

## ĐỀ 1 ÔN TẬP MÔN XÁC SUẤT THỐNG KÊ

**Câu 1.** Cho  $X, Y$  là các biến ngẫu nhiên rời rạc có phân phối xác suất đồng thời như sau

$Y \backslash X$	1	2	3
1	0,025	0,075	0,1
2	0,05	0,2	0,15
3	<b>0,1</b>	0,2	0,1

- (a) Tính  $P(X = Y)$
- (b) Tính phân phối xác suất thành phần của  $X$ .
- (c) Hỏi  $X, Y$  có độc lập không?

**Câu 2.** Cho  $X, Y$  là các biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời như sau:

$$f(x, y) = \begin{cases} c(x + y), & 0 \leq x \leq 1; 0 \leq y \leq x \\ 0, & \text{các trường hợp khác} \end{cases}$$

- (a) Tìm hằng số  $c$ .
- (b) Tìm  $f_X(x)$ .
- (c) Tính  $\mathbf{P}(\mathbf{Y} > \mathbf{0,3} | \mathbf{X} = \mathbf{0.5})$

**Câu 3** Một nhóm kỹ sư kiểm thử hiệu suất đang đánh giá thời gian phản hồi của một API nội bộ trên nền tảng dịch vụ web của công ty. Họ thực hiện đo ở 36 địa điểm khác nhau và tìm thấy thời gian phản hồi trung bình là 2,6 mili-giây. Giả sử độ lệch chuẩn thời gian phản hồi của hệ thống là 0,3 mili-giây (biết trước).

- (a) Tìm khoảng tin cậy 95% cho thời gian phản hồi trung bình của API trên toàn hệ thống.
- (b) Nếu muốn đảm bảo rằng ước lượng trung bình lệch không quá 0,05 mili-giây với độ tin cậy 95% thì cần kích thước mẫu bao nhiêu?
- (c) Thực hiện kiểm định giả thuyết với mức ý nghĩa 5% để kiểm tra tuyên bố: “Thời gian phản hồi trung bình của API là 2,5 mili-giây” Có đủ bằng chứng để bác bỏ tuyên bố đó không?

**Câu 4.** Trong một khảo sát bảo mật tại thành phố Thủ Đức, nhóm chuyên gia mạng chọn ngẫu nhiên 500 người dùng Internet tại nhà, và phát hiện 320 người trong số đó sử dụng VPN để bảo vệ thông tin cá nhân.

- (a) Tìm khoảng tin cậy 95% cho tỷ lệ thực sự người dùng Internet tại nhà ở thành phố Thủ Đức sử dụng VPN.

- (b) Giả sử bạn muốn ước lượng tỷ lệ người dùng VPN với sai số không quá  $\pm 0.04$  thì cần độ tin cậy là bao nhiêu?
- (c) Kiểm định phát biểu “Có hơn 65% người dùng Internet tại nhà ở Thủ Đức sử dụng VPN thường xuyên” ở mức ý nghĩa 1%. Liệu có đủ bằng chứng để ủng hộ cho tuyên bố đó không?

**Câu 5.** Một nhóm kỹ sư CNTT đã ghi nhận dữ liệu từ một hệ thống thông tin nội bộ của doanh nghiệp trong các đợt kiểm thử hiệu năng.

$X$  (người dùng): số lượng người dùng đồng thời truy cập vào hệ thống.

$Y$  (giây): thời gian phản hồi trung bình của hệ thống.

Số người dùng $X$	10	20	20	30	30	40	40	50	60	60
Thời gian phản hồi $Y$	0.8	1.0	1.1	1.2	1.3	1.5	1.4	1.6	1.8	2.0

- (a) Tính hệ số tương quan mẫu giữa số người dùng và thời gian phản hồi. Từ đó nhận xét mối quan hệ tuyến tính  $X$  và  $Y$  (mạnh/yếu, dương/âm).
- (b) Tìm phương trình hồi quy tuyến tính dự đoán thời gian phản hồi theo số người dùng.
- (c) Dự đoán thời gian phản hồi nếu có 45 người dùng đồng thời.

## ĐỀ 2 ÔN TẬP MÔN XÁC SUẤT THỐNG KÊ

**Câu 1.** Cho  $X, Y$  là các biến ngẫu nhiên rời rạc có phân phối xác suất đồng thời như sau

$Y \backslash X$	0	1	2
1	0,04	0,06	0,1
2	0,05	0,17	0,18
3	0,08	0,15	0,17

- (a) Tính  $P(X < Y)$
- (b) Tính phân phối xác suất thành phần của  $Y$ .
- (c) Tính  $P(X > 0 | Y > 1)$ .

**Câu 2.** Cho  $X, Y$  là các biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời như sau:

$$f(x, y) = \begin{cases} cxy, & 0 \leq x \leq 2; 0 \leq y \leq 4 \\ 0, & \text{các trường hợp khác} \end{cases}$$

- (a) Tìm hằng số  $c$ .
- (b) Tìm  $f_Y(y)$ .
- (c) Tính  $P(Y < X^2)$

**Câu 3.** Một nhóm nghiên cứu đang kiểm tra độ chính xác của một mô hình phân loại cảm xúc từ các bình luận khách hàng (positive/negative). Họ đo độ chính xác (%) của mô hình trong 7 lần thử nghiệm độc lập với kết quả:

87.2, 88.4, 86.5, 87.9, 88.1, 86.8, 87.5

Giả sử độ chính xác của mô hình phân loại cảm xúc có phân phối chuẩn.

- (a) Tìm khoảng tin cậy 99% cho độ chính xác trung bình của mô hình.
- (b) Thực hiện kiểm định giả thuyết ở mức ý nghĩa 5% để kiểm tra tuyên bố: “Mô hình có độ chính xác trung bình hơn 87%.” Liệu có đủ bằng chứng để ủng hộ cho tuyên bố đó không?

**Câu 4.** Một sàn thương mại điện tử đang khảo sát hành vi khách hàng sau khi mua hàng. Trong một mẫu ngẫu nhiên gồm 600 khách hàng đã từng mua hàng, có 390 người đã quay lại mua hàng lần thứ hai trong vòng 3 tháng.

- (a) Tìm khoảng tin cậy 90% cho tỷ lệ thực sự khách hàng quay lại mua hàng lần hai.
- (b) Giả sử bạn muốn sai số không vượt quá 1% với độ tin cậy 95% thì kích thước mẫu tối thiểu phải khảo sát là bao nhiêu?
- (c) Kiểm định tuyên bố “Có ít hơn 65% khách hàng quay lại mua hàng lần hai.” ở mức ý nghĩa 5%. Liệu có đủ bằng chứng để ủng hộ cho tuyên bố đó không?

**Câu 5.** Một nhóm phát triển AI thử nghiệm một hệ thống nhận diện khuôn mặt với các mô hình xử lý ảnh có độ phức tạp khác nhau.

$X$  (ms): Thời gian xử lý trung bình mỗi ảnh khuôn mặt (tính bằng mili giây).

$Y$  (%): Độ chính xác trung bình của mô hình trong nhận diện khuôn mặt.

Họ thu được dữ liệu sau:

Thời gian xử lý ( $X$ ms)	20	25	30	30	35	40	40	45	50	50
Độ chính xác ( $Y$ %)	78	80	82	81	84	86	85	88	90	91

- (a) Tính hệ số tương quan mẫu giữa thời gian xử lý và độ chính xác. Từ đó nhận xét mối quan hệ tuyến tính  $X$  và  $Y$  (mạnh/yếu, dương/âm).
- (b) Tìm phương trình hồi quy tuyến tính dự đoán dự đoán độ chính xác dựa trên thời gian xử lý.
- (c) Dự đoán độ chính xác nếu hệ thống xử lý mất 38 ms mỗi ảnh.