Using AR(1) model to simulate strictly stationary random sequences

Sergey Garbar
KPMI, Novgorod State University, ul. B. St. Petersburgskaya, 41 173003 Veliky Novgorod, Russia
Sergey.Garbar@novsu.ru

Abstract. We consider using AR(1) model for simulating strictly stationary random sequences and propose a method finding corresponding PDF of independent random variable in the model. Random sequences with normal, uniform, exponential and Erlang distributions are the considered examples of using that algorithm.

1. Introduction
Simulated random processes can be used to model objects and phenomena of different nature. Stationary random processes have an important role for describing time series with parameters such as the mean, variance and autocorrelation structure that do not change over time and do not follow any trends.

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

\[ Y(t) = a_1 Y(t - 1) + a_2 Y(t - 2) + \cdots + a_n Y(t - n) + bX(t) \]

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. This model can be considered to be an algorithm that transforms a sequence \( \{X(t)\} \) of independent identically distributed random variables \( X(t) \) into a sequence \( \{Y(t)\} \) with some non-zero correlation function \( R(\tau) \).

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 is a downside of that approach.

In [2] author proposes to simulate strictly stationary random sequences using autoregression coefficients 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.

Below we consider autoregression model of first order \( AR(1) \):

\[ Y(t) = aY(t - 1) + bX(t), \]

where \( a, b \geq 0, a + b = 1 \). With these limitations the simulated sequence will be stationary in wide sense. Correlation coefficient between the adjacent sequence members is \( R(1) = a \). Correlation function is exponential.
The task of this article is finding the probability density function (PDF) $f(x)$ of added random variables $X(t)$ such that the PDF of next sequence member $Y(t)$ will be a preassigned $g(y)$. It continues the work described in [3]. To examine the properties of generated sequences the joint probability density functions of adjacent sequence members are found.

2. Using AR(1) model to simulate a strictly stationary random sequence

Note that random variables $Y(t - 1)$ and $X(t)$ are independent, therefore PDF for their linear combination used in AR(1) model $Y(t) = aY(t - 1) + bX(t)$ can be found fairly straightforward. As we require strict stationarity of resulting sequence, PDFs of $Y(t - 1)$ and $Y(t)$ are the same, and we can write a relation that can be solved as an integral equation on $f(x)$

$$g(y) = \int_{-\infty}^{+\infty} g \left( \frac{y - bx}{a} \right) f(x) \frac{1}{a} \, dx.$$ 

That equation can be solved directly for some PDFs $g(y)$, but the use of characteristic function properties gives a convenient solution for some PDFs.

Let $\psi(t)$ be a characteristic function of $Y(t - 1)$ and $Y(t)$ and $\varphi(t)$ — a characteristic function of $X(t)$. As random variables $aY(t - 1)$ and $bX(t)$ are independent, their product’s characteristic function is a product of corresponding characteristic functions

$$\psi(t) = \psi(at) \cdot \varphi(bt).$$

Therefore a characteristic function $\varphi(t)$ of $X(t)$ can be expressed as

$$\varphi(bt) = \frac{\psi(t)}{\psi(at)}.$$ 

Having this expression we can write out a relation for $f(x)$ in a form

$$\frac{1}{b} \cdot f \left( \frac{x}{b} \right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{\psi(t)}{\psi(at)} \, dt.$$ 

If $g(y)$ is symmetrical relatively to some $x_0$, taking $(2y_0 - Y(t))$ instead of $Y(t)$ won’t change the PDF, but the sign of correlation coefficient $\text{corr}(Y(t - 1), 2y_0 - Y(t))$ will be changed. That way we can simulate random sequences with negative $R(1)$ in certain cases. Important example of such distribution is a uniform distribution.

2.1. Examples

Shown below are some examples of finding PDF for added random variable in AR(1) model-based algorithm for simulating strictly stationary random sequences.

2.1.1. Normal distribution is one of the most important probability distributions due to central limit theorem.

PDF of normal distribution is

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x - m)^2}{2\sigma^2}}.$$ 

Below we consider standard normal distribution with parameters $m = 0$ and $\sigma = 1$ as other their values can be attained with a linear transformation that does not alter the correlation properties.

When using AR(1) model to simulate random sequence with standard normal distribution, added random variable $X(t)$ should have PDF
which is a PDF for random variable with normal distribution with $m = 0$ and $\sigma = \frac{1+a}{\sqrt{1-a}}$.

That trivial statement can be obtained if we consider the characteristic function of the normal distribution

$$\varphi(t) = e^{-\frac{t^2}{2}}.$$

Therefore

$$\frac{1}{b} f \left( \frac{x}{b} \right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{e^{\frac{t^2}{2}}}{e^{\frac{a^2t^2}{2}}} dt.$$

That gives us the mentioned before statement.

2.1.2. Uniform distribution is denoted $X \sim U[x_1; x_2]$ for $x_1 < x_2$ and has PDF.

$$f(x) = \begin{cases} 
\frac{1}{x_2 - x_1}, & \text{при } x \in [x_1; x_2), \\
0, & \text{при } x \notin [x_1; x_2]. 
\end{cases}$$

Below we consider the standard uniform distribution with $x_1 = 0$ и $x_2 = 1$. It is considered to be a basic distribution in random variables modeling as it can be used in inverse transform sampling method.

If $a = 1/(n + 1), n \in \mathbb{N}$ then the added random variable $X(t)$ in AR(1)-based algorithm for simulating random sequences with standard uniform distribution should have PDF

$$f(x) = a \sum_{k=0}^{n} \delta \left( x - \frac{k}{n} \right) = \frac{1}{n+1} \sum_{k=0}^{n} \delta \left( x - \frac{k}{n} \right),$$

where $\delta(x)$ is Dirac’s delta function. That way each step of algorithm supposes to add an independent random variable with discrete uniform distribution in points $0, 1/n, 2/n, ..., (n-1)/n, 1$.

To find the given PDF we shall consider the characteristic function of standard uniform distribution.

$$\psi(t) = \frac{e^{it} - 1}{it}.$$

Therefore the characteristic function for $bX(t)$ is

$$\varphi(bt) = \frac{e^{ibt} - 1}{ibt} = \frac{e^{it} - 1}{e^{iat} - 1}.$$

As we only consider $a = 1/(n + 1)$

$$\varphi(bt) = \left(\frac{e^{iat}}{e^{iat} - 1}\right)^{n+1} - 1 = e^{iat} + \left(e^{iat}\right)^{n-1} + \cdots + e^{iat} + 1.$$

And PDF for $X(t)$ is
\[ \frac{1}{b} \cdot f \left( \frac{x}{b} \right) = \delta(x - an) + \delta(x - a(n-1)) + \cdots + \delta(x - a) + \delta(x). \]

That relation leads to the PDF given above.

Joint PDF of adjacent sequence members \( Y(t) \) and \( Y(t+1) \) is

\[
g(y_1, y_2) = \sum_{k=0}^{n} \delta \left( \frac{y_2}{a} - y_1 - ka \right), y_1 \in [0; 1] \text{ и } y_2 \in [0; 1],
\]

\[ 0, \ y_1 \notin [0; 1] \text{ или } y_2 \notin [0; 1]. \]

As one can see, the support of the joint PDF is a set of line segments

\[ y_2 = ay_1 + ka^2, k = 0..n, \]

therefore the simulated sequence might be not suitable for simulating some processes. In [4] the AR(1) algorithm was modified so that it allowed simulating strictly stationary random sequences with standard uniform distribution and specified value of autocorrelation function \( R(1) \) in \([-0.625, 0.625]\).

Example of support for \( a = 0.2 \) is shown on figure 1. Figure 2 shows 1000 points with the coordinates \((Y(t); Y(t+1))\), where \( Y(t) \) and \( Y(t+1) \) are the members of the simulated sequence.

![Figure 1](image1.png)  
**Figure 1.** Support of joint PDF of \( Y(t) \) and \( Y(t+1) \) for \( a = 0.2 \)

![Figure 2](image2.png)  
**Figure 2.** Location of 1000 members of simulated sequence as points with coordinates \((Y(t); Y(t+1))\) for \( a = 0.25 \)

2.1.3. **Exponential distribution.** The PDF has one parameter \( \lambda > 0 \).

\[
g(y; \lambda) = \begin{cases} \lambda e^{-\lambda y}, & y \geq 0, \\ 0, & y < 0. \end{cases}
\]

When we use AR(1) model to simulate random sequence with exponential distribution, the PDF of \( X(t) \) is
where $\delta(x)$ is Dirac’s delta function, and $g(x; \lambda b)$ is a PDF of exponential distribution with parameter $\lambda b$ ($b > 0$):

$$g(x; \lambda b) = \begin{cases} \lambda b e^{-\lambda bx}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Therefore each step of simulating algorithm supposes to add an independent random variable, which is a mixture of exponential random variable and non-random value $X = 0$ with corresponding weights $b$ and $a$.

That form of PDF can be obtained if we consider the characteristic function of exponential distribution

$$\psi(t) = \left(1 - \frac{it}{\lambda}\right)^{-1}.$$ 

Therefore

$$\frac{1}{b} f\left(\frac{x}{b}\right) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \frac{1 - ita}{1 - \frac{it}{\lambda}} \, dt = a \cdot \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \, dt + \frac{1}{2\pi} \int_{-\infty}^{+\infty} b \left(1 - \frac{it}{\lambda}\right)^{-1} e^{-itx} \, dt.$$ 

Which leads to the given above PDF.

Joint PDF of adjacent sequence members $Y(t)$ and $Y(t + 1)$ is

$$g(y_1, y_2) = \begin{cases} \lambda e^{-\lambda y_1} (a \delta(a y_1 - y_2) + \lambda b e^{\lambda (a y_1 - y_2)}), & y_1 \geq 0 \text{ and } y_2 \geq a y_1, \\ 0, & y_1 < 0 \text{ or } y_2 < a y_1. \end{cases}$$

2.1.4. **Erlang distribution** is a two parameter family of continuous probability distributions with the PDF

$$f(x; k, \lambda) = \begin{cases} \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Parameter $\lambda > 0$ is the “rate” and $k \in \mathbb{N}$ is the “shape.

The Erlang distribution with shape parameter $k = 1$ is the exponential distribution.

The chi-squared distribution with even number of degrees of freedom is a special case of the Erlang distribution.

It is a special case of the Gamma distribution. It is the distribution of a sum of $k$ independent exponential variables with mean $1/\lambda$ each.

To simulate a random sequence with the Erlang distribution using AR(1) model-based algorithm, one should on each step add a random variable with PDF

$$f(x) = a^k \delta(x) + \sum_{i=1}^{k} \binom{k}{i} a^{k-i} b^i g(x; i, \lambda b).$$

where $\delta(x)$ is Dirac’s delta function, $\binom{k}{i} = \frac{k!}{i!(k-i)!}$ are binomial coefficients, and $g(x; i, \lambda b)$ are PDFs of Erlang distribution with parameters $i$ and $\lambda b$:

$$g(x; i, \lambda b) = \begin{cases} (\lambda b)^i x^{i-1} e^{-\lambda bx}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$
Characteristic function of Erlang distribution is
\[ \psi(t) = \left(1 - \frac{it}{\lambda}\right)^{-k}. \]

Therefore
\[ \frac{1}{b} f(\frac{x}{b}) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \left(\frac{1 - \frac{it}{\lambda}}{1 - \frac{it}{\lambda}}\right)^{k} dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \left(\frac{b}{1 - \frac{it}{\lambda}} + a\right)^{k} dt. \]

As \( k \) is a positive integer, we can use the binomial formula
\[ f\left(\frac{x}{b}\right) = b \cdot \frac{1}{2\pi} \left( \int_{-\infty}^{+\infty} a^{k} e^{-itx} dt + \int_{-\infty}^{+\infty} e^{-itx} \sum_{i=1}^{k} \frac{c_{i} b^{i} a^{k-i}}{t^{i}} dt \right). \]

After the Fourier transform we get the given above PDF, which is a mixture of non-random value and \( k \) Erlang-distributed random variables.

3. Results and discussion
Formula for finding the PDF of added random variable for AR(1) model-based algorithm was found. Examples the PDFs of added random variables for AR(1) model-based algorithm were examined. Joint PDFs of adjacent members were also found.

Discussed methods of simulation are easy to implement. Computer modeling confirms the theoretical calculations.

Further work may include using inverse transformation to the members of obtained sequences to get the sequences with different PDFs and different joint PDFs of adjacent members.

4. Acknowledgments
The reported research was funded by Russian Foundation for Basic Research and Novgorod region, grant №18-41-530001.

References
[1] Ogorodnikov V A and Prigarin S M 1996 Numerical Modelling of Random Processes and Fields Algorithms and Applications Utrecht VSP p 240
[2] Kir'yanov B F and Koznov A V 1994 Protsessy avtoregressii so sluchaynymi koeffitsi-entami i ikh primenenie pri modelirovanii radiotekhnicheskikh sistem Prikladnaya matema-tika: Mezhvuz. cb. Novgorod 1 pp 3–8 (In Russian)
[3] Garbar S V 2016 Manipulating intervals of probability density function in autoregression model for simulating strictly stationary random sequences International Journal of Mathematical Models and Methods in Applied Sciences 10 pp 270-273
[4] Garbar' S V 2012 Modelirovanie stationarnykh markovskikh sluchaynykh protsessov s zadannoy plotnost'yu raspredeleniya metodom avtoregressii Vestn. Novg gos un-ta 67 pp 13-15 (In Russian)