

BÀI TẬP 5  
(Lấy mẫu lại)  
THỐNG KÊ MÁY TÍNH VÀ ỨNG DỤNG (CLC)

1. 10 con cá được chọn ngẫu nhiên từ hồ A có chiều dài (cm) đo được là

11.5   10.8   11.6   9.4   12.4   11.4   12.2   11.0   10.6   10.8

Tương tự, 8 con cá được chọn ngẫu nhiên từ hồ B có chiều dài

11.8   12.6   12.2   12.5   11.7   12.1   10.4   12.6

Dùng kiểm định hoán vị đánh giá các nhận định sau

- (a) Kỳ vọng chiều dài cá trong 2 hồ là khác nhau.
  - (b) Phương sai chiều dài cá trong 2 hồ là khác nhau.
2. Phân phối mũ  $\text{Exp}(\lambda)$  có kỳ vọng (mean) là  $\mu = \frac{1}{\lambda}$ , trung vị (median) là  $m = \frac{\ln 2}{\lambda}$  và phương sai (variance) là  $\sigma^2 = \frac{1}{\lambda^2}$ . Từ đó, cho mẫu ngẫu nhiên  $X_1, X_2, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\lambda)$ , ta có thể dùng các ước lượng sau cho  $\lambda$

$$\begin{aligned}T_1 &= \frac{1}{\bar{X}}, \\T_2 &= \frac{\ln 2}{\hat{m}}, \\T_3 &= \frac{1}{\sqrt{S^2}},\end{aligned}$$

với  $\bar{X}, \hat{m}, S^2$  lần lượt là trung bình, trung vị và phương sai mẫu.

Bảng sau là một mẫu ngẫu nhiên sinh từ phân phối mũ  $\text{Exp}(\lambda)$

1.4066	1.2917	1.4080	4.2801	1.2136	2.7461
11.1076	0.9247	5.8833	10.2513	3.8285	3.2116
0.5451	0.9896	1.1602	7.7723	0.1702	0.8907
0.2276	3.1197	11.4909	0.6475	11.2279	0.7639

- (a) Tính các giá trị ước lượng  $T_1, T_2, T_3$  cho  $\lambda$  từ mẫu dữ liệu đã cho.
- (b) Dùng kỹ thuật phân tích Bayes với phân phối tiên nghiệm phù hợp cho  $\lambda$  để ước lượng  $\lambda$ .
- (c) Dùng kỹ thuật bootstrapping, đánh giá và so sánh các phương pháp ước lượng trên ( $T_1, T_2, T_3$  và ước lượng Bayes ở Câu (b)).

3. Bảng sau là một mẫu ngẫu nhiên sinh từ phân phối mũ  $\text{Exp}(\gamma)$

23.9618	4.9055	6.0424	0.5870	4.0856	1.6503
10.1976	4.0208	25.9484	15.3954	19.5160	0.5937
11.5481	18.3895	30.4093	7.6527	9.7329	8.6130
6.2353	5.5157	9.9489	21.3850	5.1142	28.2284

(a) Dùng kiểm định hoán vị đánh giá nhận định “ $\gamma < \lambda$ ” với  $\lambda$  giá trị của tham số cho phân phối mũ ở Câu 2.

(b) Xây dựng khoảng tin cậy bootstrap 95% cho tỉ số  $l = \frac{\lambda}{\gamma}$ .

4. Từ dữ liệu “California Housing dataset” ([https://scikit-learn.org/stable/datasets/real\\_world.html#california-housing-dataset](https://scikit-learn.org/stable/datasets/real_world.html#california-housing-dataset)), dùng kĩ thuật kiểm tra chéo, chọn ra mô hình “tốt nhất” giải thích giá nhà (target) theo các đặc trưng (feature).

— HẾT —