THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

STAT6011/STAT7611 COMPUTATIONAL STATISTICS

December 11, 2021 Time: 6:30 p.m. – 8:30 p.m.

Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of the examination script.

Answer ALL TWO questions. Marks are shown in square brackets.

1. (Bayesian) Assume that the event time $T_i \overset{i.i.d.}{\sim}$ Weibull$(\theta, k), i = 1, \ldots, n$. The probability density function of the Weibull(θ, k) distribution has the form,

$$f_{\text{Weibull}}(t|\theta, k) = \theta k t^{k-1} \exp\left(-\theta t^k\right).$$

Suppose that k is known and the prior for θ is Gamma(a, b),

$$f_{\text{Gamma}}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$$

where a and b are hyperparameters.

(a) Derive the likelihood function $p(\boldsymbol{T}|\theta, k)$ for $\boldsymbol{T} = \{T_1, \ldots, T_n\}$. **[5 marks]**

(b) Derive the posterior distribution of $p(\theta|\boldsymbol{T}, k)$ (Assume here k is a constant). **[5 marks]**

(c) Calculate the marginal likelihood $p(\boldsymbol{T})$. **[10 marks]**

(d) Derive the posterior predictive density $p(\tilde{T}|\boldsymbol{T})$. How to draw predictive posterior samples of $\tilde{T}$? Describe the sampling procedure. **[10 marks]**

(e) Set $\tilde{u} = \tilde{T}^k$, and derive the closed-form distribution of $\tilde{u}|\boldsymbol{T}$. **[10 marks]**
Hints:

- For a random variable X and a monotone transformation $Y = g(X)$

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.$$

- Consider the Pareto distribution with parameters (x_m, α) for which the probability density function has the form,

$$f_{\text{Pareto}}(x|x_m, \alpha) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}}.$$

- $\frac{\Gamma(z+1)}{\Gamma(z)} = z.$

(f) (Bayesian Model Averaging) Consider three candidate models $M_j : k = k_j$ for $j = 1, 2, 3$ to fit the data $\boldsymbol{T}$ and we assign an equal weight to all candidate models, i.e., the prior model probability $P(M_j) = 1/3, \ j = 1, \ldots, 3$. For each model M_j, we assign a Gamma(a_j, b_j) prior for θ_j, i.e., $\pi(\theta_j|M_j) \sim$ Gamma(a_j, b_j) where a_j and b_j are hyperparameters. Calculate the posterior model probability of M_j, i.e., $P(M_j|\boldsymbol{T})$, for $j = 1, 2, 3$. **[10 marks]**

[Total: 50 marks]

2. (EM algorithm) Suppose that the event time $T_i \overset{i.i.d.}{\sim}$ Exp(λ), $f(T_i|\lambda) = \lambda \exp(-\lambda T_i)$. Let $\{(X_i, \Delta_i)\}_{i=1}^n$ be the observed time-to-event data, where $X_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \leq C_i)$ and C_i is the censoring time for the i-th patient.

(a) Write out the observed-data likelihood. **[5 marks]**

(b) Find the maximum likelihood estimator of λ and its variance. **[10 marks]**

(c) Derive the complete-data likelihood by treating $T_i, i = 1, \ldots, n$, as the latent variables. **[5 marks]**

(d) Derive the Q-function in the E-step of the EM algorithm by taking $T_i, i = 1, \ldots, n$, as the missing data. **[10 marks]**

(e) Write the detailed procedure for estimating λ by the EM algorithm. **[10 marks]**

(f) Given the right-censored observations

$$\{(X_i, \Delta_i)\}_{i=1}^{6} = \{(0.5, 1), (1.5, 1), (1.4, 1), (0.8, 1), (0.2, 0), (1.5, 0)\},$$

update the value of λ in one iteration using the EM algorithm in (e) by setting $\lambda^{(0)} = 0.8$ and compare your result with the MLE. **[10 marks]**

[Total: 50 marks]

********** END OF PAPER **********