UNIVERSITY OF BRISTOL

School of Mathematics

ANOMALY DETECTION

MATHM0030

[Paper code MATH–M0030J (Mock)]

January 2019 1 hours 30 minutes (this mock: 45 minutes)

This is a mock paper with one question. The actual exam will have two, and both will be used for assessment. Calculators are not permitted in this examination.

On this examination, the marking scheme is indicative and is intended only as a guide to the relative weighting of the questions.

1. (a) Describe what is meant by a "composite hypothesis" (**2 marks**)

   (b) Describe the Bonferroni correction procedure, and prove that it provides strong control of the familywise error rate regardless of any dependence between the p-values. (**10 marks**)

   (c) Recall that the Beta(α, β) distribution has density:

$$\frac{1}{\mathrm{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

   where $\alpha, \beta > 0$ and

$$\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

   and denote its cumulative distribution function by $F_{\text{Beta}}(x, \alpha, \beta)$.

   i. Let $Y_{(1)} \leq \cdots \leq Y_{(n)}$ denote the order statistics corresponding to i.i.d. replicates $Y_1, \ldots, Y_n$ of a Beta(α, β) random variable. Prove that:

$$\mathbb{P}(Y_{(k)} \leq y) = 1 - F_{\text{Binomial}}\{k-1, F_{\text{Beta}}(y, \alpha, \beta), n\},$$

   where $F_{\text{Binomial}}(x, r, n)$ is the cumulative distribution function of a Binomial random variable with success probability r and number of trials n. (**7 marks**)

   ii. Consider an $n \times m$ matrix M of independent uniform random variables on $[0, 1]$, that is,

$$M_{ij} \overset{i.i.d.}{\sim} \text{uniform}[0, 1], \quad i = 1, \ldots, n; j = 1, \ldots, m.$$

   Let $W_1, \ldots, W_n$ denote the minimum of each row, and consider the fifth smallest of these, $W_{(5)}$ (assuming $n \geq 5$). Give the probability

$$\mathbb{P}(W_{(5)} \leq w),$$

   in terms of F_{Beta} and F_{Binomial}. (**6 marks**)

   iii. A computer network has 50 computers and, in an effort to determine whether several are infected by a virus, 100 tests are applied to each. For each computer, we will record only the lowest p-value, and report an issue if the fifth smallest of these, denoted $w_{(5)}$, is unusually small. Assume all tests are independent and continuously distributed. Suppose you observe $w_{(5)} = 10^{-5}$ and you happen to know that

$$1 - F_{\text{Binomial}}\{1, F_{\text{Beta}}(10^{-5}, 1, 100), 50\} \leq 0.01.$$

   Can you reject the null hypothesis (corresponding to "no infection") at a false positive rate of 0.05? Explain your answer. (**7 marks**)

Question 1 continued...

(d) Consider a discrete time counting process $N(t)$ observed over a period $\{1, \ldots, T\}$, and denote the number of events occurring at time t by $\mathrm{d}N(t)$. We will employ the hurdle model

$$\mathrm{d}N(t) = \mathrm{d}A(t)\{\mathrm{d}B(t) + 1\},$$

where $A(t)$ and $B(t)$ are independent discrete time counting processes with $\mathrm{d}A(t) \in \{0, 1\}$, $\mathrm{d}B(t) \in \{0, 1\}$ [so that $\mathrm{d}N(t) \in \{0, 1, 2\}$]. Assume that $\mathrm{d}A(t)$ and $\mathrm{d}B(t)$ are independent sequences of independent and identically distributed Bernoulli variables with success probability μ and ν respectively, and assume a $\mathrm{Beta}(\alpha_\mu, \beta_\mu)$ prior for μ, and a $\mathrm{Beta}(\alpha_\nu, \beta_\nu)$ prior for ν.

i. Prove that the posterior distribution for $\mu \mid N(\cdot)$ is $\mathrm{Beta}\{\alpha_\mu + A(T), \beta_\mu + T - A(T)\}$, where $N(\cdot)$ denotes $N(1), \ldots, N(T)$. (**5 marks**)

ii. Prove that the posterior distribution for $\nu \mid N(\cdot)$ is $\mathrm{Beta}[\alpha_\nu + N(T) - A(T), \beta_\nu + 2A(T) - N(T)]$. (**6 marks**)

iii. Finally, give the predictive probability $\mathbb{P}\{N(T + 1) = 2 \mid N(\cdot)\}$ in terms of T, $A(T)$, $N(T)$ and the prior parameters. (**6 marks**)

End of examination.