L E C T U R E 27

Let $f(x_1, x_2, ..., x_m)$ be a joint density function of the component of m-dimensional random vector. We consider the problem of constructing a set of m-dimensional vectors corresponding to a probability distribution with a density function f, that is, constructing a sample with a given distribution. We denote by B the set of points $x \in \mathbf{R}^m$ such that $f(x_1, x_2, ..., x_m) > 0$ for $(x_1, ..., x_m) \in B$ and $f(x_1, x_2, ..., x_m) = 0$ if $(x_1, x_2, ..., x_m) \notin B$.

Consider a sequence of m-dimensional random vectors X_0, X_1,..., X_n,... that make up the Markov chain. For this Markov chain and for the Borel set $A \subset \mathbf{R}^m$, we denote by $P^{(n)}((x_1, x_2, ..., x_m), A)$ the transition probability from state $(x_1, x_2, ..., x_m)$ to the set A in exactly n steps (the condition that the set A is Borel is imposed so that it can be integrated over this set).

If for any Borel set $A \subset \mathbf{R}^m$ for any or almost any $(x_1, x_2, ..., x_m)$ we have

$$P^{(n)}(x, A) \to \int_A f(y_1, y_2, ..., y_m)\, dy_1\, dy_2 ...dy_m, \qquad \text{for} \qquad n \to +\infty \tag{36.1}$$

then realized the Markov chain, we can get a set of m-dimensional vectors (sample), corresponding to the probability distribution with the density function f. Of course, to satisfy condition (36.1), the Markov chain X_0, X_1,..., X_n,... must somehow be constructed using the function f. MCMC are used when working not only with continuous but also with discrete multidimensional distributions.

§36.1. METROPOLIS ALGORITHM.

The Markov chain X_0, X_1,..., X_n,... is sought among those, for which the transition probabilities in one step for any $(x_1, x_2, ..., x_m) \in B$ and for any Borel set $A \subset \mathbf{R}^m$ have the form:

$$P^{(1)}((x_1, x_2, ..., x_m), A) = g(x_1, x_2, ..., x_m) \cdot I_A(x_1, x_2, ..., x_m) +$$

$$+ \int_A h((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m))\, dy_1\, dy_2 ...dy_m. \tag{36.2}$$

Here $I_A(x_1, x_2, ..., x_m)$ is the indicator of a set A, that is $I_A(x_1, x_2, ..., x_m) = 1$, if $(x_1, x_2, ..., x_m) \in A$, and $I_A(x_1, x_2, ..., x_m) = 0$, if $(x_1, x_2, ..., x_m) \notin A$. It is obvious that

$$g(x_1, x_2, ..., x_m) = 1 - \int_{\mathbf{R}^m} h((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m))\, dy_1\, dy_2 ... dy_m, \tag{36.3}$$

since

$$P^{(1)}((x_1, x_2, ..., x_m), \mathbf{R}^m) = 1.$$

Function $h((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m))$ is satisfied the following reversibility condition:

$$f(x_1, x_2, ..., x_m) \cdot h((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) =$$

$$= f(y_1, y_2, ..., y_m) \cdot h((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)) \tag{36.4}$$

for any $(x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m) \in \mathbf{R}^m$.

Let's prove that from reversibility condition (36.4) follows the invariance condition:

$$\int_A f(y_1, y_2, ..., y_m)\, dy_1\, dy_2 ... dy_m = \int_{R^m} P^{(1)}(x_1, x_2, ..., x_m, A)\, f(x_1, x_2, ..., x_m)\, dx_1\, dx_2 ... dx_m \tag{36.5}$$

for any Borel set $A \subset \mathbf{R}^m$.

Really, using (36.4) and (36.3) we get:

$$\int_{R^m} P^{(1)}(x_1, x_2, ..., x_m, A)\, f(x_1, x_2, ..., x_m)\, dx_1\, dx_2 ... dx_m =$$

$$= \int_{R^m} f(x_1, x_2, ..., x_m)\, g(x_1, x_2, ..., x_m)\, I_A(x_1, x_2, ..., x_m)\ dx_1\, dx_2 ... dx_m +$$

$$+ \int_{R^m} \int_A f(x_1, x_2, ..., x_m)\, h(x_1, x_2, ..., x_m, y_1, y_2, ..., y_m)\, dy_1\, dy_2 ... dy_m\ dx_1\, dx_2 ... dx_m =$$

$$= \int_A f(x_1, x_2, ..., x_m)\, g(x_1, x_2, ..., x_m)\ dx_1\, dx_2 ... dx_m +$$

$$+ \int_{R^m} \int_A f(y_1, y_2, ..., y_m)\, h(y_1, y_2, ..., y_m, x_1, x_2, ..., x_m)\, dy_1\, dy_2 ... dy_m\, dx_1\, dx_2 ... dx_m =$$

$$= \int_A f(x_1, x_2, ..., x_m)\, g(x_1, x_2, ..., x_m)\ dx_1\, dx_2 ... dx_m +$$

$$+ \int_A \int_{R^m} f(y_1, y_2, ..., y_m)\, h(y_1, y_2, ..., y_m, x_1, x_2, ..., x_m)\ dx_1\, dx_2 ... dx_m\ dy_1\, dy_2 ... dy_m =$$

$$= \int_A f(x_1, x_2, ..., x_m)\, g(x_1, x_2, ..., x_m)\ dx_1\, dx_2 ... dx_m +$$

$$+ \int_A f(y_1, y_2, ..., y_m)\, (1 - g(y_1, y_2, ..., y_m))\ dy_1\, dy_2 ... dy_m = \int_A f(y_1, y_2, ..., y_m)\ dy_1\, dy_2 ... dy_m$$

that is (36.5).

It is interesting that the invariance condition (36.5) is the basic for validity of (36.1). In Metropolis algorithm very important is nonnegative function

$$c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)),$$

which we choose such that

$$\int_{R^m} c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m))\, dy_1\, dy_2 ... dy_m = 1 \tag{36.6}$$

for any $(x_1, x_2, ..., x_m) \in R^m$.

Next we denote

$$\alpha((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) =$$

$$\begin{cases} \min\left(1, \frac{f(y_1, y_2, ..., y_m)\, c((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m))}{f(x_1, x_2, ..., x_m)\, c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m))}\right) & \text{if } f(x_1, ..., x_m)\, c((x_1, ..., x_m), (y_1, ..., y_m)) > 0 \\ 1, & \text{if } f(x_1, ..., x_m)\, c((x_1, ..., x_m), (y_1, ..., y_m)) = 0. \end{cases}$$

We note that if $(x_1, x_2, ..., x_m) \in B$, $(y_1, y_2, ..., y_m) \notin B$ and $c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) > 0$, then $\alpha((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) = 0$.

Let us prove that the function

$$h((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) = c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m))\, \alpha((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m))$$

satisfy to reversibility condition (36.4).

We choose the points $(x_1, x_2, ..., x_m)$ and $(y_1, y_2, ..., y_m)$.

Assume that $f(x_1, x_2, ..., x_m)\, c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) = 0$. In this case the left-hand side of (36.4) equal to zero. But if $f(y_1, y_2, ..., y_m)\, c((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)) = 0$ then and the right-hand side of (36.4) equal to zero.

If $f(y_1, y_2, ..., y_m)\, c((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)) > 0$, then $\alpha((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)) = 0$ and therefore the right-hand side of (36.4) is equal to zero. Therefore, if

$$f(x_1, x_2, ..., x_m)\, c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) = 0,$$

then condition (36.4) is fulfilled.

Now let $f(x_1, x_2, ..., x_m)\, c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) > 0$. If

$$f(x_1, x_2, ..., x_m)\, c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) < f(y_1, y_2, ..., y_m)\, c((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)),$$

then

$$\alpha((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) = 1,$$

therefore we have

$$\alpha((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)) = \frac{f(x_1, x_2, ..., x_m)\, c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m))}{f(y_1, y_2, ..., y_m)\, c((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m))}.$$

Thus we have

$$f(x_1, x_2, ..., x_m)\, h((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) = f(x_1, x_2, ..., x_m)\, c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) =$$

$$= f(y_1, y_2, ..., y_m)\, c((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)) \cdot \alpha((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)) =$$

$$= f(y_1, y_2, ..., y_m)\, h((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)),$$

and condition (36.4) is fulfilled.

If

$$f(x_1, x_2, ..., x_m)\, c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) \geq f(y_1, y_2, ..., y_m)\, c((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)),$$

then

$$\alpha((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)) = 1,$$

therefore we have

$$\alpha((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) = \frac{f(y_1, y_2, ..., y_m)\, c((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m))}{f(x_1, x_2, ..., x_m)\, c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)}.$$

Thus we have

$$f(x_1, x_2, ..., x_m)\, h((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) =$$

$$= f(x_1, x_2, ..., x_m)\, c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m))\, \alpha((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) =$$

$$= f(y_1, y_2, ..., y_m)\, c((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)) =$$

$$= f(y_1, y_2, ..., y_m)\, c((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m))\, \alpha((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)) =$$

$$= f(y_1, y_2, ..., y_m)\, h((y_1, y_2, ..., y_m), (x_1, x_2, ..., x_m)),$$

and condition (36.4) is fulfilled. Therefore, $h((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m))$ defined by (36.7) satisfies the reversibility condition (36.4).

Note, that function $g(x_1, x_2, ..., x_m)$ is constructed by formula (36.7) and takes on only the values between zero and one.

Let $x^{(0)}$, $x^{(1)}$, ..., $x^{(n)}$,...is a realization of Markov chain X_0, X_1,..., X_n,.... Assume that $x^{(0)} \in B$ is fixed. In Metropolis algorithm the vector $x^{(n+1)}$ is constructed by vector $x^{(n)}$ using the following steps:

Step 1. m-dimensional vector $(y_1, y_2, ..., y_m)$ from the distribution $c(x^{(n)}, (y_1, y_2, ..., y_m))$ is simulated.

Step 2. Number u from uniformly distribution over the interval $[0, 1]$ is (from the table of random numbers).

Step 3. If $u < \alpha(x^{(n)}, (y_1, y_2, ..., y_m))$ we take $x^{(n+1)} = y$; but if $u \geq \alpha(x^{(n)}, (y_1, y_2, ..., y_m))$ we take $x^{(n+1)} = x^{(n)}$.

Not for any choice of a function $c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m))$ satisfying condition (36.6), the constructed set of m-dimensional vectors corresponds to a probability distribution with a density function $f(x_1, x_2, ..., x_m)$. For example, if $c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m)) = 0$ for any x, $y \in B$, then obviously this is not so. The reasoning below helps understand what conditions the function $c((x_1, x_2, ..., x_m), (y_1, y_2, ..., y_m))$ must satisfy.

We assume that the set B is bounded. Suppose also that this set can be divided into m-dimensional cubes A_1, A_2, ...,A_M, the length of the side of each of the cubes ε is a small positive number. Put

$$\pi_r = \int_{A_r} f(x_1, x_2, ..., x_m)\, dx_1\, dx_2\, ...dx_m, \qquad r = 1, ..., M.$$

Let $y^{(1)}$,..., $y^{(M)}$ are the centers of cubes A_1, ..., A_M. We will assume that $x^{(0)}$ is one of the points $y^{(1)}$,..., $y^{(M)}$. Further, suppose that any point y simulated in step 1, if it turned out to belong to A_r, is replaced by $y^{(r)}$. This leads to a new Markov chain with only a finite number of states. These states can be numbered with numbers $1, 2, ..., M$.

Now we show that for this new Markov chain with finite number of states probabilities (36.8) are approximately invariant.

Let $x^{(n)} = y^{(r)}$ in a Markov chain with finite number of states. New value $y^{(s)}$ obtained at Step 1 with probability approximately equals to $\varepsilon^m c(y^{(r)}, y^{(s)})$. This value is leaved with probability $\alpha(y^{(r)}, y^{(s)})$. By equality (36.7) the probability to obtain at Step 3 a new value $y^{(s)}$ approximately equals to

$$\varepsilon^m c(y^{(r)}, y^{(s)})\, \alpha(y^{(r)}, y^{(s)}) = \varepsilon^m h(y^{(r)}, y^{(s)}).$$

If a new value is not taken, then the Markov chain is leaved at state $y^{(r)}$. It is realized with probability, approximately equals to

$$1 - \sum_{s=1}^{M} \varepsilon^m h(y^{(r)}, y^{(s)}).$$

Replace this sum by integral we obtain another approximate expression for this probability

$$1 - \int_{R^m} h(y^{(r)}, y)\, dy.$$

By equation (36.3) the last expression is equal to $g(y^{(r)})$.

Thus, transition probabilities p_{rs}, that is probabilities $p^{(1)}(y^{(r)}, \{y^{(s)}\})$ for Markov chain with finite number of states approximately can be defined by the following formula:

$$p_{rs} = \begin{cases} g(y^{(r)}) + \varepsilon^m h(y^{(r)}, y^{(r)}), & \text{for} \quad r = s, \\ \varepsilon^m h(y^{(r)}, y^{(s)}), & \text{for} \quad r \neq s. \end{cases} \tag{36.9}$$

To find equations, whom approximately satisfy probabilities (36.8), we return to consideration of original Markov chain and use formula (36.5) for $A = A_s$ (invariance condition for original Markov chain is proved). For $A = A_s$ the left hand side of (36.5) equals to π_s. For the right hand side of (36.5) usin formula (36.2) and approximation $\pi_r = \varepsilon^m f(y^{(r)})$ the following chains of approximate equalities can be written:

$$\int_{R^m} \left(g(x)\, I_{A_s}(x) + \int_{A_s} h(x, y)\, dy \right) f(x)\, dx = \int_{A_s} g(x)\, f(x)\, dx + \int_{A_s} \int_{R^m} h(x, y)\, f(x)\, dx\, dy =$$

$$\varepsilon^m\, g(y^{(s)})\, f(y^{(s)}) + \int_{A_s} \sum_{r=1}^{M} h(y^{(r)}, y)\, \varepsilon^m\, f(y^{(r)})\, dy = g(y^{(s)})\, \pi_s + \varepsilon^m \sum_{r=1}^{M} h(y^{(r)}, y^{(s)})\, \pi_r = \sum_{r=1}^{M} p_{rs}\, \pi_r.$$

Thus, for probabilities π_r we obtain approximately the following system of linear equations

$$\sum_{r=1}^{M} \pi_r = 1, \quad \pi_s = \sum_{r=1}^{M} p_{rs}\, \pi_r, \quad \text{for} \quad s = 1, 2, ..., M. \tag{36.10}$$

If all transition probabilities p_{rs} for Markov chain with finite number of states are positive, then the system of equations (36.10) are satisfies the invariant probabilities π_s such that for any r and s

$$P^{(n)}(y^{(r)}, \{y^{(s)}\}) \to \pi_s \quad \text{as} \quad n \to \infty.$$

Under the positivity condition of all transition probabilities p_{rs} the system of linear equations (36.10) has a unique solution. If probabilities (36.8) satisfy to this system of linear equations, then they are invariant for Markov chain with finite number of states.

We note that these representations are approximate, not just.

Positivity of all transition probabilities p_{rs}, given by formula (36.9), guarantees by condition $c(x, y) > 0$ for any $x, y \in B$.

Example. Let f is a density function of two dimensional normal distribution

$$f(x) = \frac{1}{2\pi (det\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right), \tag{36.11}$$

where

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\,\sigma_2 \\ \rho\sigma_1\,\sigma_2 & \sigma_2^2 \end{pmatrix}$$

The sign $'$ means a transposition.

For construction of collection of vectors, which satisfy two dimensional normal distribution, it not need to use method Monte-Carlo Markov Chain. It is sufficient to construct a collection of vectors $z^{(0)}, z^{(1)}, ..., z^{(n)}, ...$ which satisfy two dimensional normal distribution with mean 0 and covariance matrix 1. Then we take a matrix

$$L = \begin{pmatrix} \sqrt{1-\rho^2}\sigma_1 & 0 \\ \rho\,\sigma_1 & \sigma_2 \end{pmatrix}$$

for which satisfied the relation $\Sigma = L'L$. We put

$$x^{(n)} = \mu + L'\, z^{(n)}, \quad n = 1, 2, ... \tag{36.12}$$

But we use density function (36.11) to illustrate Metropolis algorithm.

For each $a \geq 0$ in a plane R^2 with coordinates x_1, x_2 we can consider ellipse of dispersion

$$(x-\mu)' \Sigma^{-1} (x-\mu) = a^2,$$

where $x = (x_1, x_2)'$. Then the probability that two dimensional vector with density function (36.11) takes the values inside of this ellipse equals to

$$1 - \exp\left(-\frac{0.5a^2}{1-\rho^2}\right). \tag{36.13}$$

Consider a particular case: $\mu_1 = 0, \mu_2 = 0,\ \sigma_1 = 0.8,\ \sigma_2 = 1.2,\ \rho = 0.9.$

The probability p_1 calculated for each a by formula (36.13).

Firstly, for different N by formula (36.12) we construct a collection of points $x^{(1)}, ..., x^{(N)}$. By p_2 we denote a part of points from this collection, belonging inside of ellipse.

In the second, by issue of Metropolos algorithm, a Markov chain is realized. We use the function $c(x, y) = f_0(y - x)$, where f_0 is a density function of two dimensional normal distribution with 0 mean and covariance matrix

$$\begin{pmatrix} 0.36 & 0 \\ 0 & 0.36 \end{pmatrix}$$

The initial condition is $x^{(0)} = 0$, and the first five hundred constructed points are ignored. Thus, we consider a collection of points $x^{(500+1)}, ..., x^{(500+N)}$. By p_3 we denote the part of points from this collection, which belong to inside of ellipse of dispersion.