L E C T U R E 26

Appendix HW#5.

Problem 3. Given θ, the random variable X has a binomial distribution with $n = 3$ and probability of success θ. If the prior density of θ is

$$p(\theta) = \begin{cases} k & \text{if} \quad \frac{1}{2} < \theta < 1 \\ 0 & \text{otherwise} \end{cases}$$

what is the Bayes estimate of θ for an absolute error loss if $X = 1$.

Solution. We know that prior distribution of θ is the uniform distribution over the interval $\left(\frac{1}{2}, 1\right)$. Therefore, we have

$$p(\theta) = \begin{cases} 2, & \text{if} \quad \theta \in \left[\frac{1}{2}, 1\right] \\ 0, & \text{otherwise} \end{cases}$$

X has a binomial distribution with $n = 3$ and probability of success θ. Therefore, likelihood function has the following form:

$$f(X = 1/\theta) = \begin{cases} \binom{3}{1} \theta^1 (1-\theta)^2, & \text{if} \quad \theta \in \left[\frac{1}{2}, 1\right] \\ 0, & \text{otherwise} \end{cases}$$

Let us calculate joint density function:

$$f(X = 1, \theta) = f(X = 1/\theta) \cdot p(\theta) = \begin{cases} 6\,\theta^1 (1-\theta)^2, & \text{if} \quad \theta \in \left[\frac{1}{2}, 1\right] \\ 0, & \text{otherwise} \end{cases}$$

Calculate marginal density function. By definition, we have

$$f(X = 1) = \int_{-\infty}^{+\infty} f(X = 1, \theta)\, d\theta = \int_{1/2}^{1} f(X = 1, \theta)\, d\theta = 6 \int_{1/2}^{1} \theta (1-\theta)^2\, d\theta = \frac{5}{32}$$

Posterior density function has the form:

$$p(\theta | X = 1) = \frac{f(X = 1|\theta) \cdot p(\theta)}{f(X = 1)} = \begin{cases} \frac{6 \cdot 32}{5} \cdot \theta (1-\theta)^2 & \text{if} \quad \theta \in \left[\frac{1}{2}, 1\right] \\ 0, & \text{otherwise} \end{cases}$$

Bayes estimator for θ for absolute error loss function is the posterior median, that is

$$\int_{1/2}^{\hat\theta} p(\theta|X=1)\,d\theta = \frac{1}{2}$$

or

$$\frac{6\cdot 32}{5}\cdot\int_{1/2}^{\hat\theta}(\theta - 2\theta^2+\theta^3)\,d\theta = \frac{1}{2}.$$

We obtain the following equation

$$96\,\hat\theta^4 - 256\hat\theta^3 + 192\hat\theta^2 - 27 = 0$$

Thus we obtain

$$\hat\theta \approx 0.62065 \quad \text{and} \quad -0.31$$

and thus

$$\hat\theta = 0.62065.$$

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.