

k -Universal Random Variables with Small Sample Space

Chris Calabro

April 7, 2009

Abstract

Given $n = \prod_{i=1}^m p_i^{e_i}$, $k, l \in \mathbb{N}$ with $k \leq l$, let $e'_i = \max\{e_i, \lceil \frac{\lg l}{\lg p_i} \rceil\}$.

We show how to construct a probability space S of size $(\prod_{i=1}^m p_i^{e'_i})^k \leq (n(2l)^m)^k$ in which we can define a poly-time sampleable family of l k -universal random variables each with codomain n . We need only $O(k(\lg n + m \lg l))$ random bits to sample S .

If we settle for exact k -wise independence but only approximate uniformity, say $|\Pr(X_j = i) - \frac{1}{n}| \leq \frac{\epsilon}{n}$, then we can reduce the size of the sample space to $(2 \max\{l, \frac{n}{\epsilon}\})^k$. If ϵ is fixed, we need only $O(k \lg(n + l))$ random bits to sample S , which is slightly better than in the first construction.

First suppose $m = 1$, $n = p^e$, $e' = \max\{e, \lceil \frac{\lg l}{\lg p} \rceil\}$, $n' = p^{e'}$. Let

$$S = \{g \in \mathbb{F}_{n'}[x] \mid \deg(g) < k\},$$

and then choose $f \in_u S$. We can think of f as a uniformly random vector in $\mathbb{F}_{n'}^k$, and so our sample space has size n'^k . Define random variables

$$X'_i = f(i), \quad i \in l.$$

The i 's are distinct mod n' because $l \leq p^{e'}$. The X'_i 's are k -universal because, from Lagrange interpolation, given distinct $i_1, \dots, i_k \in \mathbb{F}_{n'}$ and (not necessarily distinct) $b_1, \dots, b_k \in \mathbb{F}_{n'}$, there is exactly 1 polynomial $g \in \mathbb{F}_{n'}[x]$ of degree $< k$ such that $\forall j \in [k] g(i_j) = b_j$, which implies that for distinct $i_1, \dots, i_k \in l$ and (not necessarily distinct) $b_1, \dots, b_k \in \mathbb{F}_{n'}$

$$\Pr(\forall j \in [k] X'_{i_j} = b_j) = n'^{-k}.$$

Our random variables are then

$$X_i = \left\lfloor \frac{X'_i n}{n'} \right\rfloor,$$

which is uniform over n since $n|n'$.

Now we handle the case where $m > 1$. For each $i \in [m]$ define a sample space S_i as above so that $|S_i| = p_i^{e'_i}$ and define random variables $Y'_{i,j} \in p_i^{e'_i}$ on S_i so that for each i , $Y_{i,j}, j \in l$ are k -universal. Then set $S = \prod_{i=1}^m S_i$, so that $|S| = (\prod_{i=1}^m p_i^{e'_i})^k$, and

$$Y_{i,j} = \left\lfloor \frac{Y'_{i,j}}{p_i^{e'_i - e_i}} \right\rfloor.$$

Then for each i , the $Y_{i,j} \in p_i^{e_i}, j \in l$ are k -universal. For $j \in l$, define X_j to be the unique solution mod n to

$$\begin{aligned} X_j &\equiv Y_{1,j} \pmod{p_1^{e_1}} \\ &\vdots \\ X_j &\equiv Y_{m,j} \pmod{p_m^{e_m}} \end{aligned}$$

guaranteed by the Chinese remainder theorem, which we will abbreviate as $CR(Y_{1,j}, \dots, Y_{m,j})$. CR is bijective and each collection $\{Y_{i,j} \mid j \in l\}, i \in [m]$ is k -universal, so the X_j are also k -universal.

For the second construction, let $r = \max\{l, \frac{n}{\epsilon}\}$ and choose a prime $p \in [r, 2r]$. Let

$$f \in_u \{g \in \mathbb{F}_p[x] \mid \deg(g) < k\} = S,$$

and for $j \in l$, define $X'_j = f(j)$ and $X_j = \lfloor \frac{X'_j n}{p} \rfloor$. Then for $i \in n$,

$$\left| Pr(X_j = i) - \frac{1}{n} \right| \leq \frac{\epsilon}{n},$$

and $|S| \leq (2 \max\{l, \frac{n}{\epsilon}\})^k$.