The numerical criterion on pseudo random numbers

Yukiko SHIMOYAMA

Department of Mathematics
Tokyo Metropolitan University
Minami-Ohsawa, Hachioji-shi, Tokyo 192-0397, Japan

In this report, we propose a numerical criterion to judge 'goodness' of a pseudo random number by achieving the following procedures.

(i) We have many sample paths of a simple random walk driven by pseudo random numbers.

(ii) From those sample paths, we have sample distributions of some random variables, that is the return number to point zero of pseudo random walk, the first hitting time, the sojourn time, and the central limit theorem, the Poisson's law.

(iii) We compute Lévy metrics between those sample distributions and the theoretical distributions.

Then our assertion is that these Lévy metrics are numerical criterions of 'goodness' of a pseudo random number.

1 Introduction

Let $\mathcal{P}^1$ be the set of all one-dimensional distributions, and set $F_{\mu}$ the distribution function for $\mu \in \mathcal{P}^1$. At the non-continuous point $a$ of $F_{\mu}(x)$, we join between $(a, F_{\mu}(a^-))$ and $(a, F_{\mu}(a^+))$ by straight line, and the graph of $F_{\mu}(x)$ comes to be a continuous curve. This curve must intersect with a line $x+y=t$ at a point, say $Q_{\mu}(t)$, and let $G_{\mu}(t)$ be the distance between $Q_{\mu}(t)$ and a point $(t, 0)$ on the x-axis.

Let $\mathcal{G}$ be the set of all of such function $G_{\mu}(t)$ for $\mu \in \mathcal{P}^1$. For any $\nu, \mu \in \mathcal{P}^1$, we define the metric $d$ on $g$ by

\[d(\mu, \nu) \equiv \sup_t |G_{\mu}(t) - G_{\nu}(t)|,\]

Obviously, each $\mu \in \mathcal{P}^1$ corresponds to $G_{\mu} \in \mathcal{G}$, and the correspondence is one to one. Hence, for any $\mu, \nu \in \mathcal{P}^1$ we can define the metric $d_L$ on $\mathcal{P}^1$ by

\[d_L(\mu, \nu) \equiv d(G_{\mu}, G_{\nu}),\]
then \((P^1, d_L)\) is a complete separable metric space. We will call the metric \(d_L\) on \(P^1\) Lévy metric.

2 Some functions of a simple random walk and their distribution

When \(\{X_n\}_n\in \mathbb{N}\) be i. i. d random variables such that

\[
P[X_n = 1] = P[X_n = -1] = \frac{1}{2}
\]

we define a simple random walk \(\{S_n\}_n\in \mathbb{N}\) by the following:

\[
S_0 \equiv 0, \quad S_n \equiv \sum_{k=1}^{n} X_k
\]

We introduce some some functions of the simple random walk (2.2) and present their theoretical distributions.

(a) Return number to the origin \(R_{2n}\),
For \(n \in \mathbb{N}\), we define a random variable \(R_{2n}\) called the return number,

\[
R_{2n} \equiv \#\{k|S_{2k} = 0, 0 < k \leq n\}.
\]

This random variable is numbers of times returning to the origin by time \(2n\), and its distribution is as follows: for \(0 \leq k \leq n\),

\[
P[R_{2n} = k] = \frac{1}{2^{2n-k}} \binom{2n-k}{n}.
\]

(b) First return time \(\tau\),
We define a random variable \(\tau\), by the first return time

\[
\tau \equiv \min\{r|S_{2r} = 0\}.
\]

Its distribution is; for \(k \geq 1\),

\[
P[\tau = 2k] = \frac{1}{2k} \binom{2k-2}{k-1} \frac{1}{2^{2k-2}}.
\]

(c) Sojourn time \(\lambda_{2n}\),
For \(n \in \mathbb{N}\) let \(\lambda_{2n}\) be the number of edges over \(x\)-axis of random walk starting from the origin, that is

\[
\lambda_{2n} = \sum_{1}^{2n} I(S_{i-1} - S_i), \quad \text{where} \quad I(x) = \begin{cases} 1 & (x > 0) \\ 0 & \text{(otherwise)} \end{cases}
\]
and it is called sojourn time. The distribution of sojourn time is as follows:

\[ P[\lambda_{2n} = 2k] = \frac{1}{2^{2n}} \binom{2n}{k} \binom{2n - 2k}{n - k}, \quad 0 \leq k \leq n. \]

(d) Central Limit Theorem (CLT)

Well-known CLT asserts that

\[
\lim_{n \to \infty} P[a \leq \frac{1}{\sqrt{n}} S_n \leq b] = N(b) - N(a), \quad a \leq b,
\]

where \( N(t) \equiv \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{s^2}{2} \right\} ds \).

By a technical reason, we must approximate \( N \) by a step function \( N^e(t) \), that is

\[
N^e(-\infty, b] \equiv \begin{cases} 0 & |b| \geq 5 \\ N(-\infty, k/100] & k/100 \leq b < (k + 1)/100 \quad \text{and} \quad |b| < 5 \\ \end{cases}
\]

where \(-500 \leq k \leq 500\). A simple computation derives that

\[
d_L(N(-\infty, *], N^e(-\infty, *)) \leq 5.641 \times 10^{-3}
\]

3 \hspace{1cm} Samples of random walk and Poisson distribution driven by pseudo random numbers

I. If \( \{a_n\} \) is the sequence of pseudo random numbers. Then put \( \tilde{X}_n \equiv (-1)^{a_n} \). Then

\[
\tilde{S}_0 \equiv 0, \quad \tilde{S}_n \equiv \sum_{k=1}^{n} \tilde{X}_k
\]

are samples of a simple random walk driven by \( \{a_n\} \).

II. Let \( b(k; n, p), 0 \leq k \leq n \), the binomial distribution with probabilities \( p \) for success, and \( q = 1 - p \) for failure result in \( k \) successes and \( n - k \) failures, i.e.

\[
b(k; n, p) = \binom{n}{k} p^k q^{n-k}
\]

Next, we put

\[
\lambda \equiv np
\]

where \( np = \lambda \) is fixed for large \( n \) and small \( p \), and

\[
b(k; n, p) \sim \frac{\lambda^n}{n!} \exp\{-\lambda\} \quad \text{as} \quad n \to \infty
\]
4 Lévy metrics by numerical experiences

I. Since the distribution functions (2.3), (2.4), and (2.5) are step functions, they are written as follows:

(4.1) \[ F(x) \equiv \sum_{k=0}^{n} a_k I_{(t_k, t_{k+1})}(x), \quad -\infty < t_0 < t_1 < \cdots < t_n < \infty. \]

Moreover ‘the distribution of pseudo random walk’ are similarly written as

(4.2) \[ G(x) \equiv \sum_{k=0}^{n} b_k I_{(t_k, t_{k+1})}(x), \quad -\infty < t_0 < t_1 < \cdots < t_n < \infty. \]

Now we can compute Lévy metric \( d_L(F, G) \) more easy.

**Lemma 4.1.** For one-dimensional distributions \( F, G \) given by (4.1, 4.2),

(4.3) \[ d_L(F, G) = \sqrt{2} \min \{ \rho(F, G), 1 \}. \]

where, \( \rho(F, G) \equiv \max \{|a_k - b_k| : 0 \leq k \leq n\}. \)

II. Using (4.3) in lemma 4.1, we measure Lévy metrics between theoretical distribution and the distribution function of samples driven by the following pseudo random numbers

(i) Sugita’s random numbers\(^1\), (ii) Mathematica’s random numbers, (iii) Rand\(^2\), (iv) Rand\(^3\),

where the number of samples are 10000 and the number of steps 10000.

**Results.**

<table>
<thead>
<tr>
<th></th>
<th>Sugita</th>
<th>Mathematica</th>
<th>Rand</th>
<th>Rand 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>return number, (2.3)</td>
<td>0.004346</td>
<td>0.0067175</td>
<td>0.00678055</td>
<td>0.00863727</td>
</tr>
<tr>
<td>first return time, (2.4)</td>
<td>0.006119</td>
<td>0.0078</td>
<td>0.0084</td>
<td>0.0072375</td>
</tr>
<tr>
<td>sojourn time, (2.5)</td>
<td>0.008019</td>
<td>0.0145775</td>
<td>0.0119898</td>
<td>0.0115034</td>
</tr>
<tr>
<td>CLT, (2.7)</td>
<td>0.011774</td>
<td>0.0103801</td>
<td>0.0105278</td>
<td>0.0770797</td>
</tr>
<tr>
<td>Poisson, (3.3)</td>
<td>0.005759</td>
<td>0.00392056</td>
<td>0.00408512</td>
<td></td>
</tr>
</tbody>
</table>

**Acknowledgement.** The author would express her thanks to Professor S. Sato, Dr. J. Hatamoto, Dr. T. Ueno, Mr. M. Tanaka, and Mr. M. Nakamura.

\(^1\) See [4].

\(^2, 3\) Standard pseudo random numbers written in C-language.
References


