Markov Chain Monte Carlo Revolution and Mixing Time Estimation

Alexander Gasnikov (Premolab MIPT)

avgasnikov@gmail.com

IV Traditional Scientific School

Zvenigorod (Moscow Region); June 19, 2012

Markov Chain Monte Carlo Method

Aim: To generate discrete random variable with distribution π.

One of the possible ways to do it: Introduce p_{ij}^0 – arbitrary symmetric ($p_{ij}^0 = p_{ji}^0$) irreducible matrix, and then we can define a Markov chain:

$$p_{ij} = a_{ij} p_{ij}^0,\ i \neq j;\ p_{ii} = 1 - \sum_{j:\ j \neq i} p_{ij},\ \text{where}$$

$a_{ij} = F\left(\pi_j / \pi_i\right)$ and $F(z)$ – arbitrary function s.t. $\dfrac{F(z)}{F(1/z)} = z$.

$$\pi_i a_{ij} p_{ij}^0 = \pi_i a_{ij} p_{ji}^0 = \pi_i F\left(\pi_j / \pi_i\right) p_{ji}^0 = \pi_j F\left(\pi_i / \pi_j\right) p_{ji}^0 = \pi_j a_{ji} p_{ji}^0.$$

For example, $\tilde{F}(z) = \min\{z, 1\}$ – (Hastings–)Metropolis algorithm. Note that for all such $F(z)$ we have $F(z) \le \tilde{F}(z)$.

Another example, Barker function $F(z) = z/(1+z)$.

Note that p_{ij}^0 is commonly taken of form $p_{ij}^0 = 1/M$, where M is a number of states, or $p_{ij}^0 = 1/(2M)$, $i \neq j$; $p_{ii}^0 = 1/2$, $i \neq j$.

At the large values of time t, according to ergodic theorem for aperiodic and irreducible Markov chains, we have that probability distribution of introduced above Markov chain is close enough to the stationary distribution, which has proved to be π. Indeed, under the mentioned above conditions we have the detailed balance condition: $\forall\ i, j \to \pi_i p_{ij} = \pi_j p_{ji}$.

The main use of this method provided the following experimental observations: *in applications this mixing time t surprisingly isn't such large*.

To estimate this mixing time there different techniques has been developed: Poincare inequality (canonical path), Cheeger inequality (conductance), Coupling technique, Coarse Ricci curvature approach e.t.c. (based on the concentration of measure phenomenon).

Examples: Top-to-random shuffle ($\sim n\log_2 n$), Riffle shuffle ($\sim \log_2 n$); Hit and Run (Fust and Fun); Ising model, Gibbs distribution, Glauber dynamic ($\sim n^{2\log_2 e/T}$, $0 < T \ll 1$); The traveling salesman problem; Simulated annealing; Cryptography.

Literature: M. Jerrum & A. Sinclair, 1996; D. Aldous Random walks on a graph, 1999; Kelbert–Sukhov Markov process, 2009; P. Diaconis // Bulletin of the AMS, 2009; D. Spielaman, 2009.

Applications to the theory of macrosystem

Example (Ehrenfests's paradox, 1907)

Dog 1 ($n_1(t)$ fleas)

Dog 2 ($n_2(t)$ fleas)

There are $M = 2n \gg 1$ fleas. At the beginning whole the fleas are situated on the Dog 1. However, it isn't important! At each time step (described Markov process is discrete on time) with the probability equals $1/M$ a random flea is choosing. This flea is jump to another dog. The process is repeated in time.

Microstate (2^M) is a way of distribution M different fleas onto 2 different dogs.

Macrostate ($M+1$) is a way of distribution M identical fleas onto 2 different dogs.

Denote by P the matrix (size $2^M \times 2^M$) of transitional probabilities of Markov chain described above on a microscopically level. Since this dynamic is reversible on time $=> P = P^T$. Thus, since P – stochastic matrix, we have

$$\left(1,...,1\right) = \left(1,...,1\right)P^T \Rightarrow \left(1,...,1\right) = \left(1,...,1\right)P.$$

Therefore we have that all the microstates are equally probable in stationary (invariant) distribution of this chain. That is mean the following

Probability of macrostate $\left(k, M-k\right)$ *in stationary distribution equals* $C_M^k 2^{-M}$.

From the Ergodic theorem and Central Limit Theorem (Moivre–Laplace, 1738):

$$\forall\ t \geq 2M \rightarrow P\left(\left|\frac{n_1\left(t\right)}{M} - \frac{1}{2}\right| \leq \frac{5}{\sqrt{M}}\right) \geq 0.99.$$

Later we'll show why in this case we have mixing time equals $\Theta\left(M\right)$.

Loschmidt's paradox (1986) for the A. Poincare theorem of return.

Mathematical expectation of time $T = \inf\left\{t > 0 :\ n_1\left(t\right) = 0, n_1\left(0\right) = n\right\}$ asymptotically (on M) equals $2^M / M$.

Google problem (L. Page and S. Brin, 1998). Goal: to find a way of ranging web-pages $\vec{p}^{\,*}$.

Directed Internet web-graph $G=(V,E)$ (vertexes are web-pages, edges are references), $M\left(\gg|V|\gg 1\right)$ – the total number of users (const), $P=(1-\alpha)E+\alpha\tilde{P}$, where $\tilde{p}_{ij}=\left|\{k:\ (i,k)\in E\}\right|^{-1}$, $i\neq j$, else = 0. Let $n_i(t)$ to be the number of users of the web-page *i* at the moment of time *t*. For the unit of time each user independently go forward one of the possible references (i,j) with probability $\alpha\ \tilde{p}_{ij}$. We assume stochastic matrix P to be irreducible. Then it can be shown that:

$$\forall\ \ q=0,...,|V|\ \ \exists\ \ \lambda_q>0,\ \ T_q=\mathrm{O}\left(Poly(M)\right):\ \ \ \forall\, t\geq T_q$$

$$P\left(\left|\frac{n_i(t)/M}{p_i^*}-1\right|\leq\frac{\lambda_q}{\sqrt{M/|V|}},\ i=0,...,q\right)\geq 0.99,$$

where $P^T\vec{p}^{\,*}=\vec{p}^{\,*}$ ($\vec{p}^{\,*}$ – is a unique solutions in class of probability distributions, find by MCMC). For the purpose of convenience we assume, that $p_1^*\geq p_2^*\geq...$

Kinetic of social inequality (V. Pareto model, 1896)

r-th inhabitant, *k*-rubles

l-th inhabitant, *m*-rubles

There are $M \gg 1$ (for example, 10 000) numbered inhabitants in the city. Inhabitant with the number *i* has $s_i(0)$ rubles at the moment of time $t = 0$.

Inhabitants are arbitrary randomly play with each other. More precisely: $\zeta N^{-1}\Delta t + o(\Delta t)$ ($\zeta > 0$) is probability of inhabitants with the arbitrary numbers r и l ($1 \le r < l \le M$) try to play one ruble according to the following rule: with probability ½ inhabitant number *l* gives (if he is not a bankrupt) one ruble to the opponent and similarly on the contrary.

Let $c_s(t)$ is the portion of inhabitants, that have exactly s rubles at the moment of time t (note, that $c_s(t)$ is a random variable). Let

$$S = \sum_{i=1}^{M} s_i(0),\ \bar{s} = S/M .$$

Then

$$\forall\ q = 0, ..., S\ \ \exists\ \lambda_q > 0,\ T_q = \mathrm{O}(M):\ \ \forall\, t \geq T_q$$

$$P\left(\left| \frac{c_s(t)}{Ce^{-s/\bar{s}}} - 1 \right| \leq \frac{\lambda_q}{\sqrt{M}},\ s = 0, ..., q \right) \geq 0.99,$$

where C is determinate from the following condition:

$$\sum_{s=0}^{S} Ce^{-s/\bar{s}} = 1, \text{ that is } C \approx 1/\bar{s} .$$

Now we are ready to consider a **general case**.

Assume that some macrosystem can stay at different states, characterized by the vector $\vec{n}$ with nonnegative integer components. Let us assume that in the system there are the following (also called "chemical") reactions:

$$\vec{n} \to \vec{n} - \vec{\alpha} + \vec{\beta}, \left(\vec{\alpha}, \vec{\beta}\right) \in J .$$

Following to the Leontovich (1934), let us introduce intensity of the reaction:

$$\lambda_{\left(\vec{\alpha},\vec{\beta}\right)}\left(\vec{n}\right) = \lambda_{\left(\vec{\alpha},\vec{\beta}\right)}\left(\vec{n} \to \vec{n} - \vec{\alpha} + \vec{\beta}\right) = M^{1-\sum_i \alpha_i} K_{\vec{\beta}}^{\vec{\alpha}} \prod_{i:\,\alpha_i > 0} n_i \cdot \ldots \cdot \left(n_i - \alpha_i + 1\right),$$

where $K_{\vec{\beta}}^{\vec{\alpha}} \geq 0$ is a constant of reaction. Note that in application it is always assume that

$$\sum\nolimits_i n_i\left(t\right) \equiv M \ \ (M \ \text{is often called scaling parameter}).$$

Thus $\lambda_{\left(\vec{\alpha},\vec{\beta}\right)}\left(\vec{n}\right)$ – is a probability of the reaction $\vec{n} \to \vec{n} - \vec{\alpha} + \vec{\beta}$ take place in the unit of time. On the macro level this is corresponds to the law of the operating mass of Guldberg–Vaage (1864).

The following theorem reflect some results of the works a) [Vedenyapin, 2000], b) and c) [Malyshev, Pirogov, Rubco, 2004].

Theorem 1. a) $\left\langle \vec{\mu}, \vec{n}(t) \right\rangle \equiv \left\langle \vec{\mu}, \vec{n}(0) \right\rangle \Leftrightarrow \vec{\mu} \perp \mathrm{Lin}\left\{ \vec{\alpha} - \vec{\beta} \right\}_{(\vec{\alpha},\vec{\beta}) \in J}$. (inv)

b) Let us assume that the following condition (unitarity) is take place:

$$\exists\ \vec{\xi} > \vec{0}:\ \forall\ \vec{\alpha} \to \sum_{\vec{\beta}:(\vec{\alpha},\vec{\beta}) \in J} K_{\vec{\beta}}^{\vec{\alpha}} \prod_j \xi_j^{\alpha_j} = \sum_{\vec{\beta}:(\vec{\alpha},\vec{\beta}) \in J} K_{\vec{\alpha}}^{\vec{\beta}} \prod_j \xi_j^{\beta_j}. \qquad \text{(U)}$$

Then the measure $\nu(\vec{n}) = \prod_i \lambda_i^{n_i} e^{-\lambda_i} / n_i!$, where $\lambda_i = \xi_i^* M$ and $\vec{\xi}^*$ arbitrary solution of (U), is invariant (stationary) measure of Markov dynamic introduced above. This measure on the set (inv) will be exponentially concentrated in a small vicinity of the most probable macrostate (which is called equilibrium of macrosystem), when parameter M is growth. To find this most probable state we have to maximize the following entropy functional on the affine set (inv):

$$E(\vec{n}) \approx -\sum_i n_i \cdot \left(\ln\left(n_i / \lambda_i \right) - 1 \right).$$

Note that (U) condition is a generalization of the detailed balance condition (well known in physics): $\exists\ \vec{\xi} > \vec{0}:\ \forall\ (\vec{\alpha}, \vec{\beta}) \in J \to K_{\vec{\beta}}^{\vec{\alpha}} \prod_j \xi_j^{\alpha_j} = K_{\vec{\alpha}}^{\vec{\beta}} \prod_j \xi_j^{\beta_j}$.

c) Let us assume that J doesn't depend on M, for the moment of time $t = 0$ and for any i there exist the following limits: $c_i(0) = \lim_{M\to\infty} n_i(0)/M > 0$. Than for arbitrary moment of time $t > 0$ and for any i there exist the following limits: $c_i(t) \overset{\text{a.s.}}{=} \lim_{M\to\infty} n_i(t)/M$. The described above limits is also called canonical scaling limits. Moreover the deterministic functions (concentrations) $c_i(t)$ satisfy to the following system of ODE:

$$\frac{dc_i}{dt} = \sum_{(\vec{\alpha},\vec{\beta})\in J} (\beta_i - \alpha_i) K_{\vec{\beta}}^{\vec{\alpha}}(\vec{c})\vec{c}^{\vec{\alpha}},\ \vec{c}^{\vec{\alpha}} = \prod_j c_j^{\alpha_j} \qquad \text{(DE)}$$

It can be shown [Batischeva, Vedenyapin, 2000], that if (U) condition is take place than all the trajectories of the system (DE) starting at (inv) remain at (inv) and converge to the unique fixed point, satisfying (U) condition. To show this it is introduced the minus entropy function: $H(\vec{c}) = \sum_i c_i \cdot \left(\ln(c_i/\xi_i) - 1\right)$ and it is shown that this function has proved to be a Lyapunov function for the dynamic (DE). This supervision was recently generalized in collaboration with Gasnikova E. V.

Theorem 2. Let invariant measure has the following representation:

$$\nu\left(\vec{n}\right) = M \exp\left(-M \cdot \left(H\left(\vec{n}/M\right) + o\left(1\right)\right)\right),\ M \gg 1,$$

where $H\left(\vec{c}\right)$ is strictly concave function. Then $H\left(\vec{c}\right)$ – is Lyapunov function for the dynamic (DE).

Scheme of the prove. Following to the work [Kalinkin, 2002], let us introduce the generating function:

$$F\left(t,\vec{s}\right) = \sum_{\vec{n}} P\left(\vec{n}\left(t\right) = \vec{n}\right) \cdot \vec{s}^{\,\vec{n}},\ \left|\vec{s}\right| \le \vec{1},\ \vec{s}^{\,\vec{n}} = s_1^{n_1} \cdot s_2^{n_2} \cdot ...,$$

which satisfies the following partial differential equation:

$$\frac{\partial F\left(t,\vec{s}\right)}{\partial t} = \sum_{\left(\vec{\alpha},\vec{\beta}\right)\in J} \left(\vec{s}^{\,\vec{\beta}} - \vec{s}^{\,\vec{\alpha}}\right) K_{\vec{\beta}}^{\vec{\alpha}} M^{1-\sum_i \alpha_i} \frac{\partial^{\alpha_1+\alpha_2+\cdots} F\left(t,\vec{s}\right)}{\partial s_1^{\alpha_1} \cdot \partial s_2^{\alpha_2} \cdot ...}. \qquad \text{(PDE)}$$

Note [Kalinkin, 2002], that if we take $\partial/\partial s_i$ from the both side of the (PDE), assuming $\vec{s} = \left(1,...,1\right)^T$, than in the limit $M \to \infty$ (assuming the existence of the limits $\lim_{M\to\infty} n_i\left(t\right)/M$, see c) Theorem 1) we obtain (DE).

Since

$$F\left(\infty,\vec{s}\right)\approx M\int e^{M\left(\left\langle\overrightarrow{\ln s},\vec{\xi}\right\rangle-H\left(\vec{\xi}\right)\right)}d\vec{\xi},$$

than

$$M^{-(\alpha_1+\alpha_2+\ldots)}\frac{\partial^{\alpha_1+\alpha_2+\ldots}F\left(\infty,\vec{s}\right)}{\partial s_1^{\alpha_1}\cdot\partial s_2^{\alpha_2}\cdot\ldots}\approx\frac{\vec{\xi}\left(\vec{s}\right)^{\vec{\alpha}}}{\vec{s}^{\,\vec{\alpha}}}C\left(\vec{s},M\right),$$

where $\vec{\xi}\left(\vec{s}\right)$ is uniquely determinate by the system

$$\overrightarrow{\ln s}=\operatorname{grad}H\left(\vec{\xi}\right),$$

and $C\left(\cdot\right)\neq 0$ doesn't depend on $\vec{\alpha}$. Hence,

$$0\equiv\sum_{\left(\vec{\alpha},\vec{\beta}\right)\in J}\left(\vec{s}^{\,\vec{\beta}}-\vec{s}^{\,\vec{\alpha}}\right)K_{\vec{\beta}}^{\vec{\alpha}}\frac{\vec{\xi}\left(\vec{s}\right)^{\vec{\alpha}}}{\vec{s}^{\,\vec{\alpha}}}=\sum_{\left(\vec{\alpha},\vec{\beta}\right)\in J}\left(e^{\left\langle\vec{\beta}-\vec{\alpha},\,\operatorname{grad}H\left(\vec{\xi}\left(\vec{s}\right)\right)\right\rangle}-1\right)K_{\vec{\beta}}^{\vec{\alpha}}\vec{\xi}\left(\vec{s}\right)^{\vec{\alpha}}\geq$$

$$\geq\sum_{\left(\vec{\alpha},\vec{\beta}\right)\in J}\left\langle\vec{\beta}-\vec{\alpha},\operatorname{grad}H\left(\vec{\xi}\left(\vec{s}\right)\right)\right\rangle K_{\vec{\beta}}^{\vec{\alpha}}\vec{\xi}\left(\vec{s}\right)^{\vec{\alpha}}=\left.\frac{dH\left(\vec{c}\right)}{dt}\right|_{\vec{c}=\vec{\xi}\left(\vec{s}\right)}$$

– full derivative of the function $H\left(\vec{c}\right)$ owing to the system (DE) in the point $\vec{\xi}\left(\vec{s}\right)$. □

Theorem 3. Let us consider a macrosystem under conditions of c) Theorem 1. Then if the system (DE) has a unique globally exponentially stable fixed point $\vec{c}^{\,*}$ on (inv) than:

- invariant measure is exponentially concentrated in a small vicinity of $M\vec{c}^{\,*}$;
- elements of correlation matrix of vector $\vec{n}(t)$ is uniformly bounded in time;
- the limits is permutable: $\lim_{M\to\infty}\lim_{t\to\infty} * = \lim_{t\to\infty}\lim_{M\to\infty} *$;
- mixing time is $\mathrm{O}\left(Poly\left(M\right)\right)$.

Counterexample. The model predator–victim (Nicolas, Prigogine, 1977):

$$V \xrightarrow{K_1} 2V,\ V+P \xrightarrow{K_2} 2P,\ P \xrightarrow{K_3} 0.$$

Hypothesis. The attractor of the determinate system (DE) (it can be as a complex set as it possible in principle) is such a set in a small vicinity of which the considered macrosystem will stay with high probability at the large values of time.

Cheeger's isoperimetric inequality approach (Fan Chung, 2005)

Let us considered the Cheeger isoperimetric inequality for the case of irreversible Markov chains, with transition probability matrix $P = \left\| p_{ij} \right\|$ and invariant measure π, corresponds to the random walks on the directed graph $G = \left(V_G, E_G \right)$:

$$h\left(G\right) = \inf_{S \subseteq V_G : \pi(S) \le 1/2} P\left(S \to \bar{S} \middle| S\right) = \inf_{S \subseteq V_G : \pi(S) \le 1/2} \frac{\sum_{(i,j) \in E_G : i \in S,\ j \in \bar{S}} \pi\left(i\right) p_{ij}}{\sum_{i \in S} \pi\left(i\right)}, \text{ (Cheeger constant)}$$

$$T\left(i, \varepsilon\right) = \Theta\left(h\left(G\right)^{-2} \left(\ln\left(\pi\left(i\right)^{-1} \right) + \ln\left(\varepsilon^{-1} \right) \right) \right), \text{ (Mixing time)}$$

$$\forall\ i \in V_G,\ t \ge T\left(i, \varepsilon\right) \to \left\| P^t\left(i, \cdot\right) - \pi\left(\cdot\right) \right\|_{TV} = \sum_j \left| P^t\left(i, j\right) - \pi\left(j\right) \right| \le \varepsilon.$$

This result can be generalized to the continuous time Markov process and as a consequence it can be applied to estimation of mixing time to equilibrium in the described above dynamic of macrosystem. Moreover, if macrosystem is fulfilled the detailed balance condition, then $h\left(G\right)^{-1} = \Theta\left(Poly\left(M\right) \right)$.

Coarse Ricci curvature approach (A. Joulin and Y. Ollivier, 2007)

Monge–Kantorovich distance:

$$W_1(\mu,\nu)=\inf_{\substack{\xi\geq 0:\int_y d\xi(x,y)=d\mu(x)\\ \int_x d\xi(x,y)=d\nu(y)}}\iint d(x,y)\,d\xi(x,y).$$

κ – Coarse Ricci curvature iff

$$\exists\ \kappa>0,\ t_0>0:\ \forall\ i,j\in V_G\rightarrow W_1\left(P^{t_0}(i,\cdot),P^{t_0}(j,\cdot)\right)\leq(1-\kappa)\,d(i,j).$$

$$W_1\left(P^t(i,\cdot),\pi(\cdot)\right)\leq\left[\kappa^{-1}\int_y d(i,y)P(i,dy)\right]\cdot(1-\kappa)^{t/t_0}.$$

For the Ehrenfests's paradox if we put

$$d\left(\begin{pmatrix}n_1\\ n_2\end{pmatrix},\begin{pmatrix}\tilde{n}_1\\ \tilde{n}_2\end{pmatrix}\right)=\left\|\begin{pmatrix}n_1\\ n_2\end{pmatrix}-\begin{pmatrix}\tilde{n}_1\\ \tilde{n}_2\end{pmatrix}\right\|_{TV}=\frac{1}{2}\left(|n_1-\tilde{n}_1|+|n_2-\tilde{n}_2|\right)=|n_1-\tilde{n}_1|$$

then $\kappa=M^{-1}$ ($t_0=1$). This fact can be generalized for the macrosystems which are fulfilled the detailed balance condition.

Numerical methods of finding equilibrium of macrosystem

Let us consider the following linear entropic programming problem:

$$-\sum_{k=1}^{m} x_k \ln\left(x_k / e\right) \to \max_{\vec{x} \in \mathrm{A} \cap \mathbb{R}_+^m};\ \mathrm{A}:\ \vec{\Lambda}\left(\vec{x}\right) = \vec{q} - \mathrm{T}\vec{x} = \vec{0}_l,\ \vec{F}\left(\vec{x}\right) = \vec{d} - G\vec{x} \geq \vec{0}_w. \quad (*)$$

To find a unique solution of (*) in [Gasnikova, 2009] it was assumed the following family of iterative algorithms, containing Bregman's algorithm, MART, GISM, algorithm of Popkov e.t.c.:

$$\begin{cases} x_k^n = \exp\left(-\sum_{p=1}^{l} t_{pk} \lambda_p^n - \sum_{q=1}^{w} g_{qk} \mu_q^n \right), \quad k = 1,\ldots,m; \\ \lambda_p^{n+1} = \lambda_p^n - \gamma_p g_p^{\lambda} \left(q_p - \sum_{k=1}^{m} t_{pk} x_k^n \right), \quad p = 1,\ldots,l; \\ \mu_q^{n+1} = \mu_q^n - \alpha_q \mu_q^n g_q^{\mu} \left(d_q - \sum_{k=1}^{m} g_{qk} x_k^n \right), \quad q = 1,\ldots,w, \end{cases} \quad (**)$$

Where the steps $\gamma_p > 0,\ \alpha_q > 0$ is sufficiently small, and rather smooth functions $\left\{g_p^{\lambda}(\cdot)\right\}_{p=1}^{l}, \left\{g_q^{\mu}(\cdot)\right\}_{q=1}^{w}$ is monotonically increasing and equal to zero in the zero point. The global convergence of process (**) was established under the condition $\exists\ \vec{z} > \vec{0}_m :\ \vec{q} - T\vec{z} = \vec{0}_l,\ \vec{d} - G\vec{z} \geq \vec{0}_w$. But sometimes it is worth to put to zero most of the components of $\left\{g_p^{\lambda}(\cdot)\right\}_{p=1}^{l}, \left\{g_q^{\mu}(\cdot)\right\}_{q=1}^{w}$, choosing them at random.

Algorithm (**) is effective especially in cases:

- T and G is sparsity;
- $l + w \ll m$.

So in this case the complexity of one step is $\mathrm{O}\left((m + l + w)\right)$ and $\mathrm{O}\left(m \cdot (l + w)\right)$ correspondently.

Literature (only for the part “Conception of equilibrium of macrosystem”)

1. *Kalinkin A. V.* Markov branching processes with interaction // Uspekhi Mat. Nauk. 2002. V. **57**:2(344). P. 23–84.

2. *Malyshev V. A.*, *Pirogov S. A.*, *Rubco A. N.* Random walks and chemical networks // Mosc. Math. J. 2004. V. 4. № 2. P. 441–453.

3. *Joulin A.*, *Ollivier Y.* Curvature, concentration and error estimates for Markov chain Monte Carlo // Ann. Prob. 2010. V. 38. № 6. P. 2418–2442. http://www.yann-ollivier.org/rech/publs/surveycurvmarkov.pdf

4. *Fan Chung* Laplacians and the Cheeger inequality for directed graphs // Annals of Combinatorics. 2005. no. 9. P. 1–19. http://math.ucsd.edu/~fan/

5. *Weidlich W.* Sociodynamics: A Systematic Approach to Mathematical Modeling in the Social Sciences. OPA, 2000.

6. Introduction to the mathematical modeling of traffic flow. Ed. A. V. Gasnikov. MCCME, 2012. http://zoneos.com/traffic/

7. *Batishcheva Ya. G.*, *Vedenyapin V. V.* The 2-nd law of thermodynamics for chemical kinetics // Matem. mod. 2005. V. 17(8). P. 106–110.