THỐNG KÊ ỨNG DỤNG

Giảng viên: TS. Bùi Thanh Hùng
Bộ môn Khoa học dữ liệu, Khoa Công nghệ thông tin
Đại học Công nghiệp thành phố Hồ Chí Minh
Email: buithanhhung@iuh.edu.vn
Website: https://sites.google.com/site/hungthanhbui1980/

Mỗi bài tập đều làm theo 2 cách: Tự tính bằng tay và viết code cho máy tính tính

Bài 1:
Lan rút các quả bóng 5 lần từ một túi 10 quả bóng trong đó có 5 quả bóng màu đỏ và 5 quả bóng màu xanh lá cây thay thế. Kiểm tra xem đây có phải là một ví dụ về các thử nghiệm Bernoulli không.

Bài 2:
Tỷ lệ người dân tham gia trên mạng lưới giao dịch có sự hiểu biết về luật giao dịch là 90%. Tại một nút giao diện có 5 người giao diện trong phạm vi luật. Gọi X là một số người am hiểu luật nhưng cố gắng tình luật trong 5 người đó.
a) Lập bảng phân phối của X.
b) Tính E (3- 5X) và V (3X-2).
c) Nếu có 30 người trong phạm vi luật, thì trung bình có bao nhiêu người hiểu biết luật nhưng cố gắng trong phạm vi? Có bao nhiêu người trong phạm vi luật làm người không hiểu luật?

Bài 3:
Một lô hàng có 8 sản phẩm loai I và 2 sản phẩm loại II. Lấy ngẫu nhiên lần lượt ra 5 sản phẩm theo phương thức hoàn lại. Gọi X là số sản phẩm loại II trong 5 sản phẩm lấy ra.
a) X có phân phối gì?
b) Tính kỳ vọng và phương sai của X.
c) Tính số sản phẩm loại II trung bình trong số sản phẩm lấy ra và tính khả năng để xảy ra điều đó.
d) Nếu lấy lần lượt ra 64 sản phẩm từ lô hàng đó (vẫn lấy theo phương thức hoàn lại) thì trung bình lấy được bao nhiêu sản phẩm loại II? Số sản phẩm loại II có khả năng xảy ra nhất là bao nhiêu?

Bài 4:
Số cuộc gọi điện thoại đến trung tâm tổng đài thường được mô tả là một biến ngẫu nhiên Poisson. Biết rằng trung bình có 10 cuộc điện thoại gọi tới trong 1 giờ.
a) Xác suất có đúng 5 cuộc điện thoại gọi tới trong 1 giờ là bao nhiêu?
b) Xác suất có 3 hoặc ít hơn 3 cuộc điện thoại gọi tới trong 1 giờ là bao nhiêu?
c) Xác suất có đúng 15 cuộc điện thoại gọi tới trong 2 giờ là bao nhiêu?
d) Xác suất có đúng 5 cuộc điện thoại gọi tới trong 30 phút là bao nhiêu?

A- PHÂN PHỐI NHỊ THỨC

Phân phối nhị thức là một phân phối **xác suất trong thống kê** tóm tắt khả năng một giá trị sẽ nhận một trong hai giá trị độc lập dưới một tập hợp tham số hoặc giả định nhất định. Các giả định cơ bản của phân phối nhị thức là chỉ có một kết quả cho mỗi thử nghiệm, rằng mỗi thử nghiệm có xác suất thành công như nhau và mỗi thử nghiệm là loại trừ lẫn nhau hoặc độc lập với nhau.

Công thức phân phối nhị thức dành cho bất kỳ biến ngẫu nhiên X, được cho bởi:

$$P(X=k)=C_n^k p^k(1-p)^{n-k}$$

Với:

- n: số lần phép thử
- k=0,1,2,…,n
- p: xác suất thành công của phép thử

Trung bình và phương sai phân phối nhị thức

Đối với phân phối nhị thức, giá trị trung bình, phương sai và độ lệch chuẩn của số lần thành công nhất định được biểu diễn bằng cách sử dụng các công thức

- Trung bình, $\mu = np$
- Phương sai, $\sigma^2 = npq$
- Độ lệch chuẩn $\sigma = \sqrt{(npq)}$

Trong đó: p là xác suất thành công; q là xác xuất thất bại được tính bởi:

$$q = 1-p$$

Các tính chất của phân phối nhị thức là:

- Có hai kết quả có thể xảy ra: đúng hoặc sai, thành công hoặc thất bại, có hoặc không (kết quả đầu ra là 2).
- Có ‘n’ số thử nghiệm độc lập hoặc một số cố định n lần thử nghiệm lặp lại.
- Xác suất thành công hay thất bại vẫn như nhau cho mỗi lần thử.
- Chỉ tính số lần thành công trong số n lần thử nghiệm độc lập.
- Mỗi phép thử là một phép thử độc lập, có nghĩa là kết quả của một phép thư không ảnh hưởng đến kết quả của phép thử khác.

B- PHÂN PHỐI BERNOULLI

Công thức Bernoulli trong xác suất

Trong n phép thử độc lập biến cố A xuất hiện đúng k lần, ký hiệu $P_n(k)$, được tính bằng công thức Bernoulli:

$$P_n(k) = C_n^k p^k q^{n-k}$$

Với:

- n phép thử độc lập
- k: số lần xuất hiện của phép thử
- p: xác suất thành công
- q = 1 – p: xác suất thất bại

Lưu ý về thử nghiệm Bernoulli:

- Thử nghiệm Bernoulli chỉ có hai kết quả có thể xảy ra.
- Hai kết quả có thể xảy ra là độc lập với nhau.
- Xác suất thành công là p và xác suất thất bại là 1 – p = q.
- Xác suất của mỗi kết quả trong mỗi thử nghiệm Bernoulli không đổi.

C- PHÂN PHỐI POISSON

Biến ngẫu nhiên rời rạc X được gọi là có **phân phối Poisson** với tham số λ, ký hiệu là X ~ P (λ), nếu X nhận các giá trị có thể có là các số nguyên không âm: 0, 1, 2, … , n và các suất tương ứng được tính theo công thức xấp xỉ Poisson.

$$P_k = P(X = k) = C_n^k p^k q^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

Trong đó e là hằng số nêpe:

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n ; e \approx 2,71828$$

Chứng minh:

Thật vậy: Do $np = \lambda \Rightarrow p = \frac{\lambda}{n}; \quad q = 1 - p = 1 - \frac{\lambda}{n}$

Đặc trưng của phân phối Poisson: Nếu X có phân phối Poisson (X ~ P (λ)) thì :

E(X) = λ ; V(X) = λ