Markov Processes Relat. Fields **10**, 629–652 (2004)

Markov Processes and **Related Fields**

On the Class of Nilpotent Markov Chains, I. The Spectrum of Covariance Operator

A.M. Alhakim[1] **, J. Kawczak**[2] **and S. Molchanov**[2]

[1] Department of Mathematics, University of Delaware
[2] Department of Mathematics, University of North Carolina, Charlotte

Received March 25, 2003, revised April 1, 2004

Abstract. We study the central limit theorem and the structure of the corresponding covariance operator for the Markov chains generated by successive (overlapping) k-tuples $(X_{n+1}, \dots, X_{n+k})$, $n = 0, 1, \dots$ formed from the i.i.d.r.v. $\{X_n\}$. The potential application of the theory includes the design of statistical tests. In particular, we present the explicit spectral analysis of the covariance matrices related to Marsaglia's k permutation test for $k = 2, 3, 4, 5$.

KEYWORDS: CLT for the nilpotent Markov chain, spectral decomposition, testing random number generators

AMS SUBJECT CLASSIFICATION: Primary 60J05, 60F05, Secondary 60F25, 62H10

1. Introduction

The goal of this paper is the analysis of one special class of Markov chains generated by overlapping frames of i.i.d.r.v. (independent identically distributed random variable) of fixed length $k \geq 1$. Such chains have finite radius of correlations, i.e. the transition generators P are nilpotent, $(P - \Pi)^k = 0$ where $\Pi = \lim_{t\to\infty} P^t$. Our attention to the subject was attracted by the papers by Marsaglia [10] and Kalman, Pincus and Singer [18, 19]; see discussion below.

We include the particular construction from [10] into general setting and we study the CLT (Central Limit Theorem) for our class of Markov chains with special attention to the spectral properties of the covariance operator. The particular case of symmetric and asymmetric Bernoulli r.v. (with direct link to [18, 19]) was the subject of a recent publication [1] of two of us. A

generalization to the multinomial case was discussed in [2]. The method of the present paper is not only more general but also much simpler.

Let $X_1,\ X_2,\ \ldots,\ X_t,\ \ldots$ be *i.i.d.r.v.* with values in a measurable space $(\mathfrak{X}, \mathscr{F}, \mu)$ and μ be a fixed probability distribution. For any integer $k \geq 1$ we consider the Markov chain $Y_t^k,\ k \geq 1$ on $\mathfrak{X}^k$ generated by a frame of width k moving along the sequence $\{X_t, t \geq 1\}$ (the sequence $\left\{Y_i^k\right\}_{i=1}^s$ can also be understood as a collection of k-dependent random variables with a special correlation structure):

$$Y_1^k = (X_1, \ldots, X_k),\ Y_2^k = (X_2, \ldots, X_{k+1}),\ \ldots\ Y_s^k = (X_s, \ldots, X_{s+k-1}). \quad (1.1)$$

This Markov Chain (MC) is ergodic and stationary with invariant measure $\mu^k(dx) = \mu(dx_1) \cdots \mu(dx_k)$ and satisfies the Döblin condition in the strongest form: for any $t > k$ and $x^k \in \mathfrak{X}^k$

$$P_{x^k}(Y_t^k \subset \Gamma) = \mu^k(\Gamma), \quad \Gamma \subset \mathscr{F}^k. \quad (\mathfrak{D})$$

Let $L^2(\mathfrak{X}^k, \mu^k)$ be the space of μ^k-square integrable functions with the centralization condition

$$\int\limits_{\mathfrak{X}^k} f(x^k)\mu^k(\mathrm{d}x^k) = 0. \quad (1.2)$$

This is the subspace of the space of all μ^k-square integrable functions that are orthogonal to the constant functions, equipped with the dot-product $(f_1 \circ f_2) = \int_{\mathfrak{X}^k} f_1 f_2 \mathrm{d}\mu^k$ and the norm $\|f\|_2 = \left(\int_{\mathfrak{X}^k} f(x^k)\, \mu^k(\mathrm{d}x^k)\right)^{1/2}$. On this space the transition operator $\mathcal{P}_{(k)}$ of Y_t^k acts according to the formula:

$$\left(\mathcal{P}_{(k)} f\right)(x_1, \ldots, x_k) = \int\limits_{\mathfrak{X}} f(x_2, \ldots, x_k, \xi)\mu\,(\mathrm{d}\xi). \quad (1.3)$$

The conjugate operator $\mathcal{P}_{(k)}^*$, with respect to the standard dot-product on L^2, is given by

$$\left(\mathcal{P}_{(k)}^* f\right)(x_1, \ldots, x_k) = \int\limits_{\mathfrak{X}} f(\eta, x_1, \ldots, x_{k-1})\mu(\mathrm{d}\eta). \quad (1.4)$$

Note. It is known (and easy to see) that

$$\|\mathcal{P}_{(k)}\|_2 = \|\mathcal{P}_{(k)}^*\|_2 = 1$$

for any $k \geq 1$.

In this paper we work only in $L^2(\mathfrak{X}^k, \mu^k)$ spaces. In the continuation of the paper (Part II) we will use the standard L^∞-norm for the general Markov

processes: $\|f\|_\infty = \sup_{x^k \in \mathfrak{X}^k} |f(x)|$. If there is no ambiguity in the display we can drop index (k) in the $\mathcal{P}$ operator and the index in the notation for the norm $\|\cdot\|$. The projector $\Pi_{(k)} = \lim_{n\to\infty} \mathcal{P}^n_{(k)} = \mathcal{P}^k_{(k)}$ on the subspace of constants is vanishing on L^2 due to centralization assumption (1.2). As a result,

$$\mathcal{P}^k_{(k)} \equiv 0, \qquad \mathcal{P}^{*k}_{(k)} \equiv 0. \tag{$\mathfrak{D}$'}$$

Probabilistically it means that the chain Y^k_t has a finite radius of correlations: the σ-algebras $\mathscr{F}_{\le t}$ and $\mathscr{F}_{\ge s}$ are independent if $s - t > k$. This (after the independence) is the strongest possible mixing condition. Any mixing coefficient $\mathfrak{m}(\tau)$ (among a dozen known) is identically zero for $\tau > k$.

The construction, presented above, gives probably the most interesting and most important class of nilpotent Markov chains. This class was used already in applications (mainly to the computational complexity and the testing of random number generators) [10, 18, 19].

Remark 1.1. Marsaglia [10] proposed in his battery of tests the following overlapping-permutation statistic. Let $X_1, \dots, X_t, \dots$ be *i.i.d.r.v.*, uniformly distributed on $\mathfrak{X} = [0,1]$ and k be an integer, say $k = 3$. For each permutation $\pi = \binom{1\ 2\ 3}{i_1\ i_2\ i_3} \in S^3$ one can calculate the local time $\tau_{i_1 i_2 i_3}(t)$ for the Markov chain $Y^3_t = \{X_t, X_{t+1}, X_{t+2}\}$, $t \ge 0$, on the tetrahedron $T_{i_1 i_2 i_3} \subset [0,1]^3$ given by the inequalities $(x_{i_1} < x_{i_2} < x_{i_3})$. Namely

$$\tau_{i_1 i_2 i_3}(t) = \sum_{i=1}^{t-2} \mathrm{I}(Y^3_t \subset T_{i_1 i_2 i_3}).$$

For large t we can use the CLT for the joint distribution of the local times $T_\pi(t)$, $\pi \in S^3$, and construct a test (or tests) based on χ^2 or Gaussian statistics.

In the recent papers [18, 19] Kalman, Pincus and Singer proposed (for the case $\mathfrak{X} = \{0,1\}$, i.e. binary sequence X_t, $t \ge 0$) the concept of computational entropy which calculation is based on the analysis of the chains Y^k_t associated with symmetric Bernoulli scheme.

The complete analysis of the CLT for the approximate entropy and the local times in the situation similar to [18,19] but more general (k-frame moving along realizations of not necessarily symmetric Bernoulli sequences) appears in [1]. In this paper we generalize the results of [1] and [2] to the case of an arbitrary distribution μ (and chains Y^k_t given by (1.1)).

Let us recall the CLT for our specific situation. If $f(x^k) \in L^2(\mathfrak{X}^k, \mu^k)$ and $(f \circ 1) = 0$, then the homological equation

$$g - \mathcal{P}g = f, \qquad (g \circ 1) = 0 \tag{1.5}$$

has a unique solution in L^2: $g(x^k) = f(x^k) + \mathcal{P}f(x^k) + \cdots + \mathcal{P}^{k-1}f(x^k)$. The normalized sum $(1/\sqrt{t})\sum_{s=1}^{t} f(Y_s^k)$ has, as a result, the representation

$$S_f(t) = \frac{1}{\sqrt{t}}\sum_{s=1}^{t-1}\big[g(Y_{s+1}^k) - \mathcal{P}g(Y_s^k)\big] + \frac{1}{\sqrt{t}}\big[g(Y_1^k) - g(Y_t^k)\big], \quad g \in L^2(\mathfrak{X}^k, \mu^k),$$

where the first term in the right part is the sum of a stationary and ergodic sequence of martingale difference and the second term tends to 0 in probability as $t \to \infty$. Billingsley's theorem in [3] gives for $S_f(t)$ the CLT

$$S_f(t) \xrightarrow{\mathscr{L}} \mathcal{N}(0, \sigma^2(f)),$$

where

$$\begin{aligned}\sigma^2(f) &= (g \circ g) - (\mathcal{P}g \circ \mathcal{P}g) = \big[(g - \mathcal{P}g) \circ (g + \mathcal{P}g)\big] \\ &= [f \circ (\mathbf{I} + 2\mathcal{P} + \ldots + 2\mathcal{P}^{k-1})f] = (f \circ \mathfrak{B}f). \end{aligned} \tag{1.6}$$

Here $\mathfrak{B}$ is a symmetric operator on $L^2(\mathfrak{X}^k, \mu^k)$ with representation

$$\mathfrak{B} = \mathbf{I} + \mathcal{P} + \mathcal{P}^* + \ldots + \mathcal{P}^{k-1} + (\mathcal{P}^*)^{k-1}. \tag{1.7}$$

Remark 1.2. Without the centralization condition $(f \circ 1) = 0$, $f \in L^2$,

$$\mathfrak{B} = \mathbf{I} + \mathcal{P} + \mathcal{P}^* + \ldots + \mathcal{P}^{k-1} + (\mathcal{P}^*)^{k-1} - (2k+1)\Pi.$$

□

We call $\mathfrak{B}$ the covariance operator for the chain Y_t^k, $t \geq 0$. In fact

$$\lim_{t\to\infty} \mathrm{Cov}\bigg(\frac{1}{\sqrt{t}}\sum_{s=1}^{t} f(Y_s^k), \frac{1}{\sqrt{t}}\sum_{s=1}^{t} h(Y_s^k)\bigg) = (\mathfrak{B}f \circ h) = (f \circ \mathfrak{B}h), \quad f, h \in L^2.$$

Although very compact, the formula above in many practical applications presents a formidable computational task. To resolve this problem we propose the decomposition of the L^2 space into the subspaces corresponding to the eigenspaces of the covariance operator, $\mathfrak{B}$.

Next, we present four theorems that fully characterize the decomposition of L^2 and, subsequently, allow for much easier calculations of the limiting covariance in the CLT.

Theorem 1.1 (Spectrum of $\mathfrak{B}$). *The spectrum of operator $\mathfrak{B}$ is discrete and depends only on k. Namely*

$$\Sigma(\mathfrak{B}) = \{0, 1, \ldots, k\} \tag{1.8}$$

and $L^2(\mathfrak{X}^k, \mu^k) = \bigoplus_{i=0}^{k} \mathcal{L}_i$, $\mathfrak{B}\psi = i\psi$ for $\psi \in \mathcal{L}_i$.

The proof is the corollary of an explicit description of the invariant subspaces $\mathcal{L}_i$, $i = 0, 1, \ldots, k$ (Theorems 1.2, 1.3, 1.4). These subspaces are infinite dimensional if $\operatorname{Dim} L^2(\mathfrak{X}, \mu) = \infty$. In the special case, when $\mathfrak{X} = \{x_1, \ldots, x_N\}$, $\mu\{x_i\} = \mu_i > 0$, $\sum_{i=1}^{N} \mu_i = 1$ (multinomial distribution of X_t, $t \geq 0$) and $\operatorname{Dim} L^2(\mathfrak{X}^k, \mu^k) = N^k - 1$, the proof of the theorem gives the dimensions of $\mathcal{L}_i$, $i = 0, 1, \ldots, k$ (which coincide with the result in [2] where these were obtained using vector methods and induction on k):

$$\begin{aligned} &\operatorname{Dim} \mathcal{L}_k = N - 1, \quad \operatorname{Dim} \mathcal{L}_{k-1} = (N-1)^2, \quad \ldots, \\ &\operatorname{Dim} \mathcal{L}_{k-i} = (N-1)^2 N^{i-1}, \quad \text{for } 1 \leq i \leq k-1, \\ &\operatorname{Dim} \mathcal{L}_0 = N^{k-1} - 1. \end{aligned} \tag{1.9}$$

Counting the constant function, this says that the dimension of the null space is N^{k-1}.

To formulate the next theorem we need to introduce the concept of a significant variable.

Definition 1.1. The variable x_s, $1 \leq s \leq k$, is *significant* for the function $f(x_1, \ldots, x_s, \ldots, x_k) \in L^2(\mathfrak{X}^k, \mu^k)$ iff $\int_{\mathfrak{X}} f(x_1, \ldots, x_s, \ldots, x_k)\mu(\mathrm{d}x_s) = 0$ (μ^{k-1}-a.e.). Hence, $f(x_1, \ldots, X_s, \ldots, x_k)$ contains a significant variable on the sth co-ordinate.

The meaning of Definition 1.1 is clear from the following construction. Let $\psi_0 = 1$, $\psi_1(x), \ldots, \psi_i(x), \ldots$ be an orthonormal system in $L^2(\mathfrak{X}, \mu)$. Then a natural orthonormal system in $L^2(\mathfrak{X}^k, \mu^k)$ is given by

$$\psi_{i_1,\ldots,i_k}(x_1, \ldots, x_k) = \psi_{i_1}(x_1) \cdots \psi_{i_k}(x_k),$$

and subject to the centralization condition $i_1 + \cdots + i_k \geq 1$. It is easy to see the variable x_s is significant for $f(x_1, \ldots, x_s, \ldots, x_k)$ iff the Fourier expansion

$$f(x_1, \ldots, x_s, \ldots, x_k) = \sum_{(i_1 \cdots i_k)} c_{i_1 \cdots i_k} \psi_{i_1 \cdots i_k}(x_1, \ldots, x_s, \ldots, x_k)$$

contains (with non-vanishing coefficient $c_{i_1,\ldots,i_s,\ldots,i_k}$) only the basis functions $\psi_{i_1,\ldots,i_s,\ldots,i_k}$, with $i_s > 0$. We use for the significant variables the capital letters, i.e. for instance the notation $f(X_1, x_2, \ldots, x_{k-1}, X_k)$ means that

$$\int_{\mathfrak{X}} f(x_1, \ldots, x_k)\mu(\mathrm{d}x_1) = \int_{\mathfrak{X}} f(x_1, \ldots, x_k)\mu(\mathrm{d}x_k) = 0.$$

The integrals over $x_2, \ldots, x_{k-1}$ of the function $f(X_1, x_2, \ldots, x_{k-1}, X_k)$ can be *trivial*, i.e. equal zero, or *nontrivial* and can be completely independent of any group of the variables from the set of $x_2, \ldots, x_{k-1}$.

Theorem 1.2 gives the spectral decomposition of $L^2(\mathfrak{X}^k, \mu^k)$ based on the orthogonal expansion. As a result it is not very efficient for the computational purposes. We need a result that links the orthogonal expansion to the eigenspaces expansion. This task is accomplished by Theorem 1.3 which describes the direct decomposition of the function $f \in L^2(\mathfrak{X}^k, \mu^k)$ into the orthogonal eigencomponents:

$$f = \sum_{i=0}^{k} f_i, \quad \mathfrak{B} f_i = i f_i \quad \text{and} \quad (\mathfrak{B} f \circ f) = \sigma^2(f) = \sum_{i=1}^{k} i \|f_i\|^2.$$

Theorem 1.4 gives the effective description of the $\mathrm{Ker}(\mathfrak{B}) = \mathcal{L}_0$ for a general Markov chain. In the last Theorem 3.1 we present a complete description of the $\mathrm{Ker}(\mathfrak{B} f \circ f)$ which is specialized to the permutation type Marsaglia test for the quality of the Pseudo Random Number Generators (see Example 3.4 in Section 3).

Theorem 1.2. *Let us consider the following subspaces in $L^2(\mathfrak{X}^k, \mu^k)$:*

$$\begin{aligned}
\mathcal{S}_k &= \{f:\ \varphi_1(X_1) + \cdots + \varphi_k(X_k),\ \varphi_i \in L^2(\mathfrak{X}, \mu),\ i = 1, 2, \ldots k\} \\
\mathcal{S}_{k-1} &= \{f:\ \varphi_1(X_1, X_2) + \cdots + \varphi_{k-1}(X_{k-1}, X_k),\ \varphi_i \in L^2(\mathfrak{X}^2, \mu^2), \\
&\qquad i = 1, 2, \ldots k-1\} \\
\mathcal{S}_{k-2} &= \{f:\ \varphi_1(X_1, x_2, X_3) + \cdots + \varphi_k(X_{k-2}, x_{k-1}, X_k),\ \varphi_i \in L^2(\mathfrak{X}^3, \mu^3), \\
&\qquad i = 1, 2, \ldots k-2\} \\
&\cdots \qquad \cdots \\
\mathcal{S}_1 &= \{f:\ \varphi_1(X_1, x_2, \cdots, x_{k-1}, X_k),\ \varphi_1 \in L^2(\mathfrak{X}^k, \mu^k)\}.
\end{aligned} \tag{1.10}$$

Then

a) *the subspaces $\mathcal{S}_i$, $i = 1, 2, \ldots, k$, are $\mathfrak{B}$-invariant and orthogonal, $L^2(\mathfrak{X}^k, \mu^k) = \bigoplus_{i=1}^{k} \mathcal{S}_i$.*

b) *For $i \geq 2$ we have additional decomposition*

$$\mathcal{S}_i = \mathcal{L}_i^{(i)} \bigoplus \mathcal{L}_i^{(0)} \tag{1.11}$$

where the components $\mathcal{L}_i^{(i)}$, $\mathcal{L}_i^{(0)}$ are $\mathfrak{B}$-eigenspaces with eigenvalues $\lambda = i$ and $\lambda = 0$. Hence, we have:

$$\begin{aligned}
\mathcal{L}_i = \mathcal{L}_i^{(i)},\ i = 2, 3, \ldots, k, \qquad & \mathcal{L}_1 = \mathcal{S}_1 \\
\mathcal{L}_0 = \bigoplus_{i=2}^{k} \mathcal{L}_i^{(0)}, \qquad & L^2(\mathfrak{X}^k, \mu^k) = \bigoplus_{i=0}^{k} \mathcal{L}_i.
\end{aligned} \tag{1.12}$$

Theorem 1.3. *Let* $\mathcal{S}_{\geqslant i} = \bigoplus_{j=i}^{k} \mathcal{S}_j$ *and in particular* $\mathcal{S}_{\geqslant k} = \mathcal{S}_k$, $\mathcal{S}_{\geqslant 1} = L^2(\mathfrak{X}^k, \mu^k)$. *For any* $i \geq 2$ *the projection* $\Pi_{\geqslant i} f$ *of* f *onto* $\mathcal{S}_{\geqslant 2}$ *is given by*

$$\begin{aligned}
\Pi_{\geqslant k} f = \Pi_k f &= \varphi_1(x_1) + \cdots + \varphi_k(x_k) \\
&= \int\limits_{\mathfrak{X}^{k-1}} f(x_1, x_2, \ldots, x_k)\mu(\mathrm{d}x_2)\mu(\mathrm{d}x_3)\ldots\mu(\mathrm{d}x_k) \\
&\quad + \cdots + \int\limits_{\mathfrak{X}^{k-1}} f(x_1, x_2, \ldots, x_k)\mu(\mathrm{d}x_1)\mu(\mathrm{d}x_2)\ldots\mu(\mathrm{d}x_{k-1}), \\
\Pi_{\geqslant k-1} &= \varphi_1(x_1, x_2) + \cdots + \varphi_{k-1}(x_{k-1}, x_k) \\
&= \int\limits_{\mathfrak{X}^{k-2}} f(x_1, x_2, \ldots, x_k)\mu(\mathrm{d}x_3)\mu(\mathrm{d}x_4)\ldots\mu(\mathrm{d}x_k) \\
&\quad + \cdots + \int\limits_{\mathfrak{X}^{k-2}} f(x_1, x_2, \ldots, x_k)\mu(\mathrm{d}x_1)\mu(\mathrm{d}x_2)\ldots\mu(\mathrm{d}x_{k-2}) \\
\cdots &\qquad \cdots \\
\Pi_{\geqslant 2} &= \varphi_1(x_1, \ldots, x_{k-1}) + \varphi_2(x_2, \ldots, x_k) \\
&= \int\limits_{\mathfrak{X}} f(x_1, x_2, \ldots, x_k)\mu(\mathrm{d}x_k) + \int\limits_{\mathfrak{X}} f(x_1, x_2, \ldots, x_k)\mu(\mathrm{d}x_1)).
\end{aligned}$$

Projections of f *onto* $\mathcal{S}_k$, $\mathcal{S}_{k-1}$, …, $\mathcal{S}_1$ *are*

$$\begin{aligned}
\Pi_k f &= \Pi_{\geqslant k} f, & \Pi_{k-1} f &= \Pi_{\geqslant (k-1)} f - \Pi_{\geqslant k} f \\
\Pi_{k-2} f &= \Pi_{\geqslant (k-2)} f - \Pi_{\geqslant (k-1)} f, & & \\
\cdots &\quad \cdots & & \\
\Pi_2 f &= \Pi_{\geqslant 2} f - \Pi_{\geqslant 3} f, & Pi_1 f &= f - \Pi_{\geqslant 2} f.
\end{aligned}$$

The following result (which is probably the new one) gives the efficient description of $\mathrm{Ker}(\mathfrak{B})$ for any Markov chain with Döblin condition $(\mathfrak{D})$.

Let us assume that $(\mathfrak{X}, \mathscr{B})$ is a measurable space, $\mathcal{P}(x, \Gamma)$ with $x \in \mathfrak{X}$ and $\Gamma \in \mathscr{B}$ is a transition operator of the homogeneous Markov chain and the condition $(\mathfrak{D})$ is satisfied for some $k_0 \geq 1$, $\delta > 0$ and probabilistic measure μ on $\mathscr{B}$

$$\mathcal{P}(k_0, x, \cdot) \geq \delta\mu(\cdot).$$

It is well known that in this case there exists a unique invariant measure $\pi(\cdot)$ and for appropriate $c > 0$ the total variation norm satisfies

$$\mathrm{Var}\big(\mathcal{P}(n, x, \cdot) - \pi(\cdot)\big) \leq c\big(1 - \delta\big)^{[n/k_0]}.$$

Let $L^2(\mathfrak{X}, \pi) = \big\{f : \int_{\mathfrak{X}} f^2 \mathrm{d}\pi < \infty,\ \int_{\mathfrak{X}} f \mathrm{d}\pi = 0\big\}$ and $\mathcal{P}^\star$ is the conjugated operator to $\mathcal{P}$ with respect to product $(f_1, f_2)_\pi = \int_{\mathfrak{X}} f_1 \bar{f}_2 \mathrm{d}\pi$.

Theorem 1.4 (Kernel of $\mathfrak{B}$). *Let $\mathfrak{B} = \mathbf{I} + \mathcal{P} + \mathcal{P}^* + \dots$ be the covariance operator of a Markov chain on $L^2(\mathfrak{X}, \mathcal{B}, \pi)$. Then*

(a) *$f \in Ker(\mathfrak{B})$ iff $f = g - \mathcal{P}g$ and $g \in Ker\,(\mathbf{I} - \mathcal{P}^*\mathcal{P})$, i.e. $g = \mathcal{P}^*\mathcal{P}g$.*

(b) *For each $f \in \mathcal{L}_0$ there exists a function $\varphi(x_1, x_2, \dots, x_{k-1}) \in L^2(\mathfrak{X}^{k-1}, \mu^k)$ such that*

$$f(x_1, x_2, \dots, x_k) = \varphi(x_1, x_2, \dots, x_{k-1}) - \varphi(x_2, x_2, \dots, x_k), \quad (\mu^k\text{-a.e.}). \tag{1.13}$$

This description of $\mathcal{L}_0$ is much more direct and efficient than the formula $\mathcal{L}_0 = \bigoplus_{i=2}^{k} \mathcal{L}_0^{(i)}$ in the second part of Theorem 1.2.

Remark 1.3. Since $(\mathbf{I} - \mathcal{P})^{-1}$ and $(\mathbf{I} - \mathcal{P})$ are bounded operators in $L^2(\cdot)$, the subspaces $\mathrm{Ker}(\mathfrak{B})$ and $\mathrm{Ker}(\mathbf{I} - \mathcal{P}^*\mathcal{P})$ are isomorphic. In particular if $\mathrm{Dim}\ \mathrm{Ker}(\mathfrak{B}) < \infty$, then in $L^2(\cdot)$

$$\dim \mathrm{Ker}(\mathfrak{B}) = \dim \mathrm{Ker}(\mathbf{I} - \mathcal{P}^*\mathcal{P}).$$

2. Proofs of the main results (Theorems 1.1–1.4)

Let us start from the proof of Theorem 1.2.

Proof of Theorem 1.2. For each basis function $\psi_{\vec{i}}(\vec{x}) = \psi_{i_1,\dots,i_k}(x_1, \dots, x_k)$ introduced above (after the Definition 1.1) let us define a *significance interval* $[l', l'']$, where $l'(\vec{i}) = \min\{s : i_s \geq 1\}$, $l''(\vec{i}) = \max\{s : i_s \geq 1\}$, and its length $l(\vec{i}) = [l''(\vec{i}) - l'(\vec{i})] + 1$. For instance, in the trigonometric basis on $[0,1]^k = T^k$, we have (here $\imath = \sqrt{-1}$)

$$\psi_{\vec{i}}(\vec{x}) = \exp[2\pi\imath(5x_3)], \quad l' = l'' = 3, \quad l = 1, \quad \psi_{\vec{i}}(\vec{x}) \in \mathcal{S}_k$$

and

$$\psi_{\vec{i}}(\vec{x}) = \exp[2\pi\imath(7x_3 - 2x_4 + x_6)], \quad l' = 3,\ l'' = 6, \quad l = 4, \quad \psi_{\vec{i}}(\vec{x}) \in \mathcal{S}_{k-3}.$$

In general, we can easily see that

$$\begin{aligned}
\mathcal{S}_k &= \mathrm{Span}\big\{\psi_{\vec{i}} : l(\vec{i}) = 1\big\},\\
\mathcal{S}_{k-1} &= \mathrm{Span}\big\{\psi_{\vec{i}} : l(\vec{i}) = 2\big\},\\
&\cdots\\
\mathcal{S}_1 &= \mathrm{Span}\big\{\psi_{\vec{i}} : l(\vec{i}) = k\big\},\\
&= \mathrm{Span}\big\{\psi_{\vec{i}} : l'(\vec{i}) = 1, l''(\vec{i}) = k\ \big\}.
\end{aligned}$$

Each basis function $\psi_{\vec{i}}$, $|\vec{i}| > 0$, belongs to exactly one of the subspaces $\mathcal{S}_i$, for $i = 1, 2, \ldots, k$. This means that $L^2(\mathfrak{X}^k, \mu^k) = \bigoplus_{i=1}^k \mathcal{S}_i$. An application of the operators $\mathcal{P}$ to each basis function either transforms it into 0 (if $i_k > 0$) or shifts it into another function with the same significance length l. The same is true for the operator $\mathcal{P}^*$. This implies that $\mathfrak{B}\mathcal{S}_i \subseteq \mathcal{S}_i$, $i = 1, 2, \ldots, k$, i.e. $\mathcal{S}_i$ are $\mathfrak{B}$-invariant subspaces of $L^2(\mathfrak{X}^k, \mu^k)$.

For the second part of Theorem 1.2 we split $\mathcal{S}_i$, $i = 2, 3, \ldots, k$, into the direct sum of two eigenspaces with eigenvalues $\lambda_i = i$ and $\lambda_0 = 0$. Namely, $\mathcal{S}_i = \mathcal{L}_i^{(i)} \oplus \mathcal{L}_i^{(0)}$, i.e. for any $f \in \mathcal{S}_i$ we have

$$f = \Pi_i^{(i)} f + \Pi_i^{(0)} f = f' + f'', \quad f' \in \mathcal{L}_i^{(i)}, \quad \mathfrak{B} f' = i f', \quad f'' \in \mathcal{L}_i^{(0)}, \quad \mathfrak{B} f'' = 0.$$

Let us start from the particular cases, $i = 1$ and $i = k$, which possess special features.

(a) *Case $i = 1$.* Let us consider

$$\mathcal{S}_1 = \{f \in L^2(\mathfrak{X}^k, \mu^k) : f = f(X_1, x_2, \ldots, x_{k-1}, X_k)\}.$$

In this case

$$\mathcal{P} f = \int_{\mathfrak{X}} f(x_2, \ldots, x_{k-1}, X_k, \xi) \mu(\mathrm{d}\xi) = 0 \tag{2.1}$$

i.e. $\mathcal{P}^l f = 0$, $l = 1, 2, \ldots, k-1$,

$$\mathcal{P}^* f = \int_{\mathfrak{X}} f(\eta, X_1, x_2, \ldots, x_{k-1}) \mu(\mathrm{d}\eta) = 0 \tag{2.2}$$

i.e. $\mathcal{P}^{*l} f = 0$, $l = 1, 2, \ldots, k-1$. Finally we get $\mathfrak{B} f = \mathbf{I} f = f$, i.e. $\mathcal{S}_1 = \mathcal{L}_1^{(1)}$ contains only eigenfunctions of $\mathfrak{B}$ with eigenvalue 1, $\mathcal{L}_1^{(0)} = \varnothing$. We cannot guarantee yet that $\mathcal{S}_1 = \mathcal{L}_1$, but this will follow from the further analysis.

(b) *Construction of $\mathcal{L}_k$.* Let $\mathcal{S}_k = \{f \in L^2(\mathfrak{X}^k, \mu^k) : f(x_1, \ldots, x_k) = f_1(X_1) + f_2(X_2) + \cdots + f_k(X_k)\}$. Of course $\int_{\mathfrak{X}} f_i(x) \mu(\mathrm{d}x) = 0$. Let us apply the operator $\mathfrak{B}$ to these functions. It is easy to see that

$$\begin{aligned} &\mathcal{P} f_1 = f_1(x_2), \quad \mathcal{P}^2 f_1 = f_1(x_3), \quad \ldots \quad \mathcal{P}^{k-1} f_1 = f_1(x_k), \\ &\mathcal{P}^* f_1 = \mathcal{P}^{*2} f_1 = \ldots = \mathcal{P}^{*(k-1)} f_1 = 0. \end{aligned} \tag{2.3}$$

In a similar way

$$\begin{aligned} &\mathcal{P} f_2 = f_2(x_3), \quad \mathcal{P}^2 f_2 = f_2(x_4), \quad \ldots \quad \mathcal{P}^{k-2} f_2 = f_2(x_k), \\ &\mathcal{P}^* f_2 = f_2(x_1), \quad \mathcal{P}^{*2} f_2 = \mathcal{P}^{*2} f_2 = \ldots = \mathcal{P}^{*(k-2)} f_2 = 0, \\ &\qquad \ldots \end{aligned} \tag{2.4}$$

This leads to the following result (in particular this indicates the $\mathfrak{B}$-invariance of $\mathcal{S}_k$, which is already known):

$$\begin{aligned}\mathfrak{B}f \;=\; & f_1(x_1)+f_1(x_2)+\cdots+f_1(x_k)\\ & +f_2(x_1)+f_2(x_2)+\cdots+f_2(x_k)\\ & +\cdots \qquad\qquad \cdots\\ & +f_k(x_1)+f_k(x_2)+\cdots+f_k(x_k).\end{aligned} \tag{2.5}$$

Let us find the eigenvalues of the restriction of $\mathfrak{B}$ on $\mathcal{S}_k$. If $\mathfrak{B}f=\lambda f$ then $\lambda(f_1(x_1)+\cdots+f_k(x_k))=\mathfrak{B}f$, i.e. for the neutral variable x we get

$$\begin{aligned}\lambda f_1(x) &= f_1(x)+f_2(x)+\cdots+f_k(x),\\ \lambda f_2(x) &= f_1(x)+f_2(x)+\cdots+f_k(x),\\ &\cdots \qquad\qquad \cdots\\ \lambda f_k(x) &= f_1(x)+f_2(x)+\cdots+f_k(x).\end{aligned} \tag{2.6}$$

This linear homogenous system has a non-zero solution only in two cases

(i) $\lambda=k$ and $f_1(x)\equiv f_2(x)\equiv\ldots\equiv f_k(x)=f$,

(ii) $\lambda=0$ and $f_1(x)+f_2(x)+\cdots+f_k(x)\equiv 0$.

It means that $\mathcal{S}_k=\mathcal{L}_k^{(k)}\oplus\mathcal{L}_k^0$ where $\mathcal{L}_k^{(k)}=\{g(x_1,\ldots,x_k): g=f(x_1)+f(x_2)+\cdots+f(x_k)\}$.

The orthogonal complement $\mathcal{L}_k^{(0)}$ of $\mathcal{L}_k^{(k)}$ in $\mathcal{S}_k$ has many equivalent descriptions. For example,

$$\begin{aligned}\mathcal{L}_k^{(0)}=\{f: f=\;&\varphi_1(x_1)-\varphi_1(x_k)+\varphi_2(x_2)-\varphi_2(x_k)+\cdots\\ &+\varphi_{k-1}(x_{k-1})-\varphi_{k-1}(x_k)\},\qquad \varphi_1,\ldots,\varphi_{k-1}\in L^2(\mathfrak{X},\mu),\end{aligned}$$

or

$$\begin{aligned}\mathcal{L}_k^{(0)}=\{f: f=\;&\varphi_1(x_1)-\varphi_1(x_2)+\varphi_2(x_2)-\varphi_2(x_3)+\cdots\\ &+\varphi_{k-1}(x_{k-1})-\varphi_{k-1}(x_k)\},\qquad \varphi_1,\ldots,\varphi_{k-1}\in L^2(\mathfrak{X},\mu).\end{aligned}$$

For any $f(x_1,\ldots,x_k)=f_1(x_1)+\cdots+f_k(x_k)$ from $\mathcal{S}_k$ the decomposition $f=f'+f''$, with $f'\in\mathcal{L}_k^{(k)}$, $f''\in\mathcal{L}_k^{(0)}$ and where $(f'\circ f'')=0$, is given by $f'=\bar f(x_1)+\cdots+\bar f(x_k)$, $\bar f(x)=\big(f_1(x)+\cdots+f_k(x)\big)/k$, and $f''=(f_1-\bar f)(x_1)+\cdots+(f_k-\bar f)(x_k)$.

The construction of the eigenspaces $\mathcal{L}_i^{(i)}$, $\mathcal{L}_i^{(0)}$ for $1<i<k$ is very similar to the case $i=k$ but with the use of two significant variables (like in $i=1$).

Let us start with the intermediate space $\mathcal{S}_i \subset L^2(\mathfrak{X}^k, \mu^k)$, $1 < i < k$:

$$\mathcal{S}_i = \big\{ f : f = f_1(X_1, x_2, \dots, x_{k-i}, X_{k-i+1}) + f_2(X_2, x_3, \dots, x_{k-i+1}, X_{k-i+2}) + \cdots + f_i(X_i, x_{i+1}, \dots, x_{k-1}, X_k) \big\}.$$

It is easy to see that

$$\begin{aligned} \mathfrak{B}f \;=\; & f_1(X_1, \dots, X_{k-i+1}) + f_1(X_2, \dots, X_{k-i+2}) + \cdots + f_1(X_i, \dots, X_k) \\ & + f_2(X_1, \dots, X_{k-i+1}) + f_2(X_2, \dots, X_{k-i+2}) + \cdots + f_2(X_i, \dots, X_k) \\ & \qquad \vdots \qquad\qquad\qquad\qquad \vdots \\ & + f_i(X_1, \dots, X_{k-i+1}) + f_i(X_2, \dots, X_{k-i+2}) + \cdots + f_i(X_i, \dots, X_k). \end{aligned} \tag{2.7}$$

The equation $\mathfrak{B}f = \lambda f$ has two types of solutions in the space $\mathcal{S}_i$:

$$\lambda = i, \qquad f_1 = f_2 = \cdots = f_i \equiv f(z_1, \dots, z_{k-i+1}),$$

and

$$\lambda = 0, \qquad (f_1 + f_2 + \cdots + f_i)(z_1, \dots, z_{k-i+1}) \equiv 0.$$

Consequently, the space can be decomposed into a direct sum of two orthogonal components in the space $\mathcal{S}_i = \mathcal{L}_i^{(i)} \oplus \mathcal{L}_i^{(0)}$:

$$\mathcal{L}_i^{(i)} = \big\{ f(x_1, \dots, x_k) : f = \varphi(x_1, \dots, x_{k-i+1}) + \varphi(x_2, \dots, x_{k-i+2}) + \cdots + \varphi(x_i, \dots, x_k) \big\}, \tag{2.8}$$

$$\mathcal{L}_i^{(0)} = \big\{ f(x_1, \dots, x_k) : f_1(x_1, \dots, x_{k-i+1}) + \cdots + f_i(x_i, \cdots, x_k), \; (f_1 + \cdots + f_i)(x_1, \dots, x_{k-i+1}) \equiv 0 \big\}. \tag{2.9}$$

For any f from $\mathcal{S}_i$ the decomposition $f = f' + f''$, with $f' \in \mathcal{L}_i^{(i)}$, $f'' \in \mathcal{L}_i^{(0)}$ and where $(f' \circ f'') = 0$, is given by

$$f' = \bar{f}(x_1, \dots, x_{k-i+1}) + \cdots + \bar{f}(x_{k-i}, \dots, x_k),$$

$$\bar{f}(x_1, \dots, x_{k-i+1}) = \frac{f_i + \cdots + f_{k-i}}{k-i}(x_1, \dots, x_{k-i+1}),$$

and

$$f'' = (f_1 - \bar{f})(x_1, \dots, x_{k-i+1}) + \cdots + (f_k - \bar{f})(x_{k-i}, \dots, x_k).$$

Finally, we can put

$$\mathcal{L}_0 = \bigoplus_{i=1}^{k} \mathcal{L}_i^{(0)}, \qquad \mathcal{L}_i = \mathcal{L}_i^{(i)}, \quad i = 1, 2, \dots, k. \tag{2.10}$$

This gives the spectral decomposition

$$L^2(\mathfrak{X}^k, \mu^k) = \bigoplus_{i=0}^{k} \mathcal{L}_i,$$

and the statement of Theorem 1.1 (except for the formulas for $\mathrm{Dim}(\mathcal{L}_i)$, $i = 0, 1, \ldots, k$, in the case when $\mathrm{Card}(\mathfrak{X}) = N$) follows directly from the above.

The dimensions of the spaces $\mathcal{L}_i$ can be easily calculated:

$$\mathrm{Dim}\,\mathcal{S}_k = (N-1)k, \quad \mathrm{Dim}\,\mathcal{L}_k^{(k)} = N-1, \quad \mathrm{Dim}\,\mathcal{L}_k^{(0)} = (N-1)(k-1).$$

If $1 < i < k$, then

$$\begin{aligned} \mathrm{Dim}\,\mathcal{S}_i &= i(N-1)^2 N^{k-i-1}, \\ \mathrm{Dim}\,\mathcal{L}_i^{(i)} &= (N-1)^2 N^{k-i-1}, \\ \mathrm{Dim}\,\mathcal{L}_i^{(0)} &= (i-1)(N-1)^2 N^{k-i-1}. \end{aligned}$$

For $i = 1$ we have

$$\begin{aligned} \mathrm{Dim}\,\mathcal{S}_1 = \mathrm{Dim}\,\mathcal{L}_1^{(1)} &= (N-1)^2 N^{k-2}, \\ \mathrm{Dim}\,\mathcal{L}_1^{(0)} &= 0. \end{aligned}$$

This implies further that

$$\mathrm{Dim}\,\mathcal{L}_0 = \sum_{i=1}^{k-1} (i-1)(N-1)^2 N^{k-i-1} + (N-1)(k-1) = N^{k-1} - 1.$$

□

Proof of Theorem 1.3. The proof of this theorem is a natural extension of the central idea contained in Theorem 1.2. For instance, after integration

$$\varphi = \int_{\mathfrak{X}} f(x_1, \ldots, x_k)\mu(\mathrm{d}x_k) + \int_{\mathfrak{X}} f(x_1, \ldots, x_k)\mu(\mathrm{d}x_1),$$

in the orthogonal expansion $f = \sum_{\vec{i}:|\vec{i}|\geq 1} c_{\vec{i}}\psi_{\vec{i}}(\vec{x})$, all basis functions $\psi_{\vec{i}}(\cdot)$ such that $i_1 > 0$ or $i_k > 0$ will disappear and, as a result, the function φ will be the projection of f onto $\mathcal{S}_{\geqslant 2}$, i.e. $\varphi = \Pi_{\geqslant 2} f$, etc. □

Proof of Theorem 1.4. If $f \in \mathcal{L}_0$, then $\sigma^2(f) = (\mathfrak{B}f \circ f) = 0$, i.e.,

$$0 = (g \circ g) - (\mathcal{P}g \circ \mathcal{P}g) = (g \circ g) - (\mathcal{P}^*\mathcal{P}g \circ g) = \left((\mathbf{I} - \mathcal{P}^*\mathcal{P})g \circ g\right).$$

The operator $\mathcal{P}^*\mathcal{P}$ is a stochastic, self-adjoint operator with the spectrum in $[-1,1]$. It means that $\mathbf{I}-\mathcal{P}^*\mathcal{P}$ is a non-negative symmetric operator and equality $((\mathbf{I}-\mathcal{P}^*\mathcal{P})g \circ g) = 0$ implies $(\mathbf{I}-\mathcal{P}^*\mathcal{P})g = 0$, or $g = \mathcal{P}^*\mathcal{P}g$. But, due to the homological equation, $f = g - \mathcal{P}g$.

If $g = g(x_1,\dots,x_k)$, then $\mathcal{P}g = \int_{\mathfrak{X}} g(x_2,\dots,x_k,z)\mu(\mathrm{d}z) = \varphi(x_2,\dots,x_k)$. Now, application of $\mathcal{P}^*$ gives $\mathcal{P}^*\mathcal{P}g = \varphi(x_1,\dots,x_{k-1})$. Finally,

$$f = \varphi(x_1,\dots,x_{k-1}) - \mathcal{P}\varphi(x_1,\dots,x_{k-1}) = \varphi(x_1,\dots,x_{k-1}) - \varphi(x_2,\dots,x_k).$$

□

3. Statistical examples and applications

Assume that for a very long sequence of elements $x_1,\dots,x_n$ from $\mathfrak{X}$ we want to test the basic hypothesis H_0 that our observations represent a sample of *i.i.d.r.v.* with distribution μ against the alternative hypothesis H_A, i.e. non-*i.i.d.r.v.*. Due to Marsaglia [10] the typical scheme of the testing based on Theorem 1.1 has the following structure.

Let us fix an integer $k \geq 1$ and a finite collection $\{f_i, i = 1,2,\dots,N\}$ of functions $f_i : \mathfrak{X}^k \longrightarrow \mathbf{R}$, $\int_{\mathfrak{X}} f_i\mu(\mathrm{d}x) = 0$. If the functions do not satisfy the second condition we have to centralize them by a constant.

For the sequence of the k-tuples $y_1 = (x_1,\dots,x_k)$, $y_2 = (x_2,\dots,x_{k+1})$, …, $y_{n-k+1} = (x_{n-k+1},\dots,x_n)$ we have to calculate the sums

$$S_i^*(n) = \frac{1}{n-k+1}\sum_{s=0}^{n-k+1} f_i(y_s), \qquad \overrightarrow{S}^*(n) = \begin{bmatrix} S_1^* \\ \vdots \\ S_N^* \end{bmatrix}$$

and find the corresponding covariance operator $\mathcal{C} = [c_{ij}]$, $c_{ij} = \lim_{n\to\infty} \mathsf{E}\, S_i^* S_j^*$. As we already know (see Theorem 1.1)

$$c_{ij} = \sum_{l=0}^{k} l\big(\Lambda_l f_i \circ \Lambda_l f_j\big)$$

where $\{\Lambda_l, l = 0,1,\dots,k\}$ are the orthogonal projectors on the invariant subspaces $\mathcal{L}_l \subset L^2(\mathfrak{X}^k,\mu^k)$.

If $\mathrm{Det}\mathcal{C} \neq 0$ (otherwise we use an appropriate subgroup of $\{f_i\}$), then

$$\mathcal{C}^{-1/2}\overrightarrow{S}^*(n) \stackrel{law}{\Longrightarrow} \mathcal{N}(0,\mathbf{I})$$

and the random variable $(\mathcal{C}^{-1}\overrightarrow{S} \circ \overrightarrow{S})$ has asymptotically a χ^2 distribution with N degrees of freedom. For a fixed confidence probability $\alpha \lessapprox 1$ one can find $(1-\alpha)$ the quantiles z'_α and z''_α such that

$$\mathsf{P}\{\chi^2_N \leq z'_\alpha\} \leq (1-\alpha)/2, \qquad \mathsf{P}\{\chi^2_N \geq z''_\alpha\} \leq (1-\alpha)/2$$

and the null hypothesis H_0 is rejected if $(\mathcal{C}^{-1}\overrightarrow{S}(n) \circ \overrightarrow{S}(n)) \notin (z'_\alpha, z''_\alpha)$.

The functions $\{f_i\}$ should be as simple as possible, otherwise in the long process of computing the sums $S_i^*(n)$ the error of the calculations can suppress the covariances. An ideal case is when we have piece-wise constant functions $\{f_i\}$ where the sums $S_i^*(n)$ can be expressed in terms of the occupation times for some subsets of $\mathfrak{X}^k$.

Example 3.1. Let $\mathfrak{X} = [0,1]$, $\mu(dx) = dx$ and we want (as in Theorem 1.1) to use the orthonormal basis $\{\psi_i(x)\}$ on $L^2(\mathfrak{X}, \mu)$ and functions $f_{i_1 \dots i_k} = \psi_{i_1}(x_1) \times \cdots \psi_{i_k}(x_k)$ from $L^2(\mathfrak{X}^k, \mu^k)$. Then the trigonometric basis $\psi_0(x) = 1/2$, $\psi_{2n}(x) = \cos(\pi n x)$, $\psi_{2n+1}(x) = \sin(\pi n x)$, $n \in \mathbf{N}$, is not good from the computational point of view, but the Haar basis (wavelets) is ideal. Let us recall the definition of the Haar wavelet functions.

Let $\psi : \mathbf{R} \to \mathbf{R}$ be defined by

$$\psi(x) = \begin{cases} 1, & \text{if } \ 0 \le x < 1/2, \\ -1, & \text{if } \ 1/2 \le x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Define $\psi_i^j : \mathbf{R} \to \mathbf{R}$ as

$$\psi_i^j(x) = \sqrt{2^j}\psi(2^j x - i), \quad j = 0, 1, \dots, \ \text{ and } \ i = 0, 1, \dots, 2^j - 1.$$

Define the vector space W^j by

$$W^j = \mathrm{Span}\{\psi_i^j : \ i = 0, 1, \dots, 2^j - 1\}$$

where Span denotes a linear span, see [4, 5].

This system usually produces an infinite number of groups of functions spanning the vector spaces W^j (since $L^2(\mathfrak{X}, \mu)$ can be infinitely dimensional).

The above Example 3.1, or more precisely its version, has applications to the testing of RNG's. There are mainly two classes of RNG's: the block RNG's, producing the word sequences $W_n = [d_{n1}, \dots, d_{nl}]$ of the fixed length l and the RNG's producing individual digits one by one. Typical RNG of the first kind uses the linear congruence of the form $W_{n+1} = aW_n + b \pmod{2^l}$. One of the commonly used generators of this type is the BSD `rand()` with $a = 1103515245$, $b = 12345$ and $l = 31$. This generator produces the 32 bit pseudo random numbers.

A famous example of the second kind is given by a generator suggested by Kirkpatrick and Stoll [8] and called R-250. This is the shift-register generator and the sequence is

$$x_n = x_{n-103} \odot x_{n-250}$$

where $\odot$ denotes "exclusive-or" operation, defined on 32-bit words. The period of this generator is about 2^{250} and it uses 250 words of state per generator.

Example 3.2. To apply the general theory for the block RNG's with a fixed length l of the successive words W_n we can introduce $\mathfrak{X} = \{x \in [0,1] : x = d_1/2 + \cdots + d_l/2^l\}$, μ to be a uniform measure on $\mathfrak{X}$ with atoms of masses of 2^{-l} (this is a distribution on the grid with the step $\Delta = 2^{-l}$). The natural basis in $L^2(\mathfrak{X}, \mu)$ consists of the first $l-1$ blocks of the wavelets:

$$\psi_{11} = \begin{cases} 1, & x \in [0, 1/2), \\ -1, & x \in [1/2, 1), \end{cases}$$

$$\psi_{21} = \begin{cases} 2^{1/2}, & x \in [0, 1/4), \\ -2^{1/2}, & x \in [1/4, 1/2), \end{cases} \qquad \psi_{22} = \begin{cases} 2^{1/2}, & x \in [1/2, 3/4), \\ -2^{1/2}, & x \in [3/4, 1), \end{cases}$$

$$\cdots \quad \cdots$$

$$\psi_{l-1,1} = \begin{cases} 2^{(l-1)/2}, & x = 0, \\ -2^{(l-1)/2}, & x = 1/2^l, \end{cases} \quad \cdots \quad \psi_{l-1,\,2^{l-1}} = \begin{cases} 2^{(l-1)/2}, & x = 1 - 2/2^l, \\ -2^{(l-1)/2}, & x = 1 - 1/2^l. \end{cases}$$

Total dimension of $L^2(\mathfrak{X}, \mu)$ is equal, of course, $1 + 2 + \cdots + 2^{l-1} = 2^l - 1$.

For the most of the practical applications we choose the fixed length typically $l \approx 30 - 60$. The result the generator produces are "almost uniform" random variables (of course this is a subject of the further study whether the random numbers are indeed almost uniform).

The following example illustrates the second class of the RNG's introduced earlier, namely the binary strings.

Example 3.3. Let us specify the general strategy for testing an output of some RNG and in the particular case of the binary string $S_t = (d_1, \ldots, d_t)$. It is natural to restrict ourselves to the generators producing digits d_t one by one, rather than blocks of digits of fixed length l (l-tuples). In this case we can apply the general theory to the space $\mathfrak{X} = \{0, 1\}$, $\mu(0) = \mu(1) = 1/2$. Typical applications include the physical generators and, say, the R-250 generator given present earlier with the initial condition that $(d_1, \ldots, d_{250})$ uniformly distributed on $\mathfrak{W}_{250}$.

The standard basis on $L^2(\mathfrak{X}, \mu)$ consists of only two functions:

$$\varphi_1 \equiv 1 \quad \text{and} \quad \varphi_2 = \chi(x) = \begin{cases} -1, & x = 0, \\ 1, & x = 1. \end{cases}$$

a) We have to start testing with the functions of one variable, namely, $\chi(x_1)$. The corresponding (leading factor) statistic is simple:

$$S^{(1)}(t) = n_0(t) - n_1(t).$$

Here $n_0(t)$ is the number of zeros in $S_t = (d_1, \ldots, d_t)$. $n_1(t)$, $n_{00}(t)$, $n_{10}(t)$, n_{01}, etc. are defined similarly. Of course,

$$\frac{S^{(1)}(t)}{\sqrt{t}} \stackrel{law}{\Longrightarrow} \mathcal{N}(0,1), \qquad \frac{[S^{(1)}(t)]^2}{t} \stackrel{law}{\Longrightarrow} \chi_1^2.$$

b) There is only one interesting function of two variables: $f(x_1, x_2) = \chi(x_1) \times \chi(x_2)$. It provides the following (second factor) statistic

$$S^{(2)}(t) = n_{00}(t) + n_{11}(t) - n_{01}(t) - n_{10}(t).$$

It is easy to see that

$$S^{(2)}(t) = n - 2R(t) + o(1).$$

Here $R(t)$ is the number of runs (sequences of the "0" and "1") in the string $S(t)$. The statistics $S^{(2)}(t)$ or $R(t)$ are sensitive to local correlations. Let us illustrate this by the following example. Consider the Markov chain over $\mathfrak{X} = \{0, 1\}$ with transition matrix $P = \left(\begin{smallmatrix} p & q \\ q & p \end{smallmatrix}\right)$, $p + q = 1$. Then, the invariant measure is uniform: $\mu(0) = \mu(1) = 1/2$ but the number of runs $R(t)$ has expectation $nq = n(1-p)$. For $p > 1/2$ we have the deficit of the runs, for $p < 1/2$ we get too many runs in comparison to the symmetric Bernoulli case when $p = 1/2$.

The statistic $R(t)$ is optimal for separating three situations: $p < 1/2$, $p = 1/2$, and $p > 1/2$.

c) There are two essential functions of three variables. The most symmetric representation of these functions has the form

$$\varphi_1(x_1, x_2, x_3) = \chi(x_1)\Big(\frac{\chi(x_2) + 1}{4}\Big)\chi(x_3),$$
$$\varphi_2(x_1, x_2, x_3) = \chi(x_1)\Big(\frac{\chi(x_2) - 1}{4}\Big)\chi(x_3).$$

Correspondingly, the asymptotically independent $\mathcal{N}(0,1)$ statistics have the following representations

$$S_1^{(3)}(t) = \frac{n_{000}(t) + n_{101}(t) - n_{100}(t) - n_{001}(t)}{2\sqrt{n}},$$
$$S_2^{(3)}(t) = \frac{n_{010}(t) + n_{111}(t) - n_{110}(t) - n_{011}(t)}{2\sqrt{n}},$$

and

$$\frac{\big[S_1^{(3)}(t)\big]^2 + \big[S_2^{(3)}(t)\big]^2}{t} \stackrel{law}{\Longrightarrow} \chi_2^2.$$

d) Subsequently, the functions with higher number of variables can be easily constructed.

In the case of asymmetric Bernoulli scheme the statistics are similar. Let again $\mathfrak{X} = \{0, 1\}$, but $\mu(0) = q$ and $\mu(1) = p$, and $p + q = 1$.

The basis in $L^2(\mathfrak{X}, \mu)$ has a form:

$$\psi_0(x) \equiv 1 \quad \text{and} \quad \psi_1(x) = \begin{cases} \sqrt{p/q}, & x = 0, \\ -\sqrt{q/p}, & x = 1. \end{cases}$$

The leading factor statistics has a simple expression

$$S^{(1)}(t) = \frac{pn_0(t) - qn_1(t)}{\sqrt{pqt}} \qquad \text{and} \quad S^{(1)}(t) \stackrel{law}{\Longrightarrow} \mathcal{N}(0, 1).$$

The second factor statistics, up to the constant factor, is equal to

$$S^{(2)}(t) = p^2 n_{00}(t) - pqn_{10}(t) - pqn_{01}(t) + q^2 n_{11}(t),$$

and so on.

Example 3.4 (Marsaglia permutation test). Let us consider the spectrum of the covariance operator $\mathbf{C}_k$ for the system of additive functionals generated by $k!$ functions

$$f_\pi(x_1, \ldots, x_k) = \mathbf{I}_\pi(x_1, \ldots, x_k) - \frac{1}{k!} \in L^2([0, 1]^k; dx_1, \ldots, dx_k),$$

$\pi \in S_k$, where S_k is a group of permutations, and $\{X_t, t \geq 1\}$ is a sequence of i.i.d.r.v. uniformly distributed on $[0, 1]$. The functionals describe the local permutations in the window of the length k, moving along the sequence $\{X_t\}$.

There are two possible approaches to the problem.

(i) Project $f_\pi(\cdot)$, $\pi \in S_k$ on the invariant subspaces $\mathcal{L}_l$ for $0 \leq l \leq k$, then diagonalize the covariance operator of the projections, and finally use the orthogonality of $\mathcal{L}_l$. Unfortunately, the functions $f_\pi(\cdot)$ are generic with respect to the decomposition

$$L^2 = \bigoplus_{i=1}^{n} \mathcal{L}_i$$

and calculations (in each subspace $\mathcal{L}_i$) are not simple.

(ii) Directly evaluate the $k! \times k!$ matrix

$$\mathbf{C}_k = \left[\mathfrak{B}_k f_\pi \circ f_{\pi'}\right], \quad \mathfrak{B}_k = \mathbf{I} + \mathcal{P} + \cdots + \mathcal{P}^{k-1} + \mathcal{P}^* + \ldots + (\mathcal{P}^*)^{k-1}.$$

Corresponding calculations (including the spectral analysis of the $\mathbf{C}_k$ matrix) are straightforward and can be done for relatively small $k = 2, 3, 4$ without much effort. Both approaches exhibit a very similar level of complexity. Let us present the final results.

Case $k = 2$

In this case we have two linearly dependent functions

$$f_1(x_1, x_2) = \mathbf{I}_{x_1<x_2} - \frac{1}{2}, \quad f_2(x_1, x_2) = \mathbf{I}_{x_1>x_2} - \frac{1}{2}, \quad f_1 + f_2 \equiv 0.$$

The covariance operator is

$$\mathbf{C}_2 = \begin{bmatrix} 1/12 & -1/12 \\ -1/12 & 1/12 \end{bmatrix}, \quad \lambda_0 = 0,\ \lambda_1 = 1/6.$$

For the testing of RNG one can use the following asymptotically Gaussian statistic:

$$S_n^{(2)} = \frac{1}{\sqrt{n}}\big[\#\{i : X_i < X_{i+1}, i = 1, \dots, n-1\} - \#\{i : X_i > X_{i+1}, i = 1, \dots, n-1\}\big] = \frac{2}{\sqrt{n}} \sum_{i=1}^{n-1} f_1(X_i, X_{i+1}).$$

Furthermore, a couple of standard results can be deduced

$$S_n^{(2)} \overset{law}{\Longrightarrow} \mathcal{N}(0, 1/3), \quad 3\big[S_n^{(2)}\big]^2 \overset{law}{\Longrightarrow} \chi_1^2.$$

Case $k = 3$

In this case calculations are a bit longer. As a final result we get

$$\mathbf{C}_3 = \frac{1}{360} \begin{bmatrix} 46 & -23 & 7 & -14 & 7 & -23 \\ -23 & 28 & -20 & 7 & -2 & 10 \\ 7 & -20 & 28 & -23 & 10 & -2 \\ -14 & 7 & -23 & 46 & -23 & 7 \\ 7 & -2 & 10 & -23 & 28 & -20 \\ -23 & 10 & -2 & 7 & -20 & 28 \end{bmatrix}$$

and the eigenvalues ordered in the descending order:

$$\lambda_1 = \frac{2+\sqrt{2}}{12}, \quad \lambda_2 = \frac{2}{15}, \quad \lambda_3 = \frac{1}{10}, \quad \lambda_4 = \frac{2-\sqrt{2}}{12}, \quad \lambda_5 = \lambda_6 = 0.$$

If $\mathfrak{e}_i$, $i = 1, 2, 3, 4$, are the normalized eigenvectors of $\mathbf{C}_3$ corresponding to $\lambda_i > 0$ for $i = 1, 2, 3, 4$, then as in $S_n^{(2)}$ we can construct the asymptotic χ_4^2 test using the statistic

$$S_n^{(3)} = \sum_{i=1}^{4} \frac{1}{\lambda_i (n-2)} \left[\sum_{s=1}^{n-2} \big(\overrightarrow{f} \circ \overrightarrow{\mathfrak{e}}_i\big)\big(Y_s^{(3)}\big) \right]^2.$$

On the other hand, it is very useful for the computational purposes to concentrate on the largest eigenvalue and the corresponding eigenvector. Thus, the focus should be on the first principal component (factor) and the statistic constructed from the linear combination of the six basic functions, $f_\pi(\cdot)$.

In this case the normalized eigenvector that corresponds to the largest eigenvalue is $\mathfrak{e}_1^T = [-1/2, \sqrt{2}/4, -\sqrt{2}/4, 1/2, -\sqrt{2}/4, \sqrt{2}/4]$. The eigenvector was determined using the Singular Value Decomposition (SVD). This method should be preferred over any other method due to the degeneracy of the covariance matrix. The first principal component accounts for more than 50% of the total variance. After including the second principal component we can explain 74% of the total variability. For the completeness of the argument we present the second eigenvector is $\mathfrak{e}_2^T = [-\sqrt{3}/3, \sqrt{3}/6, \sqrt{3}/6, -\sqrt{3}/3, \sqrt{3}/6, \sqrt{3}/6]$.

Case $k = 4$

For this case we have 24×24 variance-covariance matrix and the computations were a bit longer than before. The $\operatorname{rank}(\mathbf{C}_4) = 4! - 3! = 18$. The largest eigenvalue for the matrix $\mathbf{C}_4$ is

$$\lambda_1 = \frac{11}{180} + \frac{1}{36}\sqrt{\frac{23}{5}}\cos\left[\frac{1}{3}\arctan\left(\frac{3\sqrt{163311}}{226}\right)\right] \cong 0.114438$$

and the corresponding eigenvector has the coefficients $\mathfrak{e}_1^T = [-0.358986, -0.017021, -0.142267, -0.306704, 0.0899841, -0.0352617, 0.306704, 0.142267, -0.0352617, -0.160508, -0.324944, 0.017021, -0.017021, 0.324944, 0.160508, 0.0352617, -0.142267, -0.306704, 0.0352617, -0.0899841, 0.306704, 0.142267, 0.017021, 0.358986]$. The trace of the covariance matrix is $237/280$.

On the other hand, a direct construction of the partial sums process results in the statistic that has asymptotic distribution of χ^2_{18}.

Case $k = 5$

By direct application of Theorems 1.2 and 1.3 we were able to find the variance-covariance matrix in the case $k = 5$. As expected, based on the cases $k = 2, 3, 4$, the rank of the matrix is $\operatorname{rank}(\mathbf{C}_5) = 5! - 4! = 96$. This fact was also proved, for any k, in Theorem 3.1. The degrees of freedom that were computed directly differ, however, from the result obtained by Marsaglia (probably by numerical simulations) which is included in one of the Diehard Battery of tests for RNG, OPERM5 — "The Overlapping 5-Permutation Test". Therein, it is stated that the rank of the covariance matrix should be 99. Our calculations give exact factorization of the characteristic polynomial for $\mathbf{C}_5$ into irreducible polynomials of degrees less than or equal to 7. As a result, some of the eigenvalues are explicit (24 zeros and 54 non-zero roots of the equations of order at most 4) and remaining eigenvalues are given numerically. The largest eigenvalue is $\lambda_{\max} = 0.0314829$ and the smallest nonzero eigenvalue is $\lambda_{\min} = 0.00220416$. Also, the trace of the covariance matrix is $7253/7560$.

The complete analysis of the Marsaglia statistics and effective algorithm for the calculation of $\mathbf{C}_k^{-1}$ in the $(\mathrm{Ker}\mathbf{C}_k)^\perp$ will be the part of our next publication.

Let us apply Theorem 1.4 and Remark 1.3 to the general Marsaglia statistics (for full description of the statistic see Example 3.4). The following proposition gives the description of $\mathcal{L}_0$ for the operator $\mathfrak{B}_k$ in the case of i.i.d. uniformly distributed random variables.

Proposition 3.1. *Space $\mathcal{L}_0 = Ker(\mathfrak{B}_k) \subset L^2\left([0,1]^k, d\boldsymbol{x}\right)$ consists of functions $g = g(x_2, x_3, \ldots, x_k)$ that are constant with respect to the first variable x_1.*

Proof. The transition operator $\mathcal{P}^*\mathcal{P}$ for the corresponding Markov chain has the following structure

$$(x_1, x_2, \ldots, x_k) \xrightarrow{\mathcal{P}^*\mathcal{P}} (\xi, x_2, \ldots, x_k),$$

where ξ is a uniformly distributed on $[0,1]$ random variable. In fact, we can write each action of the operator as

$$(x_1, x_2, \ldots, x_k) \xrightarrow{\mathcal{P}} (x_1, x_2, \ldots, x_k, \xi_1) \xrightarrow{\mathcal{P}^*} (\xi_2, x_2, \ldots, x_k),$$

where $\xi_1,\ \xi_2$ are uniformly distributed on $[0,1]$ random variables. This implies that

$$\left(\mathcal{P}^*\mathcal{P}g\right)(x_1, x_2, \ldots, x_k) = \int_0^1 g(\xi, x_2, \ldots, x_k)\mathrm{d}\xi,$$

and the last integral is independent of x_1. If $g = g(x_2, x_3, \ldots, x_k)$, then $\left(\mathcal{P}^*\mathcal{P}g\right) = g$. □

Theorem 3.1 (Marsaglia permutation test). *For the Marsaglia covariance matrix $\mathbf{C}_k = (\mathfrak{B}_k f_\pi, f_{\pi'})$, as in Example 3.4, with $k \geq 2$ we have*

(*a*) $\dim(\mathbf{C}_k) = (k-1)!$,

(*b*) $\mathrm{Ker}(\mathfrak{B}_k) = (\mathbf{I} - \mathcal{P})\mathscr{G}_k^0$ *where*

$$\mathscr{G}_k^0 = \left\{\mathrm{Span}\Big(\mathbf{1}_\pi - \frac{1}{(k-1)!}\Big), \quad \pi \in S_{k-1}\right\}.$$

Proof of Theorem 3.1. Let us note first that a linear combination of the functions f_π (or $\mathbf{I}_\pi$), $\pi \in S_k$, has constant values on the tetrahedrons $T_\pi = \{x \in [0,1]^k : \mathbf{I}_\pi = 1\}$. $\mathscr{M}_k = \mathrm{Span}(f_\pi, \pi \in S_k)$ consist of the functions that are measurable with respect to algebra of subsets of $[0,1]^k$ generated by partition $\bigcup_{\pi \in S_k} T_\pi = [0,1]^k$.

All functions g in the Marsaglia space $\mathscr{M}_k$ which are independent of x_1 have a very simple structure; $g \in \mathrm{Span}(f_\pi, \pi \in S_{k-1})$. To see this, let us consider a

point $(x_2, x_3, \ldots, x_k)$ such that $x_2 < x_3 < \ldots < x_k$. We have k possibilities for inserting the variable x_1:

$$\begin{aligned} &x_1 < x_2 < x_3 < \ldots < x_k, \\ &x_2 < x_1 < x_3 < \ldots < x_k, \\ &\quad \ldots \quad \ldots \\ &x_2 < x_3 < \ldots < x_k < x_1. \end{aligned}$$

Function g is piecewise constant on T_π, $\pi \in S_k$, i.e. it has the same values on the sets $\mathscr{A}_1 = \{\mathbf{x} : x_1 < x_2 < x_3 < \ldots < x_k\}$, $\mathscr{A}_2 = \{\mathbf{x} : x_2 < x_1 < x_3 < \ldots < x_k\}$, …, $\mathscr{A}_k = \{\mathbf{x} : x_2 < x_3 < \ldots < x_k < x_1\}$. In other words, g is constant on

$$\bigcup_{i=1}^{k} \mathscr{A}_i = \{\mathbf{x} : x_2 < x_3 < \ldots < x_k\} = T_\pi, \quad \pi \in S_{k-1}.$$

This proves the first part of the theorem. The second part follows from the last statement of Theorem 1.4. □

The application of Theorems 1.2 and 1.3 in the above example demonstrates the Marsaglia procedure can be used in a quite efficient manner for the testing of RNGs that generate single digits. Moreover, the approach extends naturally to the block RNGs. Also, the complexity of the calculations for the higher values of k suggests that the simpler functions, like Haar functions, should be preferred for constructing the test statistics.

4. Conclusion

To complete our program we have to answer additional and principle questions: How to select k? How to control the remainders in CLT and the computational errors?

It is clear that among the functions $f \in L^2(\mathfrak{X}^k, \mu^k)$, $\|f\|_2 = 1$, the maximal information provides elements with maximal variance $\sigma^2(f) = (\mathfrak{B}_k f \circ f) = k$, $f \in \mathcal{L}_k$. Such functions — due to standard terminology — are the principle components of the problem. According to Theorem 1.1 for any $k \geq 1$ and for $f \in \mathcal{L}_k$

$$f(x_1, \cdots, x_k) = \varphi(x_1) + \cdots + \varphi(x_k), \quad \varphi \in L^2(\mathfrak{X}, \mu), \quad (\varphi \circ \mathbf{1}) = 0,$$

i.e. essentially, $f(\cdot)$ can be reduced to a function of one variable and the additive functional $S_k(n) = \sum_{s=0}^{n-k} f(Y_s^k)$ to $\sum_{s=0}^{n} \varphi(X_s) = S_1(n)$.

As we already mentioned above from the computational point of view (minimization of the errors, complexity of the calculations, etc.) we need to restrict ourselves to the piecewise constant functions with integer values. It

is natural to select (before centralization) $\varphi_1 = \mathbf{I}_{\Gamma_1}$, ..., $\varphi_N = \mathbf{I}_{\Gamma_N}$, where $\{\Gamma_i, i = 1, 2, \ldots, N\}$ is a partition of $\mathfrak{X}$. As a result we arrive at the classical Pearson's statistics

$$\sum_{i=1}^{N} \frac{(\nu_i - np_i)^2}{np_i} = \chi^2_{N-1}(n),$$

where $p_i = \mu(\Gamma_i)$, $\sum p_i = 1$, $\nu_i = \sum_{s=1}^{n} \mathbf{I}_{\Gamma_i}(X_s)$ and asymptotically as $n \to \infty$ has χ^2-law with $N - 1$ degrees of freedom.

The structure of the partition depends on the specific nature of $\mathfrak{X}$ and a priori statistical information.

We know, however, that χ^2-statistics is insensitive to local correlations of a random sequence and reacts mainly to the deviations from the one-dimensional theoretical distribution μ. To catch such correlations we have to work with functions f containing many significant variables. It is natural — for the second stage of testing — to consider the case of two consecutive significant variables. More precisely, let

$$f(x_1, \cdots, x_k) = \varphi(x_1, x_2) + \cdots + \varphi(x_{k-1}, x_k), \quad \varphi \perp \mathcal{L}_k, \tag{4.1}$$

i.e. $\int_{\mathfrak{X}} \varphi(x_1, x_2)\mu(\mathrm{d}x_2) \equiv \int_{\mathfrak{X}} \varphi(x_1, x_2)\mu(\mathrm{d}x_1) \equiv 0$ or, in short, $f \in \mathcal{L}_{k-1}$, $\mathfrak{B}_k f = (k-1)f$ and $(k-1)$ is the second (after $\lambda_{\max} = k$) eigenvalue.

As before the practical calculations are simpler and more precise if $\varphi(\cdot)$ is a linear combination of indicators, with the additional requirement of orthogonality to $\mathcal{L}_k$.

In general for any fixed $k > 1$ the information, which we cannot extract from the functions of the smaller number of essential variables $k_1 < k$ is provided by the functions from the eigenspace $\mathcal{L}_1$ (with $\lambda = 1$). This result was first observed for the discrete (multinomial) measure with n atoms in [2].

The last question (the remainder term for the CLT in the class of the nilpotent Markov chains or the asymptotical expansions in CLT) has a pure analytical nature. In the classical literature on this subject, see for instance [12,13,15,16], or more recent review, [6] and [14], the constants in the remainders are not effective. In the most recent work [9] estimations are effective but involved constants are very large. Consequently, the validity of the statistical inference based on the samples of order $n \sim 10^6 - 10^9$ becomes questionable.

For the nilpotent chains the situation is essentially better. We can effectively calculate in this case the higher moments of $S^*_f(n) = (1/\sqrt{n}) \sum_{s=1}^{n} f(Y^k_s)$, construct the asymptotical expansions for the characteristic function of $S^*_f(n)$ and the distribution of $S^*_f(n)$ and estimate the remainder in this expansion.

The proof of these results, the description of algorithms for the testing of RNG's based on the ideology presented above and the statistical ramifications for specific RNG's will be the subject of the second part of our paper.

Acknowledgments

The authors are grateful to a referee for carefully reading the paper and for valuable comments that enhanced the presentation of the material.

References

[1] A.M. Alhakim and S. Molchanov (2004) Some Markov chains on Abelian groups with applications. In: *Random Walks and Geometry*, Proceedings of a workshop at the Erwin Schrödinger Institute, V.A. Kaimanovich (ed.), de Gruyter, 3–33.

[2] A.M. Alhakim (2004) On the eigenvalues and eigenvectors of an overlapping Markov chain. *Probab. Theory and Relat. Fields*, **128**, 589–605.

[3] P. Billingsley (1999) *Convergence of Probability Measures.* Wiley Series in Probability and Statistics.

[4] C.K. Cui (ed.) (1992) *Wavelets: A Tutorial in Theory and Applications.* Academic Press, San Diego.

[5] I. Daubechies (1992) *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics, Philadelphia.

[6] P. Gudynas (2000) Refinements of the central limit theorem for homogeneous Markov chains. In: *Limit Theorems of Probability Theory.* Yu.V. Prokhorov, V. Statulevicius (eds.), Springer-Verlag, 167–183.

[7] T. Kato (1966) *Perturbation Theory for Linear Operators.* Springer-Verlag, New York.

[8] S. Kirkpatrick and E. Stoll (1981) A very fast shift-register sequence random number generator. *Journal of Computational Physics* **40**, 517–526.

[9] B. Mann (1996) Berry-Esseen central limit theorems for Markov chains. Ph. D. Thesis, Harvard University, Cambridge, Massachusetts.

[10] G. Marsaglia (1985) *A Current View of Random Number Generators.* Computer Science and Statistics, Elsevier Science Publisher B.V., North-Holland.

[11] G. Marsaglia *The Marsaglia Random Number CDROM, Including the DIEHARD Battery of Tests of Randomness.* Department of Statistics, Florida State University, Tallahassee, Florida.

[12] S.V. Nagayev (1957) Some limit theorems for stationary Markov chains. *Theory Probab. and Appl.* **2**, 378–406.

[13] S.V. Nagayev (1961) More exact limit theorems for homogenous Markov chains. *Theory Probab. and Appl.* **6**, 62–81.

[14] S.V. Nagayev (1996) On accuracy of approximation in central limit theorem. In: *Probability Theory and Mathematical Statistics* (St. Petersburg, 1993), Gordon and Breach, Amsterdam, 95–108.

[15] P. Ney and E. Nummelin (1987) Markov additive processes. I. Eigenvalue properties and limit theorems. *Ann. Prob.* **15** (2), 561–592.

[16] P. Ney and E. Nummelin (1987) Markov additive processes. II. Large deviations. *Ann. Prob.* **15** (2), 593–609.

[17] S.K. Park and K.W. Miller (1988) Random number generators: Good ones are hard to find. *Communications of the ACM* **31** (10), 1192–1201.

[18] S. Pincus and R.E. Kalman (1997) Not all (possibly) "random" sequences are created equal. *Proc. Natl. Acad. Sci. USA*, April 15, **94** (8), 3513–3518.

[19] S. Pincus and B.H. Singer (1996) Randomness and degrees of irregularity. *Proc. Natl. Acad. Sci. USA*, Mar. 5, **93** (5), 2083–2088.

[20] S. Tezuka (1995) *Uniform Random Numbers: Theory and Practice.* Kluwer Academic Press, Boston.