MAT7035: Computational Statistics

Midterm Test
(16:20–18:20, 14 DEC 2020)

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

(a) (Zero-truncated negative-binomial distribution) The *probability mass function* (pmf) is

$$p_x = \Pr(X = x) = c \cdot \frac{\Gamma(x+r)}{x!\Gamma(r)} \theta^r (1-\theta)^x,$$

for $x = 1, 2, \ldots, \infty$, where $r > 0$ is a known real number, $\theta \in (0, 1)$ is the parameter and c is the normalizing constant related to θ. Denote the value of c by θ before generating this zero-truncated negative-binomial distribution. **[10 marks]**

[Hint: (i) $\Gamma(r+1) = r\Gamma(r)$; (ii) The support of a negative-binomial random variable X is $\mathcal{S}_X = \{0, 1, \ldots, \infty\}$]

(b) (The standard Gumbel minimum distribution) The density function is $f(x) = \mathrm{e}^x \exp(-\mathrm{e}^x)$, where $-\infty < x < +\infty$. **[5 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 $f(x)$.

(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: $\Gamma(0.5) = \sqrt{\pi}$]

(c) Calculate the value of the acceptance probability.

3. [**15 marks**] Let X follow the finite mixture distribution with density

$$f_X(x) = \sum_{i=1}^{n} p_i f_{X_i}(x), \tag{MT.1}$$

where $X_i \sim f_{X_i}(\cdot)$ and $\{p_i\}_{i=1}^n$ are probability weights.

(a) State the *stochastic representation* (SR) method for generating one random sample from $X \sim f_X(x)$.

(b) Use the SR method to generate a random variable X following the polynomial distribution with density

$$f_X(x) = \sum_{i=1}^{n} c_i x^{i-1}, \quad 0 < x < 1,$$

where $\{c_i\}$ are positive constants such that $\sum_{i=1}^n \frac{c_i}{i} = 1$.

4. [**20 marks**] Let $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{Gamma}(\alpha, \beta)$ with density

$$f(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1}\,\mathrm{e}^{-\beta y}, \quad y > 0,$$

where $\alpha > 0$ and $\beta > 0$ are two unknown parameters, and

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}\,\mathrm{e}^{-y}\mathrm{d}y$$

is the gamma function.

(a) When α is known, find the MLE of β.

(b) When α is unknown, use Newton's method to find the MLE of α.

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

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

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 1.7 in page 10.** From

$$\begin{aligned}
1 &= \sum_{x=1}^{\infty} p_x = c \cdot \sum_{x=1}^{\infty} \frac{\Gamma(x+r)}{x!\Gamma(r)}\theta^r(1-\theta)^x \\
&= c \cdot \left\{ \theta^r + \left[\sum_{x=1}^{\infty} \frac{\Gamma(x+r)}{x!\Gamma(r)}\theta^r(1-\theta)^x \right] - \theta^r \right\} \\
&= c \cdot \left[\sum_{x=0}^{\infty} \frac{\Gamma(x+r)}{x!\Gamma(r)}\theta^r(1-\theta)^x - \theta^r \right] \\
&= c \cdot (1-\theta^r),
\end{aligned}$$

we have

$$c = \frac{1}{1-\theta^r} \quad \text{and} \qquad [\mathbf{3\ marks}]$$

and

$$p_1 = cr\theta^r(1-\theta). \qquad [\mathbf{2\ marks}]$$

The recursive identity between p_{x+1} and p_x is

$$\frac{p_{x+1}}{p_x} = \frac{c \cdot \dfrac{\Gamma(x+1+r)}{(x+1)!\Gamma(r)}\theta^r(1-\theta)^{x+1}}{c \cdot \dfrac{\Gamma(x+r)}{x!\Gamma(r)}\theta^r(1-\theta)^x} = \frac{(x+r)(1-\theta)}{x+1}. \qquad [\mathbf{2\ marks}]$$

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 $p \leftarrow \frac{(i+r)(1-\theta)}{i+1}p$, $F \leftarrow F + p$, $i \leftarrow i+1$ and go back to step 3. [**3 marks**]

(b) **This is a special case of Q1.1(e) 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

$$F(x) = 1 - \exp(-\mathrm{e}^x), \quad x \in \mathbb{R}. \qquad [\textbf{2 marks}]$$

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

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

Thus,

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

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

[**15 marks**]

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

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

[**2 marks**]

3. <u>Solution</u>. (a) **See §F.4 in page 19 of Tutorial 2.** The SR method for drawing $X \sim f_X(x)$ given by (MT.1)

Step 1: Draw $X_i = x_i \sim f_{X_i}(\cdot)$ for $i = 1, \ldots, n$ and independently draw $\mathbf{z} = \boldsymbol{z} = (z_1, \ldots, z_n)^\top \sim \text{Multinomial}(1; p_1, \ldots, p_n)$;

Step 2: Return $x = z_1 x_1 + \cdots + z_n x_n$. **[5 marks]**

(b) **See Example T2.4 of Tutorial 2.** We can write

$$f_X(x) = \sum_{i=1}^{n} \frac{c_i}{i} \cdot i x^{i-1} = \sum_{i=1}^{n} p_i f_{X_i}(x),$$

where $X_i \sim \text{Beta}(i, 1)$ or $X_i \overset{\text{d}}{=} U_i^{1/i}$ with $U_i \overset{\text{iid}}{\sim} U(0,1)$ for $i = 1, \ldots, n$. Thus, $f_X(x)$ is a mixture of n beta distributions. **[5 marks]**

The SR method for generating $X \sim f_X(x)$ is as follows:

Step 1: Draw $U_i = u_i \overset{\text{iid}}{\sim} U(0,1)$, set $x_i = u_i^{1/i}$ for $i = 1, \ldots, n$ and independently draw

$$\mathbf{z} = \boldsymbol{z} = (z_1, \ldots, z_n)^\top \sim \text{Multinomial}(1; p_1, \ldots, p_n);$$

Step 2: Return $x = z_1 x_1 + \cdots + z_n x_n$. **[5 marks]**

4. **Solution**. The likelihood function is

$$L(\alpha,\beta)=\prod_{i=1}^{n}\frac{\beta^{\alpha}}{\Gamma(\alpha)}y_i^{\alpha-1}\,\mathrm{e}^{-\beta y_i},$$

so that the log-likelihood function is

$$\ell(\alpha,\beta)=(\alpha-1)\left(\sum_{i=1}^{n}\log y_i\right)-n\bar{y}\beta+n\{\alpha\log\beta-\log\Gamma(\alpha)\},$$

where $\bar{y}=(1/n)\sum_{i=1}^{n}y_i$. **[3 marks]**

(a) When α is known, the MLE of β is

$$\hat{\beta}=\alpha/\bar{y}.$$ **[2 marks]**

(b) When α is unknown, the likelihood function of α is

$$\begin{aligned}\ell_1(\alpha) &= \ell(\alpha,\beta)|_{\beta=\alpha/\bar{y}}=c+\alpha\sum_{i=1}^{n}\log y_i-n\alpha\\ &\quad+n\alpha\log(\alpha/\bar{y})-n\log\Gamma(\alpha),\end{aligned}$$

and

$$\begin{aligned}\ell_1'(\alpha) &= \sum_{i=1}^{n}\log y_i+n\log(\alpha/\bar{y})-n\psi(\alpha),\\ \ell_1''(\alpha) &= \frac{n}{\alpha}-n\psi'(\alpha),\end{aligned}$$

where

$$\psi(\alpha)=\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}.$$ **[12 marks]**

By a one-dimensional Newton–Raphson algorithm, we have

$$\alpha^{(t+1)}=\alpha^{(t)}-\frac{\ell_1'(\alpha^{(t)})}{\ell_1''(\alpha^{(t)})}.$$ **[3 marks]**

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

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

and

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

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

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

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

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

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

$$L(\boldsymbol{\theta}|Y_{\text{obs}}, W, Z) \propto \theta_1^{n_1+W+Z_1}\theta_2^{n_2+m_1-W+Z_2}\theta_3^{n_3+Z_3}\theta_4^{n_4}\theta_5^{n_5}.$$

Taking logarithm, we obtain

$$\ell(\boldsymbol{\theta}|Y_{\text{obs}}, W, Z) = \log L(\boldsymbol{\theta}|Y_{\text{obs}}, W, Z) = (n_1+W+Z_1)\log\theta_1$$

$$+(n_2+m_1-W+Z_2)\log\theta_2 + (n_3+Z_3)\log\theta_3 + n_4\log\theta_4 + n_5\log\theta_5.$$

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

$$\begin{aligned}
\hat{\theta}_1 &= \frac{n_1+W+Z_1}{n+m_1+m_2}, & \hat{\theta}_2 &= \frac{n_2+m_1-W+Z_2}{n+m_1+m_2}, \\
\hat{\theta}_3 &= \frac{n_3+Z_3}{n+m_1+m_2}, & \hat{\theta}_i &= \frac{n_i}{n+m_1+m_2}, \quad i = 4, 5,
\end{aligned}$$

by replacing W and Z_i with $E(W|m_1, \boldsymbol{\theta})$ and $E(Z_i|m_2, \boldsymbol{\theta})$, where $n = n_1+n_2+n_3+n_4+n_5$.