L E C T U R E 25

Appendix HW#4.

Problem 4. Suppose the prior distribution of θ is uniform over the interval $(2, 5)$. Given θ, X is uniform over the interval $(0, \theta)$. What is the Bayes estimator of θ for absolute error loss if $X = 1$?

Solution. We know that prior distribution of θ is uniform distribution, density function of it is the form:

$$p(\theta) = \begin{cases} \frac{1}{3}, & \text{if} \quad \theta \in [2, 5] \\ 0, & \text{otherwise} \end{cases}$$

and X is uniform over the interval $[0, \theta]$, that is, its density function has the form:

$$f(x/\theta) = \begin{cases} \frac{1}{\theta}, & \text{if} \quad x \in [0, \theta] \\ 0, & \text{otherwise} \end{cases}$$

Let us calculate joint density function:

$$f(x, \theta) = f(x/\theta) \cdot p(\theta) = \begin{cases} \frac{1}{3\,\theta}, & \text{if} \quad \theta \in [2, 5] \quad \text{and} \quad X \in [0, \theta] \\ 0, & \text{otherwise} \end{cases}$$

Calculate marginal density function. By definition, we have

$$f(x) = \int_{-\infty}^{+\infty} f(x, \theta)\, d\theta = \int_2^5 \frac{1}{3\,\theta}\, d\theta = \frac{1}{3} \ln \theta |_2^5 = \frac{1}{3}[\ln 5 - \ln 2] = \frac{1}{3} \ln \left(\frac{5}{2}\right)$$

Therefore,

$$f(x) = \begin{cases} \frac{1}{3} \ln \left(\frac{5}{2}\right), & \text{if} \quad x \in [0, \theta] \\ 0, & \text{otherwise} \end{cases}$$

Note that since $\theta \in [2, 5]$, then x may be any number from $[0, \theta]$, then $1 \in [0, \theta]$ for any θ from $[2, 5]$.

Posterior density function has the form:

$$p(\theta | x = 1) = \frac{f(x|\theta) \cdot p(\theta)}{f(1)} = \begin{cases} \dfrac{\frac{1}{3\,\theta}}{\frac{1}{3} \ln \left(\frac{5}{2}\right)} & \text{if} \quad \theta \in [2, 5] \\ 0, & \text{otherwise} \end{cases} =$$

$$= \begin{cases} \dfrac{1}{\theta \ln\left(\dfrac{5}{2}\right)} & \text{if} \quad \theta \in [2,5] \\ 0, & \text{otherwise} \end{cases}$$

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

$$\int_2^{\hat{\theta}} \frac{1}{\theta \ln\left(\dfrac{5}{2}\right)} d\theta = \frac{\ln\hat{\theta} - \ln 2}{\ln\left(\dfrac{5}{2}\right)} = \frac{1}{2}$$

and thus

$$\hat{\theta} = \sqrt{10}.$$

Appendix HW#5.

$$5\% = 0.05 \quad \text{and} \quad 0.25\% = 0.0025.$$

Problem 4. Given θ, the random variable X has a binomial distribution with $n = 2$ and probability of success θ. If the prior density of θ is

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

what is the Bayes estimate of θ for a squared error loss if $X_1 = 1$ and $X_2 = 2$.

Solution. We know that prior distribution of θ is uniform distribution, density function of it is the form ($k = 2$):

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

We perform 2 independent experiments and obtain a sample of size 2:

$$(X_1 = 1, X_2 = 2).$$

Therefore, likelihood function has the following form:

$$P(X_1 = 1, X_2 = 2|\theta) = P(X_1 = 1|\theta) \cdot P(X_2 = 2|\theta)) =$$

$$= \binom{2}{1} \theta \cdot (1-\theta) \cdot \binom{2}{2} \theta^2 \cdot (1-\theta)^0 = 2\theta^3 \cdot (1-\theta) \propto \theta^3 - \theta^4.$$

§35.1 Examples.

The inverse transform method can be used in practice as long as we are able to get an explicit formula for $F^{-1}(y)$ in closed form. We illustrate with some examples. We use the notation $U \sim \text{unif}(0,1)$ to denote that U is a random variable with the continuous uniform distribution over the interval $(0,1)$.

1. Exponential distribution. (see my previous lectures)

2. Discrete Random Variables. Discrete inverse-transform method.

Consider a non-negative discrete random variable X with probability mass function

$$p(k) = P(X = k), \qquad k \geq 0.$$

In this case, the construction $X = F^{-1}(U)$ is explicitly given by: $X = 0$ if $U \leq p(0)$,

$$X = k, \quad \text{if} \quad \sum_{i=0}^{k-1} p(i) < U \leq \sum_{i=0}^{k} p(i), \quad k \geq 1.$$

This is known as the discrete inverse-transform method. The algorithm is easily verified directly by recalling that $P(a < U \leq b) = b - a$, for $0 \leq a < b \leq 1$; here we use

$$a = \sum_{i=0}^{k-1} p(i) < b = \sum_{i=0}^{k} p(i),$$

and so $b - a = p(k)$.

3. Bernoulli (p) and Binomial (n,p) distributions.

Suppose we want to generate a Bernoulli (p) random variable X, in which case $P(X = 0) = 1-p$ and $P(X = 1) = p$ for some $p \in (0,1)$. Then the discrete inverse-transform method yields:

Algorithm for generating a Bernoulli (p) random variable X:

i. Generate $\sim U \sim \text{unif}(0,1)$.

ii. Set $X = 0$ if $U \leq 1 - p$; $X = 1$ if $U > 1 - p$.

Suppose we want X to have a binomial (n, p) distribution, that is,

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad 0 \le k \le n.$$

One could, in principle, use the discrete inverse-transform method with these $p(k)$, but we also can note that X can be represented (in distribution) as the sum of n independent and identically distribution Bernoulli (p) random variables, Y_1, ... , Y_n;

$$X = \sum_{i=0}^{n} Y_i,$$

the number of successes out of n independent and identically distributed Bernoulli (p) trials.

Alternative algorithm for generating a binomial (n, p) random variable X:

i. Generate n independent and identically distributed random variables U_1, U_2, ... ,U_n $\sim$ unif$(0, 1)$.

ii. For each $1 \le i \le n$, set $Y_i = 0$, if $U_i \le 1 - p$; $Y_i = 1$ if $U_i > 1 - p$. (This yields n independent and identically distributed Bernoulli (p) random variables.)

iii Set $X = \sum_{i=1}^{n} Y_i$.

The advantage of this algorithm is its simplicity, we do not need to do the various computations involving the $p(k)$. On the other hand, this algorithm requires n uniforms for each copy of X versus only one uniform when using the discrete inverse-transform method. Thus we might not want to use this algorithm when n is quite large.

In fact, when n is very large, and p is small, it follows from probability course (there is a theorem lurking here), that the distribution of X is very approximately the Poisson distribution with mean np. This motivates our next example.

4. Poisson distribution with mean α:

In this case

$$p(k) = P(X = k) = e^{-\alpha} \frac{\alpha^k}{k!}, \qquad k \ge 0.$$

We could thus use the discrete inverse-transform method, but of course it involves computing (in advance) pieces like

$$\frac{\alpha^k}{k!}.$$

Here we present an alternative algorithm that makes use of properties of a Poisson process at rate α. The trick is to recall that if $\{N(t) : t \geq 0\}$ is the counting process of a Poisson process at rate α, then $N(1)$ has a Poisson distribution with mean α. Thus if we can simulate $N(1)$, then we can set $X = N(1)$ and we are done. Let $Y = N(1) + 1$, and let $t_n = X_1 + ... + X_n$ denote the n-th point of the Poisson process; the X_i are independent and identically distributed with an exponential distribution at rate α. Note that

$$Y = \min\{n \geq 1 : t_n > 1\} = \min\{n \geq 1 : X_1 + ... + X_n > 1\},$$

a stopping time.

Using the inverse transform method to generate the independent and identically distributed exponential inter revert times X_i, we can represent

$$X_i = -\frac{1}{\alpha} \ln(U_i).$$

We then can rewrite (recalling that $\ln(xy) = \ln(x) + \ln(y)$)

$$Y = \min\{n \geq 1 : \ln(U_1) + ... + \ln(U_n) < -\alpha\} =$$

$$= \min\{n \geq 1 : \ln(U_1...U_n) < -\alpha\} = \min\{n \geq 1 : U_1 \cdot ... \cdot U_n < e^{-\alpha}\}.$$

We thus can simulate Y by simply consecutively generating independent uniforms U_i and taking the product until the product first falls below $e^{-\alpha}$. The number of uniforms required yields Y. Then we get

$$X = N(1) = Y - 1$$

as our desired Poisson. Here then is the resulting algorithm:

Alternative algorithm for generating a Poisson random variable X with mean α:

i. Set $X = 0$, $P = 1$.

ii. Generate $U \sim \text{unif}(0, 1)$, set $P = UP$.

iii. If $P < e^{-\alpha}$, then stop. Otherwise, if $P \geq e^{-\alpha}$, then set $X = X + 1$ and go back to ii.

§36. THE MARKOV CHAIN MONTE CARLO (MCMC) METHODS.

In this section, we consider two Markov Chain Monte Carlo (MCMC) methods that are widely used in mathematical, and in particularly, in economics research. This is the Metropolis algorithm and Gibbs choice. We will describe both of these methods. MCMC methods are designed to simulate samples of numbers or vectors, corresponding to one-dimensional or multidimensional probability distributions. In particular, these methods are used in Bayesian statistics to study posterior distributions. The observance of the invariance condition is essential, the proofs that this condition is fulfilled is given for both methods. To justify and study the methods, the theory of Markov chains with a finite number of states is used. Using several examples, the accuracy of the considered MCMC methods is investigated. These examples include two-dimensional normal distribution, a two-dimensional exponential distribution, a mixture of two-dimensional Normal distributions. Of great importance for the development of MCMC methods are modern high-speed computers. But the more complex models are used, the more difficult it is to study it with analytical methods, and the more often it is necessary to resort to numerical methods. 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 the 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 we have

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

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.