Đề thi giữa kì môn Thống Kê Bayes

Bảo Trần

May 2022

1 Thông tin bài thi:

1. Bài thi làm trong vòng 1 tuần.
2. Chấp nhận mọi phương pháp làm bài, kể cả sử dụng code R.
3. Làm bài thi trên giấy, chụp màn hình và nộp trên course. Trong bài thi ghi rõ: Họ và Tên, MSSV. Nếu sử dụng code R thì chép hết lên giấy + chụp lại màn hình đã chạy thành công. (Điều này cần thiết vì có thể nhà trường sẽ yêu cầu nộp và chấm bản giấy)
4. Mỗi bài 3 điểm, số điểm mỗi câu được chia đều. Trừ câu 6 thì câu a) sẽ là 2 điểm và câu b) là 1 điểm.
5. Theo như phần 4 ở trên, đề không yêu cầu các bạn phải làm hết, chỉ cần làm đủ số điểm cần thiết. Khuyến khích làm càng nhiều càng tốt. Phần điểm thừa sẽ cộng sang phần điểm quá trình.
6. Các bài giống nhau dưới các mức độ sau: Giống y hệt nhau như sao chép; cùng sử dụng cách làm khác với phương pháp trong môn học với mức độ giống nhau về đáp án và trình bày ở hầu hết các bước; cùng sai với một đáp án giống hệt nhau, với cách trình bày giống hệt nhau - sẽ được tính là copy và trừ một nửa số điểm.

2 Đề thi

Câu 1 *Xét các mẫu ngẫu nhiên $X_1, X_2, \cdots, X_n$ có phân phối poisson:*

$$X_i|\theta \sim Poisson(\theta)$$

và $p(X_i|\theta) = \frac{e^{-\theta}.\theta^{x_i}}{x_i!}$.

Ta cho phân phối tiên nghiệm $\theta \sim Gamma(\alpha, \beta)$ với hàm một độ xác suất:

$$p(\theta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}\theta^{\alpha-1}e^{\frac{-\theta}{\beta}}$$

a) Ước lượng tham số θ theo maximum likelihood.

b) Tìm phân phối xác suất hậu nghiệm $\theta|X_1, X_2, \cdots, X_n$. Tìm trung bình và phương sai của phân phối hậu nghiệm này.

Câu 2 *Sử dụng lại dữ kiện ở bài 2. Cho $n = 100$, $X_i = i$. $\theta \sim Gamma(50, 100)$.*

a) Tính toán maximum likelihood và phân phối xác suất hậu nghiệm như câu 1.

b) Từ mẫu $n = 100$, $X_i = i$, ta tiếp tục cập nhật thêm 5 lần thử nữa với $X_{100+j} = 100 + j$ với $j \leq 5$. Lúc này ta có mẫu $n' = 105$. Tính maximum likelihood và so sánh với kết quả ở câu a.

Câu 3 *Một lượng hàng hoá được đem đi kiểm tra, 5 trong số đó được lựa để xem có bị hư hỏng không. Mẫu kiểm tra X có phân phối xác suất $B(5, \theta)$. Theo kinh nghiệm của những lần trước, ta biết $\theta \sim Beta(1, 9)$. Tìm ước lượng Bayes theo các hàm loss:*

a) $L(\theta, a) = (\theta - a)^2$

b) $L(\theta, a) = |\theta - a|$

Câu 4 *Một doanh nhân phải lựa chọn một vài "options" kinh doanh với mức* 100.000 *đô. a_1 - tự dùng tiền của mình để đầu tư, a_2 - chấp nhận* 60.000 *đô từ nhà đầu tư và chia họ* 45 *phần trăm lợi nhuận, a_3 - chấp nhận* 105.000 *đô từ nhà đầu tư và chia cho họ* 85 *phần trăm lợi nhuận. Lợi nhuận doanh nghiệp là 2θ với θ là số hàng bán được. Từ dữ liệu trong quá khứ, ta biết: $P(\theta = 0) = 0,2$ và*

$$P(\theta) = \frac{0,1}{3000} e^{\frac{-\theta}{3000}}$$

Các đánh giá đầu tư cho ta biết tình trạng bán hàng sắp tới sẽ tương đối tốt. Gọi x là tình trạng bán hàng tốt, đánh giá đầu tư cho X một hàm mật độ phụ thuộc vào số hàng bán được như sau:

$$p(x|\theta) = 0,6 e^{\frac{-\theta}{1000}}$$

Với các hàm loss sau:

1. $L(\theta, a_1) = 100000$, $\theta = 0$ *và* $100000 - 2\theta$, $\theta > 0$.
2. $L(\theta, a_2) = 40000$, $\theta = 0$ *và* $100000 - 1,1\theta$, $\theta > 0$.
3. $L(\theta, a_3) = -5000$, $\theta = 0$ *và* $100000 - 0,3\theta$, $\theta > 0$

Tìm quyết định hợp lí nhất trong tình huống x-bán hàng tốt ở trên.

Câu 5 *Hiện tại dịch đang bùng phát mạnh ở quốc gia T trong vòng 1 năm nay. Số lượng người tử vong do covid tại một thành phố ở quốc gia T đó trung bình theo tháng được thống kê như sau:*

$$11, 21, 35, 40, 60, 68, 61, 43, 23, 23, 21, 41$$

Gọi X_i là các biến ngẫu nhiên "tử vong do covid ở tháng thứ i. Trong bài này ta có $X_1 = 211$, $X_2 = 231$, $X_3 = 356$, ... Giả sử X_i có phân phối xác suất $Poisson(\theta)$. Trong một nghiên cứu từ trước khi bùng dịch, theo dữ liệu của quốc gia đó thì $\theta \sim \chi^2(24)$, tức là giá trị người chết trung bình hằng tháng sẽ có phân phối Chi bình phương. Tìm phân phối xác suất hậu nghiệm $\theta|X_1, .., X_{12}$.

Nhắc lại: $\theta|X_1, ..., X_{12}$ là mô hình dự báo giá trị θ - giá trị trung bình số người chết do covid hằng tháng.

Câu 6 *Một nhóm nghiên cứu đang tìm hiểu về độ dài đuôi của loài khỉ. Họ nghiên cứu trên một mẫu gồm 5 con khỉ với độ dài đuôi (theo cm) như sau:*

$$15, 18, 22, 14, 17$$

Gọi X_i là biến cố độ dài đuôi của con thứ i. Biết $X_i \sim N(\theta; 2,5)$ với θ là trung bình độ dài đuôi, $2,5$ là phương sai. Theo các dữ liệu trong quá khứ, người ta đoán được rằng $\theta \sim N(19; 3)$.

a) Ta có hai giả thuyết riêng biệt như sau: $H_0 : \theta = 19$ và $H_1 : \theta \neq 19$. Sử dụng phương pháp Bayes (tính Bayes factor) để tìm ra đưa ra kết luận rằng giả thuyết nào được chấp nhận.

b) Ta có hai giả thuyết riêng biệt như sau: $H_0 : \theta < 19$ và $H_1 : \theta \geq 19$. Sử dụng phương pháp Bayes (tính Bayes factor) để tìm ra đưa ra kết luận rằng giả thuyết nào được chấp nhận.

3 Gợi ý

Câu 1, 2, 3, 4, 6b: xem lại trong phần bài tập đã sửa vào ngày 18 tháng 5. Câu 1 và 2 có thể xem thêm ở bài 3.2.

Câu 5: Xem lại slide bài 3.2 (trên MS Team), phần phân phối Gamma và liên hệ đến các phân phối xác suất khác (từ slides 51 trở về sau).

Câu 6a: Xem lại slide bài 3.2 và bài 5 (trên MS Team), phần công thức cho bài toán kiểm định gỉa thuyết dạng điểm cho phân phối chuẩn.