MAT7035: Computational Statistics

Midterm Test
(16:20–18:20, 12/09/2019)

1. [**15 marks**] Use the inversion method to generate a random variable from the following distribution, and write down the algorithm:

(a) (Zero-truncated binomial distribution) The *probability mass function* (pmf) is $p_x = \Pr(X = x) = c \cdot \binom{m}{x}\theta^x(1-\theta)^{m-x}$, $x = 1, 2, \ldots, m$, where m is a known positive integer, $\theta \in (0, 1)$ and c is the normalizing constant related to θ. Denote the value of c by θ before generating this zero-truncated binomial distribution.

(b) (The standard Gumbel maximum distribution) The density function is $f(x) = \mathrm{e}^{-x}\exp(-\mathrm{e}^{-x})$, where $-\infty < x < +\infty$.

2. [**20 marks**] Suppose that we want to draw random samples from the target density $f(x)$ with support $\mathcal{S}_X$. Furthermore, we assume that there exist an envelope constant $c\,(\geqslant 1)$ and an envelope density $g(x)$ having the same support $\mathcal{S}_X$ such that $f(x) \leqslant cg(x)$ for all $x \in \mathcal{S}_X$.

(a) State the rejection method for generating one random sample X from $f(x)$.

(b) Using the following exponential density $g(x) = \frac{2}{3}\,\mathrm{e}^{-2x/3}$ for $x > 0$, as the envelope function to generate a random variable having the gamma density

$$f(x) = \frac{1}{\Gamma(3/2)}x^{1/2}\,\mathrm{e}^{-x}, \quad x > 0,$$

by the rejection method. [Hint: $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, $\Gamma(0.5) = \sqrt{\pi}$]

(c) Calculate the value of the acceptance probability.

3. [**15 marks**] Let $X = (X_1, X_2)^\top \sim N_2(\mu, \Sigma)$ with the joint density

$$f(x_1, x_2) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu) \right\},$$

where

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

(a) Derive the marginal density of X_1 and the conditional density of $X_2|(X_1 = x_1)$.

(b) State the conditional sampling algorithm for generating one random sample $X = (X_1, X_2)^\top$ from $f(x_1, x_2)$.

4. [**20 marks**] Let $g(x)$ be a function defined on the real line.

(a) Derive Newton's method for finding the root of $g(x) = 0$ by using the first-order Taylor approximation.

(b) Use Newton's method to calculate the numerical solution to the unique root of the equation $g(x) = 0$ on the interval $(0, \infty)$, where $g(x) = 1.95 - \mathrm{e}^{-2/x} - 2\,\mathrm{e}^{-x^4}$. The initial value is set to be $x^{(0)} = 1$. The stoping rule is: If $|x^{(t+1)} - x^{(t)}| < 10^{-5}$, stop the iteration.

5. [**30 marks**] Let $Y_{\text{obs}} = \{n_1, \ldots, n_4; m_1, m_2\}$ denote the observed frequencies and $\theta = (\theta_1, \ldots, \theta_4)^\top$ be the cell probability vector satisfying $\theta_i \geqslant 0$, $\theta_1 + \cdots + \theta_4 = 1$. Suppose that the observed-data likelihood function of θ is given by

$$L(\theta|Y_{\text{obs}}) \propto \left(\prod_{i=1}^4 \theta_i^{n_i}\right) \times (\theta_1 + \theta_2)^{m_1} \times (\theta_1 + \theta_2 + \theta_3)^{m_2}.$$

Use the EM algorithm to find the maximum likelihood estimates of θ.

=== END OF THE PAPER ===

1. <u>**Solution**</u>. (a) **Similar to Example 1.7 in page 10.** Note that 0 is truncated from the support. Thus

$$\frac{1}{c} = 1 - (1-\theta)^m \quad \text{or} \quad c = \frac{1}{1-(1-\theta)^m}.$$

We have $p_1 = \frac{m\theta(1-\theta)^{m-1}}{1-(1-\theta)^m}$ and the recursive identity between p_{x+1} and p_x is

$$\frac{p_{x+1}}{p_x} = \frac{c \cdot \binom{m}{x+1}\theta^{x+1}(1-\theta)^{m-x-1}}{c \cdot \binom{m}{x}\theta^x(1-\theta)^{m-x}} = \frac{(m-x)\theta}{(x+1)(1-\theta)}.$$

The algorithm is as follows:

Step 1: Generate $U = u$ from $U(0,1)$;

Step 2: Let $i = 1$, $p = p_1$ and $F = p$;

Step 3: If $U < F$, set $X = i$ and stop;

Step 4: Otherwise, let $p \leftarrow \frac{(m-i)\theta}{(i+1)(1-\theta)}p$, $F \leftarrow F + p$, $i \leftarrow i+1$ and go back to step 3.

(b) **This is a special case of Q1.1(f) in Assignment 1 with $\mu = 0$ and $\sigma = 1$.** The cdf of the distribution with density $f(x) = \mathrm{e}^{-x}\exp(-\mathrm{e}^{-x})$ is given by

$$F(x) = \exp(-\mathrm{e}^{-x}), \quad x \in \mathbb{R}.$$

Its inverse function is

$$F^{-1}(x) = -\log[-\log(x)], \quad x \in (0,1).$$

Thus, $F(X) \overset{d}{=} U \sim U(0,1)$ implies

$$X \overset{d}{=} F^{-1}(U) = -\log[-\log(U)].$$

The algorithm is as follows:
Step 1: Draw U from $U(0,1)$;
Step 2: Return $X = -\log[-\log(U)]$.

2. <u>**Solution**</u>. (a) THE REJECTION ALGORITHM:

Step 1. Draw $U \sim U(0,1)$ and independently draw $Y \sim g(\cdot)$;

Step 2. If $U \leqslant \frac{f(Y)}{cg(Y)}$, set $X = Y$; otherwise, go to Step 1.

(b) **This is a special case of Example 1.10 in pages 20–22 with $\theta = 2/3$.** By differentiating the ratio

$$\frac{f(x)}{g(x)} = \frac{3}{2\Gamma(3/2)} x^{1/2}\,\mathrm{e}^{-x/3}$$

with respect to x and setting the resultant derivative equal to zero, we obtain the maximal value of this ratio at $x = 3/2$. Hence

$$c = \max_{x>0} \frac{f(x)}{g(x)} = \frac{3^{3/2}\,\mathrm{e}^{-0.5}}{2^{3/2}\Gamma(3/2)},$$

and

$$\frac{f(x)}{cg(x)} = \left(\frac{2\,\mathrm{e}x}{3}\right)^{1/2} \mathrm{e}^{-x/3}.$$

On the other hand, the distribution function corresponding to the exponential density $g(x)$ is

$$G(x) = \int_0^x g(t)\,\mathrm{d}t = 1 - \frac{2}{3}\,\mathrm{e}^{-2x/3}, \quad x > 0.$$

Its inverse function is $G^{-1}(u) = -\frac{3}{2}\log(1-u)$, $0 < u < 1$.

The gamma(3/2, 1) random variable can be generated as follows:

Step 1. Draw $U_1, U_2 \overset{\text{iid}}{\sim} U(0,1)$ and set $Y = -1.5\log(U_1)$;

Step 2. If $U_2 \leqslant (2\,\mathrm{e}Y/3)^{1/2}\,\mathrm{e}^{-Y/3}$, set $X = Y$; otherwise, go to Step 1.

(c) The acceptance probability for the current rejection algorithm is

$$c^{-1} = \frac{2^{3/2}\Gamma(3/2)}{3^{3/2}\,\mathrm{e}^{-0.5}} = 0.79534.$$

3. Solution. (a) The marginal distribution of X_1 is $N(\mu_1, \sigma_{11})$. The conditional distribution of $X_2|(X_1 = x_1)$ is

$$N(\mu_2 + \sigma_{21}\sigma_{11}^{-1}(x_1 - \mu_1),\ \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12}).$$

(b) The conditional sampling algorithm is as follows:

Step 1. Draw $X_1 = x_1$ from $N(\mu_1, \sigma_{11})$;

Step 2. Draw $X_2 = x_2$ from $N(\mu_2 + \sigma_{21}\sigma_{11}^{-1}(x_1 - \mu_1),\ \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12})$.

4. **Solution**. (a) The first-order Taylor expansion of $g(x)$ around x_0 is

$$g(x) \doteq g(x_0) + (x - x_0)g'(x_0).$$

Since $g(x) = 0$, we replace x and x_0 by $x^{(t+1)}$ and $x^{(t)}$, respectively and have

$$x^{(t+1)} = x^{(t)} - \frac{g(x^{(t)})}{g'(x^{(t)})}. \tag{MT.1}$$

(b) **Similar to Example 2.2 in pages 63–64.** It is easy to verify that

$$g'(x) = -2x^{-2}\,\mathrm{e}^{-2/x} + 8x^3\,\mathrm{e}^{-x^4}.$$

Let $x^{(0)} = 1$, then from (MT.1), we obtain

$$\begin{aligned}
x^{(1)} &= x^{(0)} - \frac{g(x^{(0)})}{g'(x^{(0)})} = 1 - \frac{1.0789}{2.6724} = 0.596273,\\
x^{(2)} &= x^{(1)} - \frac{g(x^{(1)})}{g'(x^{(1)})} = x^{(1)} - \frac{0.1526}{1.2981} = 0.478749,\\
x^{(3)} &= x^{(2)} - \frac{g(x^{(2)})}{g'(x^{(2)})} = x^{(2)} - \frac{0.0370}{0.6991} = 0.425798,\\
x^{(4)} &= x^{(3)} - \frac{g(x^{(3)})}{g'(x^{(3)})} = x^{(3)} - \frac{0.0056}{0.4969} = 0.414628,\\
x^{(5)} &= x^{(4)} - \frac{g(x^{(4)})}{g'(x^{(4)})} = x^{(4)} - \frac{2.0768 \times 10^{-4}}{0.460135} = 0.414177,\\
x^{(6)} &= x^{(5)} - \frac{g(x^{(5)})}{g'(x^{(5)})} = x^{(5)} - \frac{3.2684 \times 10^{-7}}{0.458687} = 0.414176, \quad \text{and}\\
x^{(7)} &= x^{(6)} - \frac{g(x^{(6)})}{g'(x^{(6)})} = x^{(6)} - \frac{8.1334 \times 10^{-13}}{0.458685} = 0.414176.
\end{aligned}$$

Thus, $x^{(\infty)} = x^{(7)} = 0.414176$ is the unique root of $g(x) = 0$.

5. **Solution**. **Similar to Q2.4 in Assignment 2.** First, we introduce a latent random variable W to split the term $(\theta_1+\theta_2)^{m_1}$ so that the conditional predictive distribution is

$$W|(m_1,\theta) \sim \text{Binomial}\left(m_1; \frac{\theta_1}{\theta_1+\theta_2}\right),$$

and

$$E(W|m_1,\theta) = \frac{m_1\theta_1}{\theta_1+\theta_2}. \tag{MT.2}$$

Next, we introduce a latent vector $Z=(Z_1,Z_2,Z_3)^\top$ to split the term $(\theta_1+\theta_2+\theta_3)^{m_2}$ so that the conditional predictive distribution is

$$Z|(m_2,\theta) \sim \text{Multinomial}_3\left(m_2; \frac{\theta_1}{\theta_{123}}, \frac{\theta_2}{\theta_{123}}, \frac{\theta_3}{\theta_{123}}\right),$$

where $\theta_{123} \hat{=} \theta_1+\theta_2+\theta_3$ and $Z_1+Z_2+Z_3=m_2$. The conditional expectations are given by

$$E(Z_i|m_2,\theta) = \frac{m_2\theta_i}{\theta_1+\theta_2+\theta_3}, \quad i=1,2,3. \tag{MT.3}$$

Note that W is independent of Z, the complete-data likelihood function is given by

$$L(\theta|Y_{\text{obs}},W,Z) \propto \theta_1^{n_1+W+Z_1}\theta_2^{n_2+m_1-W+Z_2}\theta_3^{n_3+Z_3}\theta_4^{n_4}.$$

Taking log, we obtain

$$\ell(\theta|Y_{\text{obs}},W,Z) = \log L(\theta|Y_{\text{obs}},W,Z) = (n_1+W+Z_1)\log\theta_1$$
$$+(n_2+m_1-W+Z_2)\log\theta_2+(n_3+Z_3)\log\theta_3+n_4\log\theta_4.$$

Thus, the E-step of the EM algorithm is to compute the conditional expectations (MT.2) and (MT.3), and the M-step of the EM algorithm is to update the complete-data MLEs

$$\begin{aligned}
\hat\theta_1 &= \frac{n_1+W+Z_1}{n+m_1+m_2}, & \hat\theta_2 &= \frac{n_2+m_1-W+Z_2}{n+m_1+m_2},\\
\hat\theta_3 &= \frac{n_3+Z_3}{n+m_1+m_2}, & \hat\theta_4 &= \frac{n_4}{n+m_1+m_2},
\end{aligned}$$

by replacing W and Z_i with $E(W|m_1,\theta)$ and $E(Z_i|m_2,\theta)$, where $n = n_1 + n_2 + n_3 + n_4$.