MAST90105: Lab and Workshop 8

1 Lab

In this lab, you will learn how to use R to find a minimum of function and to generate random variables from a given CDF which does not have its inverse in closed form.

1. Function `nlm()` in R can be used to find a minimum of a nonlinear function.

   a. Define a function $f(x_1, x_2) = \exp(x_1^2 + x_2^2) - 2(\exp(x_1) + \exp(x_2))$ in R and use `nlm()` to find the minimum of this function.

   b. As you have learned, to generate a random variable Z from a continuous distribution with given CDF $F_Z(z)$, one can follow these steps:

      i. Generate $U \sim U(0, 1)$

      ii. Compute $Z = F_Z^{-1}(U)$

   The quantile function $F_Z^{-1}(q)$ might not be available in closed form and can be computed numerically. For a strictly increasing CDF $F_Z(z)$ there exists a unique solution to the equation: $F_Z(z_q) = q$. It implies the function $G_Z(z) = (F_Z(z) - q)^2$ attains its minimum (which is zero) at $z = z_q$. The point z_q at which this minimum is attained can be found using `nlm()` function in R, and one can follow these steps to generate a random variable Z in this case:

      i. Generate $U \sim U(0, 1)$

      ii. Compute $Z = \text{argmin}_z (F_Z(z) - U)^2$

   Use this approach to generate a sample of size $N = 1000$ from a continuous random variable with the PDF:

$$f_Z(z) = \begin{cases} 0.5 + z^2 + z^5, & 0 < z < 1, \\ 0, & \text{otherwise.} \end{cases}$$

   Draw a histogram of the generated data.

2. A similar approach can be used to generate a random pair (X, Y) from a continuous bivariate distribution with given joint CDF $F_{X,Y}(x, y)$. Define

$$F_{Y|X}(y|x) = \Pr(Y \leq y | X = x) = \frac{\frac{\partial}{\partial x} F_{X,Y}(x, y)}{f_X(x)},$$

   where $F_X(x)$ and $f_X(x)$ are the marginal CDF and PDF of X, respectively. Assuming $F_X(x)$ and $F_{Y|X}(y|x)$ are strictly increasing and continuously differentiable functions of x and y, respectively, one can follow these steps to generate a random pair (X, Y):

   a. Generate independent $U_X \sim U(0, 1)$ and $U_Y \sim U(0, 1)$

   b. Compute $X = F_X^{-1}(U_X)$

c. Compute $Y = F^{-1}_{Y|X}(U_Y|X)$

The inverse CDFs F^{-1}_X and $F^{-1}_{Y|X}$ can be computed numerically using `nlm()` function if they are not available in closed form.

a. **Bonus:** Show that the joint CDF of (X, Y) generated using the algorithm above is $F_{X,Y}$.

b. Use this algorithm to generate a sample of size $N = 1000$ from a continuous bivariate random variable (X, Y) with the joint PDF:

$$f_{X,Y}(x, y) = \begin{cases} x + y, & 0 < x, y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Draw a scatter plot of the generated data.

2 Workshop

1. a. A random sample $X_1, \ldots, X_n$ of size n is taken from a Poisson distribution with mean $\lambda > 0$.

   i. Show the maximum likelihood estimator of λ is $\hat{\lambda} = \bar{X}$.

   ii. Suppose with $n = 40$ we observe 5 zeros, 7 ones, 12 twos, 9 threes, 5 fours, 1 five, and 1 six. What is the maximum likelihood estimate of λ.

   b. Find the maximum likelihood estimator, $\hat{\theta}$, if $X_1, \cdots, X_n$ is a random sample from the following probability density function:

$$f(x; \theta) = (1/2) \exp(-|x - \theta|), -\infty < x < \infty, 0 < \theta < \infty$$

   This involves minimizing $\sum_{i=1}^{n} |x_i - \theta|$, which is difficult. Try $n = 5$ and a sample 6.1, -1.1, 3.2, 0.7, 1.7.Then deduce the MLE.

2. Let $f(x; \theta) = \theta x^{\theta-1}$, $0 < x < 1$, $\theta \in \Omega = \{\theta : 0 < \theta < \infty\}$ and let $X_1, \ldots, X_n$ denote a random sample from this distribution. Note that

$$\int_0^1 x\theta\, x^{\theta-1} dx = \frac{\theta}{\theta + 1}$$

   a. Sketch the p.d.f. of X for $\theta = 1/2$ and $\theta = 2$.

   b. Show that $\hat{\theta} = -n/\ln\left(\prod_{i=1}^{n} X_i\right)$ is the maximum likelihood estimator of θ.

   c. For each of the following three sets of observations from this distribution compute the maximum likelihood estimates and the methods of moments estimates.

X	Y	Z
0.0256	0.9960	0.4698
0.3051	0.3125	0.3675
0.0278	0.4374	0.5991
0.8971	0.7464	0.9513
0.0739	0.8278	0.6049
0.3191	0.9518	0.9917
0.7379	0.9924	0.1551
0.3671	0.7112	0.0710
0.9763	0.2228	0.2110
0.0102	0.8609	0.2154

($\sum_{i=1}^n \ln(x_i) = -18.2063$, $\sum_{i=1}^n \ln(y_i) = -4.5246$, $\sum_{i=1}^n \ln(z_i) = -10.42968$, $\sum_{i=1}^n x_i = 3.7401$, $\sum_{i=1}^n y_i = 7.0592$, $\sum_{i=1}^n z_i = 4.6368$.)

3. Let $X_1, \dots, X_n$ be a random sample from the exponential distribution whose p.d.f. is $f(x;\theta) = (1/\theta)\exp(-x/\theta)$, $0 < x < \infty$, $0 < \theta < \infty$.

   a. Show that $\bar{X}$ is an unbiased estimator of θ.

   b. Show that the variance of $\bar{X}$ is θ^2/n. What is a good estimate of θ if a random sample of size 5 yielded the values 3.5, 8.1, 0.9, 4.4 and 0.5?

4. Let $X_1, \dots, X_n$ be a random sample from a distribution having finite variance σ^2. Show that

$$S^2 = \sum_{i=1}^n \frac{\left(X_i - \bar{X}\right)^2}{n-1}$$

   is an unbiased estimator of σ^2. HINT: Write

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right)$$

   and compute $E(S^2)$.

5. Let $X_1, \dots, X_n$ be a random sample of size n from the distribution with p.d.f. $f(x;\theta) = (1/\theta)x^{(1-\theta)/\theta}$, $0 < x < 1$.

a. Show the mean of X is $E(X) = 1/(1+\theta)$.

b. Show the maximum likelihood estimator of θ is

$$\hat{\theta} = -\frac{1}{n}\sum_{i=1}^{n} \ln X_i$$

c. Is the MLE unbiased? (You can do this using integration by parts combined with rules about expectation of the sample mean.)

d. Show the method of moments estimator of θ is

$$\widetilde{\theta} = \frac{1-\bar{X}}{\bar{X}}.$$