On the mathematical theory of splitting and Russian roulette techniques

Viatcheslav B. Melas
St.Petersburg State University, Russia

1. Introduction

- Splitting is an universal and potentially very powerful technique for increasing efficiency of simulation studies. The idea of this technique goes back to von Neumann (see [Kan and Harris,1951]). A theory of multilevel splitting is rather well developed (see e.g.[L'Ecuryer et al.2006]). However, in the most of papers strongly restrictive assumptions on transition probabilities are imposed.
- In [Melas, 1993, 1997, 2002] and [Ermakov and Melas, 1995] a more general theory is developed. It is based on the introduction of a probability measure governing the procedures of splitting and Russian roulette. In the present talk we will describe basic ideas and results of this theory.

2. Formulation of the problem

Let random variables $X_1, X_2, \ldots$ are defined on probability space $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ and form a homogenous Markov chaim with transition function $\mathcal{P}(x, dy) = P\{X_1 \in dy | X_0 = x\}$ measurable in the forst argument and such that $\int_{\mathcal{X}} \mathcal{P}(x, dy) = 1$. We will consider Harris positively recurrent Markov chains (see, f.e., (Nummelin, 1984) for definitions).

In fact, we need only a mild restrictions that secure existence of the stationary distribution (denote it by $\pi(dx)$) and asymptotic unbiasedness of usual Monte Carlo estimators.

It is required to estimate one or several functionals of the form

$$J_i = \int_{\mathcal{X}} h_i(x)\,\pi(dx), \quad i = 1, 2, \ldots, l$$

with respect to simulation results.
In case of chains with finite state space these functionals have the form

$$J_i = \sum_{x=0}^{m} h_i(x)\,\pi_x, \quad i = 1, 2, \ldots, l.$$

3. Branching technique

Let $\{X_n\}$ be a Markov chain of the general form described above. Let η_n chains $\{X_n\}$ be simulated at the n–th step. A procedure for regulating η_n can be introduced in the following way.

Definition. A measurable function $\beta(x)$ on $(\mathcal{X}, \mathcal{A})$ with the following properties:

1. $\beta(x) \geq 0, \quad x \in \Omega,$
2. $\int_B \beta(x)\pi(dx) > 0$, if $\pi(B) > 0$

is called a branching density.

Introduce the random variable

$$r(x,y) = \lfloor \beta(y)/\beta(x) \rfloor, \text{with probability } 1 - \overline{\beta(y)/\beta(x)}$$

and

$$r(x,y) = \lfloor \beta(y)/\beta(x) \rfloor + 1, \text{with probability } \overline{\beta(y)/\beta(x)},$$

where $\lfloor a \rfloor$, with $a = \beta(y)/\beta(x)$, means the integer part of a and $\overline{a}$ means the whole part of a.
At step zero, set $\eta_0 = 1$. At each following step $(n+1 = 1, 2, \ldots)$ if $X_n^\gamma = x$, $X_{n+1}^\gamma = y$, then when $\beta(y)/\beta(x) < 1$ the simulation of path γ discontinues with probability $1 - \beta(y)/\beta(x)$.

When $\beta(y)/\beta(x) \geq 1$ we simulate $r(x,y) - 1$ additional paths, beginning from point x, where k-th steps of these chains will be denoted by index $n + k$.

Let us simulate k loops: $\{X_n^\gamma(i);\ \gamma = 1, 2, \ldots, \eta_n(i)\}$, $i = 1, 2, \ldots, k$. Let us determine estimators of the functionals J_j $(j = 1, 2, \ldots, l)$ by the formula

$$\widehat{J}_{k,\beta}(j) = \frac{1}{k}\sum_{i=1}^{k} Y_\beta(i,j) \Bigg/ \sum_{i=1}^{k} \widetilde{Y}_\beta(i,j)\ ,$$

where

$$\begin{aligned} Y_\beta(i,j) &= \sum_{n=0}^{N(i)} \sum_{\gamma=1}^{\eta_n} h_j\left(X_n^\gamma(i)\right) / \beta\left(X_n^\gamma(i)\right)\ , \\ \widetilde{Y}_\beta(i,j) &= Y_\beta(i,j) \quad \text{with} \quad h_i(x) \equiv 1. \end{aligned}$$

Note that when $\beta(x) \equiv 1$ the process reduces to direct simulation of the chain $\{X_n\}$, and the estimators $\widehat{J}_{i,\beta}$ become the known ratio estimators of the regeneration method (see Crane, Iglehart, 1974).

Denote

$$\mathcal{D}(\beta) = \left(\lim_{k\to\infty} k\,Cov\left(\widehat{J}_{k,\beta}(i),\, \widehat{J}_{k,\beta}(j)\right)\right)_{i,j=1}^{l},$$

$$T(\beta) = \lim_{k\to\infty} \sum_{i=1}^{k} T_{i,\beta}/k.$$

By virtue of Law of Large Numbers, $T(\beta)$ is the expectation of the number of steps for all chains in one loop.
Set

$$\hat{h}_j(x) = \widehat{J}_{k,\beta}(j), \quad j = 1, 2, \ldots, l,$$

under the condition that all paths begin at x
$(X_0^1(i) = x, \quad i = 1, \ldots, k)$.

4. Expression for covariances

Theorem 1. *When regularity conditions (see Melas(1993)) are valid for general Markov chains, estimators* $\widehat{J}_{k,\beta}(i)$ *are asymptotically unbiased and*

$$\begin{aligned} \mathcal{D}(\beta) &= \left(\int \beta^{-1}(x)\pi(dx)\,(d_{ij}(x) + r_\beta) \right)_{i,j=1}^{l}, \\ T(\beta) &= \int \beta(x)\,\pi(dx), \end{aligned}$$

where $d_{ij}(x) = E\hat{h}_i(x)\,\hat{h}_j(x) - E\hat{h}_i(x)\,E\hat{h}_j(x)$, E *is the expectation operator, and* r_β *is a remainder term.*

Theorem 1 permits us to introduce optimality criteria for the choice of the branching measure. To simplify the notation let us consider that $\pi(dx) = \pi(x)\,dx$. Set

$$\tau(x) = \pi(x)\,\beta(x) \Big/ \int \beta(x)\,\pi(x)\,dx \tag{1}$$

$$\tau_x = \pi_x \beta_x \Big/ \sum_x \beta_x \pi_x \quad \text{(finite chains)}. \tag{2}$$

Then $T(\tau) = 1$ for any β. Dropping the remainder term, we obtain the matrix

$$\begin{aligned} \overline{\mathcal{D}}(\tau) &= \left(\int \tau^{-1}(x)\, d_{ij}(x)\, \pi^2(x)\, dx \right), \\ \overline{\mathcal{D}}(\tau) &= \left(\sum_{x=0}^{m} \tau_x^{-1} \pi_x^2\, d_{ij}(x) \right) \quad \text{(finite chains)}. \end{aligned}$$

As an estimation accuracy criterion let us take quantities

$$trA\overline{\mathcal{D}}(\tau), \tag{3}$$

$$\det \overline{\mathcal{D}}(\tau) \tag{4}$$

where A is an arbitrary nonnegative-definite matrix. Note that when $A = I$ (where I is identity matrix) and $l = 1$, $trA\overline{\mathcal{D}}(\tau)$ is just the variance of the estimator $\widehat{J}_1$.

The probability measure $\tau(x)$ is called a design of the simulation experiment.

5. Optimal experimental designs

Consider the minimization problem for criteria (3) and (4) in the class of probability measures τ induced by branching densities. The optimal design for criterion (3) can be found with the help of the Schwarz inequality. The direct application of this inequality brings the following result.

Theorem 2. *Let the hypothesis of Theorem 1 be satisfied. Then the experimental design minimizing the value of tr $A\mathcal{D}(\tau)$ is given by formula (1), where*

$$\beta(x) = \sqrt{trA\mathcal{D}(\tau)}.$$

Theorem 3. *For finite regular Markov chains and functions $h_i(x)$ of the form $h_i(x) = \delta_{ix}$, $i = 1, \ldots, m$, there exists an experiment design minimizing* $\det \overline{\mathcal{D}}(\tau)$ *. This design is unique and satisfies the relation*

$$\tau(x) = \left(tr\, B_x B^{-1}(\tau^*)\right)^{1/2} \pi_x \Big/ \sqrt{m},$$

where

$$\begin{aligned} B_x &= (p_{xy}\delta_{yz} - p_{xy}p_{xz})_{y,z=1}^{m}, \\ x &= 0, 1, \ldots, m, \\ B(\tau) &= \sum_{x=0}^{m} \left(\pi_x^2/\tau_x\right) B_x. \end{aligned}$$

The design described in Theorem 3 will be called *D*-optimal.

In order to construct a D-optimal design the following iterative method can be used. Set

$$\tau_0(y) = \pi_y, \quad y = 0, \ldots, m,$$
$$\tau_{k+1}(y, \alpha) = (1 - \alpha)\,\tau_k(y) + \alpha\tau_{(x)}(y),$$
$$k = 0, 1, 2, \ldots$$

where

$$\tau_{(x)}(y) = \delta_{xy}$$

$$x = \arg\max_x \frac{\pi_x^2}{\tau_k^2(x)}\, tr\, B_x B^{-1}(\tau_k).$$

Set

$$\alpha_k = \arg \min_{\alpha \in [0,1]} \det B\left(\{\tau_{k+1}(y, \alpha)\}\right),$$

$$\tau_{k+1}(y) = \tau_{k+1}(y, \alpha_k), \quad y = 0, 1, \ldots, m.$$

Theorem 4. *Under the hypothesis of Theorem 3 for* $k \to \infty$

$$\det \overline{\mathcal{D}}(\tau_k) \to \det \overline{\mathcal{D}}(\tau^*),$$

where $\det \mathcal{D}(\tau^*) = \min \det \mathcal{D}(\tau)$.

Consider a random walk process on a line

$$S_1 = 0, \quad S_{n+1} = S_n + X_n, \quad k = 1, 2, \ldots,$$

where X_n, $k = 1, 2, \ldots$ are independent random variables with common density function $f(x)$, and connect it with the waiting process

$$W_1 = 0, \quad W_{n+1} = \max(0, W_n + X_n), \quad n = 1, 2, \ldots$$

It is known (see Feller, 1970) that the quantities W_n and

$$M_n = \max\{S_1, \ldots, S_n\}$$

have the same distribution. Under an additional condition on EX_1, $M_n \to M$ and $W_n \to W$ in distribution, where M and W are random variables.

Set $\pi(0) = P\{W = 0\}$, $\pi(x) = W'(x)$, $W(x) = P\{W < x\}$, $\pi(x) = 0$ for $x < 0$, $X_n = W_n$, $S = \{0\}$, $\pi(dx) = af(x)dx$, $a = 1/(1 - F(0))$, and

$$1-F(0) = \int_0^\infty f(t)dt,\ h_1(x) = 0,\ x < v, \quad h_1(x) = 1,\ x \geq v, l = 1.$$

The problem is the estimation of the functional

$$J = J_1 = \int h_1(x)\pi(x)dx.$$

Suppose that $f(x)$ satisfies the condition

$$\int_{-\infty}^{\infty} e^{\lambda_0 t} f(t)dt = 1$$

for some positive λ_0 and $EX_1 < 0$. Set $\beta(x) \equiv 1$, $\beta^*(t) = e^{\lambda_0 t}$, $0 < t < v$, $\beta^*(t) = e^{\lambda_0 v}$, $t \geq v$.

Define the relative efficiency by the formula

$$R = R_v = \frac{T_{\beta^*} \mathcal{D} \widehat{J}_{\beta^*}}{T_\beta \mathcal{D} \widehat{J}_\beta}.$$

It is known (see Feller, 1970) that as $v \to \infty$ $\theta \sim e^{-\lambda_0 t} \to 0$.

Theorem 5. *Let the conditions formulated above be satisfied and for some* $\varepsilon > 0$

$$\int_{-\infty}^{\infty} e^{2(\lambda_0+\varepsilon)t} f(t)\, dt < \infty.$$

Then for $v \to \infty$

$$R_v = O\left(\frac{1}{\theta \ln^2(1/\theta)}\right), \quad \theta \sim e^{-\lambda_0 v}.$$

6. Numerical example

To compare the branching technique (BT) and the usual regeneration technique (RT) we simulated the corresponding random walk by both methods during 100,000 loops. The magnitude v was chosen so that $\theta = 10^{-2}, 10^{-3}, \ldots, 10^{-6}, a = 1/2$. Denote by I the magnitude

$$I = \theta^2/(ETD\widehat{\theta})$$

where θ is the proper value of the estimated parameters, ET is the mean length of the simulated paths and $D\widehat{\theta}$ is the sample variance. We have obtained results presented in the following table:

$In(1/\theta)$	2	3	4	5	6
$I(RT)$	0,385	0,026	0,000	0,000	0,000
$I(BT)$	0,686	0,298	0,185	0,083	0,041

7. Example of D-optimal design

Consider a particular case of Markov chain imbedded into the random process, corresponding to the length of queue with one server and m places for waiting.

We will consider the simplest case when the input stream is a simplest stream and the time of service is an exponentially distributed random value.

Let ρ be the load of the system. Then the matrix P is of the form

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & & \\ 1-\Delta & 0 & \Delta & 0 & & & \\ 0 & 1-\Delta & 0 & \Delta & & & \\ \vdots & & & & & & \\ & & & & & 1-\Delta & 0 & \Delta \\ & & & & & 0 & 1-\Delta & \Delta \end{pmatrix},$$

$\Delta = \rho/(1+\rho)$, $\pi_1 = \Delta\pi_0$, $\pi_{i+1} = \rho^i \Delta\pi_0$, $i = 1, \dots, m-1$,
$\pi_0 = 1/(1 + \Delta + \Delta\rho + \dots + \Delta\rho^{m-1})$.
We can calculate by induction that

$$\det B(\tau) = \left(\prod_{i=1}^{m} \frac{\pi_i^2}{\tau_i} \right) \Delta^m (1-\Delta)^m.$$

From here we found that D-optimal design is $\tau^* = \{0, \frac{1}{m}, \ldots, \frac{1}{m}\}$ and

$$I(\tau) = \left(\frac{\det B(\pi)}{\det B(\tau)}\right)^{1/m} = \left(\prod_{i=1}^{m} \pi_i\right)^{-1/m} \Big/ m$$

with $\tau = \tau^*$. Note that $I(\tau)$ is a natural measure of efficiency of design τ. It means the ratio of the number of steps needed by immediate simulation for obtaining results with a given accurancy to the same number for the splitting and roulette approach. In table 1 values of the efficiency are given.

m	5	5	5	10	10	10
ρ	1/2	1/3	1/4	1/2	1/3	1/4
I	2.0	4.0	6.8	6.0	31.6	109

8. Concluding remarks

- In fact, any immediate simulation method can be improved by including the branching technique
- A representation for the variance-covariance matrix derived here allows to find an optimal branching measure
- The full explanation of the theory can be found in the book (Ermakov, Melas, 1995) and in the papers (Melas, 1993, 1997, 2002)

References

Crane, A. M., and D. L. Iglehart. 1974. Simulating stable stochastic systems. 2: Markov chains. *Journal of Association of Computing Machinery* 21 (1): 114-123.

L'Ecuyer, P., Demers, V., and B.Tuffin (2006). Splitting for rare-event simulation.In: *Proceedings of the 2006 Winter Simulation Conference*, 137–148,California Monterey.

Ermakov, S. M., and V. B. Melas. 1995. *Design and analysis of simulation experiments*. Dordrecht, London: Kluwer Academic Publishers.

Feller, W. 1970. *An introduction to probability theory and its applications*, vol. 2. 2nd ed. New York: Wiley.

Kahn, H., and T. E. Harris. (1951). Estimation of particle transmission by random sampling. *National Bureau os Standards Applied Mathematical Series* 12:27–30.

Melas, V. B. 1993. Optimal simulation design by branching technique. *Model Oriented Data Analysis*, ed. W. G. Muller, H. P. Wynn, and A. A., Zhigljavsky. Heidelberg: Physica-Verlag, 113-127.

Melas, V.B. 2002. On simulating finite Markov chains by the splitting and roulette approach. Information processes. vol 2, issue 2, pp. 228-230.

Nummelin, E. 1984. *General irreducible Markov chains and non-negative operators*. Cambridge: Cambridge University Press.

Sigman, K. 1988. Queues as Harris recurrent Markov chains. *Queueing systems* 3: 179-198.