On pseudorandom functions and asymptotic distributions

Fukuyama, Katusi

Tomokuni, Tetsuo

Monte Carlo Methods and Appl., Vol. 6, No. 3, pp. 167–174 (2000)

On pseudorandom functions and asymptotic distributions

Katusi Fukuyama

Department of Mathematics, Kobe University,
Rokko, Kobe, 657-8501 Japan
fukuyama@math.sci.kobe-u.ac.jp

Tetsuo Tomokuni

Developing division II, Megachips Co. ltd.,
Miyahara 4-5-36, Osaka, 532-0003 Japan
tomokuni@megachips.co.jp

Abstract — A simple proof of Ogawa's result on concrete construction of pseudorandom functions is presented. A construction of pseudorandom functions with symmetric stable asymptotic distribution is also given.

1. Introduction

A function on $\mathbf{R}$ is said to be pseudorandom if the limit of $\frac{1}{T}\int_0^T f(t)f(t+s)\,dt$ as $T \to \infty$ exists for all s, does not vanish at $s = 0$ and tends to 0 as $s \to \infty$. This notion was introduced by J. Bass [1] in the non-probabilistic theory of turbulence. If we consider $\mathbf{R}$ as a probability space with a flat probability measure on $[0, \infty)$, which does not really exist, this notion can be regarded as an asymptotic independence between f and shifted function $f(\cdot + a)$.

The asymptotic distribution function $F(a)$ of f is defined by the limit of $\frac{1}{T}|\{s \in [0,T] : f(t) \leq a\}|$ as $T \to \infty$ if the limit exists for all x. Here, $|\cdot|$ denotes the Lebesgue measure. If we regard f as a random variable on a real line with the flat measure on $[0, \infty)$ again, F can be regarded as the distribution function of f.

Since it is introduced for the purpose of numerical analysis, it is significant to give a concrete and efficient way to generate pseudorandom functions with gaussian asymptotic distribution function. First P. Hien [3], and later S. Ogawa [4] succeeded in constructing it in the following way.

For a sequence $\mathbf{z} = \{z_n\} \in [0,1]^{\mathbf{N}}$ and a function h on $[0,1]$ with

$$\int_0^1 h(t)\,dt = 0 \quad \text{and} \quad \int_0^1 h^2(t)\,dt < \infty, \tag{1}$$

put $q(t,\mathbf{z}) = \mathbf{1}_{[0,\infty)}(t)h(z_{[t]})$, $q_\lambda(t,\mathbf{z}) = \sqrt{\lambda}\,q(\lambda t,\mathbf{z})$, and

$$Q_\lambda^K(t,\mathbf{z}) = \int_{-\infty}^{\infty} K(s)q_\lambda(t-s,\mathbf{z})\,ds.$$

Then $Q_\lambda^K(t,\mathbf{z})$ is a pseudorandom function with gaussian asymptotic distribution function, if $\mathbf{z}$ is completely uniformly distributed (Hien [3]), or if $\mathbf{z}$ is a uniformly distributed sequence generated by an ergodic transform on $[0,1]$ (Ogawa [4]).

In this note we first give a simple proof of Ogawa's result. Next, we modify the Hien's method and give pseudorandom functions with symmetric stable asymptotic distribution. These functions are not pseudorandom in the sense of Bass, but have property that the function is nearly independent of the shifted function, that is the rough interpretation of the notion of pseudorandomness.

To state results easily, let us here modify the definition of asymptotic distribution. We say that a probability measure μ on $\mathbf{R}^n$ is an asymptotic distribution of $\mathbf{R}^n$-valued function $(f_1,\dots,f_n)$ on $\mathbf{R}$ if the distribution of $(f_1,\dots,f_n)$ on the probability space $([0,T],dt/T)$ converges to μ as $T \to \infty$, i.e.,

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T g(f_1(t),\dots,f_n(t))\,dt = \int_{\mathbf{R}^n} g(x)\,\mu(dx), \quad (g \in \mathrm{C_b}(\mathbf{R}^n)), \tag{2}$$

where $\mathrm{C_b}(\mathbf{R}^n)$ denotes the set of bounded continuous functions on $\mathbf{R}^n$. We say that a probability measure μ on $\mathbf{R}^\Pi$ is an asymptotic distribution of the system $\{f_\pi\}_{\pi\in\Pi}$ of functions on $\mathbf{R}$, if the asymptotic distribution of any finite subset of the system equals to the marginal distribution of μ.

Let us now state a version of Ogawa's theorem. Suppose that G is an ergodic transform on Lebesgue probability space $([0,1],\mathcal{B},d\omega)$, which is continuous a.e., and h is an a.e. continuous function on $[0,1]$ with (1) such that the functional central limit theorem holds for sum $\sum h(G^k\omega)$, i.e., D-valued random variables $X_n(t,\omega) = \sum_{k=1}^{[nt]} h(G^k\omega)/\sqrt{n}$ converge in law to $\sigma B(t)$, where $B(t)$ denotes the standard Brownian motion. Let $K \in \mathrm{BV_c}$, where $\mathrm{BV_c}$ denotes the class of functions of bounded variation with compact support. Let us put $\mathbf{z}_x = \{G^j x\}$ for

$x \in [0,1]$. Thanks to ergodic theorem, the set $\Omega_0 = \{x \in [0,1] \mid \mathbf{z}_x$ is uniformly distributed over $[0,1]\}$ has a full measure.

Theorem 1. *For arbitrary $x_0 \in \Omega_0$, the asymptotic distribution of the system $\{Q_\lambda^K(t,\mathbf{z}_{x_0}) : K \in \mathrm{BV_c}\}$ equals to the distribution of $\{Q_\lambda^K(t,\mathbf{z}_x) : K \in \mathrm{BV_c}\}$ on the probability space $([0,1]\times[0,1], dt\,dx)$. As $\lambda \to \infty$, it converges to the law of the gaussian system $\{G^K : K \in B\}$ with $EG^K = 0$ and $EG^KG^L = \int KL$.*

If K_1, K_2, ...are orthogonal, then G_{K_1}, G_{K_2}, ...are independent, and hence $Q_\lambda^{K_1}(t,\mathbf{z}_{x_0})$, $Q_\lambda^{K_2}(t,\mathbf{z}_{x_0})$, ...are asymptotically nearly independent. Thus we can generate nearly independent sequence of pseudorandom functions from one source $\mathbf{z}_{x_0}$.

By the above theorem we see that the asymptotic distribution function $F(a)$ of $Q_\lambda^K(t,\mathbf{z}_x)$ exists except for at most countable a, and it converges to the gaussian distribution function. Original Ogawa's theorem implicitly states the existence of $F(a)$ for all a, but there is an counterexample for that. We give such example in section 3.

Original theorem does not assume that G is continuous a.e., but we could not complete the proof without this condition. If we use the ergodic theorem directly instead of appealing to the property of uniformly distributed sequence, then we can prove the existence of $F(a)$ for all a without assuming the a.e. continuity of G. It is easily verified by putting $f = \mathbf{1}_{(-\infty,a]}$ in the proof below. But the statement must be changed to 'for almost all x, the asymptotic distribution function exists and converges to gaussian distribution function', which is weaker than the above version in view of the condition for x.

Now let us state the second theorem which extends the result by Hien. Let $X(t)$ be a symmetric stable Lévy process. Let us assume that the function h is a.e. continuous and the law of h on $([0,1], dt)$ belongs to the domain of attraction of $X(1)$, i.e., for i.i.d. $Y_1, Y_2, \ldots$, with $Y_1 \sim h$, there exists A_n such that $(Y_1 + \cdots + Y_n)/A_n$ converges in law to U. From now on we put $q_\lambda(t,\mathbf{z}) = \lambda q(\lambda t, \mathbf{z})/A_{[\lambda]}$ and define Q_λ^K as before. Denote by $BV_c[0,\infty)$ the collection of functions of bounded variation with compact support included by $[0,\infty)$.

Theorem 2. *If $\mathbf{z}$ is completely uniformly distributed over $[0,1]$, then the asymptotic distribution of the system $\{Q_\lambda^K(t,\mathbf{z}) : K \in BV_c[0,\infty)\}$ equals the distribution of the system $\{Q_\lambda^K(t,\mathbf{z}) : K \in BV_c[0,\infty)\}$ on the probability space $([0,1]\times[0,1], dt\,d\mathbf{z})$. As $\lambda \to \infty$, it converges to the*

law of $\{\int_0^\infty K(t)\,dX(t) : K \in \mathrm{BV_c}[0,\infty)\}$, *which is a system of symmetric stable random variables.*

If the supports of K's are disjoint, then the limit distribution becomes independent, and hence, the sequence of functions are asymptotically nearly independent. In this case, by the lack of square integrability of h the functions are not pseudorandom, but in the above sense, it is nearly independent of the shifted function, and hence considered to be roughly '*pseudorandom*'.

2. Proofs

Proof of Theorem 1: First we prove the 1-dimensional convergence. Assume that the support of K is contained in $(-M_0, L_0)$. Take $f \in C_b(\mathbf{R})$ arbitrarily and set $a_t = L_0 + t/\lambda$. If $t \geq L_0$, we have

$$Q_\lambda^K(t,\mathbf{z}) = \sqrt{\lambda} \int_{-M_0}^{L_0} K(s) h(z_{[\lambda(t-s)]})\,ds \quad \text{and} \quad Q_\lambda^K(t+1/\lambda,\mathbf{z}) = Q_\lambda^K(t,\theta\mathbf{z}),$$

where θ is the shift on $\mathbf{N}$, i.e, $\theta\mathbf{z} = \{z_{k+1}\}$ for $\mathbf{z} = \{z_k\}$. Thus, by decomposing the set $[0,a_n]$ into $[0,a_1]$, $[a_1,a_2]$, ..., and by changing the variables in each subinterval, we have

$$\frac{1}{a_n}\int_0^{a_n} f(Q_\lambda^K(t,\mathbf{z}))\,dt = \frac{1}{a_n}\sum_{k=0}^{n-1}\int_{a_0}^{a_1} f(Q_\lambda^K(t,\theta^k\mathbf{z}))\,dt + o(1). \tag{3}$$

Denote

$$R_\lambda(\mathbf{z}) = \int_{a_0}^{a_1} f(Q_\lambda^K(t,\mathbf{z}))\,dt = \frac{1}{\lambda}\int_0^1 f(Q_\lambda^K(a_t,\mathbf{z}))\,dt \quad \text{and} \quad R_{\lambda,G}(x) = R_\lambda(\mathbf{z}_x).$$

Since h, G are continuous a.e., $R_{\lambda,G}$ is bounded and continuous a.e. in x, and hence Riemann integrable. Since the sequence $\mathbf{z}_{x_0}$ is uniformly distributed, by Riemann integrability of $R_{\lambda,G}$, we have

$$\frac{1}{a_n}\int_0^{a_n} f(Q_\lambda^K(t,\mathbf{z}_{x_0}))\,dt = \frac{1}{a_n}\sum_{k=0}^{n-1} R_{\lambda,G}(G^k x_0) + o(1) \to \lambda\int_0^1 R_{\lambda,G}(x)\,dx.$$

Thus we have proved

$$\lim_{L\to\infty}\frac{1}{L}\int_0^L f(Q_\lambda^K(t,\mathbf{z}_{x_0}))\,dt = \int_0^1 dx\int_0^1 f(Q_\lambda^K(a_t,\mathbf{z}_x))\,dt. \tag{4}$$

Let us assume that K is right-continuous and having the left limit at each point. Since K is of bounded variation, such version of K exists. Let ν be a signed measure on $\mathbf{R}$ characterized by $K(s) = \nu((-\infty, s])$. Integration by parts yields

$$Q_\lambda^K(a_t, \mathbf{z}) = -\int_{-M_0}^{L_0} \left(\int_{-M_0}^{s} q_\lambda^K(a_t - u, \mathbf{z})\, du \right) d\nu(s). \tag{5}$$

Since we have

$$\int_{-M_0}^{s} q_\lambda(a_t - u, \mathbf{z}_x)\, du = \frac{1}{\sqrt{\lambda}} \sum_{k=[(L_0-s)\lambda+1]}^{[(L_0+M_0)\lambda]} h(G^k x) + o(1) \quad \text{as} \quad \lambda \to 0$$

for $s \in [-M_0, L_0]$ and $t \in [0,1]$, where $o(1)$ is uniform in s and t, by the functional central limit theorem, we have the following convergence in law in D space:

$$\int_{-M_0}^{s} q_\lambda(a_t - u, \mathbf{z}_x)\, du \xrightarrow{\mathcal{D}} \sigma\{B(L_0 + M_0) - B(L_0 - s)\}\ .$$

Since the discontinuity of the functional $f \mapsto \int_{-M_0}^{L_0} f(u)\, d\nu(u)$ on D is measure 0 with respect to the law of $\sigma\{B(L_0 + M_0) - B(L_0 - s)\}$, by Theorem 5.1 of [2], we have

$$\begin{aligned} Q_\lambda^K(a_t, \mathbf{z}_x) &\xrightarrow{\mathcal{D}} -\int_{-M_0}^{L_0} \sigma\{B(L_0 + M_0) - B(L_0 - s)\}\, d\nu(s) \\ &= \sigma \int_{-M_0}^{L_0} K(s)\, dB(M_0 + s). \end{aligned}$$

Since the distribution of the right hand side is gaussian with mean 0 and variance $\sigma^2 \|K\|_2^2$, we have proved the convergence of 1-dimensional distribution.

Finally, we prove the convergence of finite dimensional distribution. For $K_1, \ldots, K_n \in \mathrm{BV_c}$, and $\beta_1, \ldots, \beta_n \in \mathbf{R}$, we have $\beta_1 Q_\lambda^{K_1}(t, \mathbf{z}_x) + \cdots + Q_\lambda^{K_n}(t, \mathbf{z}_x) = Q_\lambda^K(t, \mathbf{z}_x)$ where $K = \beta_1 K_1 + \cdots + \beta_n K_n$. Thus, by using the result just we have proved, we can say that any linear combination of $Q_\lambda^{K_1}(a_t, \mathbf{z}_{x_0}), \ldots, Q_\lambda^{K_n}(a_t, \mathbf{z}_{x_0})$ on $([0,T], dt/T)$ converges in law to that of $Q_\lambda^{K_1}(t, \mathbf{z}_x), \ldots, Q_\lambda^{K_n}(t, \mathbf{z}_x)$ on $([0,1]^2, dt\, dx)$. And also the latter converges in law to the linear combination of $\sigma \int_{-M_0}^{L_0} K_1(s)\, dB(M_0+s), \ldots, \sigma \int_{-M_0}^{L_0} K_n(s)\, dB(M_0 + s)$. By the theorem of Cramér-Wold ([2] Theorem 7.7), we get the convergence of joint law. □

Proof of Theorem 2: For given t, $q_\lambda(t, \mathbf{z})$ depends only on $z_{[\lambda t]}$, and hence $Q_\lambda^K(t, \mathbf{z})$ depends only on $z_0, \ldots, z_{[\lambda(t+M_0)]}$. Thus $R_\lambda(\mathbf{z})$ can be

regarded as a function of $z_0, \ldots, z_{[\lambda(a_1+M_0)]} = z_{b_\lambda}$, where $b_\lambda = [(L_0 + M_0)\lambda + 1]$. This function on $I_\lambda = [0,1]^{b_\lambda+1}$ is bounded and continuous a.e., and hence Riemann integrable. Because $\mathbf{z}$ is completely uniformly distributed, we have

$$\frac{1}{n}\sum_{k=0}^{n-1} R_\lambda(\theta^k \mathbf{z}) \to \int_{I_\lambda} R_\lambda(z_0, \ldots, z_{b_\lambda})\, dz_0 \ldots dz_{b_\lambda} = \int_{I_\infty} R_\lambda(\mathbf{x})\, d\mathbf{x}$$

where the last integral is over $I_\infty = [0,1]^{\mathbf{N}}$ with respect to the countable product $d\mathbf{x} = dx_1\, dx_2 \ldots$ of Lebesgue measure on $[0,1]$.

Since we have

$$\int_{-M_0}^{s} q_\lambda(a_t - u, \mathbf{x})\, du = \frac{1}{A_{[\lambda]}} \sum_{k=[(L_0-s)\lambda+1]}^{[(L_0+M_0)\lambda]} h(x_i) + o(1) \quad \text{as} \quad \lambda \to 0,$$

and since $\{h(x_i)\}$ is an i.i.d. on $(I_\infty, d\mathbf{x})$ with a law belonging to the domain of attraction of the law of $X(1)$, by the functional limit theorem, we have the following convergence in law in D space:

$$\int_{-M_0}^{s} q_\lambda(a_t - u, \mathbf{z})\, du \xrightarrow{\mathcal{D}} X(L_0 + M_0) - X(L_0 - s)\ .$$

By noting that X is stochastically continuous, we can conclude the proof in the same way as that of Theorem 1. □

3. Construction of a counterexample

Suppose that $\{x_k\}$ is uniformly distributed over $[0,1]$.

First, we prove that we can take a sequence $0 = N_0 = M_0 < N_1 < M_1 < N_2 < M_2 < \cdots$ of integers and a sequence $I_0, I_1, I_2, \ldots$ of open intervals such that, if we put $H_i = \bigcup_{n=0}^{M_i} I_n$ and $J_i = \{x_n \mid n \le N_i\}$, it holds that $|I_n| \le 2^{-n}/12$, $J_i \cap (H_i \setminus H_{i-1}) = \emptyset$,

$$\frac{\#\{n \le N_i \mid x_n \in H_{i-1}\}}{N_i} \le \frac{1}{6} \quad \text{and} \quad \frac{\#\{n \le M_i \mid x_n \in H_i\}}{M_i} \ge \frac{5}{6}\ . \tag{6}$$

Let us take an open interval I_0 with a measure less than $1/12$ arbitrary. Suppose that $0 = N_0 = M_0 < N_1 < M_1 < N_2 < M_2 < \cdots < N_{i-1} < M_{i-1}$ and $I_0, I_1, I_2, \ldots, I_{M_{i-1}}$ are constructed as above. Since H_{i-1} is a finite union of the open intervals and its measure is less than $1/6$, we can take $N_i > M_{i-1}$ satisfying the first inequality of (6). Since J_i is a finite set, we can take $M_i > N_i$ satisfying

$$\frac{\#\{\, n \in (N_i, M_i] \mid x_n \notin J_i \,\}}{M_i} \geq \frac{5}{6}. \tag{7}$$

For $n \in (N_i, M_i]$, if $x_n \in J_i$ let us put $I_n = \emptyset$, otherwise let us take an open interval such that $|I_n| \leq 2^{-n}/12$, $x_n \in I_n$ and $I_n \cap J_i = \emptyset$. We clearly have $J_i \cap (H_i \setminus H_{i-1}) = \emptyset$. If $n \in (N_i, M_i]$ and $x_n \notin J_i$, we have $x_n \in I_n \subset H_i$. Thus the left-hand side of the second inequality of (6) is greater than that of (7), we see that the second inequality of (6) holds. By this we see that such sequences can be constructed inductively.

Let us put $H = \cup_{n=0}^{\infty} I_n$. Because of $J_i \subset J_{i+1}$ and $J_i \cap (H_i \setminus H_{i-1}) = \emptyset$, we have $J_i \cap (H \setminus H_{i-1}) = \emptyset$. Thus we see that the first inequality of (6) implies that of (8) below:

$$\underline{\lim}_{N\to\infty} \frac{\#\{\, n \leq N \mid x_n \in H \,\}}{N} \leq \frac{1}{6} \quad \text{and} \quad \overline{\lim}_{N\to\infty} \frac{\#\{\, n \leq N \mid x_n \in H \,\}}{N} \geq \frac{5}{6}. \tag{8}$$

The second inequality is clear. We also see that (8) is valid if we replace H by H^c, and consequently,

$$\lim_{N\to\infty} \frac{1}{N}\{\, n \leq N \mid (x_{n-1}, x_n) \in H^c \times H^c \,\} \text{ does not exist,} \tag{9}$$

since the limit infinum of the left hand side is less than $1/6$ and the limit supremum is greater than $2/3$.

Since H is a union of open intervals, H is a union of some sequence $\{O_n\}$ of disjoint open intervals. Note that $\Sigma|O_n| < \infty$. Let us take an infinitely differentiable function f such that $f(x) > 0$ for $x \in (0,1)$, and $f(x) = 0$ for $x \notin (0,1)$. For an interval $O = (a,b)$, set $f_O(x) = (b-a)f((x-a)/(b-a))$. Let us put $\tilde{h} = \Sigma_i f_{O_i}$ and $h = \tilde{h} - E\tilde{h}$. Then h and $\tilde{h}$ are infinitely differentiable functions with $\{h \leq -E\tilde{h}\} = \{\tilde{h} \leq 0\} = H^c$. Let us assume that the support of K coincides with $[0,1]$. Then, by noting $Q_1(t, \mathbf{z}) - C = \int_{t-1}^{t} K(t-s)\tilde{h}(z_{[s]})\, ds$, where $C = -E\tilde{h} \int K$, we see that $Q_1(t, \mathbf{z}) \leq C$ is equivalent to $\tilde{h}(z_{[t]}) = \tilde{h}(z_{[t]+1}) = 0$. Thus it is equivalent to $z_{[t]}, z_{[t]+1} \in H^c$. Thus we have

$$\frac{1}{L}|\{\, t \in [0, L] \mid Q_1(t, \mathbf{z}) \leq C \,\}| \sim \frac{1}{[L]}\{\, n \leq [L] \mid (x_{n-1}, x_n) \in H^c \times H^c \,\}$$

which does not converge as $L \to \infty$.

References

[1] J. Bass. Stationary Functions and Their Applications to the Theory of Turbulence, 1. Stationary Functions, *J. Math. Anal. Appl.* **47** (1974), 354–399.

[2] P. Billingsley. *Convergence of Probability Measures*, John Wiley, New York, 1968.

[3] P. P. Hien. Fonction admettant une répartition asymptotique des valuers, *C. R. Acad. Sci. Paris. Ser. A*, **267** (1968), 803–806.

[4] S. Ogawa. Pseudorandom functions whose asymptotic distributions are asymptotically gaussian, *J. Math. Anal. Appl.*, **158** (1991), 94–105.