ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
BỘ MÔN TOÁN – LÝ

ĐÁP ÁN ĐỀ THI CUỐI KỲ
MÔN XÁC SUẤT THỐNG KÊ
Học kỳ 1, năm học 2018-2019.
Ngày thi: .../.../ 2019
Thời gian làm bài: 90 phút.
Được sử dụng tài liệu giấy.

Câu 1. *(1.5 điểm)* Xác suất virus máy tính V có thể gây hại cho một tập tin bất kì là 35%. Giả sử virus V xâm nhập vào một thư mục gồm 2400 tập tin. Tính xác suất có từ 800 đến 850 tập tin bị nhiễm virus.

- ***Đáp án tham khảo:***

  Gọi X là số tập tin bị nhiễm virus trong 2400 tập tin

  Ta có: $X \sim B(n, p), \quad n = 2400, \quad p = 35\% = 0.35$

  $\Rightarrow X \sim N(\mu, \sigma), \qquad \mu = np = 840, \qquad \sigma = \sqrt{np(1-p)} = 23.367$

  $$P(800 \le X \le 850) \approx P\left(\frac{799{,}5 - \mu}{\sigma} \le Z \le \frac{850{,}5 - \mu}{\sigma}\right)$$
  $$= P\left(\frac{799{,}5 - 840}{23{,}367} \le Z \le \frac{850{,}5 - \mu}{23{,}367}\right)$$
  $$= \phi(0{,}45) - \phi(1{,}73) = 0{,}6318$$

Câu 2. *(2 điểm)* Cho X và Y là số lần phần cứng bị hỏng trong 2 phòng lab A và B trong một tháng. Phân phối đồng thời của X và Y được cho bởi bảng sau:

$P(x,y)$		x		
		0	1	2
y	0	0.52	0.20	0.04
	1	0.14	0.02	0.01
	2	0.06	0.01	0

a) Tính xác suất $P(X + Y \ge 1)$.
b) X và Y có độc lập nhau hay không? Vì sao?
c) Giả sử phòng lab A bị hỏng phần cứng trong tháng 1, tính xác suất phòng lab B cũng bị hỏng phần cứng trong tháng 1.

- ***Đáp án tham khảo:***

  a) $P(X + Y \ge 1) = 1 - P(0{,}0) = 1 - 0.52 = 0.48$

  b) $P(X = 0) = 0.72, P(Y = 0) = 0.76$

  Vì $P(X = 0, Y = 0) = 0.52 \ne P(X = 0).P(Y = 0) = 0.5472$

  nên X, Y không độc lập với nhau.

  c) $P(Y \ge 1 | X \ge 1) = \dfrac{P(X \ge 1, Y \ge)}{P(X \ge 1)} = \dfrac{0.04}{0.28}$

Câu 3. *(2 điểm)* Cho 2 biến ngẫu nhiên có hàm mật độ đồng thời

$$f(x,y) = \begin{cases} C(x-2)e^{-y} & \text{nếu } 0 \le x \le 2, y \ge 0 \\ 0 & \text{nếu } (x,y) \text{ khác} \end{cases}$$

a) Tìm *C.*
b) Tìm hàm mật độ thành phần của *X.* Tính $P(X < 1)$.
c) Tính xác suất $P(X \le 1, Y > 2)$.

- ***Đáp án tham khảo:***

a) $$\iint_{R^2} f(x,y)\,dx\,dy = 1 \Leftrightarrow \int_0^{+\infty}\int_0^2 C(x-2)e^{-y}dx\,dy = 1$$

$$\Leftrightarrow \int_0^{+\infty}\left(\left(C\frac{x^2}{2}e^{-y} - 2Ce^{-y}\right)\Big|_0^2\right)dy = 1$$

$$\Leftrightarrow \int_0^{+\infty}(-2Ce^{-y})dy = 1 \Leftrightarrow 2C\left(-e^{y}\Big|_0^{+\infty}\right) = 1$$

$$\Rightarrow C = \frac{-1}{2}$$

b) $$f_X(x) = \begin{cases} \displaystyle\int_0^{+\infty}\frac{-1}{2}(x-2)e^{-y}dy = \frac{-1}{2}(x-2) & \forall\, x \in [0,2] \\ 0 & \forall\, x \notin [0,2] \end{cases}$$

$$P(X < 1) = \int_0^1 \frac{-1}{2}(x-2)dx = \frac{3}{4}$$

c) $$P(X \le 1, Y > 2) = \int_2^{+\infty}\int_0^1 \frac{-1}{2}(x-2)e^{-y}dx\,dy$$

$$= \int_2^{+\infty}\left(\left(-\frac{x^2}{4}e^{-y} + xe^{-y}\right)\Big|_0^1\right)dy = \int_2^{+\infty}\left(-\frac{1}{4}e^{-y} + e^{-y}\right)dy = \frac{3}{4}e^2$$

Câu 4. *(1.5 điểm)* Để kiểm tra chất lượng vận chuyển hàng hóa của một công ty, người ta kiểm tra 200 hàng hóa do công ty đó vận chuyển và phát hiện có 24 hàng hóa bị hỏng trong quá trình vận chuyển. Ước lượng tỷ lệ hàng hóa bị hỏng do công ty đó vận chuyển với độ tin cậy 96%.

- ***Đáp án tham khảo:***

$$n = 200, \quad k = 24, \quad f = \frac{k}{n} = 0.12, \quad 1 - \alpha = 0.96 \Rightarrow \alpha = 0.04$$

$$\phi\left(z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} = 1 - \frac{0,04}{2} = 0,98 \xrightarrow{\text{tra bảng A4}} z_{\frac{\alpha}{2}} = 2,055$$

$$\varepsilon = z_{\frac{\alpha}{2}}.\sqrt{\frac{f(1-f)}{n}} = 2,055.\sqrt{\frac{0,12.(1-0,12)}{200}} = 0,0472$$

Khoảng ước lượng tỷ lệ: $0,12 \pm 0,047$ hoặc $[0,073; 0,167]$

Câu 5. *(2 điểm)* Để đảm bảo một máy chủ được sử dụng hiệu quả, cần ước lượng trung bình có bao nhiêu người sử dụng máy chủ đó tại cùng thời điểm. Quan sát tại 100 thời điểm khác nhau được chọn một cách ngẫu nhiên, người ta thấy trung bình có 37.7 người sử dụng máy chủ đó tại cùng thời điểm với độ lệch chuẩn σ = 9.2.

a) Ước lượng trung bình có bao nhiêu người sử dụng máy chủ đó tại cùng thời điểm với độ tin cậy là 90%.

b) Với mức ý nghĩa 1%, có thể kết luận rằng trung bình có trên 35 người sử dụng máy chủ đó tại cùng thời điểm hay không?

- ***Đáp án tham khảo:***

$n = 100, \qquad \bar{x} = 37.7, \qquad \sigma = 9.2$

a) $1 - \alpha = 90\% \Rightarrow \alpha = 0.1$

$$\phi\left(z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} = 1 - \frac{0,1}{2} = 0,95 \xrightarrow{\text{tra bảng A4}} z_{\frac{\alpha}{2}} = 1,645$$

$$\varepsilon = z_{\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}} = 1,645.\frac{9.2}{\sqrt{100}} = 1,5134$$

Khoảng ước lượng tỷ lệ: $37,7 \pm 1,5134$ hoặc $[36,1866; 39,2134]$

b) $H_0: \mu = 35, \quad H_1: \mu > 35, \quad \alpha = 0.01$

$$\phi(z_\alpha) = 1 - \alpha = 1 - 0,01 = 0,99 \xrightarrow{\text{tra bảng A4}} z_{\frac{\alpha}{2}} = 2,325$$

$$z = \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n} = \frac{37,7 - 35}{9,2}.\sqrt{100} = 2,935 > z_\alpha \Rightarrow \textit{bác bỏ } H_0, \textit{chấp nhận } H_1$$

Vậy có thể kết luận có trên 35 người sử dụng máy chủ đó tại cùng thời điểm.

Câu 6. *(1 điểm)* Một người muốn biết hiệu quả của một chương trình máy tính phụ thuộc như thế nào vào kích cỡ của dữ liệu đầu vào. Hiệu quả của chương trình được đo bằng số yêu cầu được xử lí trong một giờ. Gọi x là kích cỡ của dữ liệu đầu vào và y là số yêu cầu được xử lí trong một giờ. Chạy chương trình cho tập hợp dữ liệu có kích thước khác nhau, người đó nhận được kết quả sau:

x (gigabytes)	6	7	7	8	10	10	15
y	40	55	50	41	17	26	16

a) Viết phương trình hồi quy tuyến tính của y theo x.

b) Nếu dữ liệu đầu vào có kích cỡ là 14 gigabytes thì chương trình có thể xử lí khoảng bao nhiêu yêu cầu?

- ***Đáp án tham khảo:***

a) Phương trình hồi quy tuyến tính của y theo x: $y = 73,29 - 4,143x$

b) Nếu x = 14 (gigabytes) thì y = 15,34.