MAT7035: Computational Statistics

Midterm Test (16:20–18:20, 20 DEC 2021)

1. [**20 marks**] Use the inversion method to generate a random variable from the following distribution, and write down the algorithm:

(a) (Right-truncated Poisson distribution) Let $X \sim \text{RTP}(\lambda)$ and its *probability mass function* (pmf) be

$$p_i = \Pr(X = i) = c^{-1} \cdot \frac{\lambda^i \, \mathrm{e}^{-\lambda}}{i!}, \quad i = 0, 1, \ldots, m,$$

where $m\,(> 1)$ is a known integer, $\lambda\,(> 0)$ is the unknown parameter and c is the normalizing constant related to λ. Denote the value of c by λ before generating this right-truncated Poisson distribution. **[10 marks]**

(b) Let $X \sim f_X(x)$ and its *probability density function* (pdf) be $f_X(x) = 2(\mathrm{e}^{-x} - \mathrm{e}^{-2x})$, $x > 0$. **[10 marks]**

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 $f(x)$. **[3 marks]**

(b) Use a density, selected from the following family of exponential densities

$$g_\theta(x) = \theta\, \mathrm{e}^{-\theta x}, \quad x \geqslant 0, \quad \theta > 0,$$

as the optimal envelope function (i.e., with the largest acceptance probability) to generate a random variable having the half-normal distribution with pdf

$$f(x) = \sqrt{\frac{2}{\pi}}\, \mathrm{e}^{-x^2/2}. \quad x \geqslant 0,$$

by the rejection method. **[15 marks]**

(c) Calculate the value of the acceptance probability. **[2 marks]**

3. [**5 marks**] State the *sampling/importance resampling* (SIR) method for generating random samples of the random variable X from its pdf $f(x)$ with an importance sampling density $g(x)$.

4. [**25 marks**] Let the binary responses $y_1, \ldots, y_n$ be the corresponding realizations of independent random variables $Y_1, \ldots, Y_n$, and

$$\begin{aligned} Y_i &\sim \text{Bernoulli}(p_i), \\ p_i &= 1 - \exp[-\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})], \quad i = 1, \ldots, n, \end{aligned}$$

where $\boldsymbol{x}_i$ is a known vector of explanatory variables for subject i, and $\boldsymbol{\beta}_{p\times 1}$ is an unknown vector of parameters.

(a) Derive the score vector, the observed information matrix and the Fisher information matrix. [**20 marks**]

(b) Using the Fisher scoring algorithm to find the MLE $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ and the estimated asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$. [**5 marks**]

5. [**30 marks**] Let $Y_{\text{obs}} = \{n_1, \ldots, n_5; m_1, m_2\}$ denote the observed frequencies and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_5)^\top$ be the cell probability vector satisfying $\theta_i \geqslant 0$, $\theta_1 + \cdots + \theta_5 = 1$. Suppose that the observed-data likelihood function of $\boldsymbol{\theta}$ is given by

$$L(\boldsymbol{\theta}|Y_{\text{obs}}) \propto \left(\prod_{i=1}^5 \theta_i^{n_i}\right) \times (\theta_1 + \theta_2)^{m_1} \times (\theta_3 + \theta_4 + \theta_5)^{m_2}.$$

Use the EM algorithm to find the maximum likelihood estimates of $\boldsymbol{\theta}$.

=== END OF THE PAPER ===

1. <u>Solution</u>. (a) **Similar to Example 1.6 on page 10 and Example 1.7 on page 11.** From $1 = \sum_{i=0}^{m} p_i$, we have

$$c = \sum_{j=0}^{m} \frac{\lambda^j \,\mathrm{e}^{-\lambda}}{j!}. \qquad \textbf{[2 marks]}$$

Note that $p_0 = \mathrm{e}^{-\lambda}/c$, **[1 mark]**
and the recursive identity between p_{i+1} and p_i is

$$\frac{p_{i+1}}{p_i} = \frac{c^{-1}\cdot\dfrac{\lambda^{i+1}\,\mathrm{e}^{-\lambda}}{(i+1)!}}{c^{-1}\cdot\dfrac{\lambda^{i}\,\mathrm{e}^{-\lambda}}{i!}} = \frac{\lambda}{i+1}, \quad i \geqslant 0, \qquad \textbf{[2 marks]}$$

the algorithm is as follows:

Step 1: Generate $U = u$ from $U(0,1)$;
Step 2: Let $i \leftarrow 0$, $p \leftarrow p_0$ and $F \leftarrow p$;
Step 3: If $U \leqslant F$, set $X = i$ and stop;
Step 4: Otherwise, let $p \leftarrow p\lambda/(i+1)$, $F \leftarrow F + p$, $i \leftarrow i+1$ and go back to step 3. **[5 marks]**

(b) The cdf of the r.v. X with density $f_X(x) = 2(\mathrm{e}^{-x} - \mathrm{e}^{-2x})$ for $x > 0$ is given by

$$\begin{aligned} F(x) &= \int_0^x f_X(t)\,\mathrm{d}t = \int_0^x (2\,\mathrm{e}^{-t} - 2\,\mathrm{e}^{-2t})\,\mathrm{d}t \\ &= (-2\,\mathrm{e}^{-t} + \mathrm{e}^{-2t})\Big|_0^x = -2\,\mathrm{e}^{-x} + 2 + \mathrm{e}^{-2x} - 1 \\ &= 1 - 2\,\mathrm{e}^{-x} + (\mathrm{e}^{-x})^2 = (1 - \mathrm{e}^{-x})^2, \quad x > 0. \end{aligned} \qquad \textbf{[4 marks]}$$

From $F(x) = u$, we have

$$x = F^{-1}(u) = -\log(1 - \sqrt{u}\,), \quad u \in (0,1). \qquad \textbf{[2 marks]}$$

Thus,

$$X \overset{\mathrm{d}}{=} F^{-1}(U) \overset{\mathrm{d}}{=} -\log(1 - \sqrt{U}\,).$$

The algorithm is as follows:
Step 1: Draw $U = u$ from $U(0,1)$;
Step 2: Return $x = -\log(1 - \sqrt{u}\,)$. **[4 marks]**

2. **Solution**. (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. [**3 marks**]

(b) **See Example T1.8 in Tutorial 1.** The ratio is

$$\frac{f(x)}{g_\theta(x)} = \frac{\sqrt{\frac{2}{\pi}}\,\mathrm{e}^{-x^2/2}}{\theta\,\mathrm{e}^{-\theta x}} = \sqrt{\frac{2}{\pi}}\theta^{-1}\,\mathrm{e}^{-x^2/2+\theta x}. \qquad [\textbf{2 marks}]$$

Let

$$0 = \frac{\mathrm{d}}{\mathrm{d}x}\log\left[\frac{f(x)}{g_\theta(x)}\right] = \frac{\mathrm{d}}{\mathrm{d}x}\left(\text{constant} - \frac{x^2}{2} + \theta x\right) = -x+\theta,$$

we obtain that the maximal value of this ratio is arrived at $x = \theta$. Thus

$$c_\theta = \frac{f(\theta)}{g_\theta(\theta)} = \sqrt{\frac{2}{\pi}}\theta^{-1}\,\mathrm{e}^{\theta^2/2}, \qquad [\textbf{5 marks}]$$

and

$$c_{\text{opt}} = \min_{\theta>0} c_\theta = \min_{\theta>0}\left(\sqrt{\frac{2}{\pi}}\theta^{-1}\,\mathrm{e}^{\theta^2/2}\right).$$

Let

$$H(\theta) = \log\left(\theta^{-1}\,\mathrm{e}^{\theta^2/2}\right) = -\log\theta + \frac{\theta^2}{2},$$

and set

$$0 = H'(\theta) = -\frac{1}{\theta} + \theta,$$

we have $\theta = 1$. Hence

$$c_{\text{opt}} = \sqrt{\frac{2\,\mathrm{e}}{\pi}} \quad \text{and} \quad \frac{f(x)}{c_{\text{opt}}g_\theta(x)} = \frac{f(x)}{c_{\text{opt}}g_1(x)} = \mathrm{e}^{-x^2/2+x-0.5}. \quad [\textbf{5 marks}]$$

The rejection method:

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

Step 2. If $U_1 \leqslant \exp(-\frac{Y^2}{2} + Y - 0.5)$, return $X = Y$; otherwise, go to Step 1. [**3 marks**]

(c) The acceptance probability is $1/c_{\text{opt}} = 0.7602$. [**2 marks**]

3. **Solution**. The SIR algorithm is as follows:

Step 1: Generate $X^{(1)}, \ldots, X^{(J)} \overset{\text{iid}}{\sim} g(\cdot)$; **[2 marks]**

Step 2: Select a subset $\{X^{(k_i)}\}_{i=1}^{m}$ from $\{X^{(j)}\}_{j=1}^{J}$ via resampling *without replacement* from the discrete distribution on $\{X^{(j)}\}$ with probabilities $\{\omega_j\}$, where

$$\omega_j = \frac{f(X^{(j)})/g(X^{(j)})}{\sum_{j'=1}^{J} f(X^{(j')})/g(X^{(j')})}, \quad j = 1, \ldots, J. \quad \text{[3 marks]}$$

4. **Solution**. REMARKS: This question is very similar to Exercise 2.17 on page 119 of the Textbook.

(a) Define $F(z) = 1 - \exp(-\,\mathrm{e}^z)$ and $f(z) = F'(z) = \mathrm{e}^z \exp(-\,\mathrm{e}^z)$. The likelihood function of $\boldsymbol{\beta}$ is

$$L(\boldsymbol{\beta}) = \textstyle\prod_{i=1}^n p_i^{y_i}(1-p_i)^{1-y_i}.$$

so that the log-likelihood function is

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \Big[y_i \log(p_i) + (1-y_i)\log(1-p_i)\Big]. \qquad [\textbf{5 marks}]$$

Since

$$\frac{\partial p_i}{\partial \boldsymbol{\beta}} = \frac{\partial F(\boldsymbol{x}_i^\top \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = f(\boldsymbol{x}_i^\top \boldsymbol{\beta})\boldsymbol{x}_i,$$

the score vector, the observed information matrix, and the Fisher information matrix are respectively given by

$$\nabla \ell(\boldsymbol{\beta}) = \sum_{i=1}^n \boldsymbol{x}_i \left(\frac{y_i}{p_i} - \frac{1-y_i}{1-p_i}\right) f(\boldsymbol{x}_i^\top \boldsymbol{\beta}), \qquad [\textbf{5 marks}]$$

$$\begin{aligned} -\nabla^2 \ell(\boldsymbol{\beta}) = \boldsymbol{I}(\boldsymbol{\beta}) &= \sum_{i=1}^n \boldsymbol{x}_i \left\{ \left[\frac{y_i}{p_i^2} + \frac{1-y_i}{(1-p_i)^2}\right] f(\boldsymbol{x}_i^\top \boldsymbol{\beta}) \right. \\ &\quad \left. + \left(\frac{y_i}{p_i} - \frac{1-y_i}{1-p_i}\right) \boldsymbol{x}_i^\top \boldsymbol{\beta} \right\} f(\boldsymbol{x}_i^\top \boldsymbol{\beta}) \boldsymbol{x}_i^\top, \qquad [\textbf{5 marks}] \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{J}(\boldsymbol{\beta}) = E[-\nabla^2 \ell(\boldsymbol{\beta})] &= \sum_{i=1}^n \boldsymbol{x}_i \left(\frac{1}{p_i} + \frac{1}{1-p_i}\right) f^2(\boldsymbol{x}_i^\top \boldsymbol{\beta}) \boldsymbol{x}_i^\top \\ &= \sum_{i=1}^n \boldsymbol{x}_i \frac{f^2(\boldsymbol{x}_i^\top \boldsymbol{\beta})}{p_i(1-p_i)} \boldsymbol{x}_i^\top. \qquad [\textbf{5 marks}] \end{aligned}$$

(b) The Fisher scoring algorithm updates

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + [\boldsymbol{J}(\boldsymbol{\beta}^{(t)})]^{-1} \nabla \ell(\boldsymbol{\beta}^{(t)}). \qquad [\textbf{3 marks}]$$

The estimated asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}}) = [\boldsymbol{J}(\hat{\boldsymbol{\beta}})]^{-1}. \qquad [\textbf{2 marks}]$$

5. **Solution**. **Similar to Example T5.2 in Tutorial 5.** First, we introduce a latent random variable Z_1 to split the term $(\theta_1+\theta_2)^{m_1}$ so that the conditional predictive distribution is

$$Z_1|(m_1,\boldsymbol{\theta}) \sim \text{Binomial}\left(m_1, \frac{\theta_1}{\theta_1+\theta_2}\right),$$

and

$$E(Z_1|m_1,\boldsymbol{\theta}) = \frac{m_1\theta_1}{\theta_1+\theta_2}. \tag{MT.1}$$

[5 marks]

Next, we introduce a latent vector $Z=(Z_3,Z_4,Z_5)^\top$ to split the term $(\theta_3+\theta_4+\theta_5)^{m_2}$ so that the conditional predictive distribution is

$$Z|(m_2,\boldsymbol{\theta}) \sim \text{Multinomial}_3\left(m_2; \frac{\theta_3}{\theta_{345}}, \frac{\theta_4}{\theta_{345}}, \frac{\theta_5}{\theta_{345}}\right),$$

where $\theta_{345} \hat{=} \theta_3+\theta_4+\theta_5$ and $Z_3+Z_4+Z_5=m_2$. The conditional expectations are given by

$$E(Z_i|m_2,\boldsymbol{\theta}) = \frac{m_2\theta_i}{\theta_3+\theta_4+\theta_5}, \quad i=3,4,5. \tag{MT.2}$$

[5 marks]

Note that $Z_1 \perp\!\!\!\perp Z$, the complete-data likelihood function is given by

$$L(\boldsymbol{\theta}|Y_{\text{obs}},Z_1,Z) \propto \theta_1^{n_1+Z_1}\theta_2^{n_2+Z_2}\theta_3^{n_3+Z_3}\theta_4^{n_4+Z_4}\theta_5^{n_5+Z_5} = \prod_{i=1}^{5}\theta_i^{n_i+Z_i},$$

where $Z_2 \hat{=} m_2 - Z_1$. Taking logarithm, we obtain

$$\ell(\boldsymbol{\theta}|Y_{\text{obs}},Z_1,Z) = \log L(\boldsymbol{\theta}|Y_{\text{obs}},Z_1,Z) = \sum_{i=1}^{5}(n_i+Z_i)\log(\theta_i).$$

[10 marks]

Thus, the E-step of the EM algorithm is to compute the conditional expectations (MT.1) and (MT.2), and the M-step of the EM algorithm is to update the complete-data MLEs

$$\hat{\theta}_i = \frac{n_i+Z_i}{n+m_1+m_2}, \quad i=1,\ldots,5,$$

by replacing Z_1 and Z_i with $E(Z_1|m_1,\boldsymbol{\theta})$ and $E(Z_i|m_2,\boldsymbol{\theta})$ for $i=3,4,5$, where $n=n_1+n_2+n_3+n_4+n_5$. **[10 marks]**