Manipulating intervals of probability density function in autoregression model for simulating strictly stationary random sequences

Sergey Garbar

Abstract—We consider a problem of simulating strictly stationary random sequences. A modification of an autoregression algorithm of first order is proposed. It allows to simulate of stationary random sequences with uniform distribution. Correlation properties of resulting random sequences are examined. Value $R(1)$ of autocorrelation function must be in $[-0.625, 0.625]$. As the uniform distribution is the base for the inverse random sampling, simulated sequence can be further transformed to get the random sequence with specified distribution.

Keywords—random sequences, autoregression model, uniform distribution

I. Introduction

Simulating of random processes is used for modeling objects and phenomena of different nature, wherein stationary processes have an important role for describing time series with parameters such as the mean, variance and correlation structure not changing over time and not following any trends.

The objective of this work is to propose and examine the method for simulating the strictly stationary random sequences with uniform probability distribution. Request for strict stationarity allows for more appropriate and adequate modeling, but makes the problem more complicated and intricate. The proposed method is a modification of the autoregressive algorithm that is based on the autoregressive model.

Autoregression model of n-th order $AR(n)$

$$Y(t) = a_1Y(t-1) + a_2Y(t-2) + \ldots + a_nY(t-n) + bX(t)$$

can be used to describe wide-sense stationary random processes if the complex roots of the characteristic polynomial $\lambda^n - \sum_{i=1}^{n} a_i\lambda^{n-i} = 0$ lie inside the unit circle.

If the distribution of the added random variable $X(t)$ is normal, then the process will also be normal and strictly stationary. That happens because next member of the sequence is determined as a linear combination of previous and an added random variable, and it is known that the linear combination of normally distributed random variables is also normally distributed.

In [1] authors propose to transform strictly stationary normal sequences using probability integral transform. The necessity to inverse non-elementary cumulative distribution function of Gaussian random variable is a downside of that approach.

In [2] author proposes to simulate strictly stationary random sequences using autoregression algorithm with stochastic binary orthogonal coefficients. The downside of resulting sequences is that the next value is either independent of all the previous or equals to one of them, i.e. the joint probability function may turn up unfit to model certain processes (see [5]).

In [3] the probability density function of added random variable in $AR(1)$ model is found for some distributions.

The topic of this article is a modification of an $AR(1)$ model which allows simulating of strictly stationary random sequences with standard uniform distribution and specified value of autocorrelation function $R(1)$ in $[-0.625, 0.625]$. It continues the work described in [4]. Simulated sequence can be further transformed using inverse transform sampling to get a specified distribution.

The sections of this work are dedicated to the description and reasoning of the proposed method; analysis of the joint probability distribution function of the consequential members of the generated sequence and its correlation properties; recap of algorithm and example; using the proposed method in conjunction with inverse transform sampling method to get the distribution of sequence members different form uniform.

II. Description of method

Linear combination $Y^*(t) = Y(t-1) + b \cdot X(t)$ where $b \leq 1$ of two uniformly distributed random variables $Y(t-1)$ and $X(t)$ has the probability density function (see fig. 1)

$$g^*(y) = \begin{cases} \frac{y}{b}, & \text{if } y \in [0,1], \\ \frac{1}{b}, & \text{if } y \in (1,b], \\ \frac{1+b}{b} - \frac{y}{b}, & \text{if } y \in (b, b+1], \\ 0, & \text{if } y \notin [0, b+1]. \end{cases}$$

We define manipulating intervals of probability density function (PDF) as a linear transformation that allows to change this PDF to that of a uniform distribution in $[0.5, b+0.5]$, which then can be easily transformed to standard uniform distribution.

If $y^* \in [0; 0,5]$, then $y^{**} = (1 - y^*)$. If $y^* \in [b+0,5; b+1]$, then $y^{**} = (2b - y^* + 1)$. That way the probabilities of random variable Y^{**} belonging in $[0,1]$ and $[b, b+1]$ are changed compared to that of Y^*, and PDF $g^{**}(y)$ is changed too.

Event $Y^{**}(t) \in [0, 0.5] \cup [b+0.5, b+1]$ is impossible, and therefore $g^{**}(y) = 0$ in this set. If $y^{**} \in [0.5, 1]$ then

This work was supported in part by Project Part of the State Assignment in Field of Scientific Activity by the Ministry of Education and Science of Russian Federation, project no 1.949.2014/K.

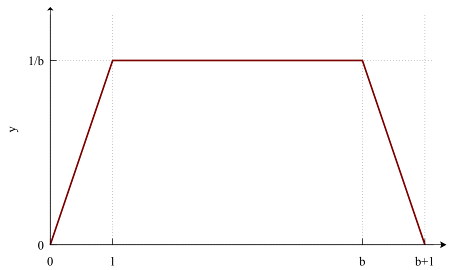

Fig. 1. Linear combination $Y(t-1)+b\cdot X(t)$ of two uniformly distributed random variables where $b \geq 1$

$g^{**}(y) = g^{*}(y) + g^{*}(1-y) = \frac{y}{b} + \frac{1-y}{b} = \frac{1}{b}$ as the probability of random variable getting into this interval is increased. Similarly if $y^{**} \in [b, b+0.5]$, then $g^{**}(y) = g^{*}(y) + g^{*}(2b-y+1) = \frac{b+1}{b} - \frac{y}{b} + \frac{b+1}{b} - \frac{2b-y+1}{b} = \frac{1}{b}$.

After applying the described transformation we obtain a random variable with PDF

$$g^{**}(y) = \begin{cases} \frac{1}{b}, & \text{if } y \in [0.5, b+0.5]\,, \\ 0, & \text{if } y \notin [0.5, b+0.5]\,. \end{cases}$$

If y^{**} is further transformed that $y = (y^{**} - \frac{0.5}{b}$, then $Y(t) = Y$ has a standard uniform distribution.

Thus the following member of the sequence either equals linear combination $(\frac{1}{b}Y(t-1) + X(t))$ of two independent uniform random variables or is a linear function of such a combination.

III. Analysis of generated sequences

A. Finding joint probability density function

Joint probability density function of $Y(t-1)$ and $Y(t)$ is

$$g(y_1, y_2) = \begin{cases} 1, & \text{if } (y_1, y_2) \in A_1, \\ 2, & \text{if } (y_1, y_2) \in A_2 \cup A_3, \\ 0, & \text{if } (y_1, y_2) \notin A_1 \cup A_2 \cup A_3, \end{cases}$$

where sets A_1, A_2 and A_3 are described as follows:
$A_1 = \Big\{(y_1, y_2) \mid y_1 \in [0,1] \wedge y_2 > \frac{0.5-y_1}{b} \wedge$
$\wedge y_2 < \frac{b-0.5+y_1}{b} \wedge y_2 > \frac{y_1-0.5}{b} \wedge y_2 < \frac{b+0.5-y_1}{b}\Big\}$,
$A_2 = \Big\{(y_1, y_2) \mid y_1 \in [0,1] \wedge y_2 > 0 \wedge y_2 < \frac{0.5-y_1}{b}\Big\}$,
$A_3 = \Big\{(y_1, y_2) \mid y_1 \in [0,1] \wedge y_2 < 1 \wedge y_2 > \frac{b+0.5-y_1}{b}\Big\}$
(see fig. 2 for the support of that joint probability density function).

This joint PDF was found the way similar to that finding the one-dimensional PDF of $g^{**}(y)$. First we found the joint probability density function of $Y(t-1)$ and $Y(t) = Y(t-1) + bX(t)$, $b \geq 1$, and then altered it according to manipulations being made.

B. Correlation properties

As we know the joint probability density function $g(y_1, y_2)$ of the two adjacent members of the simulated sequence, we can calculate the correlation coefficient $R(1)$ between them.

$$R(1) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 y_2 g(y_1, y_2) dy_2 dy_1 - E[Y(t-1)]E[Y(t)]}{\sqrt{D[Y(t-1)]D[Y(t)]}}.$$

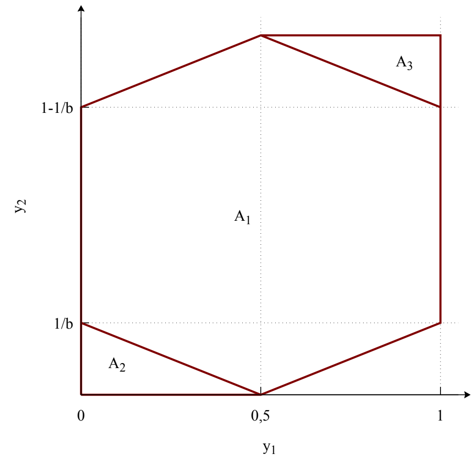

Fig. 2. Support of joint probability distribution function $g(y_1, y_2)$

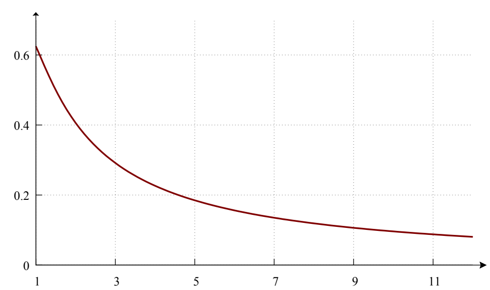

Fig. 3. Dependence of $R(1)$ on parameter b

As each sequence member has a standard uniform distribution, its expected value and dispersion are known $E[Y(t-1)] = E[Y(t)] = 0.5$, $D[Y(t-1)] = D[Y(t)] = \frac{1}{12}$, and we also know the joint probability density function $g(y_1, y_2)$. Therefore after finding the integral's value we get

$$R(1) = \left(\frac{1}{4} + \frac{1}{12b} - \frac{1}{32b^2} - \frac{1}{4}\right) \cdot 12 = \frac{1}{b} - \frac{3}{8b^2}.$$

As the value of parameter b is chosen to be greater or equal than 1, the maximum value of $R(1)$ is 0.625. See figure 3 for the dependence of $R(1)$ on b.

That dependence is easily reversible, which is useful when it is necessary to determine the value of parameter b needed for the required value of $R(1)$:

$$b = \frac{1}{2 \cdot R(1)} + \sqrt{\frac{1}{4 \cdot (R(1))^2} - \frac{3}{8 \cdot R(1)}}$$

.

As the one-dimensional PDF is symmetrical relative to the point 0.5, taking the value $(1-y)$ instead of y as the next member of the sequence won't change the uniformity of the distribution, but the sign of correlation coefficient will be changed to the opposite.

As the simulated sequence is stationary, the rest of values of autocorrelation function can be found as $R(\tau) = (R(1))^{|\tau|}$.

IV. Recap of algorithm

If $|R(1)| \in (0, 0.625]$, then the algorithm of simulating the random sequence with exponential correlation function and standard uniform distribution of its members is as follows:

1. $b := 0.5|R(1)|^{-1} + \sqrt{0.25(R(1))^{-2} - 0.375|R(1)|^{-1}}$.
2. $y_0 := U[0,1]$.
3. $y_t^* := y_{t-1} + bx_t$.
4. If $y_t^* \in [0, 0.5]$, then $y_t^{**} := 1 - y_t^*$. If $y_t^* \in [b+0.5, b+1]$, then $y_t^{**} := 2b_t^* + 1$.
5. If $R(1) > 0$, then $y_t := \frac{y_t^{**} - 0.5}{b}$, else $y_t := 1 - \frac{y_t^{**} - 0.5}{b}$.
6. Repeat steps 3–5.

Step 1 is finding the value of algorithm's parameter corresponding to the required value of $R(1)$. Step 2 is made to get the first member of the sequence. Main part of the algorithm is loop described in step 3–5. Manipulating intervals of probability density function takes place in step 4. Step 5 includes check for negativity of $R(1)$, if it is so, an additional linear transformation takes place to meet such a requirement.

Thus the next member of the sequence is derived from the previous as either linear combination of latter and an independent uniformly distributed random variable or a linear function of such a linear combination. It is denoted as y_t in step 5.

V. Example

Let us simulate a sequence with uniform distribution and $R(1) = 0.42$.

As $R(1) = 0.42 < 0.625$, we can apply manipulating intervals of PDF approach to get the desired sequence. Let us find the parameter b

$$b = \frac{0.5}{0.42} + \sqrt{\frac{0.25}{0.42^2} - \frac{0.375}{0.42}} \approx 1.9146.$$

We get the first member of the sequence y_0 as a uniform random variable in $[0,1]$, and then apply the algorithm from section IV to get the sequence of desired length.

After 10000 iterations of the described algorithm we get a sequence of numbers with sample mean 0.4951, sample variance 0.0834, and sample correlation between adjacent members 0.4241. Expected values for standard uniform distribution and chosen parameters are $0.5, 0.83(3)$ and 0.42 respectively.

When we plot the points with coordinates (y_{t-1}, y_t) we will get the visualization of the joint PDF (see fig. 4). As we can see it looks as we predicted in section III.

VI. Using inverse transform sampling

The general method for modeling a random variable with a specified distribution is the inverse transform sampling. As $Y(t)$ has a uniform distribution of $[0,1]$ and if $\tilde{X}$ has a cumulative distribution function $\tilde{F}(x)$, then the random variable $\tilde{F}^{-1}(Y(t))$ has the same distribution as $\tilde{X}$. Correlation coefficient will be greater in resulting sequence if it was greater in the original sequence [6].

Joint cumulative distribution function of the adjacent transformed variables $\tilde{X}(t-1)$ and $\tilde{X}(t)$ is

$$\tilde{F}(x_1, x_2) = G(\tilde{F}(x_1), \tilde{F}(x_2)).$$

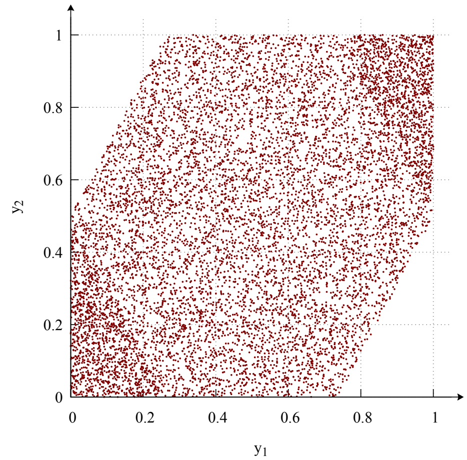

Fig. 4. Points $(y_{t-1}, y_t), t = 1, 2, \cdots, 10000$ of simulated random sequence, $R(1) = 0.42$

Therefore joint probability density function is

$$\tilde{F}(x_1, x_2) = \frac{\partial^2 \tilde{F}(x_1, x_2)}{\partial x_1 \partial x_2} = \tilde{F}(x_1)\tilde{F}(x_2)g(\tilde{F}(x_1), \tilde{F}(x_2)).$$

For the sequences obtained using described method

$$\tilde{F}(x_1, x_2) = \begin{cases} \tilde{f}(x_1) \cdot \tilde{f}(x_2), & \text{if } (x_1, x_2) \in A_1, \\ 2\tilde{f}(x_1) \cdot \tilde{f}(x_2), & \text{if } (x_1, x_2) \in A_2, \\ 0, & \text{if } (x_1, x_2) \notin A_1 \cup A_2, \end{cases}$$

where areas $\tilde{A}_1$ and $\tilde{A}_2$ are

$$\begin{aligned} \tilde{A}_1 = \Big\{ &(x_1, x_2) \mid x_1 \in [\tilde{F}^{-1}(0), \tilde{F}^{-1}(1)] \wedge \\ &\wedge x_2 > \tilde{F}^{-1}\left(\tfrac{0.5 - x_1}{b}\right) \wedge x_2 < \tilde{F}^{-1}\left(\tfrac{b - 0.5 + x_1}{b}\right) \wedge \\ &\wedge x_2 > \tilde{F}^{-1}\left(\tfrac{x_1 - 0.5}{b}\right) \wedge x_2 < \tilde{F}^{-1}\left(\tfrac{b + 0.5 - x_1}{b}\right) \Big\}, \end{aligned}$$

$$\begin{aligned} \tilde{A}_2 = \Big\{ &(x_1, x_2) \mid x_1 \in [\tilde{F}^{-1}(0), \tilde{F}^{-1}(1)] \wedge \\ &\wedge x_2 \in [\tilde{F}^{-1}(0), \tilde{F}^{-1}(1)] \wedge (x_2 < \tilde{F}^{-1}\left(\tfrac{0.5 - x_1}{b}\right) \vee \\ &\vee x_2 > \tilde{F}^{-1}\left(\tfrac{b + 0.5 - x_1}{b}\right)) \Big\}. \end{aligned}$$

So the correlation coefficient between the adjacent members of the sequence is

$$\begin{aligned} R(1) = \Bigg(&\int_{\tilde{F}^{-1}(0)}^{\tilde{F}^{-1}(0.5)} x_1 \tilde{f}(x_1) \left(\int_{\tilde{F}^{-1}(0)}^{\tilde{F}^{-1}\left(\frac{b-0.5+x_1}{b}\right)} x_2 \tilde{f}(x_2) dx_2 + \right. \\ &\left. + \int_{\tilde{F}^{-1}(0)}^{\tilde{F}^{-1}\left(\frac{0.5-x_1}{b}\right)} x_2 \tilde{f}(x_2) dx_2 \right) dx_1 + \\ &+ \int_{\tilde{F}^{-1}(0.5)}^{\tilde{F}^{-1}(1)} x_1 \tilde{f}(x_1) \left(\int_{\tilde{F}^{-1}\left(\frac{x_1-0.5}{b}\right)}^{\tilde{F}^{-1}(1)} x_2 \tilde{f}(x_2) dx_2 + \right. \\ &\left. + \int_{\tilde{F}^{-1}\left(\frac{b+0.5-x_1}{b}\right)}^{\tilde{F}^{-1}(1)} x_2 \tilde{f}(x_2) dx_2 \right) dx_1 - \\ &- E^2[\tilde{X}(t)] \Bigg) D^{-1}[\tilde{X}(t)]. \end{aligned}$$

It is important to know the maximum values of $R(1)$ that can be obtained for certain distributions. They can be found if we put $b = 1$.

VII. Conclusion

Suggested method is a modification of an algorithm based on $AR(1)$ model. It allows simulating of strictly stationary random sequences with standard uniform distribution and specified value of autocorrelation function $R(1)$ in $[-0.625, 0.625]$.

As the uniform distribution is a base for inverse transform sampling method, the proposed approach approach may be used to simulate random sequences with wide range of probability distributions, which is a merit of proposed approach. Ease of implementation is also an advantage of the proposed method.

Main disadvantage is the inapplicability for $AR(n)$ models with $n > 1$ as it this case the resulting sequence's PDF won't correspond to uniform distribution. The other disadvantage is inability to get $R(1) > 0.625$ or $R(1) < 0.625$.

Computer modeling confirms the theoretical calculations and predictions of this paper.

Further work may include application of manipulating intervals of PDF approach for simulating random fields, or use of manipulating intervals of PDF in conjunction with the autoregression algorithms with stochastic binary orthogonal coefficients for getting the autoregression model with higher order, but more suitable joint PDF than in the base method.

References

[1] V.A. Ogorodnikov and S.M. Prigarin, *Numerical Modelling of Random Processes and Fields: Algorithms and Applications*. VSP, Utrecht, the Netherlands, 1996, 240 p.

[2] B.F. Kirianov, *Fundamentals of stochastic computers and devices theory*. KAI, Kazan, USSR, 1975, 167 p. [in Russian]

[3] S.V. Garbar, *Using autoregression for modeling of Markov random processes with specified probability density* Vestn. Novg. gos. un-ta, 67, Velikiy Novrorod, Russia, 2012, p. 13-15. [in Russian]

[4] S.V. Garbar, *Manipulating intervals of probability density function for modeling sequences with uniform distribution* Vestn. Novg. gos. un-ta, 60, Velikiy Novrorod, Russia, 2010, p. 27-28 [in Russian]

[5] S.V. Garbar, *Joint probability density functions of sequences simulated using autoregressive algorithms* Review of applied and industrial maths, 2013, Vol. 20, p. 536-537 [In Russian]

[6] Devroye L., *Non-uniform random variate generation*. Springer-Verlag, New York, 1986, 806 p.