Electronic Colloquium on Computational Complexity, Comment 1 on Report No. 18 (2009)

Using more variables in the geometric generator

Yoav Tzur*

September 29, 2009

Abstract

We present an explicit construction of an ε-bias generator that outputs m bits using a seed shorter than $\frac{k}{k-1}\log m + k\log(1/\varepsilon) + k\log(k-1)$ bits, for any integer $k \geq 2$. This generator is a generalization of the geometric generator considered in [ECCC, TR09-018], which can be obtained as the special case $k = 2$.

Setting $k = \left\lceil \sqrt{\frac{\log m}{\log(1/\varepsilon)}} + 1 \right\rceil$ yields a seed of length at most $\log m + 2\sqrt{\log m \cdot \log(1/\varepsilon)} + 2\log(1/\varepsilon) + \tilde{O}(\sqrt{\log m})$. Specifically, if $\varepsilon \geq 2^{-\operatorname{poly}\log\log m}$ (e.g., if $1/\varepsilon$ is polylogarithmic in m, or even constant), the seed length is $\log m + \tilde{O}(\sqrt{\log m})$.

We use the notations and definitions of [Tzu]. We generalize the geometric generator (defined in Proposition 7 of [Tzu]) to use more variables. The construction of [Tzu] is obtained as a special case by setting $k = 2$. Recall that this construction is closely related (but not identical) to the powering construction of [AGHP].

Construction 1. *For $n, k, t \in \mathbb{N}$, set $\ell = \binom{t+k-1}{k-1}$. We define $\tilde{G}_k^{(t)} : GF(2^n)^k \to GF(2^n)^\ell$ as the mapping that on input k elements $\tilde{a}, \tilde{b}_1, \ldots, \tilde{b}_{k-1} \in GF(2^n)$, outputs all elements $\tilde{a} \cdot M(\tilde{b}_1, \ldots, \tilde{b}_{k-1})$ for M some monomial of total degree at most t.*

We first show that indeed, the number of monomials of $k-1$ variables of degree at most t is exactly $\binom{t+k-1}{k-1}$. First, the number of monomials in $k-1$ variables $b_1, \ldots, b_{k-1}$ of total degree *exactly* t can be thought of as the number of ways to choose t elements to multiply from the $k-1$ different elements $b_1, \ldots, b_{k-1}$, ignoring order, which is $\binom{t+k-2}{k-2}$. If we want all monomials of total degree *at most* t, we add the constant 1 as a k-th variable, to get $\binom{t+k-1}{k-1}$.

We now follow the proof strategy of [Tzu] to establish resilience against $GF(2^n)$-linear tests:

Proposition 2. *For every $n, k, t \in \mathbb{N}$, the generator $\tilde{G}_k^{(t)}$ is $\frac{t}{2^n}$-resilient to $GF(2^n)$-linear tests.*

*Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel. Email: yoav.tzur@weizmann.ac.il. This research was partially supported by the Israel Science Foundation (grant No. 1041/08).

ISSN 1433-8092

Proof. Fix a nontrivial $GF(2^n)$-linear combination, $\bar{c} = (\tilde{c}_1, \ldots, \tilde{c}_\ell) \in GF(2^n)^\ell$. For a seed $(\tilde{a}, \tilde{b}_1, \ldots, \tilde{b}_{k-1}) \in GF(2^n)^k$, the linear combination applied to the output of $G_k^{(t)}$ gives:

$$\left\langle \bar{c}, \tilde{G}_k^{(t)}(\tilde{a}, \tilde{b}_1, \ldots, \tilde{b}_{k-1}) \right\rangle = \sum_{i=1}^{\ell} \tilde{c}_i \cdot [\tilde{G}_k^{(t)}(\tilde{a}, \tilde{b}_1, \ldots, \tilde{b}_{k-1})]_i = \tilde{a} \cdot \sum_{i=1}^{\ell} \tilde{c}_i \cdot M_i(\tilde{b}_1, \ldots \tilde{b}_{k-1}),$$

with $(M_i)_{i=1}^{\ell}$ enumerating all monomials of $k-1$ variables of degree at most t. This defines a $(k-1)$-variate polynomial $p(\tilde{b}_1, \ldots, \tilde{b}_{k-1}) = \sum_{i=1}^{\ell} \tilde{c}_i \cdot M_i(\tilde{b}_1, \ldots \tilde{b}_{k-1})$ of degree at most t. For every fixed $(\tilde{b}_1, \ldots, \tilde{b}_{k-1}) \in GF(2^n)^{k-1}$ that is not a root of p, the expression $\tilde{a} \cdot p(\tilde{b}_1, \ldots, \tilde{b}_{k-1})$ is uniformly distributed in $GF(2^n)$, when $\tilde{a}$ is uniformly distributed in $GF(2^n)$. Thus the statistical distance between the uniform distribution and the distribution induced by the expression $\left\langle \bar{c}, \tilde{G}_k^{(t)}(\tilde{a}, \tilde{b}_1, \ldots, \tilde{b}_{k-1}) \right\rangle$ over a uniformly selected seed $(\tilde{a}, \tilde{b}_1, \ldots, \tilde{b}_{k-1})$ is at most $\Pr_{\tilde{b}_1, \ldots, \tilde{b}_{k-1}}[p(\tilde{b}_1, \ldots, \tilde{b}_{k-1}) = 0]$, which is at most $\frac{t}{2^n}$ since a (multivariate) polynomial of degree t can have at most t roots. $\square$

By Corollary 6 in [Tzu], we get that the binary version $G_k^{(t)}$ has small bias:

Corollary 3. *For every $n, k, t \in \mathbb{N}$, the generator $G_k^{(t)} : \{0,1\}^{nk} \to \{0,1\}^{n \cdot \binom{t+k-1}{k-1}}$ is a $\frac{t}{2^n}$-bias generator.*

We analyze the parameters we have obtained: using a seed of nk bits we output $m = n \cdot \binom{t+k-1}{k-1}$ bits with bias $\varepsilon = \frac{t}{2^n}$. Combining the two, we get

$$m = n \cdot \binom{\varepsilon \cdot 2^n + k - 1}{k-1} \geq \left(\frac{\varepsilon \cdot 2^n + k - 1}{k-1} \right)^{k-1} \geq 2^{(k-1)(n - \log(1/\varepsilon) - \log(k-1))},$$

using the inequality $\binom{r}{s} \geq (\frac{r}{s})^s$. This gives that $n \leq \frac{1}{k-1} \log m + \log(1/\varepsilon) + \log(k-1)$, so the seed length nk is at most $\frac{k}{k-1} \log m + k \log(1/\varepsilon) + k \log(k-1)$.

We have thus established

Theorem 4. *For every $\varepsilon \in (0,1)$ and integers $m > 0$ and $k \geq 2$, there exists an explicit ε-bias generator that generates m output bits with seed length at most $\frac{k}{k-1} \log m + k \log(1/\varepsilon) + k \log(k-1)$ bits.*

The expression $\frac{k}{k-1} \log m + k \log(1/\varepsilon)$ is minimized when $k = \sqrt{\frac{\log m}{\log(1/\varepsilon)}} + 1$. Clearly, if $\log m \leq \log(1/\varepsilon)$, then the minimal $k = 2$ (which is the original geometric generator) yields the shortest seed. However, when $\log m$ is significantly greater than $\log(1/\varepsilon)$, it is clear that a larger k would give a shorter seed (since the expression $\frac{k}{k-1}$ decreases as k increases). Since k must be an integer (and at least 2), we set $k = \left\lceil \sqrt{\frac{\log m}{\log(1/\varepsilon)}} + 1 \right\rceil$, and get a seed of length at most

$$\frac{\left\lceil \sqrt{\frac{\log m}{\log(1/\varepsilon)}} + 1 \right\rceil}{\left\lceil \sqrt{\frac{\log m}{\log(1/\varepsilon)}} \right\rceil} \cdot \log m + \left\lceil \sqrt{\frac{\log m}{\log(1/\varepsilon)}} + 1 \right\rceil \cdot \log(1/\varepsilon) + \left\lceil \sqrt{\frac{\log m}{\log(1/\varepsilon)}} + 1 \right\rceil \log \left\lceil \sqrt{\frac{\log m}{\log(1/\varepsilon)}} \right\rceil,$$

bounded by

$$\left(1+\frac{1}{\sqrt{\frac{\log m}{\log(1/\varepsilon)}}}\right)\log m+\left(\sqrt{\frac{\log m}{\log(1/\varepsilon)}}+2\right)\log(1/\varepsilon)+\tilde{O}(\sqrt{\log m}),$$

which can be simplified to

$$\log m+2\sqrt{\log m\cdot\log(1/\varepsilon)}+2\log(1/\varepsilon)+\tilde{O}(\sqrt{\log m}).$$

We have established

Corollary 5. *For every $\varepsilon \in (0,1)$ and $m \in \mathbb{N}$ there exists an explicit ε-bias generator that generates m output bits with a seed of length at most* $\log m + 2\sqrt{\log m \cdot \log(1/\varepsilon)} + 2\log(1/\varepsilon) + \tilde{O}(\sqrt{\log m})$. *Specifically, if $\varepsilon \geq 2^{-\operatorname{poly}\log\log m} \geq 2^{-\tilde{O}(\sqrt{\log m})}$, this is* $\log m + \tilde{O}(\sqrt{\log m})$.

Comparison to other generators. The standard explicit constructions of [AGHP] use a seed of length $2\log m + 2\log(1/\varepsilon)$, which is longer than the above if m is significantly greater than $1/\varepsilon$ (explicitly, if $\sqrt{\log m} > 2\sqrt{\log(1/\varepsilon)} + \operatorname{poly}\log\log m$, i.e. $m^{1-o(1)} > \varepsilon^{-4}$). We note that a construction of [NN] achieves a shorter seed when m is significantly greater than $1/\varepsilon$: they obtain $\log m + O(\log(1/\varepsilon))$; however, our construction is simpler and more natural (as are the constructions of [AGHP]).

Note that if the output length m is exponential in a "security parameter" n (for example, but not necessarily, the field size), then $\varepsilon \geq 2^{-\operatorname{poly}\log\log m}$, means $\varepsilon \geq 2^{-\operatorname{poly}\log n}$. For instance, to get 2^n bits with bias $1/\operatorname{poly}(n)$, we only need $n + \tilde{O}(\sqrt{n})$ bits of seed, as opposed to $2n$ bits in the original construction.

References

[AGHP] N. Alon, O. Goldreich, J. Hastad and R. Peralta, "Simple Constructions of Almost k-wise Independent Random Variables", *Random Structures and Algorithms*, vol. 3, pp. 289–304, 1992.

[NN] J. Naor and M. Naor, "Small-Bias Probability Spaces: Efficient Constructions and Applications", *SIAM Journal on Computing*, vol. 22, pp. 838–856, 1993.

[Tzu] Y. Tzur, "$GF(2^n)$-Linear Tests versus $GF(2)$-Linear Tests", *Electronic Colloquium on Computational Complexity*, TR 09-018, 2009.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de