STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **LI**, Number 2, June 2006

A NEW MONTE CARLO ESTIMATOR FOR SYSTEMS OF LINEAR EQUATIONS

NATALIA ROŞCA

Abstract. We propose a new Monte Carlo estimator to solve systems of linear equations. We formulate and prove some results concerning the quality and the properties of this estimator. Using this estimator, we give error bounds and construct confidence intervals for the components of the solution. We also consider numerical examples. The numerical results indicate that the proposed estimator converges faster than another two estimators from the literature.

1. **Introduction**

Let us consider the system of linear algebraic equations:

$$x = Tx + c, \tag{1}$$

where $T = (t_{ij})_{i,j=1}^{n} \in \mathbb{R}^{n\times n}$, $c = (c_1, \dots, c_n)^t \in \mathbb{R}^n$ and $I - T$ is an invertible matrix. The solution $x = (x_1, \dots, x_n)^t \in \mathbb{R}^n$ of system (1) is unique and admits the Neumann series representation:

$$x = c + Tc + T^2c + T^3c + \dots$$

or, detailed,

$$x_i = c_i + (Tc)_i + (T^2c)_i + \dots, \quad i = 1, \dots, n. \tag{2}$$

We assume that $\sum_{j=1}^{n} |t_{ij}| < 1$, $i = 1, \dots, n$, which is a sufficient condition for the convergence of Neumann series to the solution.

Received by the editors: 10.01.2006.
2000 *Mathematics Subject Classification.* 65C05.
Key words and phrases. Monte Carlo method, linear systems, estimation.

Monte Carlo methods estimate the solution of system (1), by constructing unbiased estimators for the components of the solution (see [4], [5], [10]). Let $P = (p_{ij})_{i,j=1}^{n+1} \in \mathbb{R}^{(n+1)\times(n+1)}$ be a matrix, whose elements satisfy the conditions:

1. $p_{ij} \geq 0$ such that $t_{ji} \neq 0 \Longrightarrow p_{ij} \neq 0$,
2. $\sum_{j=1}^{n} p_{ij} \leq 1, \quad i = 1, \ldots, n,$
3. $p_{i,n+1} = 1 - \sum_{j=1}^{n} p_{ij}, \quad i = 1, \ldots, n,$
4. $p_{n+1,j} = 0,\ j < n+1$,
5. $p_{n+1,n+1} = 1.$

The notation p_i is also used to denote $p_{i,n+1}$. The matrix P describes a Markov chain with the set of states $\{1, \ldots, n+1\}$, where $n+1$ is an absorbing state and p_{ij}, $i, j = 1, \ldots, n+1$, is the one step transition probability from state i to state j.

Define the weights:

$$w_{ij} = \begin{cases} \frac{t_{ji}}{p_{ij}} & \text{if } p_{ij} \neq 0 \\ 0 & \text{if } p_{ij} = 0 \end{cases}, \quad i, j = 1, \ldots, n.$$

Denote by $\gamma = (i_0, i_1, \ldots, i_k, n+1)$ a trajectory that starts at the initial state $i_0 < n+1$ and passes successfully through the sequence of states $(i_1, \ldots, i_k)$, to finally get into the absorbing state $i_{k+1} = n+1$.

Consider a vector $\alpha = (\alpha_1, \ldots, \alpha_n)$, where α_i, $i = 1, \ldots, n$, is the probability that a trajectory starts in state i, i.e.,

$$P(i_0 = i) = \alpha_i, \qquad \alpha_i \geq 0,\ i = 1, \ldots, n, \qquad \sum_{i=1}^{n} \alpha_i = 1.$$

The probability to follow trajectory γ is $P(\gamma) = \alpha_{i_0} p_{i_0 i_1} \ldots p_{i_{k-1} i_k} p_{i_k}$.

Define the estimators θ_i, $i = 1, \ldots, n$, and λ_i, $i = 1, \ldots, n$, on the space of trajectories as follows. For a trajectory $\gamma = (i_0, i_1, \ldots, i_k, n+1)$, the values of these estimators are defined as:

$$\theta_i(\gamma) = W_k(\gamma) \frac{\delta_{i_k i}}{p_{i_k}}, \qquad \lambda_i(\gamma) = \sum_{m=0}^{k} W_m(\gamma) \delta_{i_m i}, \qquad i = 1, \ldots, n,$$

where $W_m,\ m = 0, \dots, k$, are random variables whose values are:

$$\begin{aligned} W_0(\gamma) &= \frac{c_{i_0}}{\alpha_{i_0}}, \\ W_m(\gamma) &= W_{m-1}(\gamma) w_{i_{m-1} i_m} \\ &= \frac{c_{i_0}}{\alpha_{i_0}} w_{i_0 i_1} w_{i_1 i_2} \dots w_{i_{m-1} i_m}, \qquad m = 1, \dots, k. \end{aligned}$$

These values are taken with probability $P(\gamma)$ (δ_{ij} is the Kronecker symbol, i.e., $\delta_{ij} = 1$ if $i = j$ and 0 otherwise).

It is proved in [8] that θ_i and λ_i are unbiased estimators of x_i, i.e., $E(\theta_i) = E(\lambda_i) = x_i,\ i = 1, \dots, n$.

For some particular systems, the variances of the estimators θ_i and λ_i are analytically compared in [6]. In [7], the complexity of the Monte Carlo method is calculated, when certain techniques to generate the trajectories of the Markov chain are used.

2. A new estimator

Definition 1. *We define the estimator U_i, $i = 1, \dots, n$, on the space of trajectories as follows. For an arbitrary trajectory $\gamma = (i_0, i_1, \dots, i_k, n+1)$, the value of U_i is defined as:*

$$U_i(\gamma) = c_i + W_k(\gamma) \frac{t_{i i_k}}{p_{i_k}}, \qquad i = 1, \dots, n,$$

and is taken with probability $P(\gamma) = \alpha_{i_0} p_{i_0 i_1} \dots p_{i_{k-1} i_k} p_{i_k}$.

Remark 2. *The distribution of the estimator U_i, $i = 1, \dots, n$, is:*

$$U_i : \begin{pmatrix} c_i + W_k(\gamma) \frac{t_{i i_k}}{p_{i_k}} \\ \\ \alpha_{i_0} p_{i_0 i_1} \dots p_{i_{k-1} i_k} p_{i_k} \end{pmatrix}_{\substack{\gamma = (i_0, i_1, \dots, i_k, n+1) \\ i_0, i_1, \dots, i_k = 1, \dots, n}} .$$

Next, we formulate and prove some main results concerning the quality and the properties of the estimator U_i.

Theorem 3. *The expectation of U_i is equal to the component x_i of the solution of system (1), i.e.,*

$$E(U_i) = x_i, \quad i = 1, \dots, n. \tag{3}$$

In other words, U_i is an unbiased estimator of x_i, $i = 1, \ldots, n$.

Proof. We can write:

$$\begin{aligned}
E(U_i) &= \sum_{\gamma=(i_0,\ldots,i_k,n+1)} U_i(\gamma)P(\gamma) \\
&= \sum_{\gamma=(i_0,\ldots,i_k,n+1)} \left(c_i + W_k(\gamma)\frac{t_{ii_k}}{p_{i_k}}\right)P(\gamma) \\
&= \sum_{\gamma=(i_0,\ldots,i_k,n+1)} c_iP(\gamma) + \sum_{\gamma=(i_0,\ldots,i_k,n+1)} W_k(\gamma)\frac{t_{ii_k}}{p_{i_k}}P(\gamma) \\
&= c_i + \sum_{\gamma=(i_0,\ldots,i_k,n+1)} \frac{c_{i_0}}{\alpha_{i_0}} w_{i_0i_1}\ldots w_{i_{k-1}i_k}\frac{t_{ii_k}}{p_{i_k}}\alpha_{i_0}p_{i_0i_1}\ldots p_{i_{k-1}i_k}p_{i_k} \\
&= c_i + \sum_{\gamma=(i_0,\ldots,i_k,n+1)} c_{i_0}\frac{t_{i_1i_0}}{p_{i_0i_1}}\ldots\frac{t_{i_ki_{k-1}}}{p_{i_{k-1}i_k}}t_{ii_k}p_{i_0i_1}\ldots p_{i_{k-1}i_k} \\
&= c_i + \sum_{k=0}^{\infty}\sum_{i_0=1}^{n}\ldots\sum_{i_k=1}^{n} t_{ii_k}t_{i_ki_{k-1}}\ldots t_{i_1i_0}c_{i_0} \\
&= c_i + (Tc)_i + (T^2c)_i + \ldots \\
&= x_i.
\end{aligned}$$

In the last equality, we used relation (2). □

Proposition 4. *The following relationship between the estimators U_i and θ_i holds:*

$$U_i = c_i + \sum_{j=1}^{n}\theta_j t_{ij}, \qquad i = 1, \ldots, n.$$

Proof. For any trajectory γ, we can write:

$$\begin{aligned}
U_i(\gamma) &= c_i + W_k(\gamma)\frac{t_{ii_k}}{p_{i_k}} = c_i + \sum_{j=1}^{n} W_k(\gamma)\frac{\delta_{i_kj}}{p_{i_k}}t_{ij} \\
&= c_i + \sum_{j=1}^{n}\theta_j(\gamma)t_{ij}, \qquad i = 1, \ldots, n.
\end{aligned}$$

□

Theorem 5. *The following relationship between the variance of U_i and the variance of θ_i holds:*

$$Var(U_i) = \sum_{j=1}^{n} t_{ij}^2 Var(\theta_j) + \sum_{j<l} 2t_{ij}t_{il}Cov(\theta_j, \theta_l). \tag{4}$$

Proof. Using the result from Proposition 4 and some known properties of the variance, we can write:

$$\begin{aligned} Var(U_i) &= Var\Big(c_i + \sum_{j=1}^{n} \theta_j t_{ij}\Big) \\ &= \sum_{j=1}^{n} Var\Big(t_{ij}\theta_j\Big) + \sum_{j<l} 2Cov(t_{ij}\theta_j, t_{il}\theta_l) \\ &= \sum_{j=1}^{n} t_{ij}^2 Var(\theta_j) + \sum_{j<l} 2t_{ij}t_{il}Cov(\theta_j, \theta_l). \end{aligned}$$

□

Practically, to solve system (1), we generate N independent trajectories $\gamma_1, \ldots, \gamma_N$ and for each trajectory we compute the value of the estimator U_i. The values $U_i(\gamma_j)$, $j = 1, \ldots, N$, are values of the sample variables $U_{i1}, \ldots, U_{iN}$ that are independent identically distributed random variables and have the same distribution as U_i.

We use the notation $\overline{U}_{i,N}$ for the sample mean of the random variables U_{ij}, $j = 1, \ldots, N$, and $\overline{u}_{i,N}$ for its value, i.e.:

$$\overline{U}_{i,N} = \frac{\sum_{j=1}^{N} U_{ij}}{N}, \qquad \overline{u}_{i,N} = \frac{\sum_{j=1}^{N} U_i(\gamma_j)}{N}. \tag{5}$$

Proposition 6. *The estimator $\overline{U}_{i,N}$, $i = 1, \ldots, n$, has the following properties:*

$$E(\overline{U}_{i,N}) = x_i, \quad \textit{(unbiased estimator of } x_i\textit{)}, \tag{6}$$

$$\lim_{N\to\infty} Var(\overline{U}_{i,N}) = 0, \tag{7}$$

$$P(\lim_{N\to\infty} \overline{U}_{i,N} = x_i) = 1, \quad (\overline{U}_{i,N} \textit{ converges almost surely to } x_i\textit{)}. \tag{8}$$

Proof. Properties (6) and (7) can be proved using known properties of the mean and variance. For property (8), we apply the Kolmogorov theorem ([1]) to the sequence of random variables $(U_{iN})_{N\geq 1}$ that are independent identically distributed and have finite means $E(U_{iN}) = x_i < \infty$. Under these conditions, the Kolmogorov theorem asserts that relation (8) is satisfied. $\square$

Taking into account these properties, the component x_i is approximated by:

$$x_i \approx \overline{u}_{i,N} = \frac{1}{N}\sum_{j=1}^{N} U_i(\gamma_j), \quad i = 1, \dots, n. \tag{9}$$

The estimate of the solution is:

$$x_U = \left[\frac{1}{N}\sum_{j=1}^{N} U_1(\gamma_j), \dots, \frac{1}{N}\sum_{j=1}^{N} U_n(\gamma_j)\right]^t. \tag{10}$$

Similar estimates x_θ and x_λ can be obtained by replacing the estimator U_i, $i = 1, \dots, n$, by θ_i and λ_i respectively, i.e.,

$$x_\theta = \left[\frac{1}{N}\sum_{j=1}^{N} \theta_1(\gamma_j), \dots, \frac{1}{N}\sum_{j=1}^{N} \theta_n(\gamma_j)\right]^t, \tag{11}$$

$$x_\lambda = \left[\frac{1}{N}\sum_{j=1}^{N} \lambda_1(\gamma_j), \dots, \frac{1}{N}\sum_{j=1}^{N} \lambda_n(\gamma_j)\right]^t. \tag{12}$$

Remark 7. *The variance $Var(U_i)$ is in general unknown. It can be estimated using an unbiased estimation of it, given by the sample variance:*

$$\overline{\sigma}^2_{U,i} = \frac{1}{N-1}\sum_{j=1}^{N}(U_{ij} - \overline{U}_{i,N})^2. \tag{13}$$

Remark 8. *Comparing the variances of estimators U_i and θ_i can be done either analytically (using, eventually, the result from Theorem 5) or experimentally. Experimentally, we can use the same N generated trajectories γ_j, $j = 1, \dots, N$, and compute the values $\theta_i(\gamma_j)$, $j = 1, \dots, N$. Let $\theta_{i1}, \dots, \theta_{iN}$ be the corresponding sample variables. We use the same notation $\overline{\theta}_{i,N}$ for the sample mean of the random variables θ_{ij}, $j = 1, \dots, N$, and respectively for its value, i.e.,*

$$\overline{\theta}_{i,N} = \frac{\sum_{j=1}^{N}\theta_{ij}}{N}, \qquad \overline{\theta}_{i,N} = \frac{\sum_{j=1}^{N}\theta_i(\gamma_j)}{N}.$$

We estimate $Var(\theta_i)$ by the following unbiased estimator:

$$\overline{\sigma}^2_{\theta,i} = \frac{1}{N-1}\sum_{j=1}^{N}(\theta_{ij} - \overline{\theta}_{i,N})^2.$$

Comparing the variances $Var(U_i)$ and $Var(\theta_i)$ reduces to comparing their estimations $\overline{\sigma}^2_{U,i}$ and $\overline{\sigma}^2_{\theta,i}$.

3. **Error estimation**

We evaluate (estimate) the error in formula (9). One way of doing this is by using the Chebyshev inequality ([1]). We have the following main result concerning the error:

Proposition 9. *The following estimation of the error of approximation of x_i holds:*

$$P\left(\left|\overline{U}_{i,N} - x_i\right| < \frac{\sigma(U_i)}{\sqrt{N\gamma}}\right) \geq 1-\gamma, \quad \gamma \in (0,1),$$

where $\sigma(U_i)$ is the standard deviation of U_i, i.e. $\sigma^2(U_i) = Var(U_i)$.

Proof. The proof is immediately, by applying the Chebyshev inequality for the estimator $\overline{U}_{i,N}$ and choosing $\varepsilon = \frac{\sigma(U_i)}{\sqrt{N\gamma}}$. □

Another modality of estimating the error is based on the Lindeberg's limit theorem ([1]). In this case, we have the following main result:

Proposition 10. *The following estimation of the error of approximation of x_i holds:*

$$P\left(\left|\overline{U}_{i,N} - x_i\right| < \lambda\frac{\sigma(U_i)}{\sqrt{N}}\right) \approx 2\phi(\lambda) - 1, \quad \lambda > 0,$$

where

$$\phi(\lambda) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\lambda} e^{-\frac{t^2}{2}}\,dt,$$

is the Laplace function.

Proof. The proof is immediately, by applying the Lindeberg's limit theorem to the sequence of random variables $(U_{iN})_{N\geq 1}$ that are independent and identically distributed and have the same distribution as U_i. □

4. Confidence intervals

We construct confidence intervals for x_i, $i = 1, \dots, n$. We consider the confidence level $\alpha \in (0, 1)$.

Proposition 11. *A $(1-\alpha)\%$ confidence interval for x_i is:*

$$\left(\overline{U}_{i,N} - t_{N-1,1-\frac{\alpha}{2}} \frac{\overline{\sigma}_{U,i}}{\sqrt{N}}, \quad \overline{U}_{i,N} + t_{N-1,1-\frac{\alpha}{2}} \frac{\overline{\sigma}_{U,i}}{\sqrt{N}}\right). \tag{14}$$

where $\overline{U}_{i,N}$ is defined in (5), $t_{N-1,1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$-th percentile of the Student distribution with $N-1$ degrees of freedom, and $\overline{\sigma}_{U,i}$ is the sample standard deviation ($\overline{\sigma}^2_{U,i}$ is defined in (13)).

Proof. We consider the statistics:

$$T = \frac{\overline{U}_{i,N} - x_i}{\frac{\overline{\sigma}_{U,i}}{\sqrt{N}}},$$

that has the t (Student) distribution with $N-1$ degrees of freedom. We take $t_2 = t_{N-1,1-\frac{\alpha}{2}}$, $t_1 = -t_2$, i.e.,

$$F_{N-1}(t_2) = 1 - \frac{\alpha}{2}, \qquad F_{N-1}(t_1) = \frac{\alpha}{2},$$

where F_{N-1} is the distribution function of the t distribution with $N-1$ degrees of freedom. We have $P(t_1 < T < t_2) = 1 - \alpha$, which is equivalent to:

$$P\left(\overline{U}_{i,N} - t_{N-1,1-\frac{\alpha}{2}} \frac{\overline{\sigma}_{U,i}}{\sqrt{N}} < x_i < \overline{U}_{i,N} + t_{N-1,1-\frac{\alpha}{2}} \frac{\overline{\sigma}_{U,i}}{\sqrt{N}}\right) = 1 - \alpha.$$

Thus, a $(1-\alpha)\%$ confidence interval for x_i is given by (14). $\square$

5. Numerical example

We consider the system:

$$\begin{cases} x_1 = 0.1x_1 + 0.5x_2 + 0.4 \\ x_2 = 0.3x_1 + 0.1x_2 + 0.6 \end{cases}$$

with the exact solution $x = (1, 1)$.

We choose the matrix P of the following form:

$$P = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.5 & 0.1 & 0.4 \\ 0 & 0 & 1 \end{bmatrix}.$$

The matrix P describes a Markov chain with the set of states $\{1, 2, 3\}$, where state 3 is the absorbing one. As $p_{ij} = t_{ji}$, $i, j = 1, 2$, we have $w_{ij} = 1$, $i, j = 1, 2$. Since $c_1, c_2 \geq 0$ and $c_1 + c_2 = 1$, we take the vector $\alpha = c^t = (0.4, 0.6)$.

In order to get the initial state $i_0 \in \{1, 2\}$ of an arbitrary trajectory, we sample from the following discrete distribution:

$$Y_\alpha : \begin{pmatrix} 1 & 2 \\ \alpha_1 & \alpha_2 \end{pmatrix}.$$

Once the trajectory is in state $i_m = i \in \{1, 2\}$, we sample from the distribution:

$$Y_i : \begin{pmatrix} 1 & 2 & 3 \\ p_{i1} & p_{i2} & p_i \end{pmatrix},$$

described by the i-th line of matrix P, in order to determine the next state i_{m+1}. We repeat this procedure till absorbtion takes place. The sampling method is the inversion method ([2], [3]).

We generate N trajectories and we calculate the estimates x_θ, x_λ, x_U using formulas (11), (12) and (10), respectively. The following table contains: the number N of trajectories generated, the estimates x_θ, x_λ, x_U and the euclidian norm of the errors $\|x - x_\theta\|$, $\|x - x_\lambda\|$, $\|x - x_U\|$.

N	x_θ	x_λ	x_U	$\|x - x_\theta\|$	$\|x - x_\lambda\|$	$\|x - x_U\|$
5000	(0.9853 , 1.0220)	(0.9768, 0.9968)	(1.0095, 0.9978)	0.0264	0.0234	0.0098
10000	(0.9897, 1.0155)	(0.9859, 0.9948)	(1.0067, 0.9985)	0.0186	0.0150	0.0069
15000	(0.9939 , 1.0092)	(0.9875, 0.9930)	(1.0040, 0.9991)	0.0110	0.0144	0.0041
50000	(0.9945, 1.0083)	(0.9942, 0.9979)	(1.0036, 0.9992)	0.0100	0.0061	0.0037
100000	(0.9987 , 1.0020)	(0.9977, 0.9994)	(1.0009, 0.9998)	0.0024	0.0023	0.0009

The numerical results indicate that the proposed estimate x_U converges faster than the estimations x_θ and x_λ.

References

[1] Blaga, P., *Probability Theory and Mathematical Statistics*, II, Babes-Bolyai University Press, Cluj-Napoca, 1994 (In Romanian).

[2] Deak, I., *Random Number Generators and Simulation*, Akademiai Kiado, Budapest, 1990.

[3] Devroye, L., *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986.

[4] Forsythe, G.E., Leibler, R.A., *Matrix Inversion by a Monte Carlo Method*, Mathematical Tables and Other Aids to Computation 4 (1950), 127-129.

[5] Hammersley, J.M., Handscomb, D.C., *Monte Carlo Methods*, Methuen, London, 1964.

[6] Okten, G., *Solving Linear Equations by Monte Carlo Simulation*, SIAM Journal on Scientific Computing, Vol. 27, no. 2 (2005), 511-531.

[7] Roşca, N., *Monte Carlo Methods for Systems of Linear Equations*, Studia Univ. Babes-Bolyai, Mathematica, Vol. LI, no. 1 (2006), 103-114.

[8] Spanier, J., Gelbard, E.M., *Monte Carlo Principles and Neutron Transport Problems*, Addison-Wesley, 1969.

[9] Stancu, D.D., Coman, Gh., Blaga, P., *Numerical Analysis and Approximation Theory*, Vol. II, Presa Universitara Clujeana, Cluj-Napoca, 2002 (In Romanian).

[10] Wasow, W., *A Note on the Inversion of Matrices by Random Walks*, Mathematical Tables and Other Aids to Computation 6 (1952), 78-81.

BABEŞ-BOLYAI UNIVERSITY, FACULTY OF MATHEMATICS AND COMPUTER SCIENCE, STR. KOGALNICEANU 1, CLUJ-NAPOCA, ROMANIA
E-mail address: natalia@math.ubbcluj.ro