10^8	2	26	0.40	0.14	0.04	0.04
S_{12}	4.8×10^8	2	77	0.51	0.15	0.05	0.05
McL	9.0×10^8	2	24	0.16	0.05	0.04	0.04
$Sp(8,2)$	4.7×10^{10}	2	81	0.13	0.05	0.06	0.06
$O^-(10,2)$	2.5×10^{13}	2	115	0.16	0.07	0.07	0.07

In one application of particular interest (see [3] and [9]), only a very rough approximation to uniformity is required. In this situation G is a subgroup of the finite linear group $\mathrm{GL}(f,q)$ where values of f of interest might lie between, say, 10 and 100. The time required to carry out a single group operation (a matrix multiplication or inversion) is proportional to f^3, and as a consequence the number of group operations allowed in generating a random element is quite limited (in this context, $\lg |\mathrm{GL}(f,q)| \sim f^2 \lg q$ is too large). On the other hand, what is required is also quite modest. We want to be able to generate a list of elements which, with high probability, includes at least one element from each of two specified subsets of G, where it is known that each of these subsets is of size at least $|G|/(f+1)$. This will certainly be possible if we can construct a cube Z with $\|Z - U\|_{var} \leq 1/2f$, say, but presumably some much weaker condition is sufficient. The product replacement algorithm was proposed as a practical solution to this problem, but the theoretical justification remains open.

References

[1] N. Alon and J.H. Spencer, "The Probabilistic Method", (2nd ed.), Wiley, New York, 2000.

[2] L. Babai, Local expansion of vertex-transitive graphs and random generation in groups, in *Proc. 23rd ACN Symp. Theory of Comp.(1991)* (pp. 164–174).

[3] F. Cellar, C.R. Leedham-Green, S. Murray, A.C. Niemeyer and A. O'Brien, Generating random elements of a finite group, *Comm. Algebra* **23** (1995) 4931–4948.

[4] G. Cooperman, Towards a practical, theoretically sound algorithm for random generation in finite groups, (unpublished ms. posted on arXiv:math, May 2002).

[5] G. Cooperman and L. Finkelstein, Combinatorial tools for computation in group theory, in *"Groups and Computation" (L. Finkelstein and W.M. Kantor, eds.), DIMACS workshop held 1991,* Amer. Math. Soc., Providence, R.I., 1993 (pp. 53–86).

[6] P. Diaconis, "Group Representations in Probability and Statistics", Inst. Math. Statistics, Hayward, California, 1988.

[7] P. Erdős and A. Rényi, Probabilistic methods in group theory, *J. Analyse Math.* **14** (1965) 127–138.

[8] The GAP Group, *GAP—Groups, Algorithms, and Programming.* Version 4.4 5 (2005) (http://www.gap-system.org).

[9] D.F. Holt and S. Rees, An implementation of the Neumann-Praeger algorithm for the recognition of special linear groups, *Experiment. Math.* **1** (1992) 237–242.

[10] A. Lukács. Generating random elements of abelian groups, *Random Structures Algorithms* **26** (2005) 437–445.

[11] I. Pak, What do we know about the product replacement algorithm?, in "Groups and Computation III" (W.M. Kantor and A. Seress, eds.), de Gruyter, Berlin, 2001 (pp. 301–347).

[12] I. Pak, The product replacement algorithm is polynomial, *Proc. FOCS 2000* (pp. 476–485).