Bài kiểm tra 2: Thực hành lý thuyết mẫu

Ngày 18 tháng 04 năm 2023

Thời gian làm bài: 45 phút.

Note:

- Sinh viên làm bài trên R script, lưu lại với tên có dạng: "LTTK_HọTên_MSSV_Test2.R".
  Sau khi hoàn thành bài làm, copy phần code bài làm trong R sang file text .txt để backup.
  Nộp bài cả file R script và file text .txt (lưu với tên có dạng "LTTK_HọTên_MSSV_Test2.txt").
- trong quá trình làm bài kiểm tra, sinh viên có thể tham khảo tài liệu "Giới thiệu về R" (đã được giới thiệu là tài liệu tham khảo môn học).
- Dùng lệnh `help(ten_ham)` để biết cú pháp và cách sử dụng một command trong R.
- Bài làm cần trình bày như sau:

```
##
## Bai kiem tra 2 - Thuc hanh Ly thuyet Thong ke
## Nhom2 - Thu ... - tiet ....
##
## Ho ten: ......... - MSSV: ...........
##
##*****************************************

## Bai 1:

##-----------------------------------------
## Bai 2:

##-----------------------------------------
```

Bài 1. (3đ)

Trong một khu công nghiệp, người ta xây dựng hệ thống phát hiện hỏa hoạn sử dụng các cảm biến nhiệt độ, giả sử các cảm biến này hoạt động độc lập và tổng số gồm 40 cảm biến. Khả năng mỗi cảm biến kích hoạt báo động khi nhiệt độ lên đến 100^oC hoặc cao hơn là 80%. Gọi X là biến ngẫu nhiên thể hiện số lượng cảm biến kích hoạt báo động khi nhiệt độ lên đến 100^oC.

1.1. (**0,5đ**) Tìm phân phối xác suất của X
(*Viết câu mô tả trong* R để trả lời câu hỏi này - viết không dấu)

1.2. (**0,5đ**) Vẽ biểu đồ cột của hàm xác suất tương ứng của phân phối xác suất của X.
(*Hint:* dùng hàm `plot` và chọn tham số `type = 'h'`)

1.3. (**0,5đ**) Viết (các) câu lệnh để tính xác suất có ít nhất 5 cảm biến kích hoạt nhưng không quá 30 cảm biến kích hoạt, bằng hai cách

- Cách 1: lấy tổng các xác suất được cho bởi hàm xác suất.
- Cách 2: sử dụng hàm phân phối xác suất (hay hàm phân phối xác suất tích lũy) của X.

1.4. (**1,5đ**) Giả sử có mẫu ngẫu nhiên $X_1, X_2, ..., X_m$ độc lập và có cùng phân phối với X. Thực hiện mô phỏng (*simulation*) theo các bước sau:

1.4.1 Tạo vector `m.vec` chứa các cỡ mẫu khác nhau, với

$$\texttt{m.vec} = \left(10, 50, 100, 500, 1000, 5000, 10^5, 10^6, 2 \times 10^6\right).$$

1.4.2 (**0,5đ**) Tạo vector `SampleMean.vec` với độ dài bằng `length(m.vec)`.
Với mỗi $\texttt{k} \in \{1, 2, 3, ..., \texttt{length(m.vec)}\}$, ta có cỡ mẫu $m = \texttt{m.vec[k]}$:

(a.) tạo mẫu ngẫu nhiên $x_1, x_2, ..., x_m$ tương ứng với cỡ mẫu m;

(b.) tính trung bình mẫu tương ứng của $x_1, x_2, ..., x_m$ và lưu vào vị trí thứ `k` của vector `SampleMean.vec`;

1.4.3 (**0,5đ**)

- Dùng hàm `plot`, vẽ đồ thị bằng đường nét liền (chọn `lty=1`) cho `SampleMean.vec` và chọn màu `col = "blue"`.
  (sinh viên tự xác định các tham số cần thiết khác để vẽ trong hàm `plot`)
- Trên cùng `plot` vừa vẽ, hãy vẽ đường thẳng $y = \mathbb{E}(X)$ (đường thẳng song song với trục Ox của hệ tọa độ) bằng đường nét đứt (chọn `lty=4`) và chọn màu `col = "red"`.

1.4.4 Nhận xét, ta thấy:

- X là biến ngẫu nhiên có phân phối được cho theo giả thiết bài toán với $\mathbb{E}(X)$ là kỳ vọng hay trung bình của b.n.n. X,
- $X_1, X_2, ..., X_m$ là mẫu ngẫu nhiên độc lập và có cùng phân phối với X, với `SampleMean.vec` chứa các giá trị là trung bình mẫu,

(**0,5đ**) Qua phần mô phỏng, bạn có nhận xét gì về giá trị trung bình mẫu của $X_1, X_2, ..., X_m$ (so sánh với $\mathbb{E}(X)$) khi cỡ mẫu $m \to +\infty$?

Kết quả trong phần 1.4. thường được biết đến là *(Weak) Law of Large Numbers.*

Bài 2. (2,5đ)

Định lý 1. *Cho các biến ngẫu nhiên $X \sim \mathcal{N}(0;1)$ và $Y \sim \chi^2(n)$ độc lập với nhau, $n \in \mathbb{N}, n \geq 1$, thì* $W = \dfrac{X}{\sqrt{Y/n}} \sim$ *t-Student*(n).

■ Với mỗi $n \in \{10; 100; 1000\}$, hãy viết đoạn code mô phỏng định lý sau theo các bước :

2.1.(**0,5đ**) Tạo mẫu ngẫu nhiên cho X và Y tương ứng $\Rightarrow$ tạo được mẫu ngẫu nhiên cho W theo định nghĩa của W.

2.2.(**0,5đ**) Trên cùng plot, vẽ đồ thị `histogram` cho W (chọn `col = 'lightblue'`) cùng với hàm mật độ xác suất (chọn `col = 'red',lwd=2`) theo lý thuyết của W.

■ Hãy viết đoạn code thực hiện mô phỏng theo theo các bước sau:

2.3.(**0,5đ**) Trên cùng một plot, vẽ hàm mật độ xác suất theo lý thuyết cho W tương ứng với $n \in \left\{10; 100; 1000; 10^4; 10^5\right\}$ (chọn `col` lần lượt là `'blue','orange','violet','green','cyan'`).

Cũng trên plot đó, vẽ thêm hàm mật độ xác suất của biến ngẫu nhiên $Z \sim \mathcal{N}(0,1)$ (chọn `col = 'red', lty=4, lwd=2`).

(**0,5đ**) Bạn có dự đoán gì về phân phối xác suất của W khi $n \to +\infty$?
(*Hint: so sánh với phân phối của Z*)

■ **2.4.**(**0,5đ**) Hãy vẽ 3 plots (trong phần 2.1. và 2.2.) cùng với plot (trong phần 2.3.) trên cùng một figure (*Hint: sử dụng* `par( mfrow = c(2,2) )`).

Bài 3. (2đ)

Định lý 2 (Central Limit Theorem (CLT))**.** *Cho $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập và có cùng phân phối xác suất với trung bình (kỳ vọng) μ và phương sai $\sigma^2 < +\infty$. Đặt*

$$Y_n = \frac{\sqrt{n}.(\overline{X} - \mu)}{\sigma},$$

thì biến ngẫu nhiên Y_n có phân phối xác suất tiến về phân phối $\mathcal{N}(0,1)$ khi $n \to +\infty$.

Mô phỏng định lý CLT trong trường hợp $X_1, X_2, ..., X_n$ có cùng phân phối Poisson với trung bình là 25, i.e. $X_j \sim \mathcal{P}(\lambda)$ với $\lambda = 25$.
Với mỗi giá trị n, vẽ đồ thị `histogram` của Y_n cùng với đồ thị của hàm mật độ xác suất theo lý thuyết của phân phối $\mathcal{N}(0,1)$ trên cùng một `plot`.
Trên cùng một cửa sổ (hay figure), hãy vẽ đồ thị như trên, với các giá trị n lần lượt là 50, 200, 1000 và 10000. (*Hint: sử dụng* `par( mfrow = c(2,2) )`)

Remark: Lưu ý, trong câu lệnh vẽ đồ thị `histogram`, sinh viên cần tự lựa chọn giá trị cho tham số `breaks` sao cho phù hợp.

Bài 4. (2,5đ)

Cho biến ngẫu nhiên $X \sim \chi^2(15)$ và $p = 0.75$.

4.1.(1,0đ) Hãy viết hàm tên `find.Interval(p)` để tìm ước lượng cho khoảng $[a; b]$ với $a, b \in \mathbb{R}$, $a < b$ sao cho $\mathbb{P}\big(X \leq a\big) = \mathbb{P}\big(X > b\big) = \dfrac{1-p}{2}$.
Hàm `find.Interval(p)` sẽ trả về vector `(a,b)` chứa hai giá trị ước lượng cho a và b.

4.2.(1,5đ) Với a, b tìm được bởi hàm `find.Interval(p)`, thực hiện các bước sau để kiểm tra lại $\mathbb{P}(a \leq X \leq b) = p$.

4.2.1 Tạo vector `m.vec` chứa các cỡ mẫu khác nhau, với `m.vec` $= \big(10; 100; 1000; 10^4; 10^5\big)$.

4.2.2 (**0,5đ**) Tạo vector `p.hat.vec` với độ dài bằng `length(m.vec)`.
Với mỗi `k` $\in \{1, 2, 3, ..., $ `length(m.vec)` $\}$, ta có cỡ mẫu $m =$ `m.vec[k]`:

(a.) tạo mẫu ngẫu nhiên $x_1, x_2, ..., x_m \sim X$ tương ứng với cỡ mẫu m;

(b.) tính `p.hat` $= \displaystyle\sum_{j=1}^{m} \mathbb{1}_{a \leq x_j \leq b}$ và gán giá trị `p.hat` vào vị trí thứ `k` của vector `p.hat.vec`;

4.2.3 (**0,5đ**)

- Dùng hàm `plot`, vẽ đồ thị bằng đường nét liền (chọn `lty=1`) cho `p.hat.vec` và chọn màu `col = "blue"`.
(sinh viên tự xác định các tham số cần thiết khác để vẽ trong hàm `plot`)
- Trên cùng `plot` vừa vẽ, hãy vẽ đường thẳng $y = p$ (đường thẳng song song với trục Ox của hệ tọa độ) bằng đường nét đứt (chọn `lty=4`) và chọn màu `col = "red"`.

4.2.4 (**0,5đ**) Qua phần mô phỏng, bạn có nhận xét gì về giá trị `p.hat` so với p khi $m \to +\infty$?

- - - Good luck! - - -