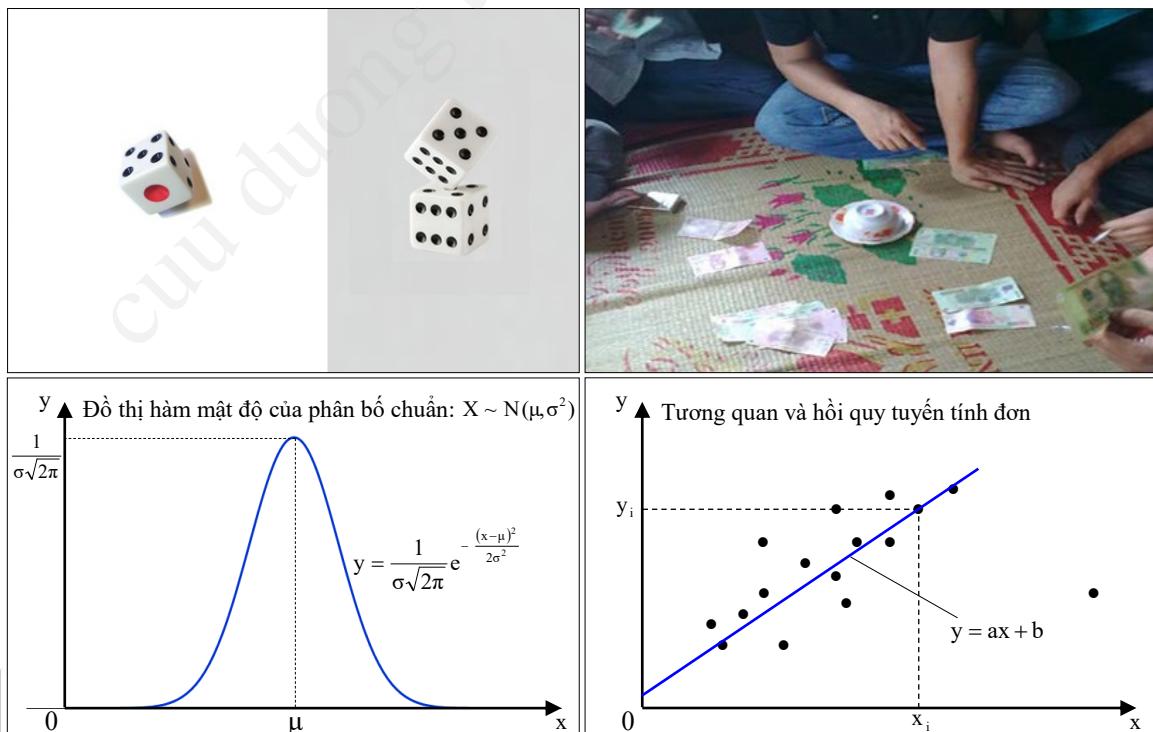


ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN



HƯỚNG DẪN GIẢI BÀI TẬP  
**XÁC SUẤT – THỐNG KÊ**



Hà Nội, 11/2013

**ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**HƯỚNG DẪN GIẢI BÀI TẬP  
XÁC SUẤT – THỐNG KÊ**

(Dành cho sinh viên ngoài khoa Toán)

**SINH VIÊN : HOÀNG VĂN TRỌNG  
NGÀNH : Địa lý tự nhiên  
ĐIỆN THOẠI : 0974 971 149  
EMAIL : hoangtronghus@gmail.com**

**Hà Nội, 11/2013**

## *Lời chia sẻ*

Hầu hết các hiện tượng trong cuộc sống đều xảy ra một cách ngẫu nhiên không thể đoán biết được. Chúng ta luôn đứng trước những lựa chọn và phải quyết định cho riêng mình. Khi lựa chọn như thế thì khả năng thành công là bao nhiêu, phương án nào là tối ưu, cơ sở của việc lựa chọn là gì? Khoa học về *Xác suất* sẽ giúp ta định lượng khả năng thành công của từng phương án để từ đó đưa ra quyết định đúng đắn.

*Thống kê* là khoa học về cách thu thập, xử lý và phân tích dữ liệu về hiện tượng rồi đưa ra kết luận có tính quy luật của hiện tượng đó. Phân tích thống kê dựa trên cơ sở của lý thuyết xác suất và có quan hệ chặt chẽ với xác suất; nó không nghiên cứu từng cá thể riêng lẻ mà nghiên cứu một tập hợp cá thể - tính quy luật của toàn bộ tổng thể. Từ việc điều tra và phân tích mẫu đại diện, có thể tạm thời đưa ra kết luận về hiện tượng nghiên cứu nhưng với khả năng xảy ra sai lầm đủ nhỏ để có thể chấp nhận được.

Trong chương trình đào tạo theo tín chỉ của các ngành ngoài khoa Toán thì *Xác suất* và *Thống kê* được gộp chung lại thành môn “*Xác suất thống kê*” với những nội dung rút gọn, đáp ứng nhu cầu về toán cho các đối tượng không chuyên. File này tập trung vào phân loại và hướng dẫn giải các dạng bài tập. Đa số các bài tập được lấy từ 3 chương đầu của giáo trình G<sub>1</sub> và 3 chương cuối của giáo trình G<sub>2</sub> (xem *Tài liệu tham khảo*). Ngoài ra, một số bài tập còn được lấy từ thực tế hoặc từ các lớp môn học khác nhau. Phần lý thuyết chỉ tóm lược nội dung chính cùng một số công thức áp dụng (xem chứng minh công thức trong giáo trình G<sub>1</sub> và G<sub>2</sub>).

Kiến thức bổ trợ cho môn học này chủ yếu là Giải tích tổ hợp (hoán vị, chỉnh hợp, tổ hợp) và tích phân của hàm một biến (xem *Phụ lục P.I*). Theo kinh nghiệm cá nhân thì phương pháp học Xác suất – Thống kê không giống những môn Đại số hoặc Giải tích mà cần hiểu kỹ lý thuyết mới dễ dàng ghi nhớ công thức và áp dụng linh hoạt vào giải bài tập cụ thể. Tuy đề thi cuối kỳ thường cho phép sử dụng tài liệu nhưng việc ghi nhớ và nắm bắt ý nghĩa các công thức sẽ giúp phản xạ tốt hơn cũng như xác định dạng bài toán chính xác hơn.

Những dòng chữ nhỏ phía cuối trang là phần giải thích và chỉ dẫn. Sau mỗi bài tập khó thường có mục “hướng dẫn” giải ở dạng khái quát. Khi cần tham khảo tài liệu này, các bạn truy cập vào “*Link download*” ở cuối file để tải về bản cập nhật mới nhất.

- Trên đây là một số kiến thức mà mình muốn chia sẻ cùng các bạn. Do thời gian hạn hẹp và sự nhận thức non kém về môn học nên chắc chắn còn nội dung nào đó viết chưa đúng hoặc chưa đầy đủ, mong các bạn thông cảm và góp ý để mình hoàn thiện thêm.

Mọi thắc mắc và góp ý xin gửi về địa chỉ email: [hoangtronghus@gmail.com](mailto:hoangtronghus@gmail.com) hoặc địa chỉ: [facebook.com/hoangtrongdl](https://facebook.com/hoangtrongdl)

Sinh viên

**Hoàng Văn Trọng**

cuuduongthancong.com

# MỤC LỤC

<b>PHẦN I: XÁC SUẤT</b>	1
<b>CHƯƠNG 1: BIẾN CỐ VÀ XÁC SUẤT CỦA BIẾN CỐ</b>	<b>1</b>
A. LÝ THUYẾT	1
1.1. Một số khái niệm cơ bản.....	1
1.2. Xác suất của biến cố .....	2
1.3. Các quy tắc tính xác suất .....	3
1.4. Công thức Bernoulli.....	3
1.5. Xác suất có điều kiện. Quy tắc nhân tổng quát.....	3
1.6. Công thức xác suất đầy đủ .....	4
1.7. Công thức Bayes.....	4
B. BÀI TẬP	4
1.1. Bài tập trong giáo trình 1 (G <sub>1</sub> ) .....	4
1.2. Nhận xét bài tập chương 1 .....	18
<b>CHƯƠNG 2: ĐẠI LƯỢNG NGẪU NHIÊN RỜI RẠC</b>	<b>20</b>
A. LÝ THUYẾT	20
2.1. Phân bố xác suất và hàm phân bố .....	20
2.2. Một số đặc trưng của đại lượng ngẫu nhiên rời rạc .....	20
2.3. Phân bố đồng thời và hệ số tương quan.....	21
2.4. Hàm của đại lượng ngẫu nhiên rời rạc.....	22
2.5. Phân bố nhị thức .....	23
2.6. Phân bố Poisson.....	23
B. BÀI TẬP	24
2.1. Bài tập trong giáo trình 1 (G <sub>1</sub> ) .....	24
2.2. Nhận xét bài tập chương 2 .....	40
<b>CHƯƠNG 3: ĐẠI LƯỢNG NGẪU NHIÊN LIÊN TỤC</b>	<b>41</b>
A. LÝ THUYẾT	41
3.1. Hàm mật độ xác suất và hàm phân bố xác suất .....	41
3.2. Một số đặc trưng của đại lượng ngẫu nhiên liên tục.....	41
3.3. Hàm của đại lượng ngẫu nhiên liên tục .....	42
3.4. Phân bố chuẩn .....	42
3.5. Phân bố mũ .....	43
3.6. Phân bố đều.....	44
B. BÀI TẬP	45
3.1. Bài tập trong giáo trình 1 (G <sub>1</sub> ) .....	45
3.2. Nhận xét bài tập chương 3 .....	63
<b>PHẦN II: THỐNG KÊ</b>	<b>64</b>
<b>CHƯƠNG 4: BÀI TOÁN ƯỚC LƯỢNG THAM SỐ</b>	<b>64</b>
A. LÝ THUYẾT	64
4.1. Một số kiến thức chuẩn bị thêm cho phần thống kê .....	64
4.2. Mẫu ngẫu nhiên và các đặc trưng của mẫu.....	66
4.3. Ước lượng điểm .....	67
4.4. Ước lượng khoảng .....	68

4.5. Số quan sát cần thiết để có sai số (hoặc độ tin cậy) cho trước .....	69
<b>B. BÀI TẬP .....</b>	<b>70</b>
4.1. Bài tập trong giáo trình 2 (G <sub>2</sub> ).....	70
4.2. Nhận xét bài tập chương 4.....	79
<b>CHƯƠNG 5: BÀI TOÁN KIỂM ĐỊNH GIÁ THIẾT.....</b>	<b>80</b>
<b>A. LÝ THUYẾT .....</b>	<b>80</b>
5.1. Kiểm định giả thiết cho giá trị trung bình .....	80
5.2. Kiểm định giả thiết cho phương sai .....	81
5.3. Kiểm định giả thiết cho tỷ lệ (hay xác suất).....	81
5.4. So sánh hai giá trị trung bình .....	82
5.5. So sánh hai phương sai.....	83
5.6. So sánh hai tỷ lệ (hay hai xác suất) .....	83
5.7. Tiêu chuẩn phù hợp Khi bình phương.....	84
5.8. Kiểm tra tính độc lập .....	85
5.9. So sánh nhiều tỷ lệ .....	85
<b>B. BÀI TẬP .....</b>	<b>86</b>
5.1. Bài tập trong giáo trình 2 (G <sub>2</sub> ).....	86
5.2. Nhận xét bài tập chương 5.....	111
<b>CHƯƠNG 6: BÀI TOÁN TƯƠNG QUAN VÀ HỒI QUY .....</b>	<b>112</b>
<b>A. LÝ THUYẾT .....</b>	<b>112</b>
6.1. Hệ số tương quan mẫu.....	112
6.2. Đường hồi quy tuyến tính thực nghiệm .....	112
<b>B. BÀI TẬP .....</b>	<b>113</b>
6.1. Bài tập trong giáo trình 2 (G <sub>2</sub> ).....	113
6.2. Nhận xét bài tập chương 6.....	115
<b>KẾT LUẬN .....</b>	<b>116</b>
<b>MỘT SỐ ĐỀ THI CUỐI KỲ .....</b>	<b>118</b>
1. Đề thi cuối kỳ II năm học 2012 – 2013 .....	118
2. Đề thi cuối kỳ I năm học 2013 – 2014 .....	126
3. Đề thi cuối kỳ II năm học 2013 – 2014.....	134
4. Đề thi cuối kỳ phụ – hè năm 2014 .....	141
5. Đề thi cuối kỳ I năm học 2014 – 2015 .....	148
6. Đề thi cuối kỳ II năm học 2014 – 2015 .....	154
7. Đề thi cuối kỳ I năm học 2015 – 2016 .....	160
8. Đề thi cuối kỳ I năm học 2015 – 2016 (dành cho sinh viên khoa Vật lý, hệ chuẩn) .....	166
9. Đề thi cuối kỳ II năm học 2015 – 2016 .....	172
10. Đề thi cuối kỳ II năm học 2015 – 2016 (dành cho sinh viên khoa Vật lý, hệ chuẩn) .....	178
11. Đề thi cuối kỳ phụ - hè năm 2016 .....	183
<b>PHỤ LỤC .....</b>	<b>189</b>
P.1. Kiến thức chuẩn bị .....	189
P.2. Tính toán chỉ số thống kê bằng máy tính bỏ túi .....	191
P.3. Tính toán xác suất thống kê bằng hàm trong Excel .....	195
P.4. Bảng tra cứu một số phân bố thường gặp .....	198
<b>TÀI LIỆU THAM KHẢO .....</b>	<b>212</b>

## PHẦN I: XÁC SUẤT<sup>1</sup>

# CHƯƠNG 1: BIẾN CỐ VÀ XÁC SUẤT CỦA BIẾN CỐ<sup>2</sup>

### A. LÝ THUYẾT

#### 1.1. Một số khái niệm cơ bản

a) *Phép thử ngẫu nhiên*:  $\xi$

- Là những hành động mà không biết trước được kết quả.

VD: tung đồng xu, gieo con xúc xắc, đánh lô đề,..

b) *Không gian mẫu*:  $\Omega$

- Là tập hợp chứa tất cả các kết quả có thể xảy ra của phép thử ngẫu nhiên.

c) *Biến cố* (hay *sự kiện*): A, B, C, D

- Là tập hợp chứa một hoặc một số phần tử có thể xảy ra của phép thử ngẫu nhiên.

$$A, B, C, D \subset \Omega$$

+ *Biến cố sơ cấp*: Là biến cố không thể chia tách nhỏ hơn được nữa.

+ *Biến cố chắc chắn*: Là biến cố luôn luôn xảy ra. Nó tương ứng với toàn bộ tập không gian mẫu  $\Omega$

+ *Biến cố rỗng* (*biến cố không thể*) là biến cố không bao giờ xảy ra. Nó tương ứng với tập con rỗng  $\emptyset$  của  $\Omega$

d) *Quan hệ giữa các biến cố*:

- *Quan hệ kéo theo*: A kéo theo B nếu A xảy ra thì B cũng xảy ra.

$$A \text{ kéo theo } B \Leftrightarrow A \subset B$$

- *Giao của hai biến cố*: là biến cố xảy ra khi cả 2 biến cố đã cho cùng xảy ra.

$$A \cap B \text{ (hay } AB\text{)}$$

- *Hợp của hai biến cố*: là biến cố xảy ra khi ít nhất 1 trong 2 biến cố đã cho xảy ra.

$$A \cup B \text{ (hay } A + B\text{)}$$

- *Biến cố đối của biến cố A*: là biến cố xảy ra khi và chỉ khi A không xảy ra.

$$\bar{A} = \{\Omega \setminus A\}$$

<sup>1</sup> Phần Xác suất thì đa số các lớp học theo giáo trình G<sub>1</sub> (xem Tài liệu tham khảo), rất ít lớp học theo giáo trình G<sub>2</sub> hoặc G<sub>4</sub>.

<sup>2</sup> So với chương 2 và chương 3 của phần Xác suất thì bài tập của chương 1 khó hơn và hay nhầm lẫn. Bài tập chương 1 thường ra vào các dạng: phép thử lặp Bernoulli, xác suất có điều kiện, công thức xác suất đầy đủ, công thức Bayes hoặc kết hợp các dạng này với nhau trong cùng một bài toán.

- Hai biến có A và B được gọi là độc lập nếu việc xảy ra A không ảnh hưởng đến việc xảy ra B và ngược lại.
- Hai biến có A và B được gọi là xung khắc nếu A và B không đồng thời xảy ra.

$$AB = \emptyset$$

e) Biểu diễn một số biến có thường gặp:

Gọi:

A = “Hiện tượng 1 xảy ra”

B = “Hiện tượng 2 xảy ra”

C = “Hiện tượng 3 xảy ra”

Thì:

$ABC$ : Cả ba hiện tượng cùng xảy ra.

$\overline{ABC}$ : Cả ba hiện tượng cùng không xảy ra.

$A \cup B \cup C$ : Có ít nhất một hiện tượng xảy ra.

$AB \cup BC \cup CA$ : Có ít nhất hai hiện tượng xảy ra.

$\overline{AB} \cup \overline{BC} \cup \overline{CA}$ : Có ít nhất hai hiện tượng không xảy ra.

$A\overline{B}\overline{C} \cup \overline{A}B\overline{C} \cup \overline{A}\overline{B}C$ : Chỉ có một hiện tượng xảy ra.

$A\overline{B}\overline{C}$ : Chỉ có hiện tượng 1 xảy ra.

## 1.2. Xác suất của biến có

Xác suất của biến có là một giá trị đo lường khả năng xuất hiện biến có đó khi thực hiện phép thử ngẫu nhiên. Ký hiệu xác suất của biến có A là:  $P(A)$

Tính chất:  $0 \leq P(A) \leq 1$

$$P(\emptyset) = 0$$

$$P(\Omega) = 1$$

a) <sup>1</sup>Định nghĩa cổ điển cho xác suất của biến có A:  $P(A) = \frac{|A|}{|\Omega|}$

Trong đó:  $|A|$  là số lượng các biến có sơ cấp có lợi cho A

$|\Omega|$  là tổng các biến có sơ cấp của không gian mẫu

Để áp dụng công thức xác suất cổ điển phải thỏa mãn đồng thời hai điều kiện sau:

{ Tổng số các biến có sơ cấp là hữu hạn.  
Các biến có sơ cấp có cùng khả năng xảy ra.

<sup>1</sup> Cách tính xác suất trong môn học này chủ yếu là theo trường phái cổ điển.

*b) Định nghĩa xác suất bằng tần suất:*

Khi tổng số các kết quả có thể xảy ra là vô hạn hoặc hữu hạn nhưng không đồng khả năng thì ta dùng định nghĩa xác suất bằng tần suất:

Thực hiện phép thử ngẫu nhiên n lần, trong điều kiện giống hệt nhau. Trong n lần đó thấy có k lần xuất hiện biến cố A thì xác suất của A được định nghĩa bởi giới hạn sau:

$$P(A) = \lim_{n \rightarrow +\infty} \frac{k}{n}$$

Trên thực tế thì  $P(A)$  được tính xấp xỉ bằng tỷ số  $\frac{k}{n}$  khi n đủ lớn.

### 1.3. Các quy tắc tính xác suất

*a) Quy tắc cộng:*

$$P(A+B) = P(A) + P(B) - P(AB)$$

$$P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC) + P(ABC)$$

Nếu A và B xung khắc thì:

$$P(A+B) = P(A) + P(B)$$

*b) Quy tắc nhân (trong trường hợp A và B độc lập):*

$$P(AB) = P(A) \cdot P(B)$$

*c) Quy tắc chuyển sang biến cố đối:*

$$P(\bar{A}) = 1 - P(A)$$

### 1.4. Công thức Bernoulli

Thực hiện phép thử ngẫu nhiên n lần một cách độc lập, trong điều kiện giống hệt nhau. Ở mỗi lần thử, xác suất của biến cố A bằng p ( $0 < p < 1$ ) thì xác suất để A xuất hiện **đúng k lần** trong n phép thử là:

$$P_k = C_n^k p^k q^{n-k} \quad (\text{với } q = 1 - p)$$

### 1.5. Xác suất có điều kiện. Quy tắc nhân tổng quát

*Khả năng để biến cố A xảy ra nếu biến cố B đã xảy ra được gọi là xác suất của A với điều kiện B. Ký hiệu:  $P(A | B)$*

$$P(A | B) = \frac{P(AB)}{P(B)}$$

Từ công thức xác suất có điều kiện ở trên, suy ra quy tắc nhân tổng quát:

$$P(AB) = P(A | B) \cdot P(B)$$

$$P(ABC) = P(A | BC) \cdot P(B | C) \cdot P(C)$$

$$P(A_1 A_2 \dots A_n) = P(A_1 | A_2 A_3 \dots A_n) \cdot P(A_2 | A_3 A_4 \dots A_n) \dots P(A_{n-1} | A_n) \cdot P(A_n)$$

## 1.6. Công thức xác suất đầy đủ

(hay *công thức xác suất toàn phần* hay *công thức xác suất tiền nghiệm*)

+ Hệ các biến cõi  $B_1, B_2, \dots, B_n$  được gọi là hệ đầy đủ nếu đồng thời thỏa mãn:

$$\begin{cases} B_1 \cup B_2 \cup \dots \cup B_n = \Omega \\ B_i B_j = \emptyset \text{ nếu } i \neq j \end{cases}$$

+ Nếu hệ biến cõi  $\{B_1, B_2, \dots, B_n\}$  là một hệ đầy đủ thì với biến cõi  $H$  bất kỳ, ta có:

$$P(H) = \sum_{i=1}^n P(H|B_i) = \sum_{i=1}^n P(H|B_i) \cdot P(B_i)$$

## 1.7. Công thức Bayes

(hay *công thức xác suất hậu nghiệm*)

Nếu hệ biến cõi  $\{B_1, B_2, \dots, B_n\}$  là một hệ đầy đủ và  $P(H) > 0$  thì:

$$P(B_k | H) = \frac{P(H|B_k)}{P(H)} = \frac{P(H|B_k) \cdot P(B_k)}{\sum_{i=1}^n P(H|B_i) \cdot P(B_i)} \quad (\text{với } 1 \leq k \leq n)$$

## B. BÀI TẬP

### 1.1. Bài tập trong giáo trình 1 ( $G_1$ )

(Mở đầu về lý thuyết xác suất và các ứng dụng, Đặng Hùng Thắng, trang 37)

#### Bài 1/37: Gieo đồng thời 2 con xúc xắc. Tìm xác suất để:

- a) Tổng số nốt là 7;
- b) Tổng số nốt là 8;
- c) Số nốt hơn kém nhau 2.

a) Xác suất để tổng số nốt bằng 7:

Tổng số kết quả có thể xảy ra là:  $6 \cdot 6 = 36$

Có 6 kết quả có tổng bằng 7 là: (1, 6); (6, 1); (2, 5); (5, 2); (3, 4); (4, 3)

$$\Rightarrow \text{Xác suất để tổng số nốt bằng 7 là: } \frac{6}{36} = \boxed{\frac{1}{6}}$$

b) Xác suất để tổng số nốt bằng 8:

Có 5 kết quả có tổng bằng 8 là: (2, 6); (6, 2); (3, 5); (5, 3); (4, 4)

$$\Rightarrow \text{Xác suất để tổng số nốt bằng 8 là: } \boxed{\frac{5}{36}}$$

c) Xác suất để số nốt hơn kém nhau 2:

Có 8 kết quả mà số nốt hơn kém nhau 2, bao gồm:

(1,3); (3,1); (2, 4); (4, 2); (3, 5); (5, 3); (4, 6); (6, 4)

$\Rightarrow$  Xác suất để số nốt hơn kém nhau 2 là:  $\frac{8}{36} = \boxed{\frac{2}{9}}$

**Bài 2/37:** Một khách sạn có 6 phòng đơn. Có 10 khách đến thuê phòng trong đó có 6 nam và 4 nữ. Người quản lý chọn ngẫu nhiên 6 người. Tìm xác suất để trong đó:

- a) Cả 6 người đều là nam;
- b) Có 4 nam và 2 nữ;
- c) Có ít nhất hai nữ.

a) Xác suất cả 6 người đều là nam<sup>1</sup>:

$$\text{Tổng số kết quả có thể xảy ra: } C_{10}^6 = 210$$

$$\text{Số kết quả thuận lợi: } C_6^6 \cdot C_4^0 = 1$$

$\Rightarrow$  Xác suất để 6 người đều là nam:  $\boxed{\frac{1}{210}}$

b) Xác suất có 4 nam và 2 nữ:

$$\frac{C_6^4 \cdot C_4^2}{210} = \frac{90}{210} = \boxed{\frac{3}{7}}$$

c) Xác suất có ít nhất 2 nữ:

$$\text{Xác suất có nhiều nhất 1 nữ: } \frac{C_4^0 \cdot C_6^6}{210} + \frac{C_4^1 \cdot C_6^5}{210} = \frac{25}{210}$$

$$\Rightarrow \text{Xác suất có ít nhất 2 nữ: } 1 - \frac{25}{210} = \frac{185}{210} = \boxed{\frac{37}{42}}$$

**Bài 3/37:** Một công ty cần tuyển 2 nhân viên, có 6 người nộp đơn trong đó có 2 nam và 4 nữ. Biết rằng khả năng được tuyển của mỗi người là như nhau.

- a) Tính xác suất để cả hai người được chọn là nữ;
- b) Tính xác suất để ít nhất một nữ được chọn;
- c) Tính xác suất để cả hai nữ được chọn nếu biết rằng có ít nhất một nữ đã được chọn;
- d) Giả sử Hoa là một trong 4 nữ. Tính xác suất để Hoa được chọn. Tính xác suất để Hoa được chọn nếu biết rằng có ít nhất một nữ được chọn.

a) Xác suất cả hai người được chọn đều là nữ:  $\frac{C_2^0 \cdot C_4^2}{C_6^2} = \frac{6}{15} = \boxed{\frac{2}{5}}$

<sup>1</sup> Xem lại kiến thức giải tích tổ hợp ở phần “Phụ lục P.I”, trang 189

b) Xác suất để ít nhất một nữ được chọn:

$$\text{Xác suất không có nữ nào được chọn: } \frac{C_2^2 \cdot C_4^0}{C_6^2} = \frac{1}{15}$$

$$\Rightarrow \text{Xác suất để ít nhất một nữ được chọn: } 1 - \frac{1}{15} = \boxed{\frac{14}{15}}$$

c) Xác suất cả 2 nữ được chọn nếu biết rằng ít nhất một nữ đã được chọn<sup>1</sup>:

A = “Cả 2 nữ được chọn”

B = “Ít nhất một nữ được chọn”

$$\Rightarrow P(A | B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} \quad (\text{Vì } A \subset B)$$

$$= \frac{2/5}{14/15} = \boxed{\frac{3}{7}}$$

d) Xác suất Hoa được chọn và xác suất Hoa được chọn nếu biết rằng ít nhất một nữ đã được chọn:

C = “Hoa được chọn”

Giả sử trong 6 người bao gồm Hoa và 5 người khác theo thứ tự: 1, 2, 3, 4, 5. Có 5 kết quả có thể xảy ra trong đó có Hoa: (Hoa, 1); (Hoa, 2); (Hoa, 3); (Hoa, 4); (Hoa, 5)

$$\Rightarrow \text{Xác suất để Hoa được chọn: } P(C) = \frac{5}{15} = \boxed{\frac{1}{3}}$$

$\Rightarrow$  Xác suất để Hoa được chọn nếu biết rằng ít nhất một nữ đã được chọn:

$$P(C | B) = \frac{P(CB)}{P(B)} = \frac{P(C)}{P(B)} = \frac{1/3}{14/15} = \boxed{\frac{5}{14}} \quad (\text{Vì } C \subset B)^2$$

**Bài 4/37: Một hòm có 9 tấm thẻ đánh số từ 1 đến 9. Chọn ngẫu nhiên ra hai tấm thẻ. Tính xác suất để tích của hai số trên hai tấm thẻ là một số chẵn.**

Tổng số kết quả có thể xảy ra:  $C_9^2 = 36$

Để tích số trên hai tấm thẻ là một số chẵn thì phải có ít nhất một trong hai tấm thẻ mang số chẵn. Nếu cả hai tấm thẻ đều mang số lẻ (1, 3, 5, 7, 9) thì tích của chúng là một số lẻ.

$$\text{Xác suất để tích 2 số là một số lẻ: } \frac{C_5^2}{36} = \frac{5}{18}$$

<sup>1</sup> Thông thường, nếu trong câu hỏi mà xuất hiện cụm từ “biết rằng”, “nếu biết”,... thì sử dụng công thức xác suất có điều kiện để giải.

<sup>2</sup> Biến cố A và C đều là con của biến cố B vì: cả 2 nữ được chọn hay mình Hoa được chọn cũng đều suy ra có ít nhất một nữ được chọn. Giao của một biến cố với biến cố con của nó thì bằng chính biến cố con đó.

$$\Rightarrow \text{Xác suất để tích 2 số là một số chẵn: } 1 - \frac{5}{18} = \boxed{\frac{13}{18}}$$

**Bài 5/37:** Ở một nước có 50 tỉnh, mỗi tỉnh có hai đại biểu Quốc hội. Người ta chọn ngẫu nhiên 50 đại biểu trong số 100 đại biểu để thành lập một Ủy ban. Tính xác suất để:

- a) Trong Ủy ban có ít nhất một đại biểu của Thủ đô;
- b) Mỗi tỉnh đều có đúng một đại biểu trong Ủy ban.

a) Xác suất trong Ủy ban có ít nhất một đại biểu của Thủ đô:

$$A = \text{"Có ít nhất một đại biểu của Thủ đô"}$$

$$\text{Xác suất để không có đại biểu của Thủ đô: } P(\bar{A}) = \frac{C_2^0 \cdot C_{98}^{50}}{C_{100}^{50}} = 0,2475$$

$$\Rightarrow \text{Xác suất có ít nhất một đại biểu của Thủ đô: } P(A) = 1 - P(\bar{A}) = \boxed{0,7525}$$

b) Xác suất mỗi tỉnh có đúng một đại biểu trong Ủy ban:

$$B = \text{"Mỗi tỉnh có đúng một đại biểu trong Ủy ban"}$$

Số cách chọn mỗi tỉnh một đại biểu:  $C_2^1 \cdot C_2^1 \dots C_2^1$  (50 số hạng)

$$\Rightarrow \text{Xác suất cần tìm: } P(B) = \frac{C_2^1 \cdot C_2^1 \dots C_2^1}{C_{100}^{50}} = \boxed{\frac{2^{50}}{C_{100}^{50}}} = 1,116 \cdot 10^{-14}$$

**Bài 6/38:** Trong tuần lễ vừa qua ở thành phố có 7 tai nạn giao thông. Xác suất để mỗi ngày xảy ra đúng một tai nạn là bao nhiêu?

Mỗi một tai nạn giao thông có thể rơi vào 1 trong 7 ngày trong tuần. Số cách xảy ra của 7 tai nạn giao thông trong tuần:  $7^7$  cách

Số cách xảy ra đúng 1 tai nạn giao thông trong mỗi ngày:  $7!$  cách

$\Rightarrow$  Xác suất để mỗi ngày xảy ra đúng một tai nạn giao thông:

$$\boxed{\frac{7!}{7^7}} = 0,00612$$

**Bài 7/38:** Một đoàn tàu có 4 toa đỗ ở một sân ga. Có 4 hành khách từ sân ga lên tàu, mỗi người độc lập với nhau chọn ngẫu nhiên một toa. Tính xác suất để một toa có 3 người, một toa có 1 người còn hai toa còn lại không có ai lên.

*Hướng dẫn: Chọn người xong rồi chọn toa. Đầu tiên chọn nhóm 3 người, tiếp theo chọn toa tàu cho nhóm này, người cuối cùng thì chọn trong các toa còn lại.*

Mỗi người có 4 lựa chọn toa tàu. Tổng số kết quả có thể xảy ra là:  $4^4 = 256$

Đầu tiên, chọn 3 trong số 4 người:  $C_4^3$  cách.

Có 4 cách chọn toa tàu cho nhóm 3 người trên. Người thứ tư có 3 cách chọn trong ba toa còn lại.

⇒ Xác suất để một toa có 3 người, một toa có 1 người và hai toa còn lại không có ai lên:

$$\frac{C_4^3 \cdot 4 \cdot 3}{256} = \frac{48}{256} = \boxed{\frac{3}{16}}$$

**Bài 8/38:** Một máy bay có ba bộ phận A, B, C với tầm quan trọng khác nhau. Máy bay sẽ rơi khi có hoặc một viên đạn trúng vào A, hoặc hai viên đạn trúng B, hoặc ba viên đạn trúng C. Giả sử các bộ phận A, B và C lần lượt chiếm 15%, 30% và 55% diện tích máy bay. Tìm xác suất để máy bay rơi nếu:

- a) Máy bay bị trúng hai viên đạn;
- b) Máy bay bị trúng ba viên đạn.

a) Xác suất máy bay rơi nếu trúng 2 viên đạn:

$$D = \text{"Máy bay rơi"}$$

Máy bay rơi khi có ít nhất 1 viên trúng bộ phận A hoặc cả hai viên trúng B:

+ Xác suất để ít nhất 1 viên trúng A (biến cố này là đối của biến cố: không viên nào trúng A):

$$1 - (0,3 + 0,55)^2 = 0,2775$$

$$+ \text{Xác suất để cả 2 viên trúng B: } 0,3^2 = 0,09$$

$$\Rightarrow \text{Xác suất để máy bay rơi: } P(D) = 0,2775 + 0,09 = \boxed{0,3675}$$

b) Xác suất máy bay rơi nếu trúng 3 viên đạn:

Máy bay không rơi chỉ khi 1 viên trúng B và 2 viên còn lại phải trúng C. Xác suất để máy bay không rơi:

$$3 \cdot (0,3 \cdot 0,55^2) = 0,27225 \text{ (có 3 cách chọn viên đạn trúng B)}$$

$$\Rightarrow \text{Xác suất để máy bay rơi: } P(D) = 1 - 0,27225 = \boxed{0,72775}$$

**Bài 9/38:** Trong một thành phố nào đó 65% dân cư thích xem đá bóng. Chọn ngẫu nhiên 12 người, hãy tính xác suất để trong đó có đúng 5 người thích xem đá bóng.

*Hướng dẫn:* Tỷ lệ người dân thích xem bóng đá là 65% nên khi chọn ngẫu nhiên một người thì xác suất để người đó thích xem bóng đá là 0,65. Chọn ngẫu nhiên 12 người tương đương với 12 phép thử lặp Bernoulli.

⇒ Xác suất để có đúng 5 người thích xem bóng đá trong 12 người được chọn ngẫu nhiên:

$$P_5(12; 0,65) = C_{12}^5 \cdot 0,65^5 \cdot (1 - 0,65)^7 = \boxed{0,0591}$$

**Bài 10/38:** Một sọt cam rất lớn được phân loại theo cách sau: Chọn ngẫu nhiên 20 quả cam làm mẫu đại diện. Nếu mẫu này không chứa quả cam hỏng nào thì sọt cam được xếp loại 1. Nếu mẫu có một hoặc hai quả hỏng thì sọt cam xếp loại 2. Trong trường hợp còn lại (có từ 3 quả hỏng trở lên) sọt cam được xếp loại 3.

Trên thực tế 3% số cam trong sọt bị hỏng. Tìm xác suất để sọt cam được xếp loại:

- a) **Loại 1;**
- b) **Loại 2;**
- c) **Loại 3.**

Chọn ngẫu nhiên 20 quả cam, xác suất chọn được quả hỏng trong mỗi lần là 0,03.

a) *Xác suất sọt cam xếp loại 1:*

Mẫu không chứa quả cam nào hỏng.

$$P_0(20; 0,03) = C_{20}^0 \cdot 0,03^0 \cdot (1 - 0,03)^{20} = 0,97^{20} = [0,5438]$$

b) *Xác suất sọt cam xếp loại 2:*

Mẫu chứa 1 hoặc 2 quả cam hỏng.

$$\begin{aligned} P_1(20; 0,03) + P_2(20; 0,03) \\ = C_{20}^1 \cdot 0,03^1 \cdot (1 - 0,03)^{19} + C_{20}^2 \cdot 0,03^2 \cdot (1 - 0,03)^{18} = [0,4352] \end{aligned}$$

c) *Xác suất sọt cam xếp loại 3:*

Mẫu chứa từ 3 quả cam hỏng trở lên.

$$\begin{aligned} 1 - [P_0(20; 0,03) + P_1(20; 0,03) + P_2(20; 0,03)] \\ 1 - (0,5438 + 0,4352) = [0,021] \end{aligned}$$

**Bài 11/38:** Một bài thi trắc nghiệm (multiple – choice test) gồm 12 câu hỏi, mỗi câu hỏi cho 5 câu trả lời, trong đó chỉ có một câu đúng. Giả sử mỗi câu trả lời đúng được 4 điểm và mỗi câu trả lời sai bị trừ 1 điểm.

Một học sinh kém làm bài bằng cách chọn hú họa một câu trả lời. Tìm xác suất để:

- a) **Anh ta được 13 điểm;**
- b) **Anh ta được điểm âm.**

Giả sử học sinh đó làm đúng x câu và sai  $(12 - x)$  câu thì số điểm đạt được là:

$$4x - (12 - x) = 5x - 12$$

a) *Xác suất để học sinh được 13 điểm:*

Ta có:  $5x - 12 = 13 \Leftrightarrow x = 5$ . Học sinh chỉ làm đúng 5 câu và sai 7 câu. Chọn hú họa 12 câu tương đương với 12 lần thử độc lập, xác suất chọn đúng 5 câu là:

$$P_5(12; 0,2) = C_{12}^5 \cdot 0,2^5 \cdot 0,8^7 = [0,0532]$$

b) Xác suất để học sinh bị điểm âm:

Ta có:  $5x - 12 < 0 \Leftrightarrow 5x < 12 \Leftrightarrow x < 2,4$ . Vậy để bị điểm âm thì học sinh chỉ làm đúng nhiều nhất 2 câu.

Xác suất để học sinh làm đúng 0, 1, 2 câu lần lượt là:

$$P_0(12; 0,2) = C_{12}^0 \cdot 0,2^0 \cdot (1-0,2)^{12} = 0,0687$$

$$P_1(12; 0,2) = C_{12}^1 \cdot 0,2^1 \cdot (1-0,2)^{11} = 0,2062$$

$$P_2(12; 0,2) = C_{12}^2 \cdot 0,2^2 \cdot (1-0,2)^{10} = 0,2835$$

$$\Rightarrow \text{Xác suất để học sinh bị điểm âm: } 0,0687 + 0,2062 + 0,2835 = \boxed{0,5584}$$

**Bài 12/39: Gieo ba con xúc xắc cân đối một cách độc lập. Tính xác suất để:**

- a) **Tổng số nốt xuất hiện là 8 nếu biết rằng ít nhất có một con ra nốt 1;**
- b) **Có ít nhất một con ra nốt 6 nếu biết rằng số nốt trên 3 con là khác nhau.**

Tổng số kết quả có thể xảy ra:  $6^3 = 216$

a) Xác suất tổng số nốt xuất hiện là 8 biết rằng ít nhất có một con ra nốt 1:

A = “Tổng số nốt xuất hiện là 8”

B = “Có ít nhất một con ra nốt 1”

Do đó: AB = “Tổng số nốt xuất hiện là 8 trong đó có một con ra nốt 1”

Số kết quả thuận lợi cho biến cố AB:

(1, 2, 5) và 5 hoán vị khác nữa

(1, 3, 4) và 5 hoán vị khác nữa

(1, 1, 6) và 2 hoán vị khác nữa

$$\Rightarrow P(AB) = \frac{6+6+3}{216} = \frac{15}{216}$$

Biến cố B là biến cố đôi của biến cố: “Không có con nào ra nốt 1”

$$\Rightarrow P(B) = 1 - \left(\frac{5}{6}\right)^3 = \frac{91}{216}$$

$$\text{Vậy: } P(A | B) = \frac{P(AB)}{P(B)} = \frac{15}{216} : \frac{91}{216} = \boxed{\frac{15}{91}}$$

b) Xác suất có ít nhất 1 con ra nốt 6 nếu biết rằng số nốt trên 3 con là khác nhau:

C = “Ít nhất một con ra nốt 6”

D = “Số nốt trên 3 con là khác nhau”

Do đó: CD = “Số nốt trên 3 con là khác nhau trong đó có 1 con ra nốt 6”

Tính số kết quả thuận lợi cho biến cố CD:

- + Chọn vị trí cho nốt 6: có 3 cách
- + Chọn nốt xuất hiện thứ hai mà khác nốt 6: có 5 cách
- + Chọn nốt xuất hiện thứ ba mà khác hai nốt trên: có 4 cách.

Suy ra số kết quả thuận lợi cho biến cỗ CD:  $3 \cdot 5 \cdot 4 = 60$  (cách)

$$\Rightarrow P(CD) = \frac{60}{216}$$

Mà:  $P(D) = \frac{A_6^3}{216} = \frac{120}{216}$  (lấy 3 con khác nhau trong số 6 con, có tính đến thứ tự)

$$\Rightarrow P(C) = \frac{P(CD)}{P(D)} = \frac{60}{216} : \frac{120}{216} = \boxed{\frac{1}{2}}$$

**Bài 13/39:** Một gia đình có hai đứa con. Tìm xác suất để cả hai đều là con trai nếu biết rằng ít nhất trong hai đứa có một đứa là trai.

A = “Cả hai đứa là con trai”

B = “Ít nhất một trong hai đứa là con trai”

Ta có:

$$\begin{aligned} P(AB) &= P(A) \quad (\text{vì } A \subset B) \\ &= 0,5^2 = 0,25 \end{aligned}$$

$P(B) = 1 - 0,5^2 = 0,75$  (B là biến cỗ đối của biến cỗ: “cả hai đứa là con gái”)

Vậy xác suất để cả hai đứa là con trai nếu biết rằng ít nhất một trong hai đứa là con trai:

$$P(A | B) = \frac{P(AB)}{P(B)} = \frac{0,25}{0,75} = \boxed{\frac{1}{3}}$$

**Bài 14/39:** Một cặp trẻ sinh đôi có thể do cùng một trứng (sinh đôi thật) hay do hai trứng khác nhau sinh ra (sinh đôi giả). Các cặp sinh đôi thật luôn có cùng giới tính. Cặp sinh đôi giả thì giới tính của mỗi đứa độc lập với nhau và có xác suất 0,5 là con trai<sup>1</sup>.

Thống kê cho thấy 34% cặp sinh đôi đều là trai, 30% cặp sinh đôi đều là gái và 36% cặp sinh đôi có giới tính khác nhau.

a) Tìm tỷ lệ cặp sinh đôi thật;

b) Tìm tỷ lệ cặp sinh đôi thật trong tổng số cặp sinh đôi cùng giới tính.

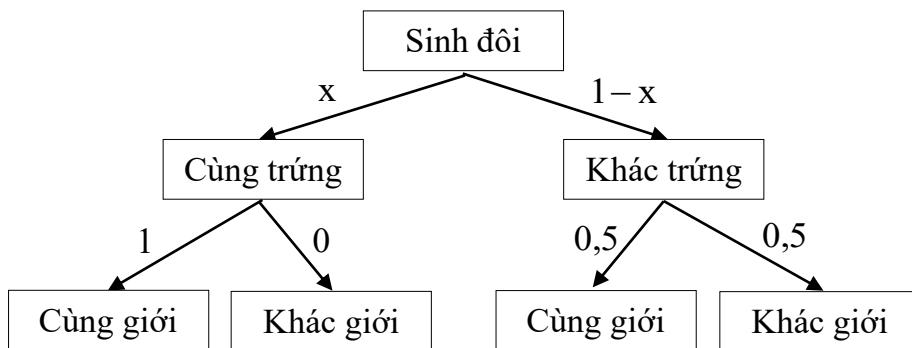
Gọi:

A = “Cặp sinh đôi thật” (cùng trứng)

B = “Cặp sinh đôi có cùng giới tính”

---

<sup>1</sup> Đối với bài dạng công thức xác suất đầy đủ thì nên vẽ sơ đồ cây để giải cho đơn giản.



a) Tìm tỷ lệ cặp sinh đôi thật:

Gọi  $x$  là tỷ lệ cặp sinh đôi thật thì  $(1 - x)$  là tỷ lệ cặp sinh đôi giả.

Theo công thức xác suất đầy đủ thì tỷ lệ các cặp sinh đôi khác giới là:

$$x \cdot 0 + (1 - x) \cdot 0.5 = 0.5(1 - x)$$

Mà theo giả thiết, có 36% cặp sinh đôi có giới tính khác nhau. Do đó:

$$0.5(1 - x) = 0.36$$

$$\Leftrightarrow 1 - x = 0.72 \Leftrightarrow x = 0.28$$

Vậy, tỷ lệ cặp sinh đôi thật:  $P(A) = 0,28$

b) Tìm tỷ lệ cặp sinh đôi thật trong tổng số cặp sinh đôi cùng giới tính:

Tỷ lệ cặp sinh đôi cùng giới tính:  $P(B) = 0,34 + 0,30 = 0,64$

$\Rightarrow$  Tỷ lệ cần tìm:

$$P(A | B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} = \frac{0,28}{0,64} = 0,4375 \quad (\text{vì } A \subset B)$$

**Bài 15/39:** Có hai chuồng thỏ. Chuồng thứ nhất có 5 con thỏ đen và 10 con thỏ trắng. Chuồng thứ hai có 3 thỏ trắng và 7 thỏ đen. Từ chuồng thứ hai ta bắt ngẫu nhiên một con thỏ cho vào chuồng thứ nhất và sau đó lại bắt ngẫu nhiên một con thỏ ở chuồng thứ nhất ra thì được một chú thỏ trắng. Tính xác suất để con thỏ trắng này là của chuồng thứ nhất.

*Hướng dẫn: Biết trước kết quả ở lần bắt thứ hai là một chú thỏ trắng. Để bài yêu cầu tìm xác suất để con thỏ trắng này có nguồn gốc của chuồng I (xác suất của một nguyên nhân nào đó dẫn đến kết quả đã biết). Áp dụng công thức Bayes.*

Gọi:

A = “Thỏ bắt ra lần thứ hai là của chuồng I”

B = “Thỏ bắt ra lần thứ hai là của chuồng II”

C = “Thỏ bắt lần thứ nhất từ chuồng II sang chuồng I là thỏ trắng”

H = “Thỏ bắt lần thứ hai là thỏ trắng”

Theo công thức xác suất đầy đủ:

$$P(H) = P(HA) + P(HB)$$

$$\text{Mà: } P(\text{HA}) = P(\text{HA} | C).P(C) + P(\text{HA} | \bar{C}).P(\bar{C})$$

(xác suất của biến cố HA còn phụ thuộc vào việc lần bắt thứ nhất từ chuồng II sang chuồng I là thỏ trắng hay đen). Do đó:

$$P(\text{HA}) = \frac{10}{16} \cdot \frac{3}{10} + \frac{10}{16} \cdot \frac{7}{10} = \frac{10}{16} = \frac{100}{160}$$

Tương tự:

$$P(\text{HB}) = P(\text{HB} | C).P(C) + P(\text{HB} | \bar{C}).P(\bar{C})$$

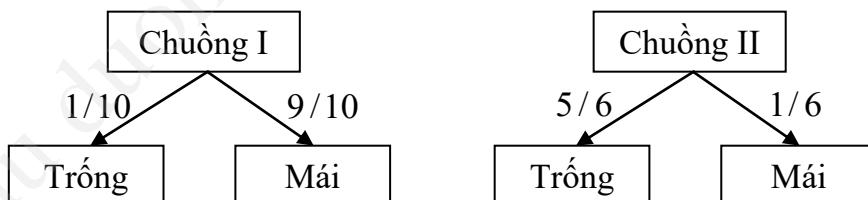
$$= \frac{1}{16} \cdot \frac{3}{10} + 0 \cdot \frac{7}{10} = \frac{3}{160}$$

$\Rightarrow$  Xác suất để con thỏ trắng bắt ở lần thứ hai là của chuồng I:

$$P(A | H) = \frac{P(\text{HA})}{P(H)} = \frac{P(\text{HA})}{P(\text{HA}) + P(\text{HB})} = \frac{\frac{100}{160}}{\frac{100}{160} + \frac{3}{160}} = \boxed{\frac{100}{103}}$$

**Bài 16/39:** Một chuồng gà có 9 con mái và 1 con trống. Chuồng gà kia có 1 con mái và 5 con trống. Từ mỗi chuồng ta bắt ngẫu nhiên ra một con đem bán. Các con gà còn lại được dồn vào một chuồng thứ ba. Nếu ta lại bắt ngẫu nhiên một con gà nữa từ chuồng này ra thì xác suất bắt được con gà trống là bao nhiêu?

*Hướng dẫn:* Xác suất bắt được gà trống ở chuồng III còn phụ thuộc vào hành động bắt trước đó ở chuồng I và chuồng II. Khi bắt ở hai chuồng I và II thì có các khả năng xảy ra: bắt được hai con trống, bắt được hai con mái, bắt được 1 trống 1 mái.



Gọi:

$A_1$  = “Bắt được con trống ở chuồng I”

$B_1$  = “Bắt được con mái ở chuồng I”

$A_2$  = “Bắt được con trống ở chuồng II”

$B_2$  = “Bắt được con mái ở chuồng II”

$H$  = “Bắt được con trống ở chuồng III”

Xác suất bắt được 2 con trống từ hai chuồng I và II:

$$P(A_1 A_2) = P(A_1).P(A_2) \quad (\text{việc bắt gà ở mỗi chuồng là độc lập với nhau})$$

$$= \frac{1}{10} \cdot \frac{5}{6} = \frac{5}{60}$$

Xác suất bắt được 2 con mồi từ hai chuồng I và II:

$$P(B_1B_2) = P(B_1) \cdot P(B_2) = \frac{9}{10} \cdot \frac{1}{6} = \frac{9}{60}$$

Xác suất bắt được 1 trống 1 mồi từ hai chuồng I và II:

$$P(A_1B_2) + P(A_2B_1) = 1 - P(A_1A_2) - P(B_1B_2) = 1 - \frac{5}{60} - \frac{9}{60} = \frac{46}{60}$$

$\Rightarrow$  Xác suất để bắt được con gà trống từ chuồng III:

$$\begin{aligned} P(H) &= P(HA_1A_2) + P(HB_1B_2) + P(HA_1B_2) + P(HA_2B_1) \\ &= \frac{4}{14} \cdot \frac{5}{60} + \frac{6}{14} \cdot \frac{9}{60} + \frac{5}{14} \cdot \frac{46}{60} = \frac{38}{105} = \boxed{0,3619} \end{aligned}$$

**Bài 17/39:** Một chiếc máy bay có thể xuất hiện ở vị trí A với xác suất  $2/3$  và ở vị trí B với xác suất  $1/3$ . Có ba phương án bố trí 4 khẩu pháo bắn máy bay như sau:

Phương án I: 3 khẩu đặt tại A, 1 khẩu đặt tại B.

Phương án II: 2 khẩu đặt tại A, 2 khẩu đặt tại B.

Phương án III: 1 khẩu đặt tại A và 3 khẩu đặt tại B.

Biết rằng xác suất bắn trúng máy bay của mỗi khẩu pháo là  $0,7$  và các khẩu pháo hoạt động độc lập với nhau, hãy chọn phương án tốt nhất.

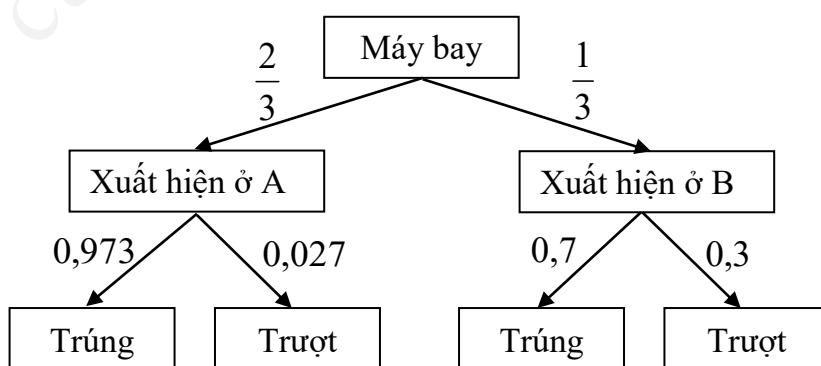
*Hướng dẫn: Phương án tốt nhất là phương án cho xác suất bắn trúng máy bay cao nhất. Ứng với mỗi phương án, áp dụng công thức xác suất đầy đủ để tính xác suất bắn trúng máy bay.*

\* Phương án I: 3 khẩu đặt tại A và 1 khẩu đặt tại B

Nếu có 3 khẩu đặt tại A thì để máy bay rơi cần ít nhất một khẩu bắn trúng. Xác suất để ít nhất một khẩu tại A bắn trúng máy bay:

$$1 - 0,3^3 = 0,973$$

(tính theo biến cố đối của biến cố: không có khẩu nào bắn trúng)



$\Rightarrow$  Xác suất để máy bay rơi trong phương án I:

$$P_1 = \frac{2}{3} \cdot 0,973 + \frac{1}{3} \cdot 0,7 = 0,882 \quad (1)$$

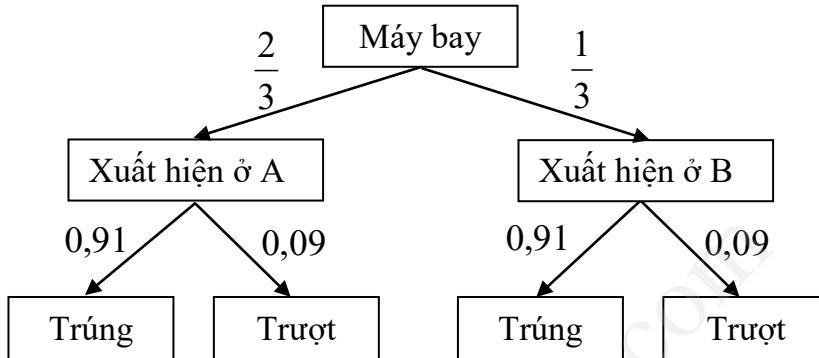
\* Phương án II: 2 khẩu đạn tại A và 2 khẩu đạn tại B

Nếu có 2 khẩu đạn tại A thì để máy bay rơi cần ít nhất một khẩu bắn trúng. Xác suất để ít nhất một khẩu tại A bắn trúng máy bay:

$$1 - 0,3^2 = 0,91$$

Tương tự, xác suất để ít nhất một khẩu tại B bắn trúng máy bay:

$$1 - 0,3^2 = 0,91$$



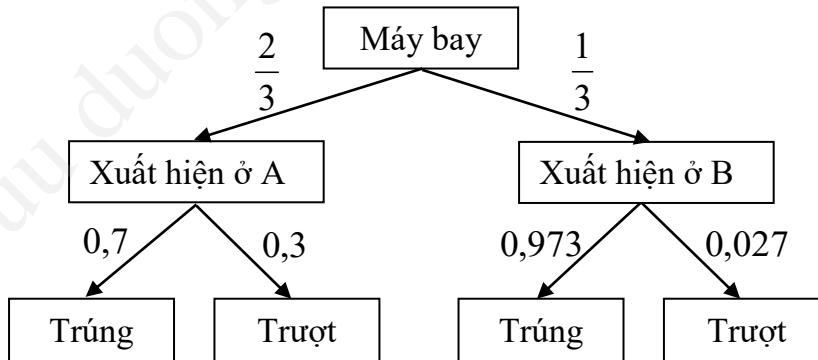
⇒ Xác suất để máy bay rơi trong phương án II:

$$P_2 = \frac{2}{3} \cdot 0,91 + \frac{1}{3} \cdot 0,91 = 0,91 \quad (2)$$

\* Phương án III: 1 khẩu đạn tại A và 3 khẩu đạn tại B

Nếu có 3 khẩu đạn tại B thì để máy bay rơi cần ít nhất một khẩu bắn trúng. Xác suất để ít nhất một khẩu tại B bắn trúng máy bay:

$$1 - 0,3^3 = 0,973$$



⇒ Xác suất để máy bay rơi trong phương án III:

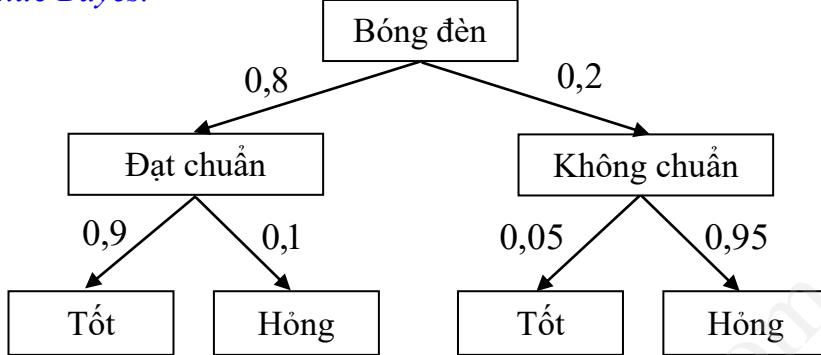
$$P_3 = \frac{2}{3} \cdot 0,7 + \frac{1}{3} \cdot 0,973 = 0,791 \quad (3)$$

Từ (1), (2) và (3) suy ra: phương án II có xác suất bắn trúng máy bay cao nhất. Chọn **phương án II** để đạt hiệu quả tốt nhất.

**Bài 18/40:** Một nhà máy sản xuất bóng đèn có tỷ lệ bóng đèn đạt tiêu chuẩn là 80%. Trước khi xuất ra thị trường, mỗi bóng đèn đều được qua kiểm tra chất lượng. Vì sự kiểm tra không thể tuyệt đối hoàn hảo nên một bóng đèn tốt có xác

số **0,9** được công nhận là **tốt** và một bóng đèn hỏng có xác suất **0,95** bị loại bỏ. Hãy tính tỷ lệ bóng đèn đạt tiêu chuẩn sau khi qua khâu kiểm tra chất lượng.

Hướng dẫn: Sau khi qua khâu kiểm tra chất lượng ta được một tỷ lệ bóng đèn tốt. Trong số những bóng đèn tốt này bao gồm cả những bóng đèn đạt chuẩn và không đạt chuẩn. Ta tính xác suất bóng đèn đạt chuẩn trong số những bóng đèn tốt. Dấu hiệu để áp dụng công thức Bayes.



Gọi:

A = “Bóng đèn đạt chuẩn”

H = “Bóng đèn được công nhận là tốt”

⇒ Tỷ lệ bóng đèn đạt tiêu chuẩn sau khi đã qua khâu kiểm tra chất lượng:

$$P(A | H) = \frac{P(HA)}{P(HA) + P(H\bar{A})} = \frac{0,8 \cdot 0,9}{0,8 \cdot 0,9 + 0,2 \cdot 0,05} = \frac{0,72}{0,73} = \boxed{0,9863}$$

**Bài 19/40:** Có 4 nhóm xạ thủ tập bắn. Nhóm thứ nhất có 5 người, nhóm thứ hai có 7 người, nhóm thứ ba có 4 người và nhóm thứ tư có 2 người. Xác suất bắn trúng đích của mỗi người trong nhóm thứ nhất, nhóm thứ hai, nhóm thứ ba và nhóm thứ tư theo thứ tự là 0,8; 0,7; 0,6 và 0,5. Chọn ngẫu nhiên một xạ thủ và xạ thủ này bắn trượt. Hãy xác định xem xạ thủ này có khả năng ở trong nhóm nào nhất.

Hướng dẫn: Xạ thủ bắn trượt có thể thuộc một trong bốn nhóm. Áp dụng công thức Bayes để kiểm tra xem xác suất xạ thủ bắn trượt này thuộc từng nhóm là bao nhiêu. Từ đó so sánh các kết quả với nhau và đưa ra kết luận.

Gọi:

A = “Xạ thủ thuộc nhóm 1”

B = “Xạ thủ thuộc nhóm 2”

C = “Xạ thủ thuộc nhóm 3”

D = “Xạ thủ thuộc nhóm 4”

H = “Xạ thủ bắn trượt”

Theo bài ra ta có:

$$P(A) = \frac{5}{18}; \quad P(B) = \frac{7}{18}; \quad P(C) = \frac{4}{18}; \quad P(D) = \frac{2}{18}$$

$$P(H | A) = 0,2; \quad P(H | B) = 0,3; \quad P(H | C) = 0,4; \quad P(H | D) = 0,5$$

+ Xác suất để xạ thủ bắn trượt này thuộc nhóm thứ nhất:

$$\begin{aligned} P(A | H) &= \frac{P(HA)}{P(HA) + P(HB) + P(HC) + P(HD)} \\ P(A | H) &= \frac{P(H | A).P(A)}{P(H | A).P(A) + P(H | B).P(B) + P(H | C).P(C) + P(H | D).P(D)} \\ P(A | H) &= \frac{0,2 \frac{5}{18}}{0,2 \frac{5}{18} + 0,3 \frac{7}{18} + 0,4 \frac{4}{18} + 0,5 \frac{2}{18}} = \frac{\frac{1}{18}}{\frac{19}{60}} = \frac{10}{57} \quad (1) \end{aligned}$$

+ Xác suất để xạ thủ bắn trượt này thuộc nhóm thứ hai:

$$\begin{aligned} P(B | H) &= \frac{P(HB)}{P(HA) + P(HB) + P(HC) + P(HD)} \\ P(B | H) &= \frac{P(H | B).P(B)}{P(H | A).P(A) + P(H | B).P(B) + P(H | C).P(C) + P(H | D).P(D)} \\ P(B | H) &= \frac{0,3 \frac{7}{18}}{0,2 \frac{5}{18} + 0,3 \frac{7}{18} + 0,4 \frac{4}{18} + 0,5 \frac{2}{18}} = \frac{\frac{7}{60}}{\frac{19}{60}} = \frac{21}{57} \quad (2) \end{aligned}$$

+ Xác suất để xạ thủ bắn trượt này thuộc nhóm thứ ba:

$$\begin{aligned} P(C | H) &= \frac{P(HC)}{P(HA) + P(HB) + P(HC) + P(HD)} \\ P(C | H) &= \frac{P(H | C).P(C)}{P(H | A).P(A) + P(H | B).P(B) + P(H | C).P(C) + P(H | D).P(D)} \\ P(C | H) &= \frac{0,4 \frac{4}{18}}{0,2 \frac{5}{18} + 0,3 \frac{7}{18} + 0,4 \frac{4}{18} + 0,5 \frac{2}{18}} = \frac{\frac{4}{45}}{\frac{19}{60}} = \frac{16}{57} \quad (3) \end{aligned}$$

+ Xác suất để xạ thủ bắn trượt này thuộc nhóm thứ tư:

$$\begin{aligned} P(D | H) &= \frac{P(HD)}{P(HA) + P(HB) + P(HC) + P(HD)} \\ P(D | H) &= \frac{P(H | D).P(D)}{P(H | A).P(A) + P(H | B).P(B) + P(H | C).P(C) + P(H | D).P(D)} \\ P(D | H) &= \frac{0,5 \frac{2}{18}}{0,2 \frac{5}{18} + 0,3 \frac{7}{18} + 0,4 \frac{4}{18} + 0,5 \frac{2}{18}} = \frac{\frac{1}{18}}{\frac{19}{60}} = \frac{10}{57} \quad (4) \end{aligned}$$

Từ (1), (2), (3) và (4) suy ra: xạ thủ đã bắn trượt này có khả năng thuộc nhóm thứ hai nhất.

**Bài 20/40:** Trong số bệnh nhân ở một bệnh viện có 50% điều trị bệnh A, 30% điều trị bệnh B và 20% điều trị bệnh C. Xác suất để chữa khỏi các bệnh A, B và C trong bệnh viện này tương ứng là 0,7; 0,8 và 0,9. Hãy tính tỷ lệ bệnh nhân được chữa khỏi bệnh A trong tổng số bệnh nhân đã được chữa khỏi bệnh.

Hướng dẫn: Tính tỷ lệ bệnh nhân được chữa khỏi bệnh bằng công thức xác suất đầy đủ. Áp dụng công thức Bayes để tính tỷ lệ bệnh nhân được chữa khỏi bệnh A trong tổng số bệnh nhân đã được chữa khỏi bệnh.

Gọi:

A = “Bệnh nhân điều trị bệnh A”

B = “Bệnh nhân điều trị bệnh B”

C = “Bệnh nhân điều trị bệnh C”

H = “Bệnh nhân được chữa khỏi bệnh”

Theo bài ra ta có:

$$P(A) = 0,5; \quad P(B) = 0,3; \quad P(C) = 0,2$$

$$P(H | A) = 0,7; \quad P(H | B) = 0,8; \quad P(H | C) = 0,9$$

⇒ Tỷ lệ bệnh nhân được chữa khỏi bệnh A trong tổng số bệnh nhân được chữa khỏi bệnh (áp dụng công thức Bayes):

$$\begin{aligned} P(A | H) &= \frac{P(HA)}{P(HA) + P(HB) + P(HC)} \\ &= \frac{P(H | A) \cdot P(A)}{P(H | A) \cdot P(A) + P(H | B) \cdot P(B) + P(H | C) \cdot P(C)} \\ &= \frac{0,7 \cdot 0,5}{0,7 \cdot 0,5 + 0,8 \cdot 0,3 + 0,9 \cdot 0,2} = \frac{0,35}{0,77} = \frac{5}{11} = \boxed{0,4545} \end{aligned}$$

## 1.2. Nhận xét bài tập chương 1

Chương 1 thường ra vào dạng bài toán sử dụng công thức xác suất đầy đủ và công thức Bayes. Còn công thức Bernoulli và xác suất có điều kiện cũng được lồng luôn vào hai dạng trên mà ít khi tách riêng thành một bài độc lập.

Cách giải:

a) Dạng bài toán sử dụng công thức xác suất đầy đủ:

+ Trước khi bài toán yêu cầu tính xác suất  $P(H)$  thì có nhiều trường hợp xảy ra. Ví dụ: trước khi tính xác suất lấy phái phẩm thì thấy có nhiều phân xưởng cùng sản xuất.

+ Ứng với mỗi trường hợp ở trên ta đặt làm một biến cõi  $B_i$ . Tất cả các biến cõi  $B_i$  ( $i$  chạy từ 1 đến  $n$ ) hợp thành một hệ đầy đủ (bao quát mọi trường hợp có thể xảy ra).

+ Tính xác suất của  $H$  ứng với từng điều kiện  $B_i$  và cộng chúng lại với nhau:  $P(H) = P(H|B_1). P(B_1) + P(H|B_2). P(B_2) + \dots + P(H|B_n). P(B_n)$

Với những bài toán đơn giản, các xác suất  $P(B_i)$  và  $P(H|B_i)$  đã được cho sẵn, chỉ cần thay vào công thức và suy ra kết quả.

Với những bài toán phức tạp, các xác suất  $P(B_i)$  có thể được tính thông qua giải tích tinh hợp (hoán vị, chỉnh hợp, tinh hợp) hoặc các xác suất  $P(H|B_i)$  có thể được tính thông qua công thức Bernoulli (ví dụ Bài 1b, đề thi cuối kỳ II năm học 2013 – 2014, trang 134).

b) *Dạng bài toán sử dụng công thức Bayes:*

+ Bài toán cho biết kết quả đã xảy ra và yêu cầu tính xác suất để kết quả này là do một hoặc một số nguyên nhân nào đó. Trong công thức Bayes đã bao hàm luôn công thức xác suất đầy đủ.

+ Ta có kết quả  $H$  đã xảy ra, mà để xảy ra  $H$  có  $n$  trường hợp:  $B_1, B_2, \dots, B_n$ . Tính xác suất để kết quả này do nguyên nhân  $B_k$  nào đó bằng cách lấy  $P(H|B_k)$  chia cho tổng các  $P(H|B_i)$  với  $i$  chạy từ 1 đến  $n$ .

+ Tương tự dạng áp dụng công thức xác suất đầy đủ, các xác suất  $P(B_i)$  và  $P(H|B_i)$  đã được cho sẵn hoặc phải tính qua nhiều bước.

*Đề thi giữa kỳ và cuối kỳ thường ra vào dạng sử dụng công thức Bayes, đây là dạng hay gây nhầm lẫn và đôi khi khó hiểu đè.*

## CHƯƠNG 2: ĐẠI LƯỢNG NGẪU NHIÊN RỜI RẠC

### A. LÝ THUYẾT

- *Đại lượng ngẫu nhiên* (ĐLNN) là đại lượng mà giá trị của nó là ngẫu nhiên, không biết trước được. Ký hiệu: X, Y, Z, ...
- *DLNN rời rạc* là đại lượng ngẫu nhiên chỉ nhận một số hữu hạn các giá trị hoặc một số vô hạn đếm được các giá trị (vô hạn đếm được như: tập số tự nhiên N).

Tập hợp các giá trị có thể có của ĐLNN X ký hiệu là:  $X(\Omega)$

#### 2.1. Phân bố xác suất và hàm phân bố

a) *Phân bố (phân phối) xác suất* của ĐLNN rời rạc X là một bảng có dạng:

X	$x_1$	$x_2$	$x_3$	.....	$x_n$
$P(X = x_i)$	$p_1$	$p_2$	$p_3$	.....	$p_n$

Trong đó:  $\left\{ \begin{array}{l} x_1, x_2, \dots, x_n \text{ là các giá trị có thể có của } X \\ p_1, p_2, \dots, p_n \text{ là các xác suất tương ứng} \\ \sum_{i=1}^n p_i = 1 \end{array} \right.$

b) *Hàm phân bố<sup>1</sup> (phân phối)* của ĐLNN rời rạc X là hàm  $F(x)$  xác định  $\forall x$  sao cho:

$$F(x) = P(X < x)$$

$$\Leftrightarrow F(x) = \begin{cases} 0 & \text{nếu } x \leq x_1 \\ p_1 + \dots + p_k & \text{nếu } x_k < x \leq x_{k+1} \quad (\text{với } 1 \leq k \leq n-1) \\ 1 & \text{nếu } x > x_n \end{cases}$$

#### 2.2. Một số đặc trưng của đại lượng ngẫu nhiên rời rạc

a) *Kỳ vọng:*

$$EX = \sum_{i=1}^n x_i p_i \quad \text{với } p_i = P(X = x_i)$$

b) *Phương sai:* là giá trị trung bình của các bình phương độ lệch của các điểm giá trị so với giá trị kỳ vọng.

$$DX = E(X - EX)^2 = EX^2 - (EX)^2 = \sum_{i=1}^n x_i^2 p_i - (EX)^2$$

c) *Độ lệch tiêu chuẩn (độ lệch chuẩn):*

$$\sigma_X = \sqrt{DX}$$

<sup>1</sup> Theo một số quan điểm khác thì hàm phân bố được định nghĩa bởi:  $F(x) = P(X \leq x)$

d) Mode:

Mode của X là giá trị  $x_0$  sao cho  $P(X = x_0)$  là lớn nhất.

### 2.3. Phân bố đồng thời và hệ số tương quan

a) Phân bố xác suất đồng thời:

Cho hai ĐLNN rời rạc X và Y với:

$$X(\Omega) = \{x_1, x_2, \dots, x_m\}$$

$$Y(\Omega) = \{y_1, y_2, \dots, y_n\}$$

Ký hiệu:  $p_{ij} = P(X = x_i; Y = y_j)$

$\Rightarrow$  Bảng phân bố xác suất đồng thời của X và Y có dạng:

X \ Y	$y_1$	$y_2$	...	$y_n$
$x_1$	$p_{11}$	$p_{12}$	...	$p_{1n}$
$x_2$	$p_{21}$	$p_{22}$	...	$p_{2n}$
...	...	...	...	...
$x_m$	$p_{m1}$	$p_{m2}$	...	$p_{mn}$

Chú ý:  $\sum_{i=1}^m \sum_{j=1}^n p_{ij} = 1$

Từ bảng phân bố xác suất đồng thời, suy ra các bảng phân bố xác suất thành phần của X và Y:

$$P(X = x_i) = \sum_{j=1}^n p_{ij} \quad \text{Ví dụ: } P(X = x_1) = p_{11} + p_{12} + \dots + p_{1n}$$

(cộng từng hàng ta được bảng phân bố xác suất của X)

$$P(Y = y_j) = \sum_{i=1}^m p_{ij} \quad \text{Ví dụ: } P(Y = y_1) = p_{11} + p_{21} + \dots + p_{m1}$$

(cộng từng cột ta được bảng phân bố xác suất của Y)

b) Đại lượng ngẫu nhiên độc lập:

Hai ĐLNN rời rạc X và Y được gọi là độc lập nếu việc biết một thông tin về giá trị của X (hoặc Y) không ảnh hưởng gì đến phân bố xác suất của Y (hoặc X). Hay:

$$P(Y = y_j | X = x_i) = P(Y = y_j)$$

$$P(X = x_i | Y = y_j) = P(X = x_i)$$

$$\Leftrightarrow P(X = x_i; Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

Do đó, X và Y độc lập  $\Leftrightarrow P_{ij} = P_i P_j \quad \forall i, j$

c) Covarian và hệ số tương quan:

+ Covarian:

$$\text{cov}(X, Y) = \sum_{i=1}^m \sum_{j=1}^n x_i y_j p_{ij} - EX \cdot EY$$

+ Hệ số tương quan:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Tính chất:  $-1 \leq \rho(X, Y) \leq 1$

Nếu X và Y độc lập thì  $\text{cov}(X, Y) = 0 \Rightarrow \rho(X, Y) = 0$

Nếu  $\rho(X, Y) = 0$  thì **chưa thể** suy ra X và Y độc lập.

Nếu  $\rho(X, Y) \neq 0$  thì suy ra X và Y phụ thuộc.

## 2.4. Hàm của đại lượng ngẫu nhiên rời rạc

Cho X và Y là 2 ĐLNN rời rạc thì ĐLNN Z xác định bởi  $Z = f(X, Y)$  cũng là một ĐLNN rời rạc.

a) Phân bố xác suất của Z:

Z	$Z_1$	$Z_2$	$Z_3$	.....	$Z_n$
$P(Z = z_i)$	$p_1$	$p_2$	$p_3$	.....	$p_n$

Trong đó:  $\begin{cases} z_i = f(x, y) \\ p_i = \sum P(X = x; Y = y) \text{ sao cho } z_i = f(x, y) \end{cases}$

b) Kỳ vọng của Z:

$$EZ = \sum z_i p_i = \sum_{i=1}^m \sum_{j=1}^n f(x_i, y_j) p_{ij}$$

Tính chất:

$$E(a) = a \quad (\text{với } a \text{ là hằng số})$$

$$E(aX) = a \cdot EX$$

$$E(X \pm Y) = EX \pm EY$$

$$E(XY) = EX \cdot EY \quad (\text{nếu } X \text{ và } Y \text{ độc lập})$$

c) Phương sai của Z:

$$DZ = \sum (z_i - EZ)^2 p_i = EZ^2 - (EZ)^2$$

Tính chất:

$$D(a) = 0 \quad (\text{với } a \text{ là hằng số})$$

$$D(aX) = a^2 DX$$

$$D(X \pm Y) = DX + DY \quad (\text{nếu } X \text{ và } Y \text{ độc lập})$$

## 2.5. Phân bố nhị thức

ĐLNN rời rạc X với  $X(\Omega) = \{0, 1, 2, \dots, n\}$  thỏa mãn:

$$P(X = k) = C_n^k p^k q^{n-k} \quad (\text{với } q = 1 - p)$$

được gọi là có phân bố nhị thức với tham số n, p.

Ký hiệu:  $X \sim B(n, p)$

Tính chất:

$$EX = np$$

$$DX = npq$$

$$\text{Mod } X = [(n + 1) p] \quad (\text{lấy phần nguyên})$$

## 2.6. Phân bố Poisson

ĐLNN rời rạc X với  $X(\Omega) = \{0, 1, 2, \dots, n\}$  thỏa mãn:

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

được gọi là có phân bố Poisson với tham số  $\lambda (\lambda > 0)$ .

Ký hiệu:  $X \sim \text{Poisson}(\lambda)$

Tính chất:

$$EX = DX = \lambda$$

$$\text{Mod } X = [\lambda]$$

$$\begin{aligned} \text{Nếu } & \begin{cases} X \sim \text{Poisson}(\lambda) \\ Y \sim \text{Poisson}(\mu) \\ X \text{ và } Y \text{ độc lập} \end{cases} & \text{thì: } Z = X + Y \sim \text{Poisson}(\lambda + \mu) \end{aligned}$$

Chú ý: Cho một hiện tượng xuất hiện trong khoảng thời gian  $\Delta t$  với cường độ c (c là giá trị trung bình của số lần xuất hiện trong khoảng thời gian  $\Delta t$ )

Nếu ĐLNN X chỉ số lần xuất hiện hiện tượng trong khoảng thời gian ( $t_1, t_2$ ) thì:

$$X \sim \text{Poisson}\left(c \cdot \frac{t_2 - t_1}{\Delta t}\right)$$

Ví dụ: Trung bình có 15 xe tải qua cầu trong khoảng thời gian 6 giờ. Gọi X là số xe tải đi qua cầu trong khoảng thời gian từ 12h đến 14h30 thì:

$$X \sim \text{Poisson}\left(15 \cdot \frac{14,5 - 12}{6}\right) \Leftrightarrow X \sim \text{Poisson}(6,25)$$

## B. BÀI TẬP

### 2.1. Bài tập trong giáo trình 1 (G<sub>1</sub>)

(Mở đầu về lý thuyết xác suất và các ứng dụng, Đặng Hùng Thắng, trang 71)

**Bài 1/71:** Một nhóm có 10 người gồm 6 nam và 4 nữ. Chọn ngẫu nhiên ra 3 người. Gọi X là số nữ trong nhóm. Hãy tìm phân bố xác suất của X và tính EX, DX, Mod X.

X = “Số nữ trong nhóm được chọn”

Các giá trị có thể có của X:  $X(\Omega) = \{0, 1, 2, 3\}$

Ta có:

$$P(X=0) = \frac{C_4^0 \cdot C_6^3}{C_{10}^3} = \frac{5}{30} ; \quad P(X=1) = \frac{C_4^1 \cdot C_6^2}{C_{10}^3} = \frac{15}{30}$$

$$P(X=2) = \frac{C_4^2 \cdot C_6^1}{C_{10}^3} = \frac{9}{30} ; \quad P(X=3) = \frac{C_4^3 \cdot C_6^0}{C_{10}^3} = \frac{1}{30}$$

+ Bảng phân bố xác suất của X:

X	0	1	2	3
P(X=x <sub>i</sub> )	$\frac{5}{30}$	$\frac{15}{30}$	$\frac{9}{30}$	$\frac{1}{30}$

+ Kỷ vọng của X:

$$EX = \sum x_i p_i \quad (\text{với } p_i = P(X=x_i))$$

$$EX = 0 \cdot \frac{5}{30} + 1 \cdot \frac{15}{30} + 2 \cdot \frac{9}{30} + 3 \cdot \frac{1}{30} = \boxed{1,2}$$

+ Phuông sai của X:

$$DX = EX^2 - (EX)^2$$

$$\text{Tính: } EX^2 = \sum x_i^2 p_i = 0^2 \cdot \frac{5}{30} + 1^2 \cdot \frac{15}{30} + 2^2 \cdot \frac{9}{30} + 3^2 \cdot \frac{1}{30} = 2$$

$$\Rightarrow DX = 2 - 1,2^2 = \boxed{0,56}$$

+ Mode của X:

$\boxed{\text{Mod } X = 1}$  vì  $P(X=1)$  là lớn nhất.

**Bài 2/71:** Cho ĐLNN X có phân bố xác suất như sau:

X	1	3	5	7	9
P	0,1	0,2	0,3	0,3	0,1

Tìm phân bố xác suất của Y = min{X, 4}.

Các giá trị có thể có của Y:

X	1	3	5	7	9
$Y = \min\{X, 4\}$	1	3	4	4	4

Do đó:  $Y(\Omega) = \{1, 3, 4\}$

$$P(Y=1) = P(X=1) = 0,1$$

$$P(Y=3) = P(X=3) = 0,2$$

$$P(Y=4) = P(X \geq 5) = 0,3 + 0,3 + 0,1 = 0,7$$

$\Rightarrow$  Bảng phân bố xác suất của Y:

Y	1	3	4
P	0,1	0,2	0,7

**Bài 3/71:** Một túi chứa 10 thẻ đỏ và 6 thẻ xanh. Chọn ngẫu nhiên ra 3 tấm thẻ (không hoàn lại).

a) Gọi X là số thẻ đỏ. Tìm phân bố xác suất của X, EX và DX.

b) Giả sử rút mỗi thẻ đỏ được 5 điểm và rút mỗi thẻ xanh được 8 điểm. Gọi Y là số điểm tổng cộng trên ba thẻ rút ra. Tìm phân bố xác suất của Y, EY, DY.

a) Tìm bảng phân bố của X, EX, DX:

Các giá trị có thể có của X:  $X(\Omega) = \{0, 1, 2, 3\}$

Tính các xác suất tương ứng:

$$P(X=0) = \frac{C_{10}^0 \cdot C_6^3}{C_{16}^3} = \frac{2}{56} ; \quad P(X=1) = \frac{C_{10}^1 \cdot C_6^2}{C_{16}^3} = \frac{15}{56}$$

$$P(X=2) = \frac{C_{10}^2 \cdot C_6^1}{C_{16}^3} = \frac{27}{56} ; \quad P(X=3) = \frac{C_{10}^3 \cdot C_6^0}{C_{16}^3} = \frac{12}{56}$$

+ Bảng phân bố xác suất của X:

X	0	1	2	3
$P(X=x_i)$	$\frac{2}{56}$	$\frac{15}{56}$	$\frac{27}{56}$	$\frac{12}{56}$

+ Kỳ vọng của X:

$$EX = \sum_{i=1}^4 x_i p_i = 0 \cdot \frac{2}{56} + 1 \cdot \frac{15}{56} + 2 \cdot \frac{27}{56} + 3 \cdot \frac{12}{56} = \frac{15}{8} = [1,875]$$

+ Phuơng sai của X:

$$DX = \sum_{i=1}^4 x_i^2 p_i - (EX)^2 = 0^2 \cdot \frac{2}{56} + 1^2 \cdot \frac{15}{56} + 2^2 \cdot \frac{27}{56} + 3^2 \cdot \frac{12}{56} - 1,875^2$$

$$DX = \frac{231}{56} - 1,875^2 = \frac{39}{64} = [0,6094]$$

b) Tìm bảng phân bố của  $Y$ ,  $EY$ ,  $DY$ :

Rút mỗi thẻ đỏ được 5 điểm và mỗi thẻ xanh được 8 điểm nên ta có quan hệ giữa  $X$  và  $Y$  như sau:

$$Y = 5X + 8(3 - X) \Leftrightarrow Y = 24 - 3X$$

Các giá trị có thể có của  $Y$ :

X	0	1	2	3
Y	24	21	18	15

+ Bảng phân bố xác suất của  $Y$ :

Y	15	18	21	24
$P(Y=y_i)$	$\frac{12}{56}$	$\frac{27}{56}$	$\frac{15}{56}$	$\frac{2}{56}$

+ Kỳ vọng của  $Y$ :

$$EY = \sum_{i=1}^4 y_i p_i = 15 \cdot \frac{12}{56} + 18 \cdot \frac{27}{56} + 21 \cdot \frac{15}{56} + 24 \cdot \frac{2}{56} = \frac{147}{8} = [18,375]$$

+ Phương sai của  $Y$ :

$$\begin{aligned} DY &= \sum_{i=1}^4 y_i^2 p_i - (EY)^2 = 15^2 \cdot \frac{12}{56} + 18^2 \cdot \frac{27}{56} + 21^2 \cdot \frac{15}{56} + 24^2 \cdot \frac{2}{56} - 18,375^2 \\ &= \frac{19215}{56} - 18,375^2 = \frac{351}{64} = [5,4844] \end{aligned}$$

**Bài 4/71: Hai xạ thủ A và B tập bắn, mỗi người bắn hai phát. Xác suất bắn trúng đích của A trong mỗi phát là 0,4 còn của B là 0,5.**

a) Gọi  $X$  là số phát trúng của A trừ đi số phát trúng của B. Tìm phân bố xác suất của  $X$ .

b) Tìm phân bố xác suất của  $Y = |X|$

a) Tìm phân bố xác suất của  $X$ :

Giả sử A và B bắn độc lập với nhau. Hai người có thể có số lần trúng là 0, 1, 2. Gọi  $Y$  là số phát trúng của A,  $Z$  là số phát trúng của B thì:  $X = Y - Z$ . Các kết quả có thể có của  $X$  là:

$$X(\Omega) = \{-2, -1, 0, 1, 2\}$$

+ Tính các xác suất tương ứng:

$$\begin{aligned} P(X = -2) &= P(Y = 0; Z = 2) = P(Y = 0) \cdot P(Z = 2) \\ &= \left(C_2^0 0,4^0 \cdot 0,6^2\right) \left(C_2^2 0,5^2 \cdot 0,5^0\right) = 0,09 \end{aligned}$$

$$\begin{aligned}
 P(X = -1) &= P(Y = 0; Z = 1) + P(Y = 1; Z = 2) \\
 &= C_2^0 0,4^0 \cdot 0,6^2 \cdot C_2^1 0,5^1 \cdot 0,5^1 + C_2^1 0,4^1 \cdot 0,6^1 \cdot C_2^2 0,5^2 \cdot 0,5^0 = 0,3 \\
 P(X = 0) &= P(Y = 0; Z = 0) + P(Y = 1; Z = 1) + P(Y = 2; Z = 2) \\
 &= C_2^0 0,4^0 \cdot 0,6^2 \cdot C_2^0 0,5^0 \cdot 0,5^2 + C_2^1 0,4^1 \cdot 0,6^1 \cdot C_2^1 0,5^1 \cdot 0,5^1 \\
 &\quad + C_2^2 0,4^2 \cdot 0,6^0 \cdot C_2^2 0,5^2 \cdot 0,5^0 = 0,37 \\
 P(X = 1) &= P(Y = 1; Z = 0) + P(Y = 2; Z = 1) \\
 &= C_2^1 0,4^1 \cdot 0,6^1 \cdot C_2^0 0,5^0 \cdot 0,5^2 + C_2^2 0,4^2 \cdot 0,6^0 \cdot C_2^1 0,5^1 \cdot 0,5^1 = 0,2 \\
 P(X = 2) &= P(Y = 2; Z = 0) \\
 &= C_2^2 0,4^2 \cdot 0,6^0 \cdot C_2^0 0,5^0 \cdot 0,5^2 = 0,04
 \end{aligned}$$

$\Rightarrow$  Bảng phân bố xác suất của X:

X	-2	-1	0	1	2
P(X = x <sub>i</sub> )	0,09	0,3	0,37	0,2	0,04

b) Tìm phân bố xác suất của Y:

Các giá trị có thể có của Y:

X	-2	-1	0	1	2
Y =  X	2	1	0	1	2

Do đó:  $Y(\Omega) = \{0, 1, 2\}$

Ta có:

$$\begin{aligned}
 P(Y = 0) &= P(X = 0) = 0,37 \\
 P(Y = 1) &= P(X = 1) + P(X = -1) = 0,2 + 0,3 = 0,5 \\
 P(Y = 2) &= P(X = 2) + P(X = -2) = 0,04 + 0,09 = 0,13
 \end{aligned}$$

$\Rightarrow$  Bảng phân bố xác suất của Y:

Y	0	1	2
P(Y = y <sub>i</sub> )	0,37	0,5	0,13

**Bài 5/71:** Khi một người đi thi lấy bằng lái xe nếu không đạt anh ta lại đăng ký thi lại cho đến khi đạt mới thôi. Gọi X là số lần anh ta dự thi. Lập phân bố xác suất của X biết rằng xác suất thi đỗ của anh ta là 1/3.

Hãy dự đoán xem trong 243 người (mỗi người đều có xác suất thi đỗ là 1/3) có bao nhiêu người thi đạt ngay lần đầu, thi đạt ở lần thứ hai, phải thi ít nhất bốn lần.

Gọi X là số lần phải thi. Các giá trị có thể có của X:

$$X(\Omega) = \{1, 2, 3, \dots, n\} \quad (\text{với } n \rightarrow +\infty)$$

+ Tính các xác suất tương ứng:

$$P(X=1) = \frac{1}{3}; \quad P(X=2) = \frac{2}{3} \cdot \frac{1}{3}$$

$$P(X=3) = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3}; \quad P(X=k) = \left(\frac{2}{3}\right)^{k-1} \cdot \frac{1}{3}$$

$\Rightarrow$  Bảng phân bố xác suất của X:

X	1	2	3	...	k
P(X = x <sub>i</sub> )	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{4}{27}$		$\left(\frac{2}{3}\right)^{k-1} \cdot \frac{1}{3}$

\* Nếu có 243 người dự thi:

+ Dự đoán số người thi đạt ngay lần đầu:  $\frac{1}{3} \cdot 243 = \boxed{81}$

+ Dự đoán số người thi đạt ở lần thứ hai:  $\frac{2}{9} \cdot 243 = \boxed{54}$

+ Dự đoán số người phải thi ít nhất bốn lần:

$$\left[1 - P(X \leq 3)\right] \cdot 243 = \left(1 - \frac{1}{3} - \frac{2}{9} - \frac{4}{27}\right) \cdot 243 = \boxed{72}$$

**Bài 6/72: Cho hai ĐLNN X và Y có phân bố xác suất như sau:**

X	0	1	2	3	4	5
P	0,15	0,3	0,25	0,2	0,08	0,02
Y	0	1	2	3	4	5
P	0,3	0,2	0,2	0,15	0,1	0,05

a) Tính EX, EY

b) Tìm P(X + Y ≤ 3) nếu X và Y độc lập.

a) Tính EX, EY:

$$+ EX = \sum_{i=1}^6 x_i p_i = 0.0,15 + 1.0,3 + 2.0,25 + 3.0,2 + 4.0,08 + 5.0,02 = \boxed{1,82}$$

$$+ EY = \sum_{j=1}^6 y_j p_j = 0.0,3 + 1.0,2 + 2.0,2 + 3.0,15 + 4.0,1 + 5.0,05 = \boxed{1,7}$$

b) Tìm P(X + Y ≤ 3):

$$P(X + Y \leq 3) = P(X = 0; Y = 0) + P(X = 0; Y = 1) + P(X = 0; Y = 2)$$

$$+ P(X = 0; Y = 3) + P(X = 1; Y = 0) + P(X = 1; Y = 1) + P(X = 1; Y = 2)$$

$$+ P(X = 2; Y = 0) + P(X = 2; Y = 1) + P(X = 3; Y = 0)$$

Vì X và Y độc lập nên:  $P(X = x_i; Y = y_j) = P(X = x_i).P(Y = y_j)$ . Do đó:

$$\begin{aligned} P(X + Y \leq 3) &= 0,15.0,3 + 0,15.0,2 + 0,15.0,2 + 0,15.0,15 + 0,3.0,3 + 0,3.0,2 \\ &\quad + 0,3.0,2 + 0,25.0,3 + 0,25.0,2 + 0,2.0,3 = \boxed{0,5225} \end{aligned}$$

**Bài 7/72:** Các ĐLNN X và Y có bảng phân bố xác suất đồng thời như sau:

X \ Y	1	2	3
1	0,12	0,15	0,03
2	0,28	0,35	0,07

a) **Chứng minh rằng X và Y độc lập.**

b) **Tìm quy luật phân bố của Z = XY**

Từ đó tính EZ và kiểm tra rằng EZ = EX.EY

a) *Chứng minh X và Y độc lập:*

$$X \text{ và } Y \text{ độc lập khi và chỉ khi}^1: P(X = x_i; Y = y_j) = P(X = x_i).P(Y = y_j) \quad \forall i, j$$

Từ bảng phân bố xác suất đồng thời của X và Y, suy ra các phân bố xác suất thành phần như sau:

$$\begin{aligned} P(X = x_i) &= \sum_{j=1}^3 p_{ij}; & P(Y = y_j) &= \sum_{i=1}^2 p_{ij} \\ \Rightarrow P(X = 1) &= 0,12 + 0,15 + 0,03 = 0,3 & P(Y = 1) &= 0,12 + 0,28 = 0,4 \\ P(X = 2) &= 0,28 + 0,35 + 0,07 = 0,7 & P(Y = 2) &= 0,15 + 0,35 = 0,5 \\ && P(Y = 3) &= 0,03 + 0,07 = 0,1 \end{aligned}$$

+ *Kiểm tra từng đăng thức:*

$$+) P(X = 1; Y = 1) = 0,12 ; \quad P(X = 1).P(Y = 1) = 0,3.0,4 = 0,12$$

$$\Rightarrow P(X = 1; Y = 1) = P(X = 1).P(Y = 1)$$

$$+) P(X = 1; Y = 2) = 0,15 ; \quad P(X = 1).P(Y = 2) = 0,3.0,5 = 0,15$$

$$\Rightarrow P(X = 1; Y = 2) = P(X = 1).P(Y = 2)$$

$$+) P(X = 1; Y = 3) = 0,03 ; \quad P(X = 1).P(Y = 3) = 0,3.0,1 = 0,03$$

$$\Rightarrow P(X = 1; Y = 3) = P(X = 1).P(Y = 3)$$

$$+) P(X = 2; Y = 1) = 0,28 ; \quad P(X = 2).P(Y = 1) = 0,7.0,4 = 0,28$$

$$\Rightarrow P(X = 2; Y = 1) = P(X = 2).P(Y = 1)$$

<sup>1</sup> Chú ý: Nếu chứng minh cho hệ số tương quan bằng 0 mà suy ra X và Y độc lập thì không đúng. Vì nếu X và Y độc lập thì suy ra hệ số tương quan bằng 0 nhưng điều ngược lại không đúng.

$$+) P(X = 2; Y = 2) = 0,35; \quad P(X = 2).P(Y = 2) = 0,7.0,5 = 0,35$$

$$\Rightarrow P(X = 2; Y = 2) = P(X = 2).P(Y = 2)$$

$$+) P(X = 2; Y = 3) = 0,07; \quad P(X = 2).P(Y = 3) = 0,7.0,1 = 0,07$$

$$\Rightarrow P(X = 2; Y = 3) = P(X = 2).P(Y = 3)$$

Ta thấy:  $P(X = x_i; Y = y_j) = P(X = x_i).P(Y = y_j) \quad \forall i, j$

Vậy, X và Y độc lập với nhau.

b) Tìm phân bố xác suất của  $Z = XY$ :

Đại lượng ngẫu nhiên Z có thể nhận các giá trị:

$$Z(\Omega) = \{1, 2, 3, 4, 6\}$$

Tính các xác suất:

$$+) P(Z = 1) = P(X = 1, Y = 1) = 0,12$$

$$+) P(Z = 2) = P(X = 1, Y = 2) + P(X = 2, Y = 1) = 0,15 + 0,28 = 0,43$$

$$+) P(Z = 3) = P(X = 1, Y = 3) = 0,03$$

$$+) P(Z = 4) = P(X = 2, Y = 2) = 0,35$$

$$+) P(Z = 6) = P(X = 2, Y = 3) = 0,07$$

$\Rightarrow$  Bảng phân bố xác suất của Z:

Z	1	2	3	4	6
$P(Z = z_i)$	0,12	0,43	0,03	0,35	0,07

$\Rightarrow$  Kỳ vọng của Z:

$$EZ = \sum_{i=1}^5 z_i p_i = 1.0,12 + 2.0,43 + 3.0,03 + 4.0,35 + 6.0,07 = [2,89] \quad (1)$$

$$\text{Mà: } EX = \sum_{i=1}^2 x_i p_i = 1.0,3 + 2.0,7 = 1,7$$

$$EY = \sum_{j=1}^3 y_j p_j = 1.0,4 + 2.0,5 + 3.0,1 = 1,7$$

$$\Rightarrow EX \cdot EY = 1,7 \cdot 1,7 = 2,89 \quad (2)$$

Từ (1) và (2) suy ra:  $EZ = EX \cdot EY$

**Bài 8/72: Số trẻ em sinh ra trong một tuần ở một làng A nào đó là một ĐLNN X có phân bố xác suất là:**

X	0	1	2	3
P	0,4	0,3	0,2	0,1

Số người chết trong một tuần ở làng A đó là một ĐLNN Y có phân bố xác suất:

Y	0	1	2	3	4
P	0,1	0,3	0,4	0,15	0,05

Giả sử rằng X và Y độc lập.

a) Tìm phân bố xác suất đồng thời của X và Y

b) Tính  $P(X > Y)$

a) Tìm phân bố xác suất đồng thời của X và Y:

$$\text{Vì } X \text{ và } Y \text{ độc lập nên: } P(X = x_i; Y = y_j) = P(X = x_i).P(Y = y_j) \quad \forall i, j$$

Ta có:

$$\begin{array}{ll}
 P(X = 0, Y = 0) = 0,4 \cdot 0,1 = 0,04 & P(X = 0, Y = 1) = 0,4 \cdot 0,3 = 0,12 \\
 P(X = 0, Y = 2) = 0,4 \cdot 0,4 = 0,16 & P(X = 0, Y = 3) = 0,4 \cdot 0,15 = 0,06 \\
 P(X = 0, Y = 4) = 0,4 \cdot 0,05 = 0,02 & \\
 \\ 
 P(X = 1, Y = 0) = 0,3 \cdot 0,1 = 0,03 & P(X = 1, Y = 1) = 0,3 \cdot 0,3 = 0,09 \\
 P(X = 1, Y = 2) = 0,3 \cdot 0,4 = 0,12 & P(X = 1, Y = 3) = 0,3 \cdot 0,15 = 0,045 \\
 P(X = 1, Y = 4) = 0,3 \cdot 0,05 = 0,015 & \\
 \\ 
 P(X = 2, Y = 0) = 0,2 \cdot 0,1 = 0,02 & P(X = 2, Y = 1) = 0,2 \cdot 0,3 = 0,06 \\
 P(X = 2, Y = 2) = 0,2 \cdot 0,4 = 0,08 & P(X = 2, Y = 3) = 0,2 \cdot 0,15 = 0,03 \\
 P(X = 2, Y = 4) = 0,2 \cdot 0,05 = 0,01 & \\
 \\ 
 P(X = 3, Y = 0) = 0,1 \cdot 0,1 = 0,01 & P(X = 3, Y = 1) = 0,1 \cdot 0,3 = 0,03 \\
 P(X = 3, Y = 2) = 0,1 \cdot 0,4 = 0,04 & P(X = 3, Y = 3) = 0,1 \cdot 0,15 = 0,015 \\
 P(X = 3, Y = 4) = 0,1 \cdot 0,05 = 0,005 &
 \end{array}$$

⇒ Bảng phân bố xác suất đồng thời của X và Y:

X \ Y	0	1	2	3	4
0	0,04	0,12	0,16	0,06	0,02
1	0,03	0,09	0,12	0,045	0,015
2	0,02	0,06	0,08	0,03	0,01
3	0,01	0,03	0,04	0,015	0,005

b) Tìm  $P(X > Y)$ :

$P(X > Y)$  bằng tổng các giá trị nằm bên dưới đường thẳng  $X = Y$ .

Xác suất cần tìm là:

$$P(X > Y) = 0,03 + 0,02 + 0,06 + 0,01 + 0,03 + 0,04 = \boxed{0,19}$$

**Bài 9/73: Cho X và Y là hai ĐLNN có phân bố xác suất đồng thời như sau:**

X \ Y	-1	1
-1	1/6	1/4
0	1/6	1/8
1	1/6	1/8

Hãy tính EX, EY, cov(X, Y) và ρ(X, Y)

Từ bảng phân bố xác suất đồng thời của X và Y, suy ra các bảng phân bố xác suất thành phần:

X	-1	0	1
P(X = x <sub>i</sub> )	10/24	7/24	7/24

Y	-1	1
P(Y = y <sub>j</sub> )	1/2	1/2

\* Kỷ vọng của X và Y:

$$EX = -1 \cdot \frac{10}{24} + 0 \cdot \frac{7}{24} + 1 \cdot \frac{7}{24} = \boxed{-\frac{1}{8}}$$

$$EY = -1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \boxed{0}$$

\* Covarian của X và Y:

$$\begin{aligned} \text{cov}(X, Y) &= EXY - EX \cdot EY = \sum x_i y_j p_{ij} - EX \cdot EY \\ &= [-1 \cdot (-1) + 0 \cdot (-1) + 1 \cdot (-1)] \frac{1}{6} + (-1) \cdot 1 \cdot \frac{1}{4} + (0 \cdot 1 + 1 \cdot 1) \cdot \frac{1}{8} - \left(-\frac{1}{8}\right) \cdot 0 = \boxed{-\frac{1}{8}} \end{aligned}$$

\* Hệ số tương quan của X và Y:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{DX \cdot DY}}$$

$$DX = EX^2 - (EX)^2 = (-1)^2 \cdot \frac{10}{24} + 0^2 \cdot \frac{7}{24} + 1^2 \cdot \frac{7}{24} - \left(-\frac{1}{8}\right)^2 = \frac{133}{192}$$

$$DY = EY^2 - (EY)^2 = (-1)^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} - 0^2 = 1$$

$$\Rightarrow \rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{DX \cdot DY}} = \frac{-\frac{1}{8}}{\sqrt{\frac{133}{192}}} = \boxed{-0,1502}$$

**Bài 10/73: Cho X và Y là hai ĐLNN có phân bố xác suất đồng thời như sau:**

X \ Y	-1	0	1
-1	4/15	1/15	4/15
0	1/15	2/15	1/15
1	0	2/15	0

a) Tìm EX, EY, cov(X,Y) và  $\rho(X,Y)$ .

b) X và Y có độc lập hay không.

a) Tìm EX, EY, cov(X, Y),  $\rho(X, Y)$ :

Từ bảng phân bố xác suất đồng thời của X và Y, suy ra các bảng phân bố xác suất thành phần như sau:

X	-1	0	1
P(X = x <sub>i</sub> )	$\frac{9}{15}$	$\frac{4}{15}$	$\frac{2}{15}$

Y	-1	0	1
P(Y = y <sub>j</sub> )	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

\* Ký vọng<sup>1</sup>:

$$EX = -1 \cdot \frac{9}{15} + 0 \cdot \frac{4}{15} + 1 \cdot \frac{2}{15} = \boxed{-\frac{7}{15}}; \quad EY = -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \boxed{0}$$

\* Covarian của X và Y:

$$\begin{aligned} \text{cov}(X, Y) &= EXY - EX \cdot EY = \sum x_i y_j p_{ij} - EX \cdot EY \\ &= (-1) \cdot (-1) \cdot \frac{4}{15} + (-1) \cdot 1 \cdot \frac{4}{15} = \boxed{0} \end{aligned}$$

\* Hệ số tương quan:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{0}{\sigma_X \cdot \sigma_Y} = \boxed{0}$$

b) X và Y có độc lập với nhau không:

Ta thấy:  $P(X = 1, Y = -1) = 0$

$$\text{Mà: } P(X = 1) \cdot P(Y = -1) = \frac{2}{5} \cdot \frac{1}{3} = \frac{2}{15}$$

$$\Rightarrow P(X = 1, Y = -1) \neq P(X = 1) \cdot P(Y = -1)$$

Vậy, X và Y **không** độc lập với nhau.

<sup>1</sup> Đáp án EX trong giáo trình G<sub>1</sub> trang 76 bị sai,  $EX \neq -1/5$

**Bài 11/73: Giả sử  $X \sim B(2; 0,4)$ ,  $Y \sim B(2; 0,7)$ ,  $X$  và  $Y$  độc lập.****a) Tìm phân bố xác suất của  $X + Y$** **b) Chứng minh rằng  $X + Y$  không có phân bố nhị thức.***a) Tìm phân bố xác suất của  $X + Y$ :*Bảng phân bố xác suất của  $X$ :  $X(\Omega) = \{0, 1, 2\}$  và  $P(X = k) = C_2^k \cdot 0,4^k \cdot 0,6^{2-k}$ 

X	0	1	2
P	0,36	0,48	0,16

Bảng phân bố xác suất của  $Y$ :  $Y(\Omega) = \{0, 1, 2\}$  và  $P(Y = k) = C_2^k \cdot 0,7^k \cdot 0,3^{2-k}$ 

Y	0	1	2
P	0,09	0,42	0,49

Đặt:  $Z = X + Y$ . Các giá trị có thể có của  $Z$ :

$$Z(\Omega) = \{0, 1, 2, 3, 4\}$$

Tính các xác suất tương ứng:

$$+ P(Z = 0) = P(X = 0, Y = 0) = 0,36 \cdot 0,09 = 0,0324$$

$$+ P(Z = 1) = P(X = 1, Y = 0) + P(X = 0, Y = 1) = 0,48 \cdot 0,09 + 0,36 \cdot 0,42 \\ = 0,1944$$

$$+ P(Z = 2) = P(X = 2, Y = 0) + P(X = 0, Y = 2) + P(X = 1, Y = 1) \\ = 0,16 \cdot 0,09 + 0,36 \cdot 0,49 + 0,48 \cdot 0,42 = 0,3924$$

$$+ P(Z = 3) = P(X = 1, Y = 2) + P(X = 2, Y = 1) = 0,48 \cdot 0,49 + 0,16 \cdot 0,42 \\ = 0,3024$$

$$+ P(Z = 4) = P(X = 2, Y = 2) = 0,16 \cdot 0,49 = 0,0784$$

 $\Rightarrow$  Bảng phân bố xác suất của  $Z$ :

Z	0	1	2	3	4
P	0,0324	0,1944	0,3924	0,3024	0,0784

*b) Chứng minh rằng  $X + Y$  không có phân bố nhị thức<sup>1</sup>:*Giả sử  $Z = X + Y$  có phân bố nhị thức  $Z \sim B(4; p)$  thì phải thỏa mãn:

$$P(Z = k) = C_4^k \cdot p^k \cdot (1-p)^{4-k} \quad \text{với cùng một giá trị xác suất } p.$$

Ta có:

$$P(Z = 0) = C_4^0 \cdot p^0 \cdot (1-p)^4 = 0,0324$$

$$(1-p)^4 = 0,0324 \Leftrightarrow p = 0,5757 \quad (1)$$

<sup>1</sup> Sử dụng phương pháp chứng minh phản chứng.

$$P(Z=4) = C_4^4 \cdot p^4 \cdot (1-p)^0 = 0,0784$$

$$p^4 = 0,0784 \Leftrightarrow p = 0,5292 \quad (2)$$

Giá trị p tại (1) và (2) khác nhau nên suy ra Z không có phân bố nhị thức.

### Bài 12/73: Cho X và Y là hai ĐLNN độc lập.

a) Giả sử  $X \sim B(1; 1/5)$ ,  $Y \sim B(2; 1/5)$ . Viết phân bố xác suất của X, Y. Từ đó tìm phân bố xác suất của  $X + Y$ . Kiểm tra rằng  $X + Y \sim B(3; 1/5)$

b) Giả sử  $X \sim B(1; 1/2)$ ,  $Y \sim B(2; 1/5)$ . Tìm phân bố xác suất của  $X + Y$ . Chứng minh rằng  $X + Y$  không có phân bố nhị thức.

a) Tìm phân bố xác suất của X, Y,  $X + Y$ . Chứng minh rằng  $X + Y \sim B(3; 1/5)$ :

+ Các giá trị có thể có của X:  $X(\Omega) = \{0, 1\}$

$$P(X=0) = C_1^0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^1 = \frac{4}{5}; \quad P(X=1) = C_1^1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^0 = \frac{1}{5}$$

$\Rightarrow$  Bảng phân bố xác suất của X:

X	0	1
P	$\frac{4}{5}$	$\frac{1}{5}$

+ Các giá trị có thể có của Y:  $Y(\Omega) = \{0, 1, 2\}$

$$P(Y=0) = C_2^0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^2 = \frac{16}{25}; \quad P(Y=1) = C_2^1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^1 = \frac{8}{25}$$

$$P(Y=2) = C_2^2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^0 = \frac{1}{25}$$

$\Rightarrow$  Bảng phân bố xác suất của Y:

Y	0	1	2
P	$\frac{16}{25}$	$\frac{8}{25}$	$\frac{1}{25}$

+ Các giá trị có thể có của  $Z = X + Y$  là:  $Z(\Omega) = \{0, 1, 2, 3\}$

$$P(Z=0) = P(X=0, Y=0) = \frac{4}{5} \cdot \frac{16}{25} = \frac{64}{125} \quad (\text{vì X và Y độc lập})$$

$$P(Z=1) = P(X=1, Y=0) + P(X=0, Y=1) = \frac{1}{5} \cdot \frac{16}{25} + \frac{4}{5} \cdot \frac{8}{25} = \frac{48}{125}$$

$$P(Z=2) = P(X=0, Y=2) + P(X=1, Y=1) = \frac{4}{5} \cdot \frac{1}{25} + \frac{1}{5} \cdot \frac{8}{25} = \frac{12}{125}$$

$$P(Z=3) = P(X=1, Y=2) = \frac{1}{5} \cdot \frac{1}{25} = \frac{1}{125}$$

⇒ Bảng phân bố xác suất của Z:

Z	0	1	2	3
P	$\frac{64}{125}$	$\frac{48}{125}$	$\frac{12}{125}$	$\frac{1}{125}$

+ Giả sử  $Z \sim B(3; 1/5)$ . Ta có:

$$P(Z=0) = \left(\frac{4}{5}\right)^3 = \frac{64}{125}; \quad P(Z=1) = C_3^1 \cdot \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^2 = \frac{48}{125}$$

$$P(Z=2) = C_3^2 \cdot \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^1 = \frac{12}{125}; \quad P(Z=3) = \left(\frac{1}{5}\right)^3 = \frac{1}{125}$$

Các giá trị xác suất tìm được hoàn toàn phù hợp với bảng phân bố xác suất ở trên. Vậy  $Z = X + Y \sim B(3; 1/5)$ .

b) Tìm phân bố xác suất của  $X + Y$ . Chứng minh  $X + Y$  không có phân bố nhị thức:

+ Các giá trị có thể có của  $Z = X + Y$ :

$$Z(\Omega) = \{0, 1, 2, 3\}$$

$$P(Z=0) = P(X=0, Y=0) = \frac{1}{2} \cdot \frac{16}{25} = \frac{16}{50}$$

$$P(Z=1) = P(X=1, Y=0) + P(X=0, Y=1)$$

$$= \frac{1}{2} \cdot \frac{16}{25} + \frac{1}{2} \cdot \frac{8}{25} = \frac{24}{50}$$

$$P(Z=2) = P(X=1, Y=1) + P(X=0, Y=2)$$

$$= \frac{1}{2} \cdot \frac{8}{25} + \frac{1}{2} \cdot \frac{1}{25} = \frac{9}{50}$$

$$P(Z=3) = P(X=1, Y=2) = \frac{1}{2} \cdot \frac{1}{25} = \frac{1}{50}$$

⇒ Bảng phân bố xác suất của Z:

Z	0	1	2	3
P	$\frac{16}{50}$	$\frac{24}{50}$	$\frac{9}{50}$	$\frac{1}{50}$

Giả sử Z có phân bố nhị thức  $Z \sim B(3; p)$  thì phải thỏa mãn:

$$P(Z=k) = C_3^k \cdot p^k (1-p)^{3-k} \quad \text{với cùng một giá trị xác suất } p.$$

Ta có:

$$P(Z=0) = (1-p)^3 = \frac{16}{50} \Leftrightarrow p = 0,316 \quad (1)$$

$$P(Z=3) = p^3 = \frac{1}{50} \Leftrightarrow p = 0,27144 \quad (2)$$

Giá trị p tại (1) và (2) khác nhau nên suy ra  $Z = X + Y$  không có phân bố nhị thức

**Bài 13/74:** Trong một thành phố nhỏ, trung bình một tuần có 2 người chết. Tính xác suất để:

a) Không có người nào chết trong vòng một ngày.

b) Có ít nhất 3 người chết trong vòng hai ngày.

a) Xác suất không có người nào chết trong vòng một ngày<sup>1</sup>:

Trung bình một tuần có 2 người chết  $\Rightarrow$  trung bình một ngày có  $2/7$  người chết.

Gọi X là số người chết trong vòng một ngày thì:  $X \sim \text{Poisson}(2/7)$

$$\Rightarrow P(X=0) = e^{-\frac{2}{7}} \cdot \frac{\left(\frac{2}{7}\right)^0}{0!} = \boxed{0,7515}$$

b) Xác suất có ít nhất 3 người chết trong vòng hai ngày:

Trung bình một tuần có 2 người chết  $\Rightarrow$  trung bình hai ngày có  $4/7$  người chết.

Gọi Y là số người chết trong vòng hai ngày thì:  $Y \sim \text{Poisson}(4/7)$ . Xác suất cần tìm:

$$\begin{aligned} P(Y \geq 3) &= 1 - [P(Y=0) + P(Y=1) + P(Y=2)] \\ &= 1 - e^{-\frac{4}{7}} \cdot \frac{(4/7)^0}{0!} - e^{-\frac{4}{7}} \cdot \frac{(4/7)^1}{1!} - e^{-\frac{4}{7}} \cdot \frac{(4/7)^2}{2!} \\ &= 1 - 0,5647 - 0,3227 - 0,0922 = \boxed{0,0204} \end{aligned}$$

**Bài 14/74:** Tại một trạm kiểm soát giao thông trung bình một phút có 2 xe ôtô đi qua.

a) Tìm xác suất để có đúng 6 xe đi qua trong vòng ba phút.

b) Tính xác suất để trong khoảng thời gian t phút có ít nhất 1 xe ôtô đi qua.

Xác định t để xác suất này bằng 0,99.

a) Xác suất có đúng 6 xe đi qua trong vòng ba phút:

Trung bình một phút có 2 ôtô đi qua  $\Rightarrow$  trung bình 3 phút có 6 ôtô đi qua.

Gọi X là số xe đi qua trong vòng ba phút. Ta có:  $X \sim \text{Poisson}(6)$

$$\Rightarrow P(X=6) = e^{-6} \cdot \frac{6^6}{6!} = \boxed{0,1606}$$

<sup>1</sup> Thông thường, khi xuất hiện giá trị trung bình trong một khoảng thời gian  $\Delta t$  thì áp dụng phân bố Poisson.

b) Xác suất để trong t phút có ít nhất 1 ôtô đi qua. Tìm t để cho xác suất bằng 0,99:

Trung bình một phút có 2 ôtô đi qua  $\Rightarrow$  trung bình t phút có  $2t$  ôtô đi qua.

Gọi Y là số xe đi qua trong vòng t phút. Ta có:  $Y \sim \text{Poisson}(2t)$

$\Rightarrow$  Xác suất có ít nhất 1 ôtô đi qua trong vòng t phút:

$$1 - P(Y = 0) = 1 - e^{-2t} \cdot \frac{(2t)^0}{0!} = \boxed{1 - e^{-2t}}$$

Xác suất trên bằng 0,99 suy ra:  $1 - e^{-2t} = 0,99$

$$e^{-2t} = 0,01 \Leftrightarrow t = \frac{\ln 0,01}{-2} = \frac{-4,605}{-2} = \boxed{2,3026} \text{ (phút)}$$

### Bài 15/74: Tại một nhà máy nào đó trung bình một tháng có 2 tai nạn lao động

a) Tính xác suất để trong khoảng thời gian ba tháng xảy ra nhiều nhất là 3 tai nạn.

b) Tính xác suất để trong ba tháng liên tiếp, mỗi tháng xảy ra nhiều nhất một tai nạn.

a) Xác suất trong ba tháng xảy ra nhiều nhất 3 tai nạn lao động:

Trung bình một tháng có 2 tai nạn lao động  $\Rightarrow$  trung bình ba tháng có 6 tai nạn.

Gọi X là số tai nạn lao động trong vòng ba tháng thì  $X \sim \text{Poisson}(6)$

$\Rightarrow$  Xác suất cần tìm:

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

$$= e^{-6} \frac{6^0}{0!} + e^{-6} \frac{6^1}{1!} + e^{-6} \frac{6^2}{2!} + e^{-6} \frac{6^3}{3!} = e^{-6} (1 + 6 + 18 + 36) = \boxed{0,1512}$$

b) Xác suất để trong ba tháng liên tiếp, mỗi tháng xảy ra nhiều nhất 1 tai nạn:

Giả sử tai nạn lao động của mỗi tháng là độc lập với nhau.

Gọi Y là số tai nạn trong một tháng thì  $Y \sim \text{Poisson}(2)$ . Xác suất để mỗi tháng có nhiều nhất 1 tai nạn lao động là:

$$P(Y \leq 1) = e^{-2} \frac{2^0}{0!} + e^{-2} \frac{2^1}{1!} = 0,406$$

$\Rightarrow$  Xác suất trong ba tháng liên tiếp, mỗi tháng có nhiều nhất 1 tai nạn<sup>1</sup>:

$$0,406^3 = \boxed{0,0669}$$

### Bài 16/74: Một trạm cho thuê xe taxi có 3 chiếc xe. Hàng ngày trạm phải nộp thuế 8 USD cho 1 chiếc xe (dù xe đó có được thuê hay không). Mỗi chiếc xe được cho thuê với giá 20 USD. Giả sử số yêu cầu thuê xe của trạm trong 1 ngày là ĐLNN X có phân bố Poisson với tham số $\lambda = 2,8$ .

<sup>1</sup> Đáp án trong giáo trình G1 trang 78 bị nhầm lẫn vì  $Y \sim \text{Poisson}(2)$  chứ không phải  $Y \sim \text{Poisson}(1)$

a) Gọi Y là số tiền thu được trong một ngày của trạm (nếu không có ai thuê thì số tiền thu được là -24 USD). Tìm phân bố xác suất của Y. Từ đó tính số tiền trung bình thu được của trạm trong 1 ngày.

b) Giải bài toán trên trong trường hợp trạm có 4 chiếc xe.

c) Trạm nên có 3 hay 4 chiếc xe.

a) Tìm phân bố xác suất của Y, tính EY:

X là số yêu cầu thuê xe trong 1 ngày:  $X \sim \text{Poisson}(2,8)$

Y là số tiền thu được trong 1 ngày. Ta có quan hệ giữa X và Y như sau:

$$+ Y = 20X - 24 \quad \text{nếu } X < 3.$$

$$+ Y = 60 - 24 = 36 \quad \text{nếu } X \geq 3.$$

Các giá trị có thể có của Y:

X	0	1	2	$\geq 3$
Y	-24	-4	16	36

+ Tính các xác suất tương ứng:

$$P(Y = -24) = P(X = 0) = e^{-2,8} \cdot \frac{2,8^0}{0!} = 0,0608$$

$$P(Y = -4) = P(X = 1) = e^{-2,8} \cdot \frac{2,8^1}{1!} = 0,1703$$

$$P(Y = 16) = P(X = 2) = e^{-2,8} \cdot \frac{2,8^2}{2!} = 0,2384$$

$$P(Y = 36) = P(X \geq 3) = 1 - 0,0608 - 0,1703 - 0,2384 = 0,5305$$

$\Rightarrow$  Bảng phân bố xác suất của Y:

Y	-24	-4	16	36
P	0,0608	0,1703	0,2384	0,5305

$\Rightarrow$  Số tiền trung bình trạm thu được trong 1 ngày:

$$EY = -24 \cdot 0,0608 - 4 \cdot 0,1703 + 16 \cdot 0,2384 + 36 \cdot 0,5305 = [20,773] \text{ (USD)}$$

b) Tìm phân bố xác suất của Y và EY trong trường hợp trạm có 4 chiếc xe:

Ta có quan hệ giữa X và Y như sau:

$$+ Y = 20X - 32 \quad \text{nếu } X < 4.$$

$$+ Y = 80 - 32 = 48 \quad \text{nếu } X \geq 4.$$

Các giá trị có thể có của Y:

X	0	1	2	3	$\geq 4$
Y	-32	-12	8	28	48

+ Tính các xác suất tương ứng:

$$P(Y = -32) = P(X = 0) = e^{-2,8} \cdot \frac{2,8^0}{0!} = 0,0608$$

$$P(Y = -12) = P(X = 1) = e^{-2,8} \cdot \frac{2,8^1}{1!} = 0,1703$$

$$P(Y = 8) = P(X = 2) = e^{-2,8} \cdot \frac{2,8^2}{2!} = 0,2384$$

$$P(Y = 28) = P(X = 3) = e^{-2,8} \cdot \frac{2,8^3}{3!} = 0,2225$$

$$P(Y = 48) = P(X \geq 4) = 1 - 0,0608 - 0,1703 - 0,2384 - 0,2225 = 0,3080$$

⇒ Bảng phân bố xác suất của Y:

Y	-32	-12	8	28	48
P	0,0608	0,1703	0,2384	0,2225	0,308

⇒ Số tiền trung bình trạm thu được trong 1 ngày:

$$\begin{aligned} EY &= -32 \cdot 0,0608 - 12 \cdot 0,1703 + 8 \cdot 0,2384 + 28 \cdot 0,2225 + 48 \cdot 0,308 \\ &= 18,932 \quad (\text{USD}) \end{aligned}$$

c) Ta thấy, khi trạm có 3 xe thì số tiền trung bình thu được trong ngày lớn hơn khi trạm có 4 xe. Vậy trạm nên có 3 chiếc xe.

## 2.2. Nhận xét bài tập chương 2

Chương 2 liên quan đến phân bố xác suất của đại lượng ngẫu nhiên rời rạc. Từ bảng phân bố xác suất tính các đặc trưng của ĐLNN (kỳ vọng, phương sai, độ lệch chuẩn; covarian, hệ số tương quan của hai ĐLNN).

Lập bảng phân bố xác suất của ĐLNN rời rạc. Tùy trường hợp thông tin để bài cho mà các xác suất  $p_i$  được tìm bằng cách:

- + Sử dụng giải tích tổ hợp
- + Sử dụng các phân bố rời rạc: nhị thức, Poisson.
- + Thông qua một hoặc một số ĐLNN rời rạc khác (hàm của ĐLNN rời rạc)

Nhìn chung, các bài tập của chương này không phức tạp. Chỉ cần ghi nhớ các công thức và áp dụng để tính toán.

# CHƯƠNG 3: ĐẠI LƯỢNG NGẪU NHIÊN LIÊN TỤC

## A. LÝ THUYẾT

➤ *DLNN X được gọi là DLNN liên tục nếu:*

- + Tập hợp tất cả các giá trị có thể có của X lấp đầy một khoảng của trục số (hoặc nhiều khoảng hoặc toàn bộ trục số)
- + Với mọi số thực a thì:  $P(X = a) = 0$

### 3.1. HÀM MẬT ĐỘ XÁC SUẤT VÀ HÀM PHÂN BỐ XÁC SUẤT

a) *Hàm mật độ xác suất:*

Hàm số  $f(x)$  xác định  $\forall x \in |R$  được gọi là hàm mật độ xác suất của ĐLNN liên tục X nếu thỏa mãn đồng thời các điều kiện sau:

- +  $f(x) \geq 0 \quad \forall x \in |R$
- +  $\int_{-\infty}^{+\infty} f(x)dx = 1$
- +  $P(a < X < b) = P(a \leq X \leq b) = \int_a^b f(x)dx$

b) *Hàm phân bố xác suất:*

Hàm số  $F(x)$  xác định  $\forall x \in |R$  gọi là hàm phân bố xác suất của ĐLNN liên tục X và được xác định bởi:

$$F(x) = P(X < x)$$

Tính chất:

- +  $0 \leq F(x) \leq 1$
- +  $F(x)$  là hàm không giảm: nếu  $x_2 > x_1$  thì  $F(x_2) \geq F(x_1)$
- +  $F(x)$  là hàm liên tục
- + 
$$\begin{cases} \lim_{x \rightarrow +\infty} F(x) = 1 \\ \lim_{x \rightarrow -\infty} F(x) = 0 \end{cases}$$
- +  $F(x) = \int_{-\infty}^x f(t) dt; \quad f(x) = F'(x) \quad (t \text{ được thay cho } x \text{ để tránh trùng biển})$

### 3.2. MỘT SỐ ĐẶC TRƯNG CỦA ĐẠI LƯỢNG NGẪU NHIÊN LIÊN TỤC

a) *Kỳ vọng:*

$$EX = \int_{-\infty}^{+\infty} xf(x)dx$$

b) Phuong sai:

$$DX = \int_{-\infty}^{+\infty} x^2 f(x) dx - (EX)^2$$

c) Độ lệch chuẩn:

$$\sigma_x = \sqrt{DX}$$

d) Mode X:

Mode của X là giá trị  $x_0$  sao cho  $f(x)$  đạt max.

Kí hiệu: Mod X

e) Median:

Giá trị m được gọi là median của X nếu:

$$P(X < m) = P(X > m) = \frac{1}{2} \Leftrightarrow F(m) = \frac{1}{2}$$

### 3.3. Hàm của đại lượng ngẫu nhiên liên tục

Cho X là một ĐLNN liên tục thì  $Y = g(X)$  được gọi là hàm của X.

a) Hàm mật độ xác suất và hàm phân bố xác suất của Y:

+ Hàm phân bố xác suất:

$$F_Y(x) = P(Y < x) = P(g(X) < x)$$

+ Hàm mật độ xác suất:

$$f_Y(x) = F'_Y(x)$$

b) Kỳ vọng:

$$EY = \int_{-\infty}^{+\infty} g(x) f(x) dx \quad (\text{với } f(x) \text{ là hàm mật độ của X})$$

c) Phuong sai:

$$DY = \int_{-\infty}^{+\infty} g^2(x) f(x) dx - (EY)^2$$

### 3.4. Phân bố chuẩn

a) Phân bố chuẩn tắc (hay chuẩn hóa):

ĐLNN liên tục Z được gọi là có phân bố chuẩn tắc nếu Z có hàm mật độ:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \text{Kí hiệu: } Z \sim N(0, 1)$$

$$\text{Hàm phân bố xác suất của Z: } \Phi(x) = P(Z < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

Vì hàm  $e^{-\frac{t^2}{2}}$  không có nguyên hàm nên tích phân trên được tính bằng phương pháp số (sử dụng máy tính). Tra cứu các giá trị của hàm  $\Phi(x)$  trong bảng Phụ lục P.4.3

### Tính chất:

$$P(Z < a) = \Phi(a)$$

$$P(Z > a) = 1 - \Phi(a) = \Phi(-a)$$

$$P(a < Z < b) = \Phi(b) - \Phi(a)$$

$$EZ = \text{Mod } Z = \text{Median } Z = 0$$

$$DZ = 1$$

### *b) Phân bố chuẩn:*

ĐLNN X được gọi là có phân bố chuẩn với tham số  $\mu$  và  $\sigma^2$  ( $\sigma > 0$ ) nếu ĐLNN  $Z = \frac{X - \mu}{\sigma}$  có phân bố chuẩn tắc. Kí hiệu:  $X \sim N(\mu, \sigma^2)$

#### \* *Hàm mật độ của X:*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

#### \* *Hàm phân bố của X:*

$$F(x) = P(X < x) = P\left(Z < \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

### Tính chất:

$$P(X < a) = P\left(Z < \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P(X > a) = 1 - P(X < a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P(a < X < b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$EX = \text{Mod } X = \text{Median } X = \mu$$

$$DX = \sigma^2$$

### **3.5. Phân bố mũ**

ĐLNN liên tục X được gọi là có phân bố mũ với tham số  $\lambda$  ( $\lambda > 0$ ) nếu X có hàm mật độ:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (\text{khi } x > 0) \\ 0 & (\text{khi } x \leq 0) \end{cases}$$

a) *Hàm phân bố:*

$$+ F(x) = \int_{-\infty}^x \lambda e^{-\lambda t} dt = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \quad (\text{nếu } x > 0)$$

$$+ F(x) = 0 \quad (\text{nếu } x \leq 0)$$

b) *Kỳ vọng:*

$$EX = \int_{-\infty}^{+\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

c) *Phương sai:*

$$DX = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx - (EX)^2 = \frac{1}{\lambda^2}$$

d) *Độ lệch chuẩn:*

$$\sigma_x = \sqrt{DX} = \frac{1}{\lambda}$$

### 3.6. Phân bố đều

ĐLNN liên tục X được gọi là có phân bố đều trên đoạn  $[a, b]$  nếu X có thể nhận bất kì giá trị nào trên  $[a, b]$  với xác suất như nhau và không nhận giá trị nào bên ngoài đoạn  $[a, b]$ .

a) *Hàm mật độ:*

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{khi } x \in [a, b] \\ 0 & \text{khi } x \notin [a, b] \end{cases}$$

b) *Hàm phân bố:*

$$F(x) = \int_{-\infty}^x f(x) dx = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a}$$

c) *Kỳ vọng:*

$$EX = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$$

d) *Phương sai:*

$$DX = \int_a^b \frac{x^2}{b-a} dx - (EX)^2 = \frac{(b-a)^2}{12}$$

e) *Độ lệch chuẩn:*  $\sigma_x = \sqrt{DX} = \frac{b-a}{2\sqrt{3}}$

f) Mode X:

Là giá trị  $x_i$  bất kì sao cho  $x_i \in [a, b]$

g) Median X:

$$m = \frac{a + b}{2}$$

h) Xác suất để X rơi vào đoạn  $[\alpha, \beta]$ :

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} \frac{1}{b-a} dx = \frac{\beta - \alpha}{b-a} \quad \text{với } \alpha, \beta \in [a, b]$$

Xác suất này chỉ phụ thuộc vào độ dài đoạn  $[\alpha, \beta]$  và tỉ lệ thuận với độ dài đoạn đó.

## B. BÀI TẬP

### 3.1. Bài tập trong giáo trình 1 (G1)

(Mở đầu về lý thuyết xác suất và các ứng dụng, Đặng Hùng Thắng, trang 102)

**Bài 1/102: Cho ĐLNN liên tục X có hàm mật độ**

$$f(x) = \begin{cases} cx^2(1-x) & \text{khi } x \in [0, 1] \\ 0 & \text{khi } x \notin [0, 1] \end{cases}$$

a) Tìm hằng số c.

b) Tìm mod.

c) Tìm  $P(0,4 < X < 0,6)$ .

a) Tìm hằng số c:

Vì  $f(x)$  là hàm mật độ của ĐLNN liên tục X nên:

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x) dx = 1 &\Leftrightarrow \int_0^1 cx^2(1-x) dx = 1 \Leftrightarrow c \int_0^1 (x^2 - x^3) dx = 1 \\ &\Leftrightarrow c \left[ \frac{x^3}{3} - \frac{x^4}{4} \right]_0^1 = 1 \Leftrightarrow c \left( \frac{1}{3} - \frac{1}{4} \right) = 1 \Leftrightarrow c \frac{1}{12} = 1 \Leftrightarrow \boxed{c = 12} \end{aligned}$$

b) Tìm Mode X:

Mod của X là số  $x_0$  sao cho:  $f(x_0) \geq f(x) \quad \forall x \in (x_0 - \delta; x_0 + \delta)$

Ta có:  $f(x) = 12x^2(1-x)$  với  $x \in [0, 1]$

$$\text{Giải phương trình: } f'(x) = 0 \Leftrightarrow 24x - 36x^2 = 0 \Leftrightarrow 12x(2 - 3x) \Leftrightarrow \begin{cases} x = 0 \\ x = \frac{2}{3} \end{cases}$$

Bảng xét dấu  $f'(x)$ :

x	0	$2/3$	
f'(x)	-	0	+
f(x)	-	0	-

↗ CD ↗

Vậy:  $\text{Mod } X = \frac{2}{3}$

c) Tính  $P(0,4 < X < 0,6)$ :

$$\begin{aligned} P(0,4 < X < 0,6) &= \int_{0,4}^{0,6} f(x) dx = \int_{0,4}^{0,6} 12x^2(1-x)dx = 12 \left( \frac{x^3}{3} - \frac{x^4}{4} \right) \Big|_{0,4}^{0,6} \\ &= (4x^3 - 3x^4) \Big|_{0,4}^{0,6} = 4 \cdot 0,6^3 - 3 \cdot 0,6^4 - 4 \cdot 0,4^3 + 3 \cdot 0,4^4 = [0,296] \end{aligned}$$

**Bài 2/102:** Cho ĐLNN X có phân bố đều trên  $[1, 2]$ . Tìm  $P(2 < X^2 < 5)$ .

$$P(2 < X^2 < 5) = P(\sqrt{2} < X < \sqrt{5}) = P(\sqrt{2} < X < 2)$$

$$= \int_{\sqrt{2}}^2 \frac{1}{2-1} dx = x \Big|_{\sqrt{2}}^2 = 2 - \sqrt{2} = [0,5858]$$

**Bài 3/102:** Cho ĐLNN X có phân bố đều trên  $[-1, 3]$ . Tìm  $P(X^2 < 2)$

Hàm mật độ xác suất của X là:

$$f(x) = \begin{cases} \frac{1}{3-(-1)} = \frac{1}{4} & \text{khi } x \in [-1, 3] \\ 0 & \text{khi } x \notin [-1, 3] \end{cases}$$

Xác suất cần tìm:

$$P(X^2 < 2) = P(-\sqrt{2} < X < \sqrt{2}) = P(-1 < X < \sqrt{2})$$

$$= \int_{-1}^{\sqrt{2}} \frac{1}{4} dx = \frac{1}{4} x \Big|_{-1}^{\sqrt{2}} = \frac{\sqrt{2} + 1}{4} = [0,6036]$$

**Bài 4/102:** Cho ĐLNN X có hàm mật độ

$$f(x) = \begin{cases} kx^2 & \text{khi } 0 \leq x \leq 3 \\ 0 & \text{khác} \end{cases}$$

- a) Tìm hằng số k
- b) Tính  $P(X > 2)$
- c) Tìm median.
- d) Xác định a để  $P(X < a) = 3/4$

a) Tìm hằng số  $k$ :

Vì  $f(x)$  là hàm mật độ của ĐLNN liên tục  $X$  nên suy ra:

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x) dx = 1 &\Leftrightarrow \int_0^3 kx^2 dx = 1 \Leftrightarrow k \left. \frac{x^3}{3} \right|_0^3 = 1 \Leftrightarrow k \frac{27}{3} = 1 \\ &\Leftrightarrow k = \boxed{\frac{1}{9}} \end{aligned}$$

b) Tính  $P(X > 2)$ :

$$P(X > 2) = \int_2^{+\infty} f(x) dx = \int_2^3 \frac{1}{9} x^2 dx = \frac{1}{27} x^3 \Big|_2^3 = \frac{27 - 8}{27} = \frac{19}{27} = \boxed{0,7037}$$

c) Tìm median của  $X$ :

Med của  $X$  là số  $m$  sao cho  $P(X < m) = 1/2$

$$\begin{aligned} P(X < m) = \frac{1}{2} &\Leftrightarrow \int_0^m \frac{1}{9} x^2 dx = \frac{1}{2} \Leftrightarrow \frac{1}{27} x^3 \Big|_0^m = \frac{1}{2} \Leftrightarrow \frac{m^3}{27} = \frac{1}{2} \\ &\Leftrightarrow m = \sqrt[3]{\frac{27}{2}} = \frac{3}{\sqrt[3]{2}} = \boxed{2,3811} \end{aligned}$$

d) Xác định  $a$  để  $P(X < a) = 3/4$ :

$$\begin{aligned} P(X < a) = \frac{3}{4} &\Leftrightarrow \int_0^a \frac{1}{9} x^2 dx = \frac{3}{4} \Leftrightarrow \frac{1}{27} x^3 \Big|_0^a = \frac{3}{4} \Leftrightarrow \frac{a^3}{27} = \frac{3}{4} \Leftrightarrow a^3 = \frac{81}{4} \\ &\Leftrightarrow a = \sqrt[3]{\frac{81}{4}} = \boxed{2,7257} \end{aligned}$$

**Bài 5/103: Cho ĐLNN  $X$  có hàm mật độ**

$$f(x) = \begin{cases} \frac{3}{4}x(2-x) & \text{khi } 0 \leq x \leq 2 \\ 0 & \text{khác} \end{cases}$$

a) Vẽ đồ thị của  $f(x)$

b) Tìm  $P(X > 1,5)$  và  $P(0,9 < X < 1,1)$ .

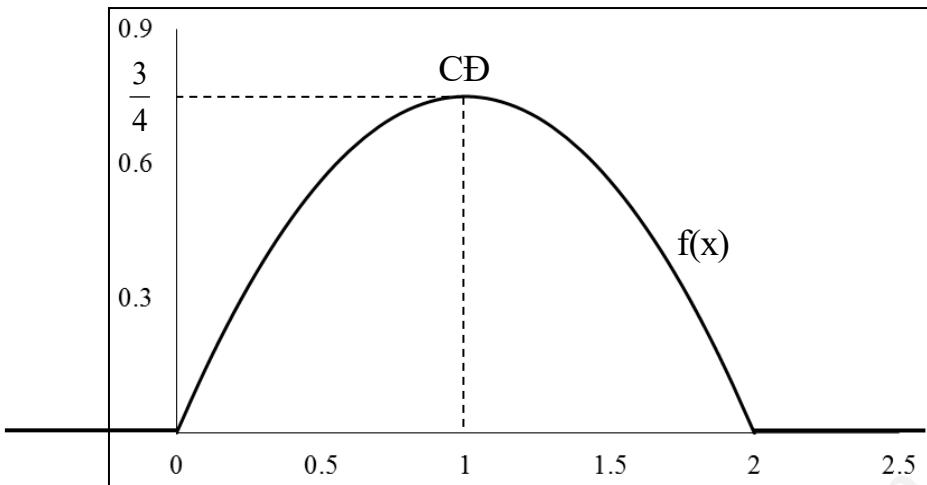
a) Vẽ đồ thị của hàm số  $f(x)$ :

Hàm số  $f(x)$  xác định trên  $\mathbb{R}$  và chỉ nhận giá trị  $> 0$  trên đoạn  $[0, 2]$ .

Tìm điểm cực đại của hàm số:

$$f'(x) = 0 \Leftrightarrow 2 - 2x = 0 \Leftrightarrow x = 1. \text{ Do đó: } \text{CD}\left(1, \frac{3}{4}\right)$$

⇒ Đồ thị hàm số  $f(x)$  như sau:



b) Tìm  $P(X > 1,5)$  và  $P(0,9 < X < 1,1)$ :

$$\begin{aligned} P(X > 1,5) &= \int_{1,5}^{+\infty} f(x) dx = \int_{1,5}^2 \frac{3}{4} x(2-x) dx = \frac{3}{4} \left( x^2 - \frac{x^3}{3} \right) \Big|_{1,5}^2 \\ &= \frac{3}{4} \left( 2^2 - \frac{2^3}{3} - 1,5^2 + \frac{1,5^3}{3} \right) = \boxed{0,1563} \end{aligned}$$

$$\begin{aligned} P(0,9 < X < 1,1) &= \int_{0,9}^{1,1} \frac{3}{4} x(2-x) dx = \frac{3}{4} \left( x^2 - \frac{x^3}{3} \right) \Big|_{0,9}^{1,1} \\ &= \frac{3}{4} \left( 1,1^2 - \frac{1,1^3}{3} - 0,9^2 + \frac{0,9^3}{3} \right) = \boxed{0,1495} \end{aligned}$$

**Bài 6/103:** Tuổi thọ của một loài côn trùng nào đó là một ĐLNN X (tính bằng tháng) với hàm mật độ

$$f(x) = \begin{cases} kx^2(4-x) & \text{khi } 0 \leq x \leq 4 \\ 0 & \text{khác} \end{cases}$$

- a) Xác định k và vẽ đồ thị của  $f(x)$
- b) Tìm mod của X
- c) Tìm xác suất để côn trùng chết trước khi nó được 1 tháng tuổi.

a) Xác định k và vẽ đồ thị của hàm số  $f(x)$ :

+ Vì  $f(x)$  là hàm mật độ của X nên suy ra:

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \Leftrightarrow \int_0^4 kx^2(4-x) dx = 1 \Leftrightarrow k \left( \frac{4x^3}{3} - \frac{x^4}{4} \right) \Big|_0^4 = 1$$

$$\Leftrightarrow k \left( 4 \cdot \frac{4^3}{3} - \frac{4^4}{4} \right) = 1 \Leftrightarrow k \frac{64}{3} = 1 \Leftrightarrow \boxed{k = \frac{3}{64}}$$

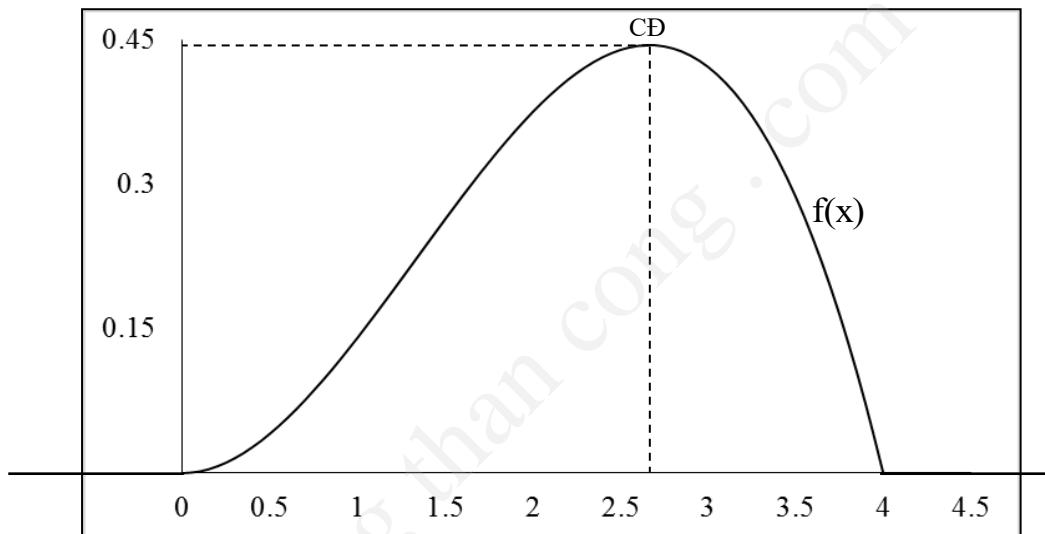
+ Vẽ đồ thị của hàm số  $f(x)$ :

Hàm số  $f(x)$  xác định trên  $\mathbb{R}$  và chỉ nhận giá trị  $> 0$  trên đoạn  $[0, 4]$ .

Tìm điểm cực đại của hàm số:

$$f'(x) = 0 \Leftrightarrow \frac{3}{8}x - \frac{9}{64}x^2 = 0 \Leftrightarrow \begin{cases} x = 0 \\ x = \frac{8}{3} \end{cases} \Rightarrow \text{CD} \left( \frac{8}{3}, \frac{4}{9} \right)$$

$\Rightarrow$  Đồ thị hàm số  $f(x)$  như sau:



b) Tìm Mod của X:

Mod của X là giá trị  $x_0$  sao hàm  $f(x)$  đạt max quanh lân cận của điểm  $x_0$ :

$$\Rightarrow \boxed{\text{Mod } X = \frac{8}{3}} \quad (\text{theo điểm CD ở câu a})$$

c) Xác suất cón trùng chết trước khi nó được 1 tháng tuổi:

$$\begin{aligned} P(X < 1) &= \int_{-\infty}^1 f(x) dx = \int_0^1 \frac{3}{64}x^2(4-x) dx = \frac{3}{64} \left( \frac{4x^3}{3} - \frac{x^4}{4} \right) \Big|_0^1 \\ &= \frac{3}{64} \left( \frac{4}{3} - \frac{1}{4} \right) = \frac{39}{768} = \frac{13}{256} = \boxed{0,0508} \end{aligned}$$

**Bài 7/103: Cho ĐLNN X có hàm mật độ**

$$f(x) = \begin{cases} \frac{x}{4} + \frac{1}{2} & \text{khi } -2 \leq x \leq 0 \\ -\frac{x}{4} + \frac{1}{2} & \text{khi } 0 \leq x \leq 2 \\ 0 & \text{khác} \end{cases}$$

Tìm kỳ vọng và phương sai của X.

\* Kỳ vọng của X:

$$\begin{aligned} EX &= \int_{-\infty}^{+\infty} x f(x) dx = \int_{-2}^0 x \left( \frac{x}{4} + \frac{1}{2} \right) dx + \int_0^2 x \left( -\frac{x}{4} + \frac{1}{2} \right) dx \\ &= \left( \frac{x^3}{12} + \frac{x^2}{4} \right) \Big|_{-2}^0 + \left( -\frac{x^3}{12} + \frac{x^2}{2} \right) \Big|_0^2 = -\frac{1}{3} + \frac{1}{3} = \boxed{0} \end{aligned}$$

\* Phương sai của X:

$$\begin{aligned} DX &= EX^2 - (EX)^2 \\ EX^2 &= \int_{-2}^0 x^2 \left( \frac{x}{4} + \frac{1}{2} \right) dx + \int_0^2 x^2 \left( -\frac{x}{4} + \frac{1}{2} \right) dx \\ &= \left( \frac{x^4}{16} + \frac{x^3}{6} \right) \Big|_{-2}^0 + \left( -\frac{x^4}{16} + \frac{x^3}{6} \right) \Big|_0^2 = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \\ \Rightarrow DX &= \frac{2}{3} - 0^2 = \boxed{\frac{2}{3}} \end{aligned}$$

**Bài 8/103: Cho ĐLNN X có hàm mật độ**

$$f(x) = \begin{cases} kx & \text{khi } 0 \leq x \leq 1 \\ k & \text{khi } 1 \leq x \leq 4 \\ 0 & \text{khác} \end{cases}$$

a) Tìm hằng số k

b) Tìm kỳ vọng, phương sai và median.

a) *Tìm hằng số k:*

Vì  $f(x)$  là hàm mật độ của X nên suy ra:

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \Leftrightarrow \int_0^1 kx dx + \int_1^4 k dx = 1 \Leftrightarrow k \frac{x^2}{2} \Big|_0^1 + kx \Big|_1^4 = 1$$

$$\Leftrightarrow \frac{k}{2} + 4k - k = 1 \Leftrightarrow \frac{k}{2} + 3k = 1 \Leftrightarrow \boxed{k = \frac{2}{7}}$$

b) Tìm kỳ vọng, phương sai và median:

\* Kỳ vọng của X:

$$EX = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^1 \frac{2}{7} x^2 dx + \int_1^4 \frac{2}{7} x dx = \frac{2}{7} \left( \frac{1}{3} + \frac{15}{2} \right) = \frac{47}{21} = \boxed{2,2381}$$

\* Phương sai của X:

$$EX^2 = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^1 \frac{2}{7} x^3 dx + \int_1^4 \frac{2}{7} x^2 dx = \frac{2}{7} \left( \frac{1}{4} + \frac{63}{3} \right) = \frac{85}{14}$$

$$\Rightarrow DX = EX^2 - (EX)^2 = \frac{85}{14} - \left( \frac{47}{21} \right)^2 = \frac{937}{882} = \boxed{1,0624}$$

\* Median của X:

$$\text{Median của } X \text{ là số } m \text{ sao cho: } P(X < m) = \frac{1}{2}$$

$$\text{Ta có: } \int_0^1 \frac{2}{7} x dx = \frac{x^2}{7} \Big|_0^1 = \frac{1}{7} < \frac{1}{2}. \text{ Do đó } m \in (1, 4), \text{ suy ra:}$$

$$\int_0^1 \frac{2}{7} x dx + \int_1^m \frac{2}{7} dx = \frac{1}{2} \Leftrightarrow \int_1^m \frac{2}{7} dx = \frac{1}{2} - \frac{1}{7}$$

$$\frac{2}{7} x \Big|_1^m = \frac{5}{14} \Leftrightarrow \frac{2}{7} m - \frac{2}{7} = \frac{5}{14} \Leftrightarrow \frac{2m}{7} = \frac{9}{14} \Leftrightarrow m = \frac{9}{4} = \boxed{2,25}$$

**Bài 9/103: Trọng lượng của một con gà 6 tháng tuổi là một ĐLNN X (đơn vị kg) có hàm mật độ**

$$f(x) = \begin{cases} k(x^2 - 1) & \text{khi } 2 \leq x \leq 3 \\ 0 & \text{khác} \end{cases}$$

Tìm trọng lượng trung bình của con gà 6 tháng tuổi và độ lệch tiêu chuẩn.

Vì  $f(x)$  là hàm mật độ của  $X$  nên suy ra:

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \Leftrightarrow \int_2^3 k(x^2 - 1) dx = 1 \Leftrightarrow k \left( \frac{x^3}{3} - x \right) \Big|_2^3 = 1$$

$$\Leftrightarrow k \left( 9 - 3 - \frac{8}{3} + 2 \right) = 1 \Leftrightarrow k \frac{16}{3} = 1 \Leftrightarrow k = \frac{3}{16}$$

+ Trọng lượng trung bình của gà:

$$\begin{aligned} EX &= \int_{-\infty}^{+\infty} x f(x) dx = \int_2^3 \frac{3}{16} x (x^2 - 1) dx = \frac{3}{16} \left( \frac{x^4}{4} - \frac{x^2}{2} \right) \Big|_2^3 \\ &= \frac{3}{16} \left( \frac{3^4}{4} - \frac{3^2}{2} - \frac{2^4}{4} + \frac{2^2}{2} \right) = \frac{3}{16} \cdot \frac{55}{4} = \frac{165}{64} = \boxed{2,578} \text{ (kg)} \end{aligned}$$

+ Độ lệch tiêu chuẩn của trọng lượng gà:

$$\begin{aligned} EX^2 &= \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_2^3 \frac{3}{16} (x^4 - x^2) dx = \frac{3}{16} \left( \frac{x^5}{5} - \frac{x^3}{3} \right) \Big|_2^3 \\ &= \frac{3}{16} \left( \frac{3^5}{5} - \frac{3^3}{3} - \frac{2^5}{5} + \frac{2^3}{3} \right) = \frac{3}{16} \cdot \frac{538}{15} = \frac{269}{40} \\ DX &= EX^2 - (EX)^2 = \frac{269}{40} - \left( \frac{165}{64} \right)^2 = 0,0783 \end{aligned}$$

⇒ Độ lệch tiêu chuẩn cần tìm:

$$\sigma_x = \sqrt{DX} = \boxed{0,27977} \text{ (kg)}$$

**Bài 10/104: Diện tích của một chiếc lá của một loài cây nào đó là một ĐLNN X (đơn vị cm<sup>2</sup>) với hàm mật độ**

$$f(x) = \begin{cases} kx^2(x-2)^2 & \text{khi } 0 \leq x \leq 2 \\ 0 & \text{khác} \end{cases}$$

a) Xác định k và vẽ đồ thị của f(x)

b) Tìm kỳ vọng và phương sai của X.

a) Xác định k và vẽ đồ thị của f(x):

+ Vì f(x) là hàm mật độ của X nên suy ra:

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x) dx &= 1 \Leftrightarrow \int_0^2 kx^2(x-2)^2 dx = 1 \Leftrightarrow k \left( \frac{x^5}{5} - x^4 + \frac{4x^3}{3} \right) \Big|_0^2 = 1 \\ &\Leftrightarrow k \left( \frac{32}{5} - 16 + \frac{32}{3} \right) = 1 \Leftrightarrow k \frac{16}{15} = 1 \Leftrightarrow \boxed{k = \frac{15}{16}} \end{aligned}$$

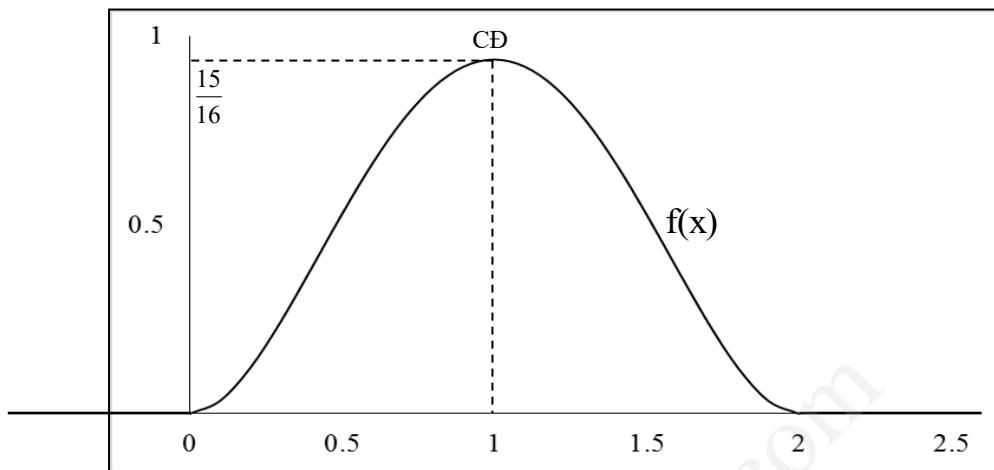
+ Đồ thị của hàm số f(x):

Tìm điểm cực đại của f(x):

$$f'(x) = 0 \Leftrightarrow (x^4 - 4x^3 + 4x^2)' = 0 \Leftrightarrow 4x^3 - 12x^2 + 8x = 0 \Leftrightarrow \begin{cases} x = 0 \\ x = 1 \\ x = 2 \end{cases}$$

$$\Rightarrow CD \left( 1, \frac{15}{16} \right)$$

$\Rightarrow$  Đồ thị hàm số  $f(x)$  như sau:



b) Kỳ vọng và phương sai của  $X$ :

+ Kỳ vọng của X:

$$\begin{aligned} EX &= \int_{-\infty}^{+\infty} xf(x) dx = \int_0^2 \frac{15}{16} (x^5 - 4x^4 + 4x^3) dx = \frac{15}{16} \left( \frac{x^6}{6} - \frac{4x^5}{5} + x^4 \right) \Big|_0^2 \\ &= \frac{15}{16} \left( \frac{2^6}{6} - \frac{4 \cdot 2^5}{5} + 2^4 \right) = \boxed{1} \text{ (cm}^2\text{)} \end{aligned}$$

+ Phương sai của X:

$$\begin{aligned} EX^2 &= \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^2 \frac{15}{16} (x^6 - 4x^5 + 4x^4) dx \\ &= \frac{15}{16} \left( \frac{x^7}{7} - \frac{4x^6}{6} + \frac{4x^5}{5} \right) \Big|_0^2 = \frac{15}{16} \left( \frac{2^7}{7} - \frac{4 \cdot 2^6}{6} + \frac{4 \cdot 2^5}{5} \right) = \frac{8}{7} \end{aligned}$$

$$\Rightarrow DX = EX^2 - (EX)^2 = \frac{8}{7} - 1^2 = \boxed{\frac{1}{7}}$$

**Bài 11/104: Cho ĐLNN X có hàm mật độ**

$$f(x) = \begin{cases} kx^{-3/2} & \text{khi } x \geq 1 \\ 0 & \text{khi } x < 1 \end{cases}$$

a) Tìm  $k$  và hàm phân bố  $F(x)$

b) Tìm hàm mật độ của  $Y = 1/X$

c) Tính  $P(0,1 < Y < 0,2)$

a) Tìm k và hàm phân bố  $F(x)$ :

+ Vì  $f(x)$  là hàm mật độ của X nên suy ra:

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \Leftrightarrow \int_0^{+\infty} kx^{-3/2} dx = 1 \Leftrightarrow k \left( \frac{x^{-1/2}}{-1/2} \right) \Big|_0^{+\infty} = 1$$

$$\Leftrightarrow -2k(0-1) = 1 \Leftrightarrow 2k = 1 \Leftrightarrow k = \boxed{\frac{1}{2}}$$

+ Hàm phân bố của X có dạng:  $F(x) = P(X < x)$

Nếu  $x < 1$ :  $P(X < x) = 0$

$$\text{Nếu } x \geq 1: P(X < x) = \int_1^x \frac{1}{2} t^{-3/2} dt = \frac{1}{2} \left[ \frac{t^{-1/2}}{-1} \right]_1^x = 1 - \frac{1}{\sqrt{x}}$$

Vậy: 
$$F(x) = \begin{cases} 1 - \frac{1}{\sqrt{x}} & \text{khi } x \geq 1 \\ 0 & \text{khi } x < 1 \end{cases}$$

b) Tìm hàm mật độ của  $Y = \frac{1}{X}$ :

+ Tìm hàm phân bố của Y:

$$F_Y(x) = P(Y < x) = P\left(\frac{1}{X} < x\right) = P\left(X > \frac{1}{x}\right) = 1 - P\left(X < \frac{1}{x}\right)$$

- Nếu  $\frac{1}{x} < 1 \Leftrightarrow \begin{cases} x > 1 \\ x < 0 \end{cases}$  thì:  $P\left(X < \frac{1}{x}\right) = 0 \Leftrightarrow 1 - P\left(X < \frac{1}{x}\right) = 1$

- Nếu  $\frac{1}{x} \geq 1 \Leftrightarrow \begin{cases} x > 0 \\ x \leq 1 \end{cases}$  thì:  $1 - P\left(X < \frac{1}{x}\right) = 1 - \int_1^{\frac{1}{x}} \frac{1}{2} t^{-3/2} dt = 1 + t^{-\frac{1}{2}} \Big|_1^{\frac{1}{x}} = 1 + \sqrt{x} - 1 = \sqrt{x}$

Hàm phân bố của Y là:

$$F_Y(x) = \begin{cases} \sqrt{x} & \text{khi } x \in (0,1] \\ 1 & \text{khi } x \notin (0,1] \end{cases}$$

$\Rightarrow$  Hàm mật độ của Y:

$$f_Y(x) = F_Y'(x) = \begin{cases} \frac{1}{2\sqrt{x}} & \text{khi } x \in (0,1] \\ 0 & \text{khi } x \notin (0,1] \end{cases}$$

c) Tính  $P(0,1 < Y < 0,2)$ :

$$P(0,1 < Y < 0,2) = \int_{0,1}^{0,2} \frac{1}{2\sqrt{x}} dx = \sqrt{x} \Big|_{0,1}^{0,2} = \sqrt{0,2} - \sqrt{0,1} = \frac{\sqrt{2}-1}{\sqrt{10}} = \boxed{0,131}$$

### Bài 12/104: ĐLNN X có hàm mật độ

$$f(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{khi } |x| \leq 1 \\ 0 & \text{khác} \end{cases}$$

Tìm kỳ vọng và phương sai của  $Y = 2X^2$

+ Kỳ vọng của Y:

$$\begin{aligned} EY &= \int_{-\infty}^{+\infty} g(x) f(x) dx \quad (\text{trong đó } g(x) \text{ là hàm của } Y \text{ theo } X) \\ &= \int_{-1}^1 2x^2 \frac{3}{4}(1-x^2) dx = \frac{3}{2} \int_{-1}^1 (x^2 - x^4) dx = \frac{3}{2} \left( \frac{x^3}{3} - \frac{x^5}{5} \right) \Big|_{-1}^1 \\ &= \frac{3}{2} \left( \frac{1}{3} - \frac{1}{5} + \frac{1}{3} - \frac{1}{5} \right) = \boxed{\frac{2}{5}} \end{aligned}$$

+ Phương sai của Y:

$$\begin{aligned} EY^2 &= \int_{-\infty}^{+\infty} g^2(x) f(x) dx = \int_{-1}^1 \frac{3}{4} 4x^4 (1-x^2) dx = 3 \int_{-1}^1 (x^4 - x^6) dx \\ &= 3 \left( \frac{x^5}{5} - \frac{x^7}{7} \right) \Big|_{-1}^1 = 3 \left( \frac{1}{5} - \frac{1}{7} + \frac{1}{5} - \frac{1}{7} \right) = \frac{12}{35} \\ \Rightarrow DY &= EY^2 - (EY)^2 = \frac{12}{35} - \left( \frac{2}{5} \right)^2 = \boxed{\frac{32}{175}} \end{aligned}$$

### Bài 13/104: Bán kính X của một vòng tròn có phân bố đều trên đoạn $[0, a]$ . Tìm kỳ vọng và độ lệch tiêu chuẩn của diện tích A của vòng tròn.

Hàm mật độ của X:

$$f(x) = \begin{cases} \frac{1}{a} & \text{khi } x \in [0, a] \\ 0 & \text{khi } x \notin [0, a] \end{cases}$$

ĐLNN liên tục Y là diện tích của vòng tròn nên Y là hàm của X:

$$Y = \pi X^2$$

+ Kỳ vọng của Y:

$$EY = \int_{-\infty}^{+\infty} g(x) f(x) dx \quad (\text{trong đó } g(x) \text{ là hàm của } Y \text{ theo } X)$$

$$= \int_0^a \frac{1}{a} \pi x^2 dx = \frac{\pi}{a} \int_0^a x^2 dx = \frac{\pi}{a} \left( \frac{x^3}{3} \right) \Big|_0^a = \frac{\pi}{a} \cdot \frac{a^3}{3} = \boxed{\frac{\pi a^2}{3}}$$

+ Độ lệch tiêu chuẩn của Y:

$$EY^2 = \int_{-\infty}^{+\infty} g^2(x) f(x) dx = \int_0^a \frac{1}{a} \pi^2 x^4 dx = \frac{\pi^2}{a} \int_0^a x^4 dx = \frac{\pi^2}{a} \left( \frac{x^5}{5} \right) \Big|_0^a = \frac{\pi^2 a^4}{5}$$

$$DY = EY^2 - (EY)^2 = \frac{\pi^2 a^4}{5} - \left( \frac{\pi a^2}{3} \right)^2 = \frac{4\pi^2 a^4}{45}$$

$$\Rightarrow \sigma_Y = \sqrt{DY} = \boxed{\frac{2}{3\sqrt{5}} \pi a^2}$$

**Bài 14/104: Cho ĐLNN X có hàm mật độ**

$$f(x) = \begin{cases} 3x^2 & \text{khi } x \in [0,1] \\ 0 & \text{khi } x \notin [0,1] \end{cases}$$

Xét  $Y = 2\sqrt{X}$ . Tìm:

a)  $P(0,5 < Y < 1,5)$

b)  $P(Y > 1)$ .

a) Tính  $P(0,5 < Y < 1,5)$ :

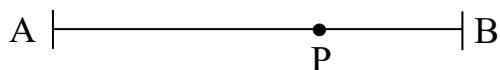
$$P(0,5 < Y < 1,5) = P(0,5 < 2\sqrt{X} < 1,5) = P\left(\left(\frac{0,5}{2}\right)^2 < X < \left(\frac{1,5}{2}\right)^2\right)$$

$$= P(0,0625 < X < 0,5625) = \int_{0,0625}^{0,5625} 3x^2 dx = x^3 \Big|_{0,0625}^{0,5625} = 0,5625^3 - 0,0625^3 = \boxed{0,1777}$$

b) Tính  $P(Y > 1)$ :

$$P(Y > 1) = P(2\sqrt{X} > 1) = P(X > 0,25) = \int_{0,25}^1 3x^2 dx = x^3 \Big|_{0,25}^1 = 1 - 0,25^3 = \boxed{0,9844}$$

**Bài 15/104: Một đoạn thẳng AB dài 10cm bị gãy ngẫu nhiên ở mọi điểm P. Hai đoạn AP và PB được sử dụng để làm hai cạnh của một hình chữ nhật. Tìm kỳ vọng và độ lệch tiêu chuẩn của diện tích hình chữ nhật.**



Hoàng Văn Trọng – 0974.971.149

Gọi X là độ dài đoạn AP. X là ĐLNN liên tục có phân bố đều trên đoạn [0, 10].  
Hàm mật độ của X là:

$$f(x) = \begin{cases} \frac{1}{10} & \text{khi } x \in [0,10] \\ 0 & \text{khi } x \notin [0,10] \end{cases}$$

Gọi Y là diện tích hình chữ nhật được tạo nên từ hai đoạn AP và PB thì Y là hàm của X:

$$Y = X(10 - X)$$

+ Kỳ vọng của diện tích HCN:

$$EY = \int_{-\infty}^{+\infty} g(x)f(x) dx \quad (\text{trong đó } g(x) \text{ là hàm của } Y \text{ theo } X)$$

$$\begin{aligned} &= \int_0^{10} \frac{1}{10} x(10 - x) dx = \frac{1}{10} \int_0^{10} (10x - x^2) dx = \frac{1}{10} \left( 5x^2 - \frac{x^3}{3} \right) \Big|_0^{10} = \frac{1}{10} \left( 500 - \frac{1000}{3} \right) \\ &= \frac{50}{3} = \boxed{16,67} \quad (\text{cm}^2) \end{aligned}$$

+ Độ lệch tiêu chuẩn của diện tích HCN:

$$\begin{aligned} EY^2 &= \int_{-\infty}^{+\infty} g^2(x)f(x) dx = \int_0^{10} \frac{1}{10} x^2 (10 - x)^2 dx = \frac{1}{10} \int_0^{10} (100x^2 - 20x^3 + x^4) dx \\ &= \frac{1}{10} \left( \frac{100x^3}{3} - 5x^4 + \frac{x^5}{5} \right) \Big|_0^{10} = \frac{1000}{10} \left( \frac{100}{3} - 50 + \frac{100}{5} \right) = \frac{1000}{3} \end{aligned}$$

$$DY = EY^2 - (EY)^2 = \frac{1000}{3} - \left( \frac{50}{3} \right)^2 = \frac{500}{9}$$

$$\Rightarrow \sigma_Y = \sqrt{DY} = \sqrt{\frac{500}{9}} = \boxed{\frac{10\sqrt{5}}{3}} \quad (\text{cm}^2)$$

**Bài 16/105:** Một người hàng ngày đi bộ từ nhà đến nơi làm việc với quãng đường 600m với vận tốc đều v (m/giây). Biết rằng thời gian đi bộ của người đó là một ĐLNN phân bố đều trong khoảng từ 6 phút đến 10 phút:

- a) Tìm kỳ vọng và độ lệch tiêu chuẩn của vận tốc v
- b) Tìm median của vận tốc v.

Gọi X là thời gian đi bộ (phút) của người đó tới nơi làm việc. Hàm mật độ của X:

$$f(x) = \begin{cases} \frac{1}{4} & \text{khi } x \in [6,10] \\ 0 & \text{khi } x \notin [6,10] \end{cases}$$

Gọi Y là vận tốc đều (m/s) trên quãng đường từ nhà tới nơi làm việc, ta có:

$$Y = \frac{600}{60X} = \frac{10}{X}$$

a) Kỳ vọng và độ lệch tiêu chuẩn của Y:

+ Kỳ vọng của Y:

$$EY = \int_{-\infty}^{+\infty} g(x) f(x) dx \quad (\text{với } g(x) \text{ là hàm của } Y \text{ theo } X)$$

$$= \int_6^{10} \frac{1}{4} \frac{10}{x} dx = \frac{5}{2} \int_6^{10} \frac{dx}{x} = \frac{5}{2} \ln x \Big|_6^{10} = \frac{5}{2} \ln \frac{10}{6} = [1,277] \text{ (m/s)}$$

+ Độ lệch tiêu chuẩn của Y:

$$EY^2 = \int_{-\infty}^{+\infty} g^2(x) f(x) dx = \int_6^{10} \frac{1}{4} \left( \frac{10}{x} \right)^2 dx = 25 \int_6^{10} \frac{dx}{x^2} = 25 \left( -\frac{1}{x} \right) \Big|_6^{10} = 25 \left( \frac{1}{6} - \frac{1}{10} \right) = \frac{5}{3}$$

$$DY = EY^2 - (EY)^2 = \frac{5}{3} - \left( \frac{5}{2} \ln \frac{10}{6} \right)^2 = 0,03577$$

$$\Rightarrow \sigma_Y = \sqrt{DY} = [0,1891] \text{ (m/s)}$$

**Bài 17/105:** Trọng lượng của một con bò là một ĐLNN phân bố chuẩn với kỳ vọng là 250kg và độ lệch tiêu chuẩn 40kg. Tìm xác suất để một con bò có trọng lượng:

- a) Nặng hơn 300kg;
- b) Nhẹ hơn 175kg;
- c) Trong khoảng 260kg đến 270kg.

Theo giả thiết:  $X \sim N(250; 40^2)$

a) Xác suất để con bò nặng hơn 300 kg:

$$\begin{aligned} P(X > 300) &= 1 - P(X < 300) = 1 - P\left(\frac{X - 250}{40} < \frac{300 - 250}{40}\right) \\ &= 1 - P\left(Z < \frac{50}{40}\right) \quad \text{với } Z = \frac{X - 250}{40} \text{ là đại lượng có phân bố chuẩn tắc.} \\ &= 1 - P(Z < 1,25) = 1 - \Phi(1,25) = 1 - 0,8944 = [0,1056] \end{aligned}$$

b) Xác suất để con bò nhẹ hơn 175 kg:

$$\begin{aligned} P(X < 175) &= P\left(\frac{X - 250}{40} < \frac{175 - 250}{40}\right) = \Phi\left(\frac{175 - 250}{40}\right) = \Phi(-1,875) \\ &= 1 - \Phi(1,875) = 1 - 0,9697 = [0,0303] \end{aligned}$$

c) Xác suất để con bò nằm trong khoảng 260 đến 270 kg:

$$\begin{aligned} P(260 < X < 270) &= P\left(\frac{260 - 250}{40} < \frac{X - 250}{40} < \frac{270 - 250}{40}\right) \\ &= \Phi\left(\frac{270 - 250}{40}\right) - \Phi\left(\frac{260 - 250}{40}\right) = \Phi(0,5) - \Phi(0,25) = 0,6915 - 0,5987 \\ &= \boxed{0,0928} \end{aligned}$$

**Bài 18/105:** Thời gian từ nhà đi đến trường của sinh viên Bình là một ĐLNN X có phân bố chuẩn. Biết rằng 65% số ngày Bình đến trường mất hơn 20 phút còn 8% số ngày mất hơn 30 phút.

- a) Tìm thời gian trung bình và độ lệch tiêu chuẩn của thời gian đến trường.
- b) Nếu Bình xuất phát từ nhà trước giờ vào học 25 phút thì xác suất để Bình muộn học là bao nhiêu.
- c) Bình cần phải xuất phát trước giờ vào học bao nhiêu phút để khả năng bị muộn học là bé hơn 0,02.

Theo giả thiết:

$$X \sim N(\mu, \sigma^2)$$

$$P(X > 20) = 0,65;$$

$$P(X > 30) = 0,08$$

a) Tìm thời gian trung bình và độ lệch tiêu chuẩn của thời gian đến trường:

Ta có:

$$\begin{aligned} + P(X > 20) = 0,65 &\Leftrightarrow P(X < 20) = 0,35 \Leftrightarrow P\left(\frac{X - \mu}{\sigma} < \frac{20 - \mu}{\sigma}\right) = 0,35 \\ &\Leftrightarrow \Phi\left(\frac{20 - \mu}{\sigma}\right) = 0,35 \end{aligned}$$

Tra bảng phân bố chuẩn tắc (bảng Phụ lục P.4.3, trang 202) ta có:

$$\Phi(0,385) = 0,65 \Rightarrow \Phi(-0,385) = 0,35$$

$$\Rightarrow \frac{20 - \mu}{\sigma} = -0,385 \quad (1)$$

$$\begin{aligned} + P(X > 30) = 0,08 &\Leftrightarrow P(X < 30) = 0,92 \Leftrightarrow P\left(\frac{X - \mu}{\sigma} < \frac{30 - \mu}{\sigma}\right) = 0,92 \\ &\Leftrightarrow \Phi\left(\frac{30 - \mu}{\sigma}\right) = 0,92 \end{aligned}$$

Tra bảng phân bố chuẩn ta có:  $\Phi(1,406) = 0,92$

$$\Rightarrow \frac{30-\mu}{\sigma} = 1,406 \quad (2)$$

Từ (1) và (2) ta có hệ hai phương trình bậc nhất 2 ẩn:

$$\begin{cases} \frac{20-\mu}{\sigma} = -0,385 \\ \frac{30-\mu}{\sigma} = 1,406 \end{cases} \Leftrightarrow \begin{cases} -\mu + 0,385\sigma = -20 \\ -\mu - 1,406\sigma = -30 \end{cases} \Leftrightarrow \boxed{\begin{cases} \mu = 22,15 \\ \sigma = 5,58 \end{cases}}$$

Vậy, thời gian trung bình là 22,15 phút và độ lệch tiêu chuẩn là 5,58 phút.

b) Xác suất muộn học nếu Bình xuất phát trước giờ vào học 25 phút:

Bình bị muộn học khi đi lâu hơn 25 phút. Xác suất muộn học là:

$$\begin{aligned} P(X > 25) &= 1 - P(X < 25) = 1 - P\left(\frac{X - 22,15}{5,58} < \frac{25 - 22,15}{5,58}\right) \\ &= 1 - \Phi\left(\frac{25 - 22,15}{5,58}\right) = 1 - \Phi(0,51) = 1 - 0,695 = \boxed{0,305} \end{aligned}$$

c) Xuất phát trước giờ vào học bao nhiêu phút để xác suất muộn học bé hơn 0,02:

Giả sử Bình xuất phát trước giờ vào học t phút. Xác suất bị muộn học là:

$$P(X > t) = 1 - P(X < t) = 1 - P\left(\frac{X - 22,15}{5,58} < \frac{t - 22,15}{5,58}\right) = 1 - \Phi\left(\frac{t - 22,15}{5,58}\right)$$

Xác suất trên nhỏ hơn 0,02 nên ta có:

$$1 - \Phi\left(\frac{t - 22,15}{5,58}\right) < 0,02 \Leftrightarrow \Phi\left(\frac{t - 22,15}{5,58}\right) > 0,98$$

Tra bảng phân bố chuẩn ta thấy:  $\Phi(2,055) = 0,98$ . Do đó:

$$\frac{t - 22,15}{5,58} > 2,055 \Leftrightarrow t - 22,15 > 11,4669 \Leftrightarrow \boxed{t > 33,62}$$

Vậy, để khả năng bị muộn học bé hơn 0,02 thì Bình phải xuất phát trước giờ vào học nhiều hơn 33,62 phút.

**Bài 19/105:** Chiều dài của một loài cây là một ĐLNN có phân bố chuẩn. Trong một mẫu gồm 640 cây có 25 cây thấp hơn 18m và 110 cây cao hơn 24m.

- a) Tìm kỳ vọng và độ lệch tiêu chuẩn của chiều cao của cây,
- b) Ước lượng số cây có độ cao trong khoảng từ 16m đến 20m trong mẫu nói trên.

Gọi X là chiều cao cây:  $X \sim N(\mu, \sigma^2)$

Ta có:  $P(X < 18) = 25/640 = 0,0391$ ;  $P(X > 24) = 110/640 = 0,1719$

a) Tìm kỳ vọng và độ lệch tiêu chuẩn của chiều cao cây:

$$+ P(X < 18) = 0,0391 \Leftrightarrow P\left(\frac{X-\mu}{\sigma} < \frac{18-\mu}{\sigma}\right) = 0,0391 \Leftrightarrow \Phi\left(\frac{18-\mu}{\sigma}\right) = 0,0391$$

Tra bảng phân bố chuẩn ta thấy:  $\Phi(1,761) = 0,9609 \Rightarrow \Phi(-1,761) = 0,0391$

$$\Rightarrow \frac{18-\mu}{\sigma} = -1,761 \quad (1)$$

$$+ P(X > 24) = 0,1719 \Leftrightarrow P(X < 24) = 0,8281 \Leftrightarrow P\left(\frac{X-\mu}{\sigma} < \frac{24-\mu}{\sigma}\right) = 0,8281$$

$$\Leftrightarrow \Phi\left(\frac{24-\mu}{\sigma}\right) = 0,8281. \text{ Tra bảng ta thấy: } \Phi(0,947) = 0,8281$$

$$\Rightarrow \frac{24-\mu}{\sigma} = 0,947 \quad (2)$$

Từ (1) và (2) ta có hệ hai phương trình bậc nhất 2 ẩn:

$$\begin{cases} \frac{18-\mu}{\sigma} = -1,761 \\ \frac{24-\mu}{\sigma} = 0,947 \end{cases} \Leftrightarrow \begin{cases} -\mu + 1,761\sigma = -18 \\ -\mu - 0,947\sigma = -24 \end{cases} \Leftrightarrow \boxed{\begin{cases} \mu = 21,9 \\ \sigma = 2,216 \end{cases}}$$

Vậy, kỳ vọng của chiều cao cây là 21,9 m và độ lệch tiêu chuẩn của chiều cao cây là 2,216 m.

b) *Uớc lượng số cây có độ cao trong khoảng từ 16m đến 20m trong mẫu trên:*

Tỷ lệ cây có chiều cao từ 16m đến 20m:

$$\begin{aligned} P(16 < X < 20) &= P\left(\frac{16-21,9}{2,216} < \frac{X-21,9}{2,216} < \frac{20-21,9}{2,216}\right) \\ &= \Phi\left(\frac{20-21,9}{2,216}\right) - \Phi\left(\frac{16-21,9}{2,216}\right) = \Phi(-0,857) - \Phi(-2,662) \\ &= 1 - \Phi(0,857) - 1 + \Phi(2,662) = \Phi(2,662) - \Phi(0,857) = 0,9961 - 0,8042 \\ &= 0,1919 \end{aligned}$$

$\Rightarrow$  Số cây ước lượng:  $640 \cdot 0,1919 = 122,8$ . Vậy số cây ước lượng là 123 cây.

**Bài 20/105: Cho  $X$  là ĐLNN có phân bố mũ với tham số  $\lambda = 2$ . Tìm kỳ vọng và độ lệch tiêu chuẩn của  $e^{-X}$ .**

Theo bài ra ta có hàm mật độ của  $X$ :

$$f(x) = \begin{cases} 2e^{-2x} & \text{khi } x \geq 0 \\ 0 & \text{khi } x < 0 \end{cases}$$

+ Kỳ vọng của Y = e<sup>-X</sup>

$$EY = \int_{-\infty}^{+\infty} g(x) f(x) dx \quad (\text{với } g(x) \text{ là hàm của } Y \text{ theo } X)$$

$$= \int_0^{+\infty} 2e^{-2x} e^{-x} dx = 2 \int_0^{+\infty} e^{-3x} dx = -\frac{2}{3} e^{-3x} \Big|_0^{+\infty} = -\frac{2}{3} (0 - 1) = \boxed{\frac{2}{3}}$$

+ Độ lệch tiêu chuẩn của Y:

$$\begin{aligned} EY^2 &= \int_{-\infty}^{+\infty} g^2(x) f(x) dx = \int_0^{+\infty} 2e^{-2x} e^{-2x} dx = 2 \int_0^{+\infty} e^{-4x} dx = -\frac{1}{2} (e^{-4x}) \Big|_0^{+\infty} \\ &= -\frac{1}{2} (0 - 1) = \frac{1}{2} \end{aligned}$$

$$DY = EY^2 - (EY)^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

$$\Rightarrow \sigma_Y = \sqrt{DY} = \boxed{0,2357}$$

**Bài 21/105: Cho X là ĐLNN có phân bố mũ với tham số  $\lambda = 1$  và  $Y = 2X^2$ . Tính:**

a)  $P(2 < Y < 18)$

b)  $P(Y < 4)$

Theo bài ra ta có hàm mật độ của X:

$$f(x) = \begin{cases} e^{-x} & \text{khi } x \geq 0 \\ 0 & \text{khi } x < 0 \end{cases}$$

a) *Tính  $P(2 < Y < 18)$ :*

$$P(2 < Y < 18) = P(2 < 2X^2 < 18) = P(1 < X < 3)$$

$$= \int_1^3 e^{-x} dx = -e^{-x} \Big|_1^3 = \frac{1}{e} - \frac{1}{e^3} = \boxed{0,3181}$$

b) *Tính  $P(Y < 4)$ :*

$$P(Y < 4) = P(2X^2 < 4) = P(X < \sqrt{2}) = \int_0^{\sqrt{2}} e^{-x} dx$$

$$= -e^{-x} \Big|_0^{\sqrt{2}} = 1 - \frac{1}{e^{\sqrt{2}}} = \boxed{0,7569}$$

### 3.2. Nhận xét bài tập chương 3

Chương 3 liên quan đến phân bố xác suất của đại lượng ngẫu nhiên liên tục. Từ hàm mật độ xác suất tính các đặc trưng của ĐLNN (kỳ vọng, phương sai, độ lệch chuẩn, mode, median).

Bài toán thường cho trước hàm mật độ và yêu cầu tìm các đặc trưng. Nếu không cho trước hàm mật độ thì có thể suy ra từ thông tin đề bài cho hoặc thông qua một đại lượng ngẫu nhiên liên tục khác.

Một số dạng bài tập chính của chương 3:

- + Dạng bài về phân bố chuẩn (dạng quan trọng nhất của chương này, đặc biệt phân bố chuẩn còn được sử dụng rất nhiều trong phần Thống kê): yêu cầu sử dụng thành thạo cách tra bảng phân bố chuẩn tắc (bảng *Phụ lục P.4.3*, trang 202), tìm xác suất để ĐLNN X rơi vào một khoảng nào đó, ...

- + Dạng bài về phân bố mũ.

- + Dạng bài về phân bố đều: dạng này đơn giản nên thường không ra trong đề thi.

- + Dạng bài về phân bố khác: cho hàm mật độ có chứa tham số c. Yêu cầu tìm c, mod, median, tính xác suất để ĐLNN rơi vào một khoảng nào đó.

- + Dạng bài về hàm của ĐLNN: cho hàm mật độ của ĐLNN liên tục X, tìm kỳ vọng và phương sai của ĐLNN liên tục Y = g(X). Không nên tìm hàm mật độ của Y để suy ra các đặc trưng mà tính trực tiếp bằng định lý:

- Kỳ vọng:

$$EY = \int_{-\infty}^{+\infty} g(x) f(x) dx \quad (\text{với } f(x) \text{ là hàm mật độ của } X)$$

- Phương sai:

$$DY = \int_{-\infty}^{+\infty} g^2(x) f(x) dx - (EY)^2$$

## PHẦN II: THỐNG KÊ

### CHƯƠNG 4: BÀI TOÁN ƯỚC LƯỢNG THAM SỐ

#### A. LÝ THUYẾT

##### 4.1. Một số kiến thức chuẩn bị thêm cho phần thống kê

Bản chất của các công thức trong phần thống kê như: ước lượng tham số, kiểm định giả thiết, tương quan và hồi quy,... đều dựa trên cơ sở của lý thuyết xác suất. Thống kê là điều tra một mẫu đại diện cho toàn bộ tổng thể nên kết luận được rút ra có thể đúng, có thể sai với một xác suất nào đó. Trong mục 4.1 này sẽ giới thiệu thêm về định lý giới hạn trung tâm và một số phân bố của đại lượng ngẫu nhiên liên tục được sử dụng nhiều trong phần thống kê.

a) *Định lý giới hạn trung tâm*<sup>1</sup>:

Cho  $X_1, X_2, \dots, X_n$  là dãy các đại lượng ngẫu nhiên độc lập có cùng phân bố (không nhất thiết phải cùng phân bố chuẩn, chỉ cần cùng kỳ vọng  $\mu$  và phương sai  $\sigma^2$ ) thì đại lượng ngẫu nhiên  $\bar{X}$  sẽ hội tụ tới phân bố chuẩn với kỳ vọng  $\mu$  và phương sai  $\sigma^2/n$  khi  $n \rightarrow +\infty$ .

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ khi } n \rightarrow +\infty \quad \text{trong đó: } \bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Người ta thường lấy  $n \geq 30$  thì  $\bar{X}$  có phân bố chuẩn với các tham số như trên.

b) *Phân bố Student*<sup>2</sup>:

ĐLNN X được gọi là có phân bố Student nếu hàm mật độ của X có dạng:

$$f(x, k) = \frac{1}{\sqrt{k\pi}} \frac{G\left(\frac{k+1}{2}\right)}{G\left(\frac{k}{2}\right)} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}$$

Trong đó:  $\begin{cases} -\infty < x < +\infty \\ k: \text{số bậc tự do của phân bố Student} \\ G(t) \text{ là hàm Gamma: } G(t) = \int_0^\infty x^{t-1} e^{-x} dx \end{cases}$

$\Rightarrow$  Ứng với mỗi bậc tự do k, ta có một hàm mật độ  $f(x, k)$  tương ứng:

$$+ k = 1 \Rightarrow f(x, 1) = 0,3183 (1 + x^2)^{-1}$$

$$+ k = 2 \Rightarrow f(x, 2) = 0,36754 \left(1 + \frac{x^2}{2}\right)^{-3/2}$$

<sup>1</sup> Ngoài định lý này còn có một số định lý giới hạn khác có ý nghĩa thực tiễn: xấp xỉ phân bố phân bố nhị thức bằng phân bố Poisson, xấp xỉ phân bố nhị thức bằng phân bố chuẩn (xem giáo trình G1 trang 165 và 168)

<sup>2</sup> Xem thêm hàm mật độ cho biết, lưu ý tính chất của các phân bố này.

$$+ k \rightarrow \infty \Rightarrow f(x, k) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (\text{phân bố Student tiến tới phân bố chuẩn tắc})$$

\* Tính chất:

Cho  $X_1, X_2, \dots, X_n$  là dãy các величин độc lập có cùng phân bố chuẩn  $N(\mu, \sigma^2)$  thì величин ngẫu nhiên:

$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \quad \text{có phân bố Student với } (n - 1) \text{ bậc tự do.}$$

Trong đó:  $\bar{X} = \frac{\sum X_i}{n}$   
 $s = \sqrt{\frac{n-1}{n} \sum \frac{(X_i - \bar{X})^2}{n}}$  là phương sai hiệu chỉnh của mẫu

Từ tính chất này, suy ra bài toán ước lượng và kiểm định giả thiết trong trường hợp biết phân bố chuẩn nhưng chưa biết phương sai lý thuyết.

c) *Phân bố Khi bình phương:*

ĐLNN  $X$  được gọi là có phân bố Khi bình phương nếu hàm mật độ của  $X$  có dạng:

$$f(x, k) = \frac{2^{1-\frac{k}{2}} x^{k-1} e^{-\frac{x^2}{2}}}{G\left(\frac{k}{2}\right)}$$

Trong đó:  $\begin{cases} 0 \leq x < +\infty \\ k: số bậc tự do của phân bố Khi bình phương \end{cases}$   
 $G(t)$  là hàm Gamma:  $G(t) = \int_0^\infty x^{t-1} e^{-x} dx$

\* Tính chất:

+ Cho  $X_1, X_2, \dots, X_n$  là dãy các величин độc lập có cùng phân bố chuẩn  $N(0, 1)$  thì величин ngẫu nhiên:

$$Q = \sum_{i=1}^n X_i^2 \quad \text{có phân bố Khi bình phương với } n \text{ bậc tự do: } Q \sim \chi^2(n)$$

+ Nếu các ĐLNN  $X_1, X_2, \dots, X_n$  độc lập và có cùng phân bố chuẩn  $N(\mu, \sigma^2)$  thì  
 $DLNN Q = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2}$  có phân bố Khi bình phương với  $(n - 1)$  bậc tự do.

$$(với \bar{X} = \frac{\sum X_i}{n})$$

d) *Phân bố Fisher:*

ĐLNN X được gọi là có phân bố Fisher nếu hàm mật độ của X có dạng:

$$f(x, m, n) = \frac{1}{B\left(\frac{m}{2}; \frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}$$

Trong đó:  $\begin{cases} 0 \leq x < +\infty \\ m, n: \text{bậc tự do của phân bố Fisher} \\ B(m, n) \text{ là hàm Beta: } B(m, n) = \int_0^1 x^{m-1} (1-x)^{n-1} dx \end{cases}$

\* Tính chất:

+ Cho  $X_1, \dots, X_m; Y_1, \dots, Y_n$  là dãy các đại lượng ngẫu nhiên độc lập có cùng phân bố chuẩn tắc  $N(0, 1)$  thì đại lượng ngẫu nhiên:

$$F = \frac{\frac{1}{m} \sum_{i=1}^m X_i^2}{\frac{1}{n} \sum_{j=1}^n Y_j^2} \text{ có phân bố Fisher với } m; n \text{ bậc tự do.}$$

+ Nếu các ĐLNN  $X_1, \dots, X_m$  độc lập và có cùng phân bố chuẩn  $N(\mu_1, \sigma_1^2)$ ; các ĐLNN  $Y_1, \dots, Y_n$  độc lập và có cùng phân bố chuẩn  $N(\mu_2, \sigma_2^2)$  thì:

$$F = \frac{\hat{s}_1^2/\sigma_1^2}{\hat{s}_2^2/\sigma_2^2} \text{ có phân bố Fisher với } (m-1); (n-1) \text{ bậc tự do.}$$

Trong đó:  $\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i; \quad \bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j$

$$\hat{s}_1^2 = \frac{m}{m-1} \sum_{i=1}^m \frac{(X_i - \bar{X})^2}{m} \text{ là phương sai hiệu chỉnh của } X.$$

$$\hat{s}_2^2 = \frac{n}{n-1} \sum_{j=1}^n \frac{(Y_j - \bar{Y})^2}{n} \text{ là phương sai hiệu chỉnh của } Y.$$

## 4.2. Mẫu ngẫu nhiên và các đặc trưng của mẫu

a) *Mẫu ngẫu nhiên:*

*Tổng thể* của đối tượng nghiên cứu theo một dấu hiệu nào đó trong một phạm vi nhất định là tập hợp tất cả các phần tử nằm trong phạm vi đó và mang dấu hiệu nghiên cứu. Ví dụ: khi nghiên cứu chiều cao cây trong khu rừng A thì *tổng thể* là tất cả các cây trong khu rừng A đó.

*Mẫu* là một số phần tử của *tổng thể* được chọn ra. Ví dụ: khi nghiên cứu chiều cao cây trong khu rừng A thì *mẫu* là những cây trong ô tiêu chuẩn được chọn ra.

- + Số lượng các phần tử của mẫu nhỏ hơn số lượng các phần tử của tổng thể.
- + Mẫu là thông tin không đầy đủ về đại lượng ngẫu nhiên nên các kết luận được rút ra từ việc nghiên cứu mẫu có thể đúng hoặc sai với một xác suất nào đó.
- + Tiến hành n quan sát độc lập về ĐLNN X ta được tập giá trị ( $x_1, x_2, \dots, x_n$ ). Số n được gọi là cỡ mẫu (hay kích thước mẫu). Có thể coi mẫu ngẫu nhiên cỡ n là n ĐLNN độc lập cùng phân bố với X và ( $x_1, x_2, \dots, x_n$ ) là giá trị cụ thể mà các ĐLNN ( $X_1, X_2, \dots, X_n$ ) nhận. Việc coi như thế để có thể áp dụng các định lý giới hạn (như định lý giới hạn trung tâm) trong một số bài toán ước lượng hoặc kiểm định giả thiết.
- + Mẫu số liệu được biểu diễn dưới dạng điểm hoặc dạng khoảng.

*b) Các đặc trưng của mẫu:*

\* *Kỳ vọng mẫu:*

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (\bar{X} \text{ cũng là một đại lượng ngẫu nhiên})$$

Khi n đủ lớn ( $n \geq 30$ ) thì  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$  với  $\mu$  và  $\sigma^2$  lần lượt là kỳ vọng và phương sai lý thuyết của đại lượng ngẫu nhiên X.

\* *Phương sai mẫu:*

$$s^2 = \sum_{i=1}^n \frac{X_i^2}{n} - \bar{X}^2$$

\* *Phương sai hiệu chỉnh của mẫu:*

$$\hat{s}^2 = \frac{n}{n-1} s^2$$

### 4.3. Ước lượng điểm

- + Ước lượng điểm cho kỳ vọng EX là:  $\bar{X}$
- + Ước lượng điểm **không chêch** cho phương sai DX là:  $\hat{s}^2$
- + Ước lượng điểm cho xác suất  $P(A)$  là:  $p^* = \frac{m}{n}$  (m là số lần xảy ra A, n là cỡ mẫu)
- + Ước lượng điểm cho Median:

- Nếu mẫu số liệu cho ở dạng điểm thì sắp xếp số liệu theo chiều tăng dần và ước lượng điểm cho median là:

$$\Rightarrow \text{Nếu } n \text{ lẻ: } \text{Med}^* = x_{\frac{n+1}{2}}$$

$$\Rightarrow \text{Nếu } n \text{ chẵn: } \text{Med}^* = \frac{1}{2} \left( x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right)$$

- Nếu mẫu số liệu cho ở dạng khoảng: giả sử số liệu thứ n/2 nằm ở khoảng  $(x_k, x_{k+1})$  và trong khoảng này có  $m_k$  quan sát, ước lượng điểm cho median là<sup>1</sup>:

$$\text{Med}^* = x_k + \frac{x_{k+1} - x_k}{m_k} \left( \frac{n}{2} - \sum_{i=1}^{k-1} m_i \right)$$

#### 4.4. Ước lượng khoảng

Khoảng ( $f_1(X_1, X_2, \dots, X_n); f_2(X_1, X_2, \dots, X_n)$ ) được gọi là khoảng ước lượng cho tham số  $a$  với độ tin cậy  $1 - \alpha$  nếu  $a$  thuộc vào khoảng trên với xác suất  $1 - \alpha$  (trong đó  $\alpha$  là mức ý nghĩa, biểu thị khả năng xảy ra sai sót của kết luận khi ước lượng).

$$P(f_1 < a < f_2) = 1 - \alpha$$

+ Xác suất để  $a$  nằm ngoài khoảng  $(f_1, f_2)$  bằng  $\alpha$

+ Khoảng  $(f_1, f_2)$  gọi là khoảng tin cậy.

+ Giá trị  $\frac{f_2 - f_1}{2}$  gọi là sai số (hay độ chính xác) của ước lượng.

a) Ước lượng khoảng cho kỳ vọng, mức ý nghĩa  $\alpha$ :

\* Trường hợp đã biết phương sai lý thuyết ( $DX = \sigma^2$ ),  $X$  chuẩn hoặc  $n \geq 30$ :

$$EX \in \left( \bar{X} - u\left(\frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}}; \bar{X} + u\left(\frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}} \right)$$

+ Trong đó:  $u(\alpha/2) = \Phi^{-1}(1 - \alpha/2)$

+ Sai số của ước lượng:  $\varepsilon = u\left(\frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}}$

\* Trường hợp chưa biết phương sai lý thuyết,  $X$  phải có phân bố chuẩn:

$$EX \in \left( \bar{X} - t_{n-1}\left(\frac{\alpha}{2}\right) \frac{s}{\sqrt{n-1}}; \bar{X} + t_{n-1}\left(\frac{\alpha}{2}\right) \frac{s}{\sqrt{n-1}} \right)$$

+ Sai số của ước lượng:  $\varepsilon = t_{n-1}\left(\frac{\alpha}{2}\right) \frac{s}{\sqrt{n-1}}$

\* Trường hợp chưa biết phương sai lý thuyết, chưa biết  $X$  chuẩn nhưng  $n \geq 30$ :

$$EX \in \left( \bar{X} - u\left(\frac{\alpha}{2}\right) \frac{s}{\sqrt{n-1}}; \bar{X} + u\left(\frac{\alpha}{2}\right) \frac{s}{\sqrt{n-1}} \right)$$

+ Sai số của ước lượng:  $\varepsilon = u\left(\frac{\alpha}{2}\right) \frac{s}{\sqrt{n-1}}$

---

<sup>1</sup> Công thức ước lượng điểm cho median thực chất là công thức nội suy tuyến tính.

b) *Uớc lượng khoảng cho phuong sai của X có phân bố chuẩn, mức ý nghĩa  $\alpha$ :*

$$DX \in \left( \frac{ns^2}{\chi_{n-1}^2(\alpha/2)}; \frac{ns^2}{\chi_{n-1}^2(1-\alpha/2)} \right)$$

$$+ \text{Sai số của ước lượng: } \varepsilon = \frac{1}{2} \left( \frac{ns^2}{\chi_{n-1}^2(\alpha/2)} + \frac{ns^2}{\chi_{n-1}^2(1-\alpha/2)} \right)$$

c) *Uớc lượng khoảng cho tỷ lệ (hay xác suất) với n lớn, p không quá gần 0 hoặc 1; mức ý nghĩa  $\alpha$ :*

Đặt:  $p^* = \frac{m}{n}$ . Khoảng ước lượng cho tỷ lệ là:

$$p \in \left( p^* - u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p^*(1-p^*)}{n}}; p^* + u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p^*(1-p^*)}{n}} \right)$$

$$+ \text{Sai số của ước lượng: } \varepsilon = u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p^*(1-p^*)}{n}}$$

#### 4.5. Số quan sát cần thiết để có sai số (hoặc độ tin cậy) cho trước

Trong mỗi liên hệ giữa sai số, độ tin cậy và cỡ mẫu; yêu cầu tìm một thông số khi biết hai thông số còn lại.

*Ví dụ:* Tìm n để sai số ước lượng khoảng cho kỳ vọng nhỏ hơn 0,2 (trường hợp X có phân bố chuẩn, đã biết phuong sai lý thuyết), mức ý nghĩa  $\alpha$ :

+ Sai số ước lượng khoảng cho kỳ vọng là:

$$\varepsilon = u\left(\frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}}$$

+ Sai số nhỏ hơn 0,2 nên:

$$u\left(\frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}} < 0,2 \Leftrightarrow n > \left( \frac{u(\alpha/2)}{0,2} \sigma \right)^2$$

## B. BÀI TẬP

### 4.1. Bài tập trong giáo trình 2 (G<sub>2</sub>)

(Xác suất thống kê, Đào Hữu Hồ, trang 156)

**Bài 1/156:** Điều tra doanh số hàng tháng của 100 hộ kinh doanh một ngành nào đó ta thu được số liệu sau:

Doanh số X (triệu đồng)	10,1	10,2	10,4	10,5	10,7	10,8	10,9	11	11,3	11,4
Số hộ	2	3	8	13	25	20	12	10	6	1

a) Chỉ ra Median mẫu.

b) Với độ tin cậy 95%, có thể nói doanh số trung bình/tháng của các hộ nằm trong khoảng nào? Giả thiết X tuân theo luật chuẩn.

c) Ước lượng tỷ lệ % các hộ có doanh số/tháng ≥ 11 triệu. Với độ tin cậy 99% tỷ lệ này thấp nhất là bao nhiêu?

a) Chỉ ra Median mẫu:

Cỡ mẫu chẵn ( $n = 100$ ) và được cho dưới dạng điểm nêu median mẫu là:

$$\text{Med}^* = \frac{1}{2} \left( x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right) = \frac{1}{2} (x_{50} + x_{51})$$

Ta thấy, phần tử thứ 50 và 51 của mẫu đều có giá trị bằng 10,7. Do đó:

$$\text{Med}^* = \frac{1}{2} (10,7 + 10,7) = [10,7]$$

b) Ước lượng khoảng cho  $EX$ , mức ý nghĩa  $\alpha = 0,05$  và  $X$  tuân theo luật chuẩn:

Dạng bài ước lượng khoảng cho kỳ vọng, trường hợp chưa biết phương sai lý thuyết,  $X$  có phân bố chuẩn. Khoảng ước lượng như sau:

$$EX \in \left( \bar{X} - t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s}{\sqrt{n-1}}; \quad \bar{X} + t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s}{\sqrt{n-1}} \right)$$

Sử dụng máy tính bỏ túi<sup>1</sup> ta được: (xem hướng dẫn sử dụng máy tính bỏ túi để tính toán các chỉ số thống kê tại Phụ lục P.2.1, trang 191)

$$\bar{X} = 10,74$$

$$s = 0,2604$$

⇒ Ước lượng khoảng cho kỳ vọng:

$$EX \in \left( 10,74 - t_{99} (0,025) \frac{0,2604}{\sqrt{99}}; \quad 10,74 + t_{99} (0,025) \frac{0,2604}{\sqrt{99}} \right)$$

<sup>1</sup> Được phép sử dụng máy tính bỏ túi để tính toán các đại lượng thống kê như: kỳ vọng mẫu, phương sai mẫu, phương sai hiệu chỉnh mẫu,...

$$EX \in \left( 10,74 - 1,984 \cdot \frac{0,2604}{\sqrt{99}}; \quad 10,74 + 1,984 \cdot \frac{0,2604}{\sqrt{99}} \right)$$

$$EX \in (10,69; \quad 10,79)$$

Vậy với độ tin cậy 95%, có thể nói doanh số trung bình/tháng của các hộ nằm trong khoảng (10,69; 10,79) triệu đồng.

c) *Uớc lượng khoảng cho tỷ lệ hộ có doanh số/tháng ≥ 11 triệu, mức ý nghĩa 0,01:*

$$\text{Tỷ lệ mẫu: } p^* = \frac{10+6+1}{100} = 0,17$$

⇒ Khoảng ước lượng cho tỷ lệ:

$$p \in \left( p^* - u \left( \frac{\alpha}{2} \right) \sqrt{\frac{p^*(1-p^*)}{n}}; \quad p^* + u \left( \frac{\alpha}{2} \right) \sqrt{\frac{p^*(1-p^*)}{n}} \right)$$

$$p \in \left( 0,17 - u(0,005) \sqrt{\frac{0,17 \cdot 0,83}{100}}; \quad 0,17 + u(0,005) \sqrt{\frac{0,17 \cdot 0,83}{100}} \right)$$

$$p \in \left( 0,17 - 2,58 \sqrt{\frac{0,17 \cdot 0,83}{100}}; \quad 0,17 + 2,58 \sqrt{\frac{0,17 \cdot 0,83}{100}} \right)$$

$$p \in (0,0731; \quad 0,2669)$$

Vậy với độ tin cậy 99%, tỷ lệ hộ có doanh số/tháng ≥ 11 triệu đồng nằm trong khoảng (0,0731; 0,2669) và tỷ lệ này thấp nhất là 7,31%.

**Bài 2/157: Điều tra 365 điểm trồng lúa của một huyện ta được các số liệu sau:**

Năng suất X (tạ/ha)	25	30	33	34	35	36	37	39	40
Số điểm trồng lúa	6	13	38	74	106	85	30	10	3

a) Chỉ ra Median mẫu, giá trị trung bình mẫu?

b) Với độ tin cậy 95% năng suất lúa trung bình của huyện thấp nhất và cao nhất là bao nhiêu? (giả thiết năng suất lúa là biến ngẫu nhiên chuẩn).

c) Tỷ lệ % số điểm trồng lúa có năng suất cao hơn 35 tạ/ha? Tỷ lệ này thấp nhất là bao nhiêu với độ tin cậy 99%.

a) Chỉ ra Median mẫu và trung bình mẫu:

Cỡ mẫu lẻ ( $n = 365$ ) và được cho dưới dạng điểm nên median mẫu là:

$$\text{Med}^* = x_{\frac{n+1}{2}} = x_{183} = \boxed{35}$$

Giá trị trung bình mẫu:

$$\bar{X} = \frac{1}{n} \sum x_i m_i = \frac{1}{365} \left( 25 \cdot 6 + 30 \cdot 13 + 33 \cdot 38 + 34 \cdot 74 + 35 \cdot 106 + 36 \cdot 85 + 37 \cdot 30 + 39 \cdot 10 + 40 \cdot 3 \right) = \boxed{34,79}$$

b) *Năng suất lúa trung bình của huyện thấp nhất và cao nhất với độ tin cậy 95%:*

Dạng bài toán ước lượng khoảng cho giá trị trung bình, trường hợp biết phân bố chuẩn nhưng chưa biết phương sai lý thuyết. Khoảng ước lượng là:

$$EX \in \left( \bar{X} - t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s}{\sqrt{n-1}}; \quad \bar{X} + t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s}{\sqrt{n-1}} \right)$$

Sử dụng máy tính bỏ túi:

$$\bar{X} = 34,79 \quad (\text{theo câu a})$$

$$s = 2,152$$

$\Rightarrow$  Uớc lượng khoảng cho kỳ vọng:

$$EX \in \left( 34,79 - t_{364}(0,025) \frac{2,152}{\sqrt{364}}; \quad 34,79 + t_{364}(0,025) \frac{2,152}{\sqrt{364}} \right)$$

$$EX \in \left( 34,79 - 1,967 \cdot \frac{2,152}{\sqrt{364}}; \quad 34,79 + 1,967 \cdot \frac{2,152}{\sqrt{364}} \right)$$

$$EX \in (34,57; \quad 35,01)$$

Vậy với độ tin cậy 95%, năng suất lúa trung bình của huyện thấp nhất là 34,57 tạ/ha và cao nhất là 35,01 tạ/ha

c) *Tỷ lệ % số điểm trồng lúa có năng suất cao hơn 35 tạ/ha? Tỷ lệ này thấp nhất là bao nhiêu với độ tin cậy 99%:*

Dạng bài toán ước lượng điểm cho tỷ lệ và ước lượng khoảng cho tỷ lệ.

+ *Uớc lượng điểm cho tỷ lệ % số điểm trồng lúa có năng suất cao hơn 35 tạ/ha:*

$$p^* = \frac{85+30+10+3}{365} = \frac{128}{365} \approx 0,3507 = 35,07(\%)$$

+ *Khoảng ước lượng cho tỷ lệ số điểm trồng lúa có năng suất cao hơn 35 tạ/ha:*

$$p \in \left( p^* - u \left( \frac{\alpha}{2} \right) \sqrt{\frac{p^*(1-p^*)}{n}}; \quad p^* + u \left( \frac{\alpha}{2} \right) \sqrt{\frac{p^*(1-p^*)}{n}} \right)$$

$$p \in \left( 0,3507 - u(0,005) \sqrt{\frac{128 \cdot 237}{365^3}}; \quad 0,3507 + u(0,005) \sqrt{\frac{128 \cdot 237}{365^3}} \right)$$

$$p \in \left( 0,3507 - 2,58 \cdot \sqrt{\frac{128 \cdot 237}{365^3}}; \quad 0,3507 + 2,58 \cdot \sqrt{\frac{128 \cdot 237}{365^3}} \right)$$

$$p \in (0,2863; \quad 0,4151)$$

Vậy với độ tin cậy 99%, tỷ lệ trên thấp nhất là 28,63%.

**Bài 3/157:** Có một khu rừng có diện tích rất lớn. Căn cứ vào kết quả điều tra ngẫu nhiên trên 31 ô, mỗi ô có diện tích trên 0,1 ha được giá trị trung bình mẫu (thể tích gỗ trung bình trên mỗi ô) và sai số tiêu chuẩn trên mỗi ô là  $\bar{X} = 10,2 \text{ m}^3$ ,  $s = 1,45 \text{ m}^3$ . Hãy ước lượng số quan sát cần thiết để sai số ước lượng không vượt quá  $0,4 \text{ m}^3$  với độ tin cậy 95%?

Nếu muốn sai số không vượt quá  $0,5$  thì cần điều tra bổ sung hay không?

Nếu muốn sai số không vượt quá  $0,5$  và số quan sát không vượt quá cỡ mẫu đã điều tra thì độ tin cậy của ước lượng là bao nhiêu?

Hãy chỉ ra ước lượng khoảng đó. Giả thiết  $X$  tuân theo luật chuẩn.

\* *Ước lượng số quan sát cần thiết để sai số không vượt quá  $0,4 \text{ m}^3$ ; độ tin cậy 95%:*

Sai số của ước lượng trong trường hợp biết phân bố chuẩn, chưa biết phương sai lý thuyết là:

$$\varepsilon = t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s}{\sqrt{n-1}}$$

Sai số không vượt quá  $0,4$  nên:

$$t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s}{\sqrt{n-1}} \leq 0,4 \Leftrightarrow \sqrt{n-1} \geq \frac{t_{n-1}(\alpha/2)}{0,4} s$$

$$\Leftrightarrow n \geq \left( \frac{t_{n-1}(\alpha/2)}{0,4} s \right)^2 + 1 \Leftrightarrow n \geq \left( \frac{t_{n-1}(0,025)}{0,4} 1,45 \right)^2 + 1 \Leftrightarrow [n \geq 56]$$

<sup>1</sup>Vậy để sai số không vượt quá  $0,4 \text{ m}^3$  thì số quan sát tối thiểu bằng 56 ô.

\* *Nếu muốn sai số không vượt quá  $0,5$  thì cần điều tra bổ sung hay không:*

Sai số không vượt quá  $0,5$  hay:  $n \geq \left( \frac{t_{n-1}(0,025)}{0,5} 1,45 \right)^2 + 1 \Leftrightarrow [n \geq 36]$

Do đó, cần điều tra bổ sung thêm 5 ô nữa.

\* *Tìm độ tin cậy nếu sai số không vượt quá  $0,5$  và số quan sát không vượt quá cỡ mẫu ( $n = 31$ ) đã điều tra:*

Ta có:

$$t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s}{\sqrt{n-1}} \leq 0,5 \Leftrightarrow t_{30}(\alpha/2) \leq \frac{0,5\sqrt{n-1}}{s} \Leftrightarrow t_{30}(\alpha/2) \leq \frac{0,5\sqrt{30}}{1,45}$$

$$\Leftrightarrow t_{30}(\alpha/2) \leq 1,8887 \Leftrightarrow \alpha/2 \geq 0,0344 \Leftrightarrow [1-\alpha \leq 0,9312]$$

Vậy, độ tin cậy  $\leq 93,12\%$

\* *Chỉ ra ước lượng khoảng cho giá trị trung bình, độ tin cậy 95%:*

Ước lượng khoảng cho giá trị trung bình, trường hợp biết phân bố chuẩn nhưng chưa biết phương sai lý thuyết:

<sup>1</sup> Giải bất phương trình bằng cách thử các giá trị  $n$  tương ứng. Có thể lấy gần đúng  $t_{30}(0,025)$  và tính ra  $n = 56$ .

$$EX \in \left( \bar{X} - t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s}{\sqrt{n-1}}; \quad \bar{X} + t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s}{\sqrt{n-1}} \right)$$

$$EX \in \left( \bar{X} - t_{30} (0,025) \frac{s}{\sqrt{n-1}}; \quad \bar{X} + t_{30} (0,025) \frac{s}{\sqrt{n-1}} \right)$$

$$EX \in \left( 10,2 - 2,042 \cdot \frac{1,45}{\sqrt{30}}; \quad 10,2 + 2,042 \cdot \frac{1,45}{\sqrt{30}} \right)$$

$$EX \in (9,66; \quad 10,74)$$

Vậy với độ tin cậy 95%, thể tích gỗ trung bình trên mỗi ô nằm trong khoảng (9,66; 10,74)

**Bài 4/157:** Với độ tin cậy 95% hãy ước lượng mức xăng tiêu hao trung bình cho một loại ôtô chạy từ A đến B, nếu chạy thử 30 lần trên đoạn đường này người ta ghi nhận được lượng xăng tiêu hao như sau:

Lượng xăng (lít)	(9,6 – 9,8]	(9,8 – 10]	(10 – 10,2]	(10,2 – 10,4]	(10,4 – 10,6]
Số lần tương ứng	3	5	10	8	4

Biết rằng lượng xăng hao phí là đại lượng ngẫu nhiên chuẩn.

Dạng bài toán ước lượng khoảng cho giá trị trung bình, trường hợp biết phân bố chuẩn nhưng chưa biết phương sai lý thuyết.

Khoảng ước lượng:

$$EX \in \left( \bar{X} - t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s}{\sqrt{n-1}}; \quad \bar{X} + t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s}{\sqrt{n-1}} \right)$$

Sử dụng máy tính ta được:

$$\bar{X} = 10,1333$$

$$s = 0,2328$$

⇒ Khoảng ước lượng:

$$EX \in \left( 10,1333 - t_{29} (0,025) \frac{0,2328}{\sqrt{29}}; \quad 10,1333 + t_{29} (0,025) \frac{0,2328}{\sqrt{29}} \right)$$

$$EX \in \left( 10,1333 - 2,045 \cdot \frac{0,2328}{\sqrt{29}}; \quad 10,1333 + 2,045 \cdot \frac{0,2328}{\sqrt{29}} \right)$$

$$EX \in (10,045; \quad 10,222)$$

Vậy với độ tin cậy 95%, mức xăng tiêu hao trung bình của ôtô khi chạy từ A đến B nằm trong khoảng (10,045; 10,222) lít.

**Bài 5/158:** Để ước lượng số cá trong hồ, người ta đánh bắt 2000 con đánh dấu rồi thả xuống. Vài ngày sau, ta lại đánh bắt 400 con thì thấy 80 con được đánh dấu. Với độ tin cậy 95%, số cá trong hồ có bao nhiêu con?

Dạng bài toán ước lượng khoảng cho tỷ lệ (hay xác suất).

$$\text{Tỷ lệ mẫu: } p^* = \frac{80}{400} = 0,2$$

⇒ Khoảng ước lượng cho tỷ lệ cá được đánh dấu:

$$p \in \left( p^* - u \left( \frac{\alpha}{2} \right) \sqrt{\frac{p^*(1-p^*)}{n}}; \quad p^* + u \left( \frac{\alpha}{2} \right) \sqrt{\frac{p^*(1-p^*)}{n}} \right)$$

$$p \in \left( 0,2 - u(0,025) \sqrt{\frac{0,2 \cdot 0,8}{400}}; \quad 0,2 + u(0,025) \sqrt{\frac{0,2 \cdot 0,8}{400}} \right)$$

$$p \in \left( 0,2 - 1,96 \cdot \sqrt{\frac{0,2 \cdot 0,8}{400}}; \quad 0,2 + 1,96 \cdot \sqrt{\frac{0,2 \cdot 0,8}{400}} \right)$$

$$p \in (0,1608; \quad 0,2392)$$

⇒ Khoảng ước lượng cho tổng số cá (N) trong hồ:

$$N \in \left( \frac{2000}{0,2392}; \quad \frac{2000}{0,1608} \right) \Leftrightarrow N \in (8362; \quad 12438)$$

Vậy với độ tin cậy 95%, tổng số cá trong hồ nằm trong khoảng (8362; 12438) con.

**Bài 6/158:** Một xí nghiệp đưa ra thị trường một sản phẩm mới. Để xem đánh giá của người tiêu dùng đối với loại sản phẩm mới này như thế nào, người ta phát cho mỗi người mua hàng đó một phiếu thăm dò và yêu cầu gửi lại cho xí nghiệp chậm nhất là sau 3 tháng. Vì điều kiện thời gian nên xí nghiệp không thể hỏi ý kiến của tất cả khách hàng trong cả nước, cho nên họ chỉ gửi phiếu thăm dò cho khách hàng ở Hà Nội. Kết quả sau 3 tháng xí nghiệp nhận được 300 phiếu thăm dò, trong đó có 90 phiếu tỏ ra là thích loại sản phẩm này (cả về chức năng và giá cả). Hãy ước lượng tỷ lệ thực khách hàng thích loại sản phẩm này?

Với độ tin cậy 95% tỷ lệ đó cao nhất là bao nhiêu?

Muốn nhận được ước lượng khoảng cho tỷ lệ thực với độ chính xác là 0,03 thì cần thăm dò thêm bao nhiêu phiếu nữa.

Với mẫu  $n = 300$ , ước lượng khoảng có độ chính xác là 0,0436 thì độ tin cậy của kết luận về ước lượng khoảng là bao nhiêu?

\* Ước lượng khoảng cho tỷ lệ thực khách hàng thích sản phẩm mới đưa ra. Với độ tin cậy 95%, tỷ lệ đó cao nhất là bao nhiêu:

$$\text{Tỷ lệ mẫu: } p^* = \frac{90}{300} = 0,3$$

⇒ Khoảng ước lượng:

$$p \in \left( p^* - u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p^*(1-p^*)}{n}}, \quad p^* + u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p^*(1-p^*)}{n}} \right)$$

$$p \in \left( 0,3 - u(0,025) \sqrt{\frac{0,3 \cdot 0,7}{300}}, \quad 0,3 + u(0,025) \sqrt{\frac{0,3 \cdot 0,7}{300}} \right)$$

$$p \in \left( 0,3 - 1,96 \sqrt{\frac{0,3 \cdot 0,7}{300}}, \quad 0,3 + 1,96 \sqrt{\frac{0,3 \cdot 0,7}{300}} \right)$$

$$p \in (0,2481; \quad 0,3519)$$

Vậy với độ tin cậy 95%, tỷ lệ thực khách hàng thích sản phẩm mới nằm trong khoảng (0,2481; 0,3519) và tỷ lệ này cao nhất là 35,19%.

\* Muốn độ chính xác của ước lượng là 0,03 thì cần thăm dò thêm bao nhiêu phiếu:

Độ chính xác của ước lượng là 0,03 hay:

$$u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p^*(1-p^*)}{n}} \leq 0,03 \Leftrightarrow 1,96 \sqrt{\frac{0,3 \cdot 0,7}{n}} \leq 0,03$$

$$\Leftrightarrow n \geq \frac{1,96^2 \cdot 0,3 \cdot 0,7}{0,03^2}$$

$$\Leftrightarrow n \geq 897$$

Vậy để độ chính xác là 0,03 thì cần thăm dò thêm tối thiểu 597 phiếu nữa.

\* Tìm độ tin cậy của ước lượng khoảng biết độ chính xác là 0,0436:

Ta có:

$$u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p^*(1-p^*)}{n}} = 0,0436 \Leftrightarrow u\left(\frac{\alpha}{2}\right) \sqrt{\frac{0,3 \cdot 0,7}{300}} = 0,0436$$

$$\Leftrightarrow u\left(\frac{\alpha}{2}\right) = 1,648 \Leftrightarrow \frac{\alpha}{2} = 0,05 \Leftrightarrow \alpha = 0,1 \Leftrightarrow [1 - \alpha = 0,9]$$

Vậy độ tin cậy của ước lượng là 90%.

**Bài 7/158: Để nghiên cứu tuổi thọ của một loại bóng đèn, người ta thắp thử 100 bóng và có số liệu sau:**

Tuổi thọ X (giờ)	1010 – 1030	1030 – 1050	1050 – 1070	1070 – 1090	1090 – 1110
Số bóng tương ứng	2	3	8	13	25
Tuổi thọ X (giờ)	1110 – 1130	1130 – 1150	1150 – 1170	1170 – 1190	1190 – 1210
Số bóng tương ứng	20	12	10	6	1

Sau khi cải tiến kỹ thuật người ta lại thắp thử 100 bóng. Kết quả như sau:

Tuổi thọ Y (giờ)	1150	1160	1170	1180	1190	1200
Số bóng	10	15	20	30	15	10

a) Hãy chỉ ra ước lượng điểm và ước lượng khoảng ( $\alpha = 0,05$ ) cho tuổi thọ trung bình (EX, EY) và bình phương độ lệch của tuổi thọ bóng đèn (DX, DY) trước và sau khi cải tiến).

b) Với độ tin cậy 95% có thể nói việc cải tiến kỹ thuật đã làm tăng tuổi thọ trung bình của bóng đèn lên ít nhất bao nhiêu giờ.

c) Nếu ước lượng khoảng cho EX có độ chính xác là 6,05 thì độ tin cậy tương ứng là bao nhiêu.

d) Nếu muốn ước lượng khoảng cho EX với độ tin cậy 95%, độ chính xác là 5 thì cần quan sát thêm bao nhiêu bóng đèn nữa.

Giả sử X và Y đều tuân theo luật chuẩn.

a) Chỉ ra các ước lượng điểm và ước lượng khoảng cho EX, EY, DX, DY:

\* Đối với đại lượng ngẫu nhiên X (tuổi thọ bóng đèn trước khi cải tiến):

+ Ước lượng điểm cho EX: sử dụng máy tính

$$\bar{X} = 1111,4$$

+ Ước lượng điểm cho DX (ước lượng chênh):

$$s_x^2 = 1402,04$$

+ Ước lượng khoảng cho EX (trường hợp biết phân bố chuẩn nhưng chưa biết phương sai lý thuyết):

$$EX \in \left( \bar{X} - t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s_x}{\sqrt{n_x - 1}}; \quad \bar{X} + t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s_x}{\sqrt{n_x - 1}} \right)$$

$$EX \in \left( 1111,4 - t_{99}(0,025) \sqrt{\frac{1402,04}{99}}; \quad 1111,4 + t_{99}(0,025) \sqrt{\frac{1402,04}{99}} \right)$$

$$EX \in \left( 1111,4 - 1,984 \sqrt{\frac{1402,04}{99}}; \quad 1111,4 + 1,984 \sqrt{\frac{1402,04}{99}} \right)$$

$$EX \in (1103,9; \quad 1118,9)$$

+ Ước lượng khoảng cho DX:

$$DX \in \left( \frac{n_x s_x^2}{\chi_{n-1}^2(\alpha/2)}; \quad \frac{n_x s_x^2}{\chi_{n-1}^2(1-\alpha/2)} \right)$$

$$DX \in \left( \frac{100 \cdot 1402,04}{\chi_{99}^2(0,025)}; \quad \frac{100 \cdot 1402,04}{\chi_{99}^2(0,975)} \right)$$

$$DX \in \left( \frac{100.1402,04}{128,42}; \frac{100.1402,04}{73,36} \right) \Leftrightarrow DX \in (1091,76; 1911,18)$$

\* Đối với đại lượng ngẫu nhiên Y (tuổi thọ bóng đèn sau khi cải tiến):

+ Uớc lượng điểm cho EY:

$$\bar{Y} = 1175,5$$

+ Uớc lượng điểm cho DY (uớc lượng chêch):

$$s_Y^2 = 204,75$$

+ Uớc lượng khoảng cho EY (trường hợp biết phân bố chuẩn nhưng chưa biết phương sai lý thuyết):

$$EY \in \left( \bar{Y} - t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s_Y}{\sqrt{n_Y - 1}}; \bar{Y} + t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s_Y}{\sqrt{n_Y - 1}} \right)$$

$$EY \in \left( 1175,5 - t_{99}(0,025) \sqrt{\frac{204,75}{99}}; 1175,5 + t_{99}(0,025) \sqrt{\frac{204,75}{99}} \right)$$

$$EY \in \left( 1175,5 - 1,984 \sqrt{\frac{204,75}{99}}; 1175,5 + 1,984 \sqrt{\frac{204,75}{99}} \right)$$

$$EY \in (1172,6; 1178,4)$$

+ Uớc lượng khoảng cho DY:

$$DY \in \left( \frac{n_Y s_Y^2}{\chi_{n-1}^2(\alpha/2)}; \frac{n_Y s_Y^2}{\chi_{n-1}^2(1-\alpha/2)} \right)$$

$$DY \in \left( \frac{100.204,75}{\chi_{99}^2(0,025)}; \frac{100.204,75}{\chi_{99}^2(0,975)} \right)$$

$$DY \in \left( \frac{100.204,75}{128,42}; \frac{100.204,75}{73,36} \right) \Leftrightarrow DY \in (159,44; 279,1)$$

b) Việc cải tiến kỹ thuật đã làm tăng tuổi thọ trung bình của bóng đèn lên ít nhất bao nhiêu giờ, với độ tin cậy 95%:

Dạng bài ước lượng khoảng cho sự khác nhau giữa hai giá trị trung bình<sup>1</sup>:

⇒ Khoảng ước lượng:

$$\left( \bar{Y} - \bar{X} - u \left( \frac{\alpha}{2} \right) \sqrt{\frac{s_Y^2}{n_Y - 1} + \frac{s_X^2}{n_X - 1}}; \bar{Y} - \bar{X} + u \left( \frac{\alpha}{2} \right) \sqrt{\frac{s_Y^2}{n_Y - 1} + \frac{s_X^2}{n_X - 1}} \right)$$

<sup>1</sup> Sinh viên ngoài khoa Toán thì bỏ qua dạng bài này.

$$\Leftrightarrow \left( 1175,5 - 1111,4 - u(0,025) \sqrt{\frac{204,75}{99} + \frac{1402,04}{99}}; \right.$$

$$\left. 1175,5 - 1111,4 + u(0,025) \sqrt{\frac{204,75}{99} + \frac{1402,04}{99}} \right) \Leftrightarrow [56,2; 72]$$

Vậy với độ tin cậy 95%, việc cải tiến kỹ thuật đã làm tăng tuổi thọ trung bình của bóng đèn lên ít nhất 56,2 giờ.

c) *Tìm độ tin cậy của ước lượng khoảng cho EX nếu độ chính xác là 6,05:*

Độ chính xác của ước lượng khoảng cho giá trị trung bình, trường hợp biết phân bố chuẩn nhưng chưa biết phương sai lý thuyết:

$$\varepsilon = t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s_x}{\sqrt{n-1}}$$

Độ chính xác là 6,05 nên:

$$t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s_x}{\sqrt{n-1}} = 6,05 \Leftrightarrow t_{99} \left( \frac{\alpha}{2} \right) \sqrt{\frac{1402,04}{99}} = 6,05 \Leftrightarrow t_{99} \left( \frac{\alpha}{2} \right) = 1,6077$$

$$\Leftrightarrow \frac{\alpha}{2} = 0,0555 \Leftrightarrow \alpha = 0,111 \Leftrightarrow [1 - \alpha = 0,889] \quad \text{Vậy độ tin cậy là 88,9%}.$$

d) *Cần quan sát thêm bao nhiêu bóng đèn nữa nếu muốn ước lượng khoảng có độ tin cậy là 95% và độ chính xác là 5:*

$$\text{Ta có: } t_{n-1} \left( \frac{\alpha}{2} \right) \frac{s_x}{\sqrt{n-1}} = 5 \Leftrightarrow t_{99} (0,025) \frac{\sqrt{1402,04}}{\sqrt{n-1}} = 5$$

(có thể lấy  $t_{n-1}(\alpha/2) = t_{99}(0,025)$  vì giá trị này không thay đổi nhiều khi  $n > 100$ )

$$\Leftrightarrow 1,984 \frac{\sqrt{1402,04}}{\sqrt{n-1}} = 5 \Leftrightarrow n = \left( \frac{1,984 \cdot \sqrt{1402,04}}{5} \right)^2 + 1$$

$$\Leftrightarrow [n \approx 222] \quad \text{Vậy, cần quan sát thêm 122 bóng đèn nữa.}$$

## 4.2. Nhận xét bài tập chương 4

Chương 4 nói riêng cũng như phần thống kê nói chung, chỉ cần áp dụng công thức và tính toán cẩn thận. Các câu hỏi của chương 4 thường rõ ràng và có thể xác định được ngay dạng cần áp dụng. Những dạng cần nhớ của chương này:

- + Ước lượng điểm cho kỳ vọng, phương sai, tỷ lệ: thường chỉ dùng kết quả làm trung gian cho bài toán ước lượng khoảng.
- + Ước lượng khoảng cho kỳ vọng: chú ý phân biệt 3 trường hợp
- + Ước lượng khoảng cho phương sai.
- + Ước lượng khoảng cho tỷ lệ.
- + Tìm thông số chưa biết trong mối liên hệ giữa cỡ mẫu, sai số và độ tin cậy.

# CHƯƠNG 5: BÀI TOÁN KIỂM ĐỊNH GIÁ THIẾT<sup>1</sup>

## A. LÝ THUYẾT

Kiểm định giả thiết là bài toán chọn 1 trong 2 quyết định khi thông tin không đầy đủ, được thu thập từ mẫu. Trên cơ sở của lý thuyết xác suất (các định lý giới hạn) ta đi tìm một miền S sao cho: nếu mẫu thuộc S thì ta bác bỏ giả thiết H ban đầu, ngược lại thì chấp nhận H. Miền S được gọi là miền bác bỏ (hay miền tiêu chuẩn). Gọi  $\alpha$  là mức ý nghĩa của bài toán - thể hiện khả năng xảy ra sai sót của phép kiểm định.

### 5.1. Kiểm định giả thiết cho giá trị trung bình

Có 3 bài toán:

$$H: \mu = \mu_0 \quad | \quad K: \begin{cases} \mu \neq \mu_0 \\ \mu > \mu_0 \\ \mu < \mu_0 \end{cases} \quad \text{với mức ý nghĩa } \alpha$$

a) Trường hợp đã biết phương sai lý thuyết ( $DX = \sigma^2$ ), X chuẩn hoặc  $n \geq 30$ :

Ba miền bác bỏ tương ứng<sup>2</sup>:

$$S_1 = \left\{ \frac{|\bar{X} - \mu_0|}{\sigma} \sqrt{n} \geq u(\alpha/2) \right\}$$

$$S_2 = \left\{ \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n} \geq u(\alpha) \right\}$$

$$S_3 = \left\{ \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n} \leq -u(\alpha) \right\}$$

b) Trường hợp chưa biết phương sai lý thuyết, X có phân bố chuẩn:

Ba miền bác bỏ tương ứng:

$$S_1 = \left\{ \frac{|\bar{X} - \mu_0|}{s} \sqrt{n-1} \geq t_{n-1}(\alpha/2) \right\}$$

$$S_2 = \left\{ \frac{\bar{X} - \mu_0}{s} \sqrt{n-1} \geq t_{n-1}(\alpha) \right\}$$

$$S_3 = \left\{ \frac{\bar{X} - \mu_0}{s} \sqrt{n-1} \leq -t_{n-1}(\alpha) \right\}$$

(với s là phương sai mẫu)

<sup>1</sup> Chương 5 rất quan trọng và không thể không có trong các đề thi cuối kỳ!

<sup>2</sup> Miền bác bỏ xảy ra tương ứng với bất đẳng thức trong dấu {...} xảy ra.

c) Trường hợp chưa biết phương sai lý thuyết, chưa biết  $X$  chuẩn nhưng  $n \geq 30$ :

Ba miền bác bỏ tương ứng:

$$S_1 = \left\{ \frac{|\bar{X} - \mu_0|}{s} \sqrt{n-1} \geq u(\alpha/2) \right\}$$

$$S_2 = \left\{ \frac{|\bar{X} - \mu_0|}{s} \sqrt{n-1} \geq u(\alpha) \right\}$$

$$S_3 = \left\{ \frac{|\bar{X} - \mu_0|}{s} \sqrt{n-1} \leq -u(\alpha) \right\}$$

## 5.2. Kiểm định giả thiết cho phương sai

Có 3 bài toán:

$$H: \sigma^2 = \sigma_0^2 \quad | \quad K: \begin{cases} \sigma^2 \neq \sigma_0^2 \\ \sigma^2 > \sigma_0^2 \\ \sigma^2 < \sigma_0^2 \end{cases} \quad \text{với mức ý nghĩa } \alpha$$

Tương ứng với 3 miền bác bỏ:

$$S_1 = \left\{ \frac{ns^2}{\sigma_0^2} \geq \chi_{n-1}^2(\alpha/2) \right\} \cup \left\{ \frac{ns^2}{\sigma_0^2} \leq \chi_{n-1}^2(1-\alpha/2) \right\}$$

$$S_2 = \left\{ \frac{ns^2}{\sigma_0^2} \geq \chi_{n-1}^2(\alpha) \right\}$$

$$S_3 = \left\{ \frac{ns^2}{\sigma_0^2} \leq \chi_{n-1}^2(1-\alpha) \right\}$$

## 5.3. Kiểm định giả thiết cho tỷ lệ (hay xác suất)

Có 3 bài toán:

$$H: p = p_0 \quad | \quad K: \begin{cases} p \neq p_0 \\ p > p_0 \\ p < p_0 \end{cases} \quad \text{với mức ý nghĩa } \alpha$$

Tương ứng với 3 miền bác bỏ:

$$S_1 = \left\{ \frac{|p^* - p_0|}{\sqrt{p_0(1-p_0)}} \sqrt{n} \geq u(\alpha/2) \right\} \quad \text{với } p^* = \frac{m}{n} \text{ là tỷ lệ mẫu}$$

$$S_2 = \left\{ \frac{p^* - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} \geq u(\alpha) \right\}$$

$$S_3 = \left\{ \frac{p^* - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} \leq -u(\alpha) \right\}$$

#### 5.4. So sánh hai giá trị trung bình

So sánh hai giá trị trung bình  $\mu_1$  và  $\mu_2$  của hai đại lượng ngẫu nhiên X và Y

Có 3 bài toán:

$$H: \mu_1 = \mu_2 \quad | \quad K: \begin{cases} \mu_1 \neq \mu_2 \\ \mu_1 > \mu_2 \\ \mu_1 < \mu_2 \end{cases} \quad \text{với mức ý nghĩa } \alpha$$

a) Trường hợp đã biết hai phương sai lý thuyết ( $DX = \sigma_1^2$ ,  $DY = \sigma_2^2$ ), X và Y chuẩn hoặc cả hai có cỡ mẫu  $\geq 30$ :

Ba miền bác bỏ tương ứng:

$$S_1 = \left\{ \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq u(\alpha/2) \right\}$$

$$S_2 = \left\{ \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq u(\alpha) \right\} \quad S_3 = \left\{ \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq -u(\alpha) \right\}$$

b) Trường hợp chưa biết hai phương sai lý thuyết nhưng X và Y chuẩn và  $DX = DY$ :

Tính:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

Ba miền bác bỏ tương ứng:

$$S_1 = \left\{ |t| \geq t_{n_1+n_2-2}(\alpha/2) \right\}$$

$$S_2 = \left\{ t \geq t_{n_1+n_2-2}(\alpha) \right\}$$

$$S_3 = \left\{ t \leq -t_{n_1+n_2-2}(\alpha) \right\}$$

c) Trường hợp chưa biết hai phương sai lý thuyết, chưa biết  $X$  và  $Y$  chuẩn nhưng hai cỡ mẫu  $n_1, n_2 \geq 30$ :

Ba miền bác bỏ tương ứng:

$$S_1 = \left\{ \frac{\left| \bar{X} - \bar{Y} \right|}{\sqrt{\frac{s_1^2}{n_1-1} + \frac{s_2^2}{n_2-1}}} \geq u(\alpha/2) \right\}$$

$$S_2 = \left\{ \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n_1-1} + \frac{s_2^2}{n_2-1}}} \geq u(\alpha) \right\}$$

$$S_3 = \left\{ \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n_1-1} + \frac{s_2^2}{n_2-1}}} \leq -u(\alpha) \right\}$$

## 5.5. So sánh hai phương sai

Có 3 bài toán:

$$H: \sigma_1^2 = \sigma_2^2 \quad | \quad K: \begin{cases} \sigma_1^2 \neq \sigma_2^2 \\ \sigma_1^2 > \sigma_2^2 \\ \sigma_1^2 < \sigma_2^2 \end{cases} \quad \text{với mức ý nghĩa } \alpha$$

Ta gộp lại thành một bài toán:

$$H: \sigma_1^2 = \sigma_2^2 \quad | \quad K: \sigma_1^2 > \sigma_2^2 \quad \text{với mức ý nghĩa } \alpha$$

Lập tỷ số:

$$F = \frac{\hat{s}_1^2}{\hat{s}_2^2} \text{ nếu } \hat{s}_1^2 > \hat{s}_2^2 \quad \text{hoặc} \quad F = \frac{\hat{s}_2^2}{\hat{s}_1^2} \text{ nếu } \hat{s}_2^2 > \hat{s}_1^2$$

$\Rightarrow$  Miền bác bỏ:

$$S = \left\{ F \geq F_{n_1-1; n_2-1}(\alpha) \right\} \text{ nếu } \hat{s}_1^2 > \hat{s}_2^2$$

$$\text{hoặc } S = \left\{ F \geq F_{n_2-1; n_1-1}(\alpha) \right\} \text{ nếu } \hat{s}_2^2 > \hat{s}_1^2$$

## 5.6. So sánh hai tỷ lệ (hay hai xác suất)

Có 3 bài toán:

$$H: p_1 = p_2 \quad | \quad K: \begin{cases} p_1 \neq p_2 \\ p_1 > p_2 \\ p_1 < p_2 \end{cases} \quad \text{với mức ý nghĩa } \alpha$$

Gọi  $m_1$  và  $m_2$  lần lượt là số lần xảy ra biến cố A ứng với mẫu X và mẫu Y.

Tính:

$$u = \frac{\frac{m_1 - m_2}{n_1 - n_2}}{\sqrt{\frac{m_1 + m_2}{n_1 + n_2} \left(1 - \frac{m_1 + m_2}{n_1 + n_2}\right) \frac{n_1 + n_2}{n_1 n_2}}}$$

Ba miền bắc bỏ tương ứng:

$$S_1 = \{ |u| \geq u(\alpha/2) \}$$

$$S_2 = \{ u \geq u(\alpha) \}$$

$$S_3 = \{ u \leq -u(\alpha) \}$$

### 5.7. Tiêu chuẩn phù hợp Khi bình phương

“Tiêu chuẩn phù hợp Khi bình phương” dùng để kiểm tra xem đại lượng ngẫu nhiên X có phân bố F(x) nào đó hay không.

Bài toán:

$$H: X \text{ có phân bố } F(x) \mid K: X \text{ không có phân bố } F(x) \quad (\text{mức ý nghĩa } \alpha)$$

Chia miền giá trị quan sát được từ mẫu thành k khoảng, trong mỗi khoảng có  $m_i$  quan sát ( $m_i \geq 5$ , nếu không thỏa mãn thì tiến hành gộp khoảng)

\* Trường hợp phân bố  $F(x)$  đã biết (đã cho sẵn các xác suất  $p_i$ , ứng với khoảng thứ i)

Miền bắc bỏ của bài toán là:

$$S = \left\{ \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n \geq \chi_{k-1}^2(\alpha) \right\}$$

\* Trường hợp phân bố  $F(x)$  chưa biết và phụ thuộc tham số:

Tính các xác suất  $p_i$  thông qua các tham số, trong đó các tham số được tính xấp xỉ bằng ước lượng điểm:

+ Nếu kiểm tra xem X có phân bố nhị thức B (k, p) hay không thì ước lượng điểm cho tham số p là:  $p^* = \frac{\bar{X}}{n}$

+ Nếu kiểm tra xem X có phân bố Poisson ( $\lambda$ ) thì ước lượng điểm cho tham số  $\lambda$ :

$$\lambda^* = \bar{X}$$

+ Nếu kiểm tra xem X có phân bố chuẩn N ( $\mu, \sigma^2$ ) hay không thì ước lượng điểm cho hai tham số  $\mu$  và  $\sigma^2$  là:

$$\begin{cases} \mu^* = \bar{X} \\ \sigma^2 * = \hat{s}^2 = \frac{n}{n-1} s^2 \end{cases}$$

+ Nếu kiểm tra xem X có phân bố mū với tham số λ hay không thì ước lượng điểm cho tham ẩn λ là:

$$\lambda^* = \frac{1}{\bar{X}}$$

+ Nếu kiểm tra xem X có phân bố đều trên đoạn [a, b] hay không thì ước lượng điểm cho hai tham ẩn a và b là:

$$\begin{cases} a^* = \bar{X} - \hat{s}\sqrt{3} \\ b^* = \bar{X} + \hat{s}\sqrt{3} \end{cases}$$

Sau khi ước lượng xong các tham ẩn, tính các xác suất  $p_i$  dựa trên tham ẩn.

⇒ Miền bắc bỏ của bài toán là:

$$S = \left\{ \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n \geq \chi_{k-r-1}^2(\alpha) \right\} \quad (\text{với } r \text{ là số lượng tham ẩn})$$

## 5.8. Kiểm tra tính độc lập

Bài toán:

$$H: X \text{ và } Y \text{ độc lập} \mid K: X \text{ và } Y \text{ phụ thuộc} \quad (\text{mức ý nghĩa } \alpha)$$

Mẫu X có r khoảng, mẫu Y có s khoảng. Gọi  $n_{ij}$  là số quan sát sao cho X thuộc khoảng i và Y thuộc khoảng j

Tính:

$$n = \sum_{i=1}^r \sum_{j=1}^s n_{ij} \quad (\text{tổng tất cả các giá trị của mẫu hay cỡ mẫu})$$

$$n_{i\bullet} = \sum_{j=1}^s n_{ij} \quad (\text{tổng các giá trị theo hàng thứ } i)$$

$$n_{\bullet j} = \sum_{i=1}^r n_{ij} \quad (\text{tổng các giá trị theo cột thứ } j)$$

$$\chi^2 = n \left( \sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_{i\bullet} n_{\bullet j}} - 1 \right)$$

⇒ Miền bắc bỏ:

$$S = \left\{ \chi^2 \geq \chi_{(r-1),(s-1)}^2(\alpha) \right\}$$

## 5.9. So sánh nhiều tỷ lệ

“So sánh nhiều tỷ lệ” là trường hợp đặc biệt của “kiểm tra tính độc lập” khi một trong hai khoảng bằng 2 (giả sử  $r = 2$ ).

Sắp xếp giá trị của mẫu X thành hai hàng: biến cố A xảy ra và biến cố A không xảy ra. Giữ nguyên các cột ứng với các khoảng giá trị của Y.

Bài toán:

$$\left\{ \begin{array}{l} H: p_1 = p_2 = \dots = p_n \\ K: \text{các tỷ lệ không như nhau} \\ \text{Mức ý nghĩa } \alpha \end{array} \right.$$

$$\text{Tính: } \chi^2 = n \left( \sum_{i=1}^2 \sum_{j=1}^s \frac{n_{ij}^2}{n_{i\bullet} n_{\bullet j}} - 1 \right)$$

$\Rightarrow$  Miền bác bỏ:

$$S = \left\{ \chi^2 \geq \chi^2_{s-1}(\alpha) \right\}$$

## B. BÀI TẬP

### 5.1. Bài tập trong giáo trình 2 (G<sub>2</sub>)

(Xác suất thống kê, Đào Hữu Hò, trang 197)

**Bài 1/197:** Để nghiên cứu tuổi thọ của một loại bóng đèn, người ta thắp thử 100 bóng và có số liệu sau:

Tuổi thọ X (giờ)	1010 – 1030	1030 – 1050	1050 – 1070	1070 – 1090	1090 – 1110
Số bóng tương ứng	2	3	8	13	25
Tuổi thọ X (giờ)	1110 – 1130	1130 – 1150	1150 – 1170	1170 – 1190	1190 – 1210
Số bóng tương ứng	20	12	10	6	1

Sau khi cải tiến kỹ thuật người ta lại thắp thử 100 bóng. Kết quả như sau:

Tuổi thọ Y (giờ)	1150	1160	1170	1180	1190	1200
Số bóng	10	15	20	30	15	10

Giả sử X và Y đều tuân theo luật chuẩn. Hãy kiểm định giả thiết  $EX = EY$  với đối thiết  $EX < EY$ ,  $\alpha = 0,05$ .

*Hướng dẫn: Dạng bài toán so sánh hai giá trị trung bình, trường hợp biết phân bố chuẩn nhưng chưa biết phương sai lý thuyết (có thể dùng trường hợp cỡ mẫu lớn).*

Bài toán:

$$H : EX = EY \quad | \quad K : EX < EY \quad \text{mức ý nghĩa } \alpha = 0,05$$

Miền bác bỏ:

$$S = \left\{ t \leq -t_{n_1+n_2-2}(\alpha) \right\} \quad \text{với: } t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

(trong đó  $n_1, n_2, s_1^2, s_2^2$  lần lượt là cỡ mẫu và phương sai mẫu của X và Y)

Sử dụng máy tính ta có:

$$\bar{X} = 1111,4$$

$$s_1^2 = 1402,04$$

$$\bar{Y} = 1175,5$$

$$s_2^2 = 204,75$$

Do đó:

$$t = \frac{1111,4 - 1175,5}{\sqrt{\frac{100 \cdot 1402,04 + 100 \cdot 204,75}{100 + 100}}} \sqrt{\frac{100 \cdot 100(100 + 100 - 2)}{100 + 100}} = -15,91$$

Tra bảng của phân bố Student ta được:

$$-t_{n_1+n_2-2}(\alpha) = -t_{198}(0,05) = -1,653. \text{ Suy ra: } t < -t_{n_1+n_2-2}(\alpha)$$

Do đó, miền bác bỏ đã xảy ra. Ta bác bỏ H và chấp nhận K.

Vậy với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng việc cải tiến kỹ thuật đã nâng cao tuổi thọ của bóng đèn.

**Bài 2/197: Trọng lượng các bao gạo là đại lượng ngẫu nhiên chuẩn  $N(50; 0,01)$ . Có nhiều ý kiến khách hàng phản ánh là trọng lượng bị thiếu. Một nhóm thanh tra đã cân ngẫu nhiên 25 bao gạo trong kho, kết quả như sau:**

Trọng lượng bao gạo (kg)	48 – 48,5	48,5 – 49	49 – 49,5	49,5 – 50	50 – 50,5
Số bao	2	5	10	6	2

Hãy xem ý kiến khách hàng có đúng không bằng cách kiểm tra giả thiết  $\mu = 50$  và đối thiết  $\mu < 50$ ,  $\alpha = 0,05$ .

*Hướng dẫn: Dạng bài toán kiểm định giả thiết cho giá trị trung bình, trường hợp biết phân bố chuẩn nhưng chưa biết phương sai lý thuyết.*

Bài toán:

$$H: \mu = 50 \quad | \quad K: \mu < 50 \quad \text{với mức ý nghĩa } \alpha = 0,05$$

Miền bác bỏ:

$$S = \left\{ \frac{\bar{X} - \mu_0}{s} \sqrt{n-1} \leq -t_{n-1}(\alpha) \right\}$$

Sử dụng máy tính ta có:

$$\bar{X} = 49,27 ; s = 0,5192$$

Suy ra:

$$\frac{\bar{X} - \mu_0}{s} \sqrt{n-1} = \frac{49,27 - 50}{0,5192} \sqrt{24} = -6,888$$

Tra bảng của phân bố Student ta được:  $-t_{n-1}(\alpha) = -t_{24}(0,05) = -1,711$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng trọng lượng trung bình của các bao gạo đã bị thiêu, ý kiến của khách hàng là đúng.

**Bài 3/197:** Trong điều kiện chăn nuôi bình thường, lượng sữa trung bình của một con bò là 14kg/ngày. Nghi ngờ điều kiện chăn nuôi kém đi làm cho lượng sữa giảm xuống, người ta điều tra ngẫu nhiên 25 con và tính được lượng sữa trung bình của một con trong một ngày là 12,5 và độ lệch tiêu chuẩn  $s = 2,5$ . Với mức ý nghĩa  $\alpha = 0,05$  hãy kết luận điều nghi ngờ nói trên. Giả thiết lượng sữa bò là một đại lượng ngẫu nhiên chuẩn.

*Hướng dẫn: Dạng bài toán kiểm định giả thiết cho giá trị trung bình, trường hợp biết phân bố chuẩn nhưng chưa biết phương sai lý thuyết.*

Bài toán:

$$H: \mu = 14 \quad | \quad K: \mu < 14 \quad \text{với mức ý nghĩa } \alpha = 0,05$$

Miền bác bỏ:

$$S = \left\{ \frac{\bar{X} - \mu_0}{s} \sqrt{n-1} \leq -t_{n-1}(\alpha) \right\}$$

Tính:

$$\frac{\bar{X} - \mu_0}{s} \sqrt{n-1} = \frac{12,5 - 14}{2,5} \sqrt{24} = -2,939$$

Tra bảng của phân bố Student ta được:  $-t_{n-1}(\alpha) = -t_{24}(0,05) = -1,711$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng lượng sữa đã giảm xuống – điều nghi ngờ trên là đúng.

**Bài 4/197:** Một máy sản xuất tự động với tỷ lệ chính phẩm 98%. Sau một thời gian hoạt động, người ta nghi ngờ tỷ lệ trên đã bị giảm. Kiểm tra ngẫu nhiên 500 sản phẩm thấy có 28 phế phẩm. Với  $\alpha = 0,05$  hãy kiểm tra xem chất lượng làm việc của máy có còn được như trước hay không?

*Hướng dẫn: Dạng bài toán kiểm định giả thiết cho tỷ lệ.*

Bài toán:

$$H: p = 0,98 \quad | \quad K: p < 0,98 \quad \text{với mức ý nghĩa } \alpha = 0,05$$

Tỷ lệ chính phẩm của mẫu:

$$p^* = \frac{500 - 28}{500} = 0,944$$

Miền bác bỏ:  $S = \left\{ \frac{p^* - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} \leq -u(\alpha) \right\}$

Tính:

$$\frac{p^* - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} = \frac{0,944 - 0,98}{\sqrt{0,98 \cdot 0,02}} \sqrt{500} = -5,75$$

Tra bảng của phân bố chuẩn tắc ta được:  $-u(\alpha) = -u(0,05) = -1,645$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy, với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng tỷ lệ chính phẩm đã giảm xuống – điều nghi ngờ là đúng.

**Bài 5/198:** Để so sánh trọng lượng trung bình của trẻ sơ sinh ở thành thị và nông thôn, người ta cân thử trọng lượng của 10 000 cháu và thu được kết quả sau đây:

Vùng	Số cháu được cân	Trọng lượng trung bình	Độ lệch tiêu chuẩn mẫu
Nông thôn	8000	3 kg	0,3 kg
Thành thị	2000	3,2 kg	0,2 kg

Với mức ý nghĩa  $\alpha = 0,05$  có thể coi trọng lượng trung bình của trẻ sơ sinh ở thành thị cao hơn ở nông thôn hay không? (giả thiết trọng lượng trẻ sơ sinh là đại lượng ngẫu nhiên chuẩn)

*Hướng dẫn: Dạng bài toán so sánh hai giá trị trung bình, trường hợp chưa biết hai phương sai lý thuyết, biết phân bố chuẩn.*

Gọi  $\mu_1$  và  $\mu_2$  lần lượt là trọng lượng trung bình của trẻ sơ sinh ở nông thôn và thành thị, ứng với đại lượng ngẫu nhiên X và Y. Bài toán:

$$H: \mu_1 = \mu_2 \quad | \quad K: \mu_1 < \mu_2 \quad \text{với mức ý nghĩa } \alpha = 0,05$$

Miền bác bỏ:

$$S = \{ t \leq -t_{n_1+n_2-2}(\alpha) \}$$

Trong đó:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

Tính:

$$t = \frac{3 - 3,2}{\sqrt{\frac{8000 \cdot 0,3^2 + 2000 \cdot 0,2^2}{8000 + 2000}}} \sqrt{\frac{8000 \cdot 2000 (8000 + 2000 - 2)}{8000 + 2000}} = -28,28$$

Tra bảng của phân bố Student ta được:  $-t_{n_1+n_2-2}(\alpha) = -t_{9998}(0,05) = -1,645$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy, với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng trọng lượng trẻ sơ sinh ở thành thị cao hơn ở nông thôn.

**Bài 6/198:** Thông kê số tai nạn lao động tại hai xí nghiệp có các số liệu sau:

Xí nghiệp	Số công nhân	Số tai nạn lao động
I	200	20
II	800	120

Với mức ý nghĩa  $\alpha = 0,05$  hãy kết luận xem chất lượng công tác bảo vệ an toàn lao động tại hai xí nghiệp trên có khác nhau không?

Hướng dẫn: Dạng bài toán so sánh hai tỷ lệ. Xí nghiệp nào có tỷ lệ tai nạn ít hơn thì chất lượng công tác bảo vệ an toàn lao động tốt hơn.

Bài toán:

$$H: p_1 = p_2 \mid K: p_1 \neq p_2 \quad \text{với mức ý nghĩa } \alpha = 0,05$$

Miền bác bỏ:

$$S = \{ | u | \geq u(\alpha/2) \}$$

Trong đó:

$$u = \frac{\frac{m_1 - m_2}{n_1 - n_2}}{\sqrt{\frac{m_1 + m_2}{n_1 + n_2} \left(1 - \frac{m_1 + m_2}{n_1 + n_2}\right) \frac{n_1 + n_2}{n_1 n_2}}}$$

(với  $m_1$  và  $m_2$  lần lượt là số tai nạn ở xí nghiệp I và xí nghiệp II)

$$\text{Tính: } u = \frac{\frac{20}{200} - \frac{120}{800}}{\sqrt{\frac{140}{1000} \left(1 - \frac{140}{1000}\right) \frac{1000}{200.800}}} = -1,823$$

Tra bảng của phân bố chuẩn ta được:  $u(0,025) = 1,96$

Do đó, miền bác bỏ không xảy ra. Ta chấp nhận H và bác bỏ K.

Vậy, với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng công tác bảo vệ an toàn tại hai xí nghiệp là như nhau.

**Bài 7/198:** Trồng cùng một giống lúa trên hai thửa ruộng như nhau và bón hai loại phân khác nhau. Đến ngày thu hoạch ta có kết quả như sau: thửa thứ nhất lấy mẫu 1000 bông lúa thấy số hạt trung bình mỗi bông  $\bar{X} = 70$  hạt và  $s_x = 10$ ; thửa thứ hai lấy mẫu 500 bông thấy số hạt trung bình trên mỗi bông  $\bar{Y} = 72$  hạt và  $s_y = 20$ .

Hỏi sự khác nhau giữa  $\bar{X}$  và  $\bar{Y}$  là ngẫu nhiên hay bản chất ( $\alpha = 0,05$ )?

Hướng dẫn: Dạng bài toán so sánh hai giá trị trung bình, trường hợp chưa biết phân bố chuẩn, chưa biết phương sai lý thuyết nhưng cỡ mẫu lớn. Nếu phép kiểm định cho kết luận  $\mu_1 = \mu_2$  thì sự khác nhau trên là ngẫu nhiên, ngược lại nếu phép kiểm định cho kết luận  $\mu_1 \neq \mu_2$  thì sự khác nhau là do bản chất.

Bài toán:

$$H: \mu_1 = \mu_2 \mid K: \mu_1 \neq \mu_2 \quad \text{với mức ý nghĩa } \alpha = 0,05$$

Miền bác bỏ:

$$S = \left\{ \frac{\overline{|X - Y|}}{\sqrt{\frac{s_x^2}{n_x - 1} + \frac{s_y^2}{n_y - 1}}} \geq u(\alpha/2) \right\}$$

Ta có:

$$\frac{\overline{|X - Y|}}{\sqrt{\frac{s_x^2}{n_x - 1} + \frac{s_y^2}{n_y - 1}}} = \frac{|70 - 72|}{\sqrt{\frac{10^2}{999} + \frac{20^2}{499}}} = 2,106$$

Tra bảng của phân bố chuẩn tắc ta được:  $u(\alpha/2) = u(0,025) = 1,96$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ  $H$  và chấp nhận  $K$ . Vậy với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng sự khác nhau giữa  $\overline{X}$  và  $\overline{Y}$  là do bản chất.

**Bài 8/199: Điều tra doanh số hàng tháng của 100 hộ kinh doanh một ngành nào đó ta thu được số liệu sau:**

Doanh số X (triệu đồng)	10,1	10,2	10,4	10,5	10,7	10,8	10,9	11	11,3	11,4
Số hộ	2	3	8	13	25	20	12	10	6	1

Hãy kiểm tra giả thiết cho rằng số liệu đã cho phù hợp với phân bố chuẩn?  
(mức ý nghĩa  $\alpha = 0,05$ )

*Hướng dẫn: Dạng bài “tiêu chuẩn phù hợp Khi bình phương” vì kiểm tra xem mẫu số liệu đã cho có phân bố  $F(x)$  nào đó hay không. Trong trường hợp này,  $F(x)$  chưa biết và phụ thuộc tham số, các tham số được tính bằng ước lượng điểm.*

Bài toán:

$$\begin{cases} H: X \text{ có phân bố chuẩn} \\ K: X \text{ không có phân bố chuẩn} \\ \text{Mức ý nghĩa } \alpha = 0,05 \end{cases}$$

Từ mẫu số liệu, sử dụng máy tính ta được:

$$\overline{X} = 10,74$$

$$\hat{s}^2 = 0,0685$$

Do đó, ước lượng điểm cho các tham số của phân bố chuẩn  $N(\mu, \sigma^2)$ :

$$\begin{cases} \mu^* = \overline{X} \\ \sigma^2* = \hat{s}^2 \end{cases} \Leftrightarrow \begin{cases} \mu^* = 10,74 \\ \sigma^2* = 0,0685 \Rightarrow \sigma^* = 0,2617 \end{cases}$$

Từ mẫu số liệu trên, ta tiến hành gộp khoảng để thỏa mãn điều kiện  $m_i \geq 5 \forall i$ . Số khoảng sau khi gộp là  $k = 8$ .

Doanh số X (triệu đồng)	< 10,3	10,3 – 10,45	10,45 – 10,6	10,6 – 10,75	10,75 – 10,85	10,85 – 10,95	10,95 – 11,15	> 11,15
Số hộ	5	8	13	25	20	12	10	7

Tính các xác suất  $p_i$  của phân bố chuẩn:

$$p_1 = \Phi\left(\frac{10,3 - 10,74}{0,2617}\right) = \Phi(-1,6813) = 0,0464$$

$$p_2 = \Phi\left(\frac{10,45 - 10,74}{0,2617}\right) - \Phi\left(\frac{10,3 - 10,74}{0,2617}\right) = 0,1339 - 0,0464 = 0,0875$$

$$p_3 = \Phi\left(\frac{10,6 - 10,74}{0,2617}\right) - \Phi\left(\frac{10,45 - 10,74}{0,2617}\right) = 0,2963 - 0,1339 = 0,1624$$

$$p_4 = \Phi\left(\frac{10,75 - 10,74}{0,2617}\right) - \Phi\left(\frac{10,6 - 10,74}{0,2617}\right) = 0,5152 - 0,2963 = 0,2189$$

$$p_5 = \Phi\left(\frac{10,85 - 10,74}{0,2617}\right) - \Phi\left(\frac{10,75 - 10,74}{0,2617}\right) = 0,6629 - 0,5152 = 0,1477$$

$$p_6 = \Phi\left(\frac{10,95 - 10,74}{0,2617}\right) - \Phi\left(\frac{10,85 - 10,74}{0,2617}\right) = 0,7889 - 0,6629 = 0,126$$

$$p_7 = \Phi\left(\frac{11,15 - 10,74}{0,2617}\right) - \Phi\left(\frac{10,95 - 10,74}{0,2617}\right) = 0,9414 - 0,7889 = 0,1525$$

$$p_8 = \Phi(+\infty) - \Phi\left(\frac{11,15 - 10,74}{0,2617}\right) = 1 - 0,9414 = 0,0586$$

⇒ Miền bắc bỏ của bài toán:

$$S = \left\{ \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n \geq \chi_{k-r-1}^2 (\alpha) \right\} \quad (\text{với } r = 2 \text{ là số lượng tham số})$$

Tính:

$$\frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n = \frac{1}{100} \left( \frac{5^2}{0,0464} + \frac{8^2}{0,0875} + \frac{13^2}{0,1624} + \frac{25^2}{0,2189} + \frac{20^2}{0,1477} + \frac{12^2}{0,126} + \frac{10^2}{0,1525} + \frac{7^2}{0,0586} \right) - 100 = 5,09$$

Tra bảng của phân bố Khi bình phương ta được:

$$\chi_{k-r-1}^2 (\alpha) = \chi_{8-2-1}^2 (0,05) = \chi_5^2 (0,05) = 11,07$$

Do đó miền bắc đã không xảy ra, ta chấp nhận H và bác bỏ K. Vậy, với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng số liệu trên phù hợp với phân bố chuẩn.

**Bài 9/199: Điều tra 365 điểm trồng lúa của một huyện ta được các số liệu sau:**

Năng suất X (tạ/ha)	25	30	33	34	35	36	37	39	40
Số điểm trồng lúa	6	13	38	74	106	85	30	10	3

Hãy kiểm tra giả thiết cho rằng số liệu đã cho được rút ra từ đại lượng ngẫu nhiên chuẩn ( $\alpha = 0,05$ )

*Hướng dẫn: Dạng bài “tiêu chuẩn phù hợp Khi bình phương” vì kiểm tra xem mẫu số liệu đã cho có phân bố  $F(x)$  nào đó hay không. Trong trường hợp này,  $F(x)$  chưa biết và phụ thuộc tham số, các tham số được tính bằng các ước lượng điểm (tương tự bài 8/199)*

Bài toán:

$$\begin{cases} H: X \text{ có phân bố chuẩn} \\ K: X \text{ không có phân bố chuẩn} \\ \text{Mức ý nghĩa } \alpha = 0,05 \end{cases}$$

Sử dụng máy tính ta có:

$$\bar{X} = 34,79$$

$$\hat{s}^2 = 4,6439$$

Do đó, ước lượng điểm cho các tham số của phân bố chuẩn  $N(\mu, \sigma^2)$ :

$$\begin{cases} \mu^* = \bar{X} \\ \sigma^2* = \hat{s}^2 \end{cases} \Leftrightarrow \begin{cases} \mu^* = 34,79 \\ \sigma^2* = 4,6439 \Rightarrow \sigma^* = 2,155 \end{cases}$$

Từ mẫu số liệu trên, ta tiến hành gộp khoảng để thỏa mãn điều kiện  $m_i \geq 5 \forall i$ . Số khoảng sau khi gộp là  $k = 8$ .

Năng suất X (tạ/ha)	< 27,5	27,5 – 31,5	31,5 – 33,5	33,5 – 34,5	34,5 – 35,5	35,5 – 36,5	35,5 – 38	> 38
Số điểm trồng lúa	6	13	38	74	106	85	30	13

Tính các xác suất  $p_i$  của phân bố chuẩn:

$$p_1 = \Phi\left(\frac{27,5 - 34,79}{2,155}\right) = \Phi(-3,383) = 0,0004$$

$$p_2 = \Phi\left(\frac{31,5 - 34,79}{2,155}\right) - \Phi\left(\frac{27,5 - 34,79}{2,155}\right) = 0,0634 - 0,0004 = 0,063$$

$$p_3 = \Phi\left(\frac{33,5 - 34,79}{2,155}\right) - \Phi\left(\frac{31,5 - 34,79}{2,155}\right) = 0,2747 - 0,0634 = 0,2113$$

$$p_4 = \Phi\left(\frac{34,5 - 34,79}{2,155}\right) - \Phi\left(\frac{33,5 - 34,79}{2,155}\right) = 0,4465 - 0,2747 = 0,1718$$

$$p_5 = \Phi\left(\frac{35,5 - 34,79}{2,155}\right) - \Phi\left(\frac{34,5 - 34,79}{2,155}\right) = 0,6291 - 0,4465 = 0,1826$$

$$p_6 = \Phi\left(\frac{36,5 - 34,79}{2,155}\right) - \Phi\left(\frac{35,5 - 34,79}{2,155}\right) = 0,7863 - 0,6291 = 0,1572$$

$$p_7 = \Phi\left(\frac{38 - 34,79}{2,155}\right) - \Phi\left(\frac{36,5 - 34,79}{2,155}\right) = 0,9318 - 0,7863 = 0,1455$$

$$p_8 = \Phi(+\infty) - \Phi\left(\frac{38 - 34,79}{2,155}\right) = 1 - 0,9318 = 0,0682$$

⇒ Miền bắc bộ của bài toán:

$$S = \left\{ \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n \geq \chi_{k-r-1}^2(\alpha) \right\} \quad (\text{với } r = 2 \text{ là số lượng tham số})$$

Tính:

$$\frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n = \frac{1}{365} \left( \frac{6^2}{0,0004} + \frac{13^2}{0,063} + \frac{38^2}{0,2113} + \frac{74^2}{0,1718} + \dots + \frac{106^2}{0,1826} + \frac{85^2}{0,1572} + \frac{30^2}{0,1455} + \frac{13^2}{0,0682} \right) - 365 = 313,2$$

Tra bảng của phân bố Khi bình phương ta được:

$$\chi_{k-r-1}^2(\alpha) = \chi_{8-2-1}^2(0,05) = \chi_5^2(0,05) = 11,07$$

Do đó miền bắc bộ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy, với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng số liệu đã cho không được rút ra từ đại lượng ngẫu nhiên chuẩn.

**Bài 10/199: Số con của 2000 phụ nữ thủ đô dưới 25 tuổi được cho ở bảng sau:**

Số con X	0	1	2	3	4
Số phụ nữ	1090	650	220	30	10

Với mức ý nghĩa  $\alpha = 0,05$  có thể xem X tuân theo luật Poisson hay không?

*Hướng dẫn: Dạng bài “tiêu chuẩn phù hợp Khi bình phương” vì kiểm tra xem mẫu số liệu đã cho có tuân theo phân bố  $F(x)$  nào đó hay không.*

Bài toán:

$$\begin{cases} H: X \text{ có phân bố Poisson} \\ K: X \text{ không có phân bố Poisson} \\ \text{Mức ý nghĩa } \alpha = 0,05 \end{cases}$$

Sử dụng máy tính ta có:

$$\bar{X} = 0,61$$

Do đó, ước lượng điểm cho tham số  $\lambda$  của phân bố Poisson là:  $\lambda^* = \bar{X} = 0,61$

Tính các xác suất  $p_i$  từ tham số vừa tìm được:

$$P(X = 0) = e^{-0,61} \frac{0,61^0}{0!} = 0,5434$$

$$P(X = 1) = e^{-0,61} \frac{0,61^1}{1!} = 0,3314$$

$$P(X = 2) = e^{-0,61} \frac{0,61^2}{2!} = 0,1011$$

$$P(X = 3) = e^{-0,61} \frac{0,61^3}{3!} = 0,0206$$

$$P(X \geq 4) = 1 - P(X < 4) = 0,0035$$

$\Rightarrow$  Miền bác bỏ của bài toán:

$$S = \left\{ \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n \geq \chi_{k-r-1}^2(\alpha) \right\}$$

(với  $k = 5$  là số khoảng và  $r = 1$  là số lượng tham số)

Tính:

$$\frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n = \frac{1}{2000} \left( \frac{1090^2}{0,5434} + \frac{650^2}{0,3314} + \frac{220^2}{0,1011} + \frac{30^2}{0,0206} + \frac{10^2}{0,0035} \right) - 2000 = 6,154$$

Tra bảng của phân bố Khi bình phương ta được:

$$\chi_{k-r-1}^2(\alpha) = \chi_{5-1-1}^2(0,05) = \chi_3^2(0,05) = 7,81$$

Do đó miền bác bỏ không xảy ra, ta chấp nhận  $H$  và bác bỏ  $K$ . Vậy, với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng  $X$  tuân theo luật Poisson.

**Bài 11/199<sup>1</sup>:** Cùng một loại hạt giống đem xử lý theo hai phương án khác nhau. Kết quả quan sát chiều cao cây con của mỗi phương án được cho dưới đây:

<b>Phương án I</b>	<b>39,2</b>	<b>29</b>	<b>28,5</b>	<b>33,5</b>	<b>41,7</b>	<b>37,2</b>
	<b>37,3</b>	<b>27,7</b>	<b>23,4</b>	<b>33,4</b>	<b>29,2</b>	<b>35,6</b>
<b>Phương án II</b>	<b>20,8</b>	<b>33,8</b>	<b>28,6</b>	<b>23,4</b>	<b>22,7</b>	<b>30,9</b>
	<b>31</b>	<b>27,4</b>	<b>19,5</b>	<b>29,6</b>	<b>23,2</b>	<b>18,7</b>
	<b>20,7</b>	<b>17,6</b>	<b>29,4</b>	<b>27,7</b>	<b>25,5</b>	<b>14,5</b>

<sup>1</sup> Bài 11, bài 12 và bài 20 thuộc dạng so sánh hai giá trị trung bình, trường hợp phi tham số. Sinh viên ngoài khoa Toán không cần ôn tập dạng bài này.

Hãy dùng tiêu chuẩn phi tham số để kiểm tra xem hai phương án xử lý có ảnh hưởng đến sinh trưởng chiều cao cây con hay không ( $\alpha = 0,05$ )

Hướng dẫn: Dạng bài so sánh hai giá trị trung bình, trường hợp phi tham số (không biệt phân bố chuẩn, không biết phương sai lý thuyết, cỡ mẫu nhỏ). Vì hai mẫu độc lập (không được cho theo từng cặp) nên áp dụng tiêu chuẩn Mann – Whitney để xét tính thuần nhất của mẫu ( $EX = EY$  hay không).

Gọi X là chiều cao cây con ở phương án I, Y là chiều cao cây con ở phương án II

Bài toán:

- { H: Phương án xử lý không ảnh hưởng đến chiều cao cây con ( $EX = EY$ )
- { K: Phương án xử lý có ảnh hưởng đến chiều cao cây con ( $EX \neq EY$ )
- Mức ý nghĩa  $\alpha = 0,05$

- Gộp các phần tử của cả hai mẫu thành một và tính hạng (rank) của từng phần tử  $m_i$  trong mẫu gồm  $(n_1 + n_2)$  phần tử. Mẫu khi gộp thành một và sắp xếp theo chiều tăng dần như sau:

Phần tử $m_i$	14,5	17,6	18,7	19,5	20,7	20,8	22,7	23,2	23,4	23,4
STT	1	2	3	4	5	6	7	8	9	10
Phần tử $m_i$	25,5	27,4	27,7	27,7	28,5	28,6	29	29,2	29,4	29,6
STT	11	12	13	14	15	16	17	18	19	20
Phần tử $m_i$	30,9	31	33,4	33,5	33,8	35,6	37,2	37,3	39,2	41,7
STT	21	22	23	24	25	26	27	28	29	30

- Tính hạng của từng phần tử:

Phần tử $m_i$	14,5	17,6	18,7	19,5	20,7	20,8	22,7	23,2	23,4	
Rank	1	2	3	4	5	6	7	8	9,5	
Phần tử $m_i$	25,5	27,4	27,7		28,5	28,6	29	29,2	29,4	29,6
Rank	11	12	13,5		15	16	17	18	19	20
Phần tử $m_i$	30,9	31	33,4	33,5	33,8	35,6	37,2	37,3	39,2	41,7
Rank	21	22	23	24	25	26	27	28	29	30

+ Tổng hạng của các phần tử trong mẫu X:

$$\begin{aligned}
 R_1 &= \text{rank}(23,4) + \text{rank}(27,7) + \text{rank}(28,5) + \text{rank}(29) + \text{rank}(29,2) \\
 &\quad + \text{rank}(33,4) + \text{rank}(33,5) + \text{rank}(35,6) + \text{rank}(37,2) + \text{rank}(37,3) \\
 &\quad + \text{rank}(39,2) + \text{rank}(41,7) \\
 &= 9,5 + 13,5 + 15 + 17 + 18 + 23 + 24 + 26 + 27 + 28 + 29 + 30 = 260
 \end{aligned}$$

Tính các giá trị U, EU, DU:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 = 12 \cdot 18 + \frac{12 \cdot 13}{2} - 260 = 34$$

$$EU = \frac{n_1 n_2}{2} = \frac{12 \cdot 18}{2} = 108$$

$$\begin{aligned} DU &= \frac{n_1 n_2 (n_1 + n_2 + 1)}{12} = \frac{12 \cdot 18 (12 + 18 + 1)}{12} = 558 \\ \Rightarrow \text{Miền bắc bỏ: } S &= \left\{ \frac{|U - EU|}{\sqrt{DU}} \geq u(\alpha/2) \right\} \end{aligned}$$

Ta có:

$$\frac{|U - EU|}{\sqrt{DU}} = \frac{|34 - 108|}{\sqrt{558}} = 3,133$$

Tra bảng của phân bố chuẩn tắc ta được:  $u(\alpha/2) = u(0,025) = 1,96$

Do đó miền bắc bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng phương án xử lý đã ảnh hưởng đến chiều cao con.

**Bài 12/199:** Giả sử ta muốn xác định xem hiệu quả của chế độ ăn kiêng đối việc giảm trọng lượng như thế nào. 20 người quá béo đã thực hiện chế độ ăn kiêng, trọng lượng của từng người trước khi ăn kiêng (X kg) và sau khi ăn kiêng (Y kg) được cho như sau:

X	80	78	85	70	90	78	92	88	75		
Y	75	77	80	70	84	74	85	82	80		
X	75	63	72	89	76	77	71	83	78	82	90
Y	65	62	71	83	72	82	71	79	76	83	81

Dùng tiêu chuẩn phi tham số kiểm tra xem chế độ ăn kiêng có tác dụng làm giảm trọng lượng hay không? ( $\alpha = 0,05$ )

*Hướng dẫn: Dạng bài so sánh hai giá trị trung bình, trường hợp phi tham số (không biết phân bố chuẩn, không biết phương sai lý thuyết, cỡ mẫu nhỏ). Vì hai mẫu là phụ thuộc (được cho theo từng cặp) nên áp dụng tiêu chuẩn Wilcoxon để xét xem EX có bằng EY hay không. Sinh viên ngoài khoa Toán thì không cần ôn tập dạng bài này!*

Bài toán:

$$\begin{cases} H: \text{Chế độ ăn kiêng không ảnh hưởng đến trọng lượng (EX = EY)} \\ K: \text{Chế độ ăn kiêng có ảnh hưởng đến trọng lượng (EX} \neq EY) \\ \text{Mức ý nghĩa } \alpha = 0,05 \end{cases}$$

+ Tính  $d_i = (x_i - y_i)$  và lấy các giá trị  $d_i \neq 0$  sắp xếp theo chiều tăng dần, ta được:

Giá trị $d_i$	-5	-5	-1	1	1	1	2	4	4
STT	1	2	3	4	5	6	7	8	9
Giá trị $d_i$	4	5	5	6	6	6	7	9	10
STT	10	11	12	13	14	15	16	17	18

Gọi  $n$  là số  $d_i \neq 0$ . Suy ra:  $n = 18$

+ Tính hạng của  $|d_i|$ :

Giá trị $ d_i $	1	2	4	5	6	7	9	10
Rank	2,5	5	7	10,5	14	16	17	18

+ Tính T, ET, DT:

$$T = \sum_{d_i > 0} \text{Rank}(|d_i|) = 3.2,5 + 5 + 3.7 + 2.10,5 + 3.14 + 16 + 17 + 18 = 147,5$$

$$ET = \frac{n(n+1)}{4} = \frac{18.19}{4} = 85,5$$

$$DT = \frac{n(n+1)(2n+1)}{24} = \frac{18.19.37}{24} = 527,25$$

⇒ Miền bắc bỏ:

$$S = \left\{ \frac{|T - ET|}{\sqrt{DT}} \geq u(\alpha/2) \right\}$$

Ta có:

$$\frac{|T - ET|}{\sqrt{DT}} = \frac{|147,5 - 85,5|}{\sqrt{527,25}} = 2,7$$

Tra bảng của phân bố chuẩn tắc ta được:  $u(\alpha/2) = u(0,025) = 1,96$

Do đó miền bắc bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng chế độ ăn kiêng đã ảnh hưởng đến trọng lượng. Mà dựa vào mẫu ta thấy  $\bar{X} < \bar{Y}$  nên có thể cho rằng chế độ ăn kiêng có tác dụng làm giảm trọng lượng.

### Bài 13/200: Dùng ba phương án xử lý hạt giống, kết quả cho như sau:

Kết quả	Phương án I	Phương án II	Phương án III
Số hạt mọc	360	603	490
Số hạt không mọc	40	97	180

a) Các phương án xử lý có tác dụng như nhau đối với tỷ lệ nảy mầm hay không? (mức ý nghĩa  $\alpha = 0,05$ )

b) Tìm phương án xử lý tốt nhất ( $\alpha = 0,05$ )

Hướng dẫn: Câu a thuộc dạng bài so sánh nhiều tỷ lệ, nếu tỷ lệ nảy mầm bằng nhau thì các phương án xử lý giống có tác dụng như nhau. Câu b thuộc dạng bài so sánh hai tỷ lệ (chọn ra 2 tỷ lệ cao nhất và so sánh chúng với nhau trước).

a) Các phương án xử lý có tác dụng như nhau đối với tỷ lệ này mầm hay không:

Bài toán:

$\begin{cases} H: \text{Các tỷ lệ nảy mầm là như nhau.} \\ K: \text{Các tỷ lệ nảy mầm không như nhau.} \\ \text{Mức ý nghĩa } 0,05 \end{cases}$

Miền bắc bỏ:

$$S = \left\{ \chi^2 \geq \chi^2_{s-1} (\alpha) \right\} \quad (\text{với } s \text{ là số phương án xử lý hạt giống, } s = 3)$$

$$\text{Tính: } \chi^2 = n \left( \sum_{i=1}^2 \sum_{j=1}^3 \frac{n_{ij}^2}{n_{i\bullet} n_{\bullet j}} - 1 \right)$$

$$\chi^2 = 1770 \left( \frac{360^2}{400.1453} + \frac{603^2}{700.1453} + \frac{490^2}{670.1453} + \frac{40^2}{400.317} + \frac{97^2}{700.317} + \frac{180^2}{670.317} - 1 \right) = 61,39$$

Tra bảng của phân bố Khi bình phương ta được:

$$\chi^2_{s-1}(\alpha) = \chi^2_2(0,05) = 5,99$$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng các phương án xử lý khác nhau thì có tác dụng khác nhau đối với tỷ lệ này mầm.

b) Chọn phương án xử lý tốt nhất:

Các tỷ lệ mẫu của 3 phương án là:

$$p_1^* = \frac{360}{400} = 0,9$$

$$p_2^* = \frac{603}{700} = 0,8614$$

$$p_3^* = \frac{490}{670} = 0,7313$$

Ta thấy:  $p_1^* > p_2^* > p_3^*$ . Trước hết, so sánh tỷ lệ này mầm của phương án I và II

Bài toán:

$$H: p_1 = p_2 \quad | \quad K: p_1 > p_2 \quad (\text{mức ý nghĩa } \alpha = 0,05)$$

Miền bác bỏ:

$$S = \{ u \geq u(\alpha) \}$$

Trong đó:

$$u = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_2}}{\sqrt{\frac{m_1 + m_2}{n_1 + n_2} \left( 1 - \frac{m_1 + m_2}{n_1 + n_2} \right) \frac{n_1 + n_2}{n_1 n_2}}} = \frac{\frac{360}{400} - \frac{603}{700}}{\sqrt{\frac{360 + 603}{1100} \left( 1 - \frac{360 + 603}{1100} \right) \frac{1100}{400 \cdot 700}}} \\ = 1,864$$

Tra bảng ta được:  $u(\alpha) = u(0,05) = 1,645$ . Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy, với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng phương án I cho tỷ lệ này mầm tốt hơn. Kết hợp với tỷ lệ mẫu, suy ra phương án I là tốt nhất.

**Bài 14/200:** Nghiên cứu sự ảnh hưởng của hoàn cảnh gia đình đối với tình hình phạm tội của trẻ em ở tuổi vị thành niên qua 148 em nhỏ người ta thu được kết quả sau:

Tình trạng phạm tội \ Hoàn cảnh gia đình	Bố hoặc mẹ đã chết	Bố mẹ ly hôn	Còn cả bố mẹ
Không phạm tội	20	25	13
Phạm tội	29	43	18

Với mức ý nghĩa  $\alpha = 0,05$  có thể kết luận là hoàn cảnh gia đình của trẻ em độc lập với phạm tội hay không?

*Hướng dẫn: Áp dụng dạng bài toán kiểm tra tính độc lập (hoặc so sánh nhiều tỷ lệ, vì số khoảng của tình trạng phạm tội bằng 2)*

Bài toán:

$$\left\{ \begin{array}{l} H: \text{Phạm tội độc lập với hoàn cảnh gia đình} \\ K: \text{Phạm tội phụ thuộc vào hoàn cảnh gia đình} \\ \text{Mức ý nghĩa } 0,05 \end{array} \right.$$

Miền bác bỏ:

$$S = \left\{ \chi^2 \geq \chi_{s-1}^2 (\alpha) \right\} \quad (\text{với } s \text{ là các dạng hoàn cảnh gia đình, } s = 3)$$

Tính:

$$\begin{aligned} \chi^2 &= n \left( \sum_{i=1}^2 \sum_{j=1}^3 \frac{n_{ij}^2}{n_{i \cdot} \cdot n_{\cdot j}} - 1 \right) \\ &= 148 \left( \frac{20^2}{49.58} + \frac{25^2}{68.58} + \frac{13^2}{31.58} + \frac{29^2}{49.90} + \frac{43^2}{68.90} + \frac{18^2}{31.90} - 1 \right) = 0,32 \end{aligned}$$

Tra bảng của phân bố Khi bình phương ta được:

$$\chi_{s-1}^2 (\alpha) = \chi_2^2 (0,05) = 5,99$$

Do đó miền bác bỏ không xảy ra, ta chấp nhận H và bác bỏ K. Vậy với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng tỷ lệ phạm tội độc lập với hoàn cảnh gia đình.

**Bài 15/201:** Theo dõi sự phụ thuộc giữa màu mắt và màu tóc ở 124 phụ nữ ở một nước châu Âu ta có kết quả sau:

Màu mắt \ Màu tóc	Vàng nâu	Nâu	Đen	Vàng hoe
Màu mắt				
Xanh	25	9	3	7
Xám	13	17	10	7
Nâu mực	7	13	8	5

Với  $\alpha = 0,05$  hãy kiểm tra giả thiết cho rằng màu của tóc và màu của mắt độc lập với nhau.

Hướng dẫn: Dạng bài toán kiểm tra tính độc lập với số khoảng màu mắt ( $r = 3$ ) và số khoảng màu tóc ( $s = 4$ ).

Bài toán:

$$\begin{cases} H: \text{Màu tóc và màu mắt độc lập với nhau} \\ K: \text{Màu tóc và màu mắt phụ thuộc với nhau} \\ \text{Mức ý nghĩa } 0,05 \end{cases}$$

Miền bác bỏ:

$$S = \left\{ \chi^2 \geq \chi_{(r-1)(s-1)}^2(\alpha) \right\}$$

(với  $r = 3$  là số khoảng màu mắt và  $s = 4$  là số khoảng màu tóc)

Tính:

$$\begin{aligned} \chi^2 &= n \left( \sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_{i\bullet} \cdot n_{\bullet j}} - 1 \right) \\ &= 124 \left( \frac{25^2}{45.44} + \frac{9^2}{39.44} + \frac{3^2}{21.44} + \frac{7^2}{19.44} + \frac{13^2}{45.47} + \frac{17^2}{39.47} + \right. \\ &\quad \left. + \frac{10^2}{21.47} + \frac{7^2}{19.47} + \frac{7^2}{45.33} + \frac{13^2}{39.33} + \frac{8^2}{21.33} + \frac{5^2}{19.33} - 1 \right) = 15,07 \end{aligned}$$

Tra bảng của phân bố Khi bình phương ta được:

$$\chi_{(r-1)(s-1)}^2(\alpha) = \chi_6^2(0,05) = 12,59$$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ  $H$  và chấp nhận  $K$ . Vậy với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng màu tóc và màu mắt phụ thuộc với nhau.

**Bài 16/201:** Để xác định thời vụ phun thuốc diệt sâu có lợi nhất, tổ bảo vệ cây trồng đã theo dõi các lứa sâu trong từng thời kỳ và đếm số sâu non mới nở bắt được. Kết quả ghi ở bảng sau:

Thời kỳ theo dõi	Tháng 1	Tháng 2	Tháng 3	Tháng 4	Tháng 5
Số sâu non mới nở bắt được	62	28	70	75	15
Tổng số sâu bắt được	488	392	280	515	185

a) Tỷ lệ sâu non mới nở trong các thời kỳ quan sát khác nhau có ý nghĩa hay không? ( $\alpha = 0,05$ )

b) Xác định thời kỳ có tỷ lệ sâu mới nở cao nhất. Ước lượng tỷ lệ này. Để ước lượng có sai số không vượt quá 0,02 cần bắt bao nhiêu sâu? ( $\alpha = 0,05$ )

Hướng dẫn: Câu a thuộc dạng bài so sánh nhiều tỷ lệ. Câu b, trước hết so sánh hai thời kỳ có tỷ lệ mẫu lớn nhất; nếu kết quả chọn thời kỳ có tỷ lệ mẫu nhỏ hơn thì lại tiếp tục làm bài toán so sánh hai tỷ lệ khác.

Từ bảng tần suất ở trên, suy ra:

Thời kỳ theo dõi	Tháng 1	Tháng 2	Tháng 3	Tháng 4	Tháng 5
Số sâu non mới nở bắt được	62	28	70	75	15
<b>Số sâu khác</b>	426	364	210	440	170

a) Tỷ lệ sâu non mới nở trong các thời kỳ quan sát khác nhau có ý nghĩa hay không:

Bài toán:  $\begin{cases} H: \text{Tỷ lệ sâu non trong các thời kỳ là như nhau.} \\ K: \text{Tỷ lệ sâu non trong các thời kỳ không như nhau.} \\ \text{Mức ý nghĩa } \alpha = 0,05 \end{cases}$

Miền bác bỏ:

$$S = \left\{ \chi^2 \geq \chi_{s-1}^2(\alpha) \right\} \quad (\text{với } s = 5 \text{ là số lượng thời kỳ})$$

$$\text{Tính: } \chi^2 = n \left( \sum_{i=1}^2 \sum_{j=1}^5 \frac{n_{ij}^2}{n_{i\bullet} \cdot n_{\bullet j}} - 1 \right)$$

$$\chi^2 = 1860 \left( \frac{62^2}{488.250} + \frac{28^2}{392.250} + \frac{70^2}{280.250} + \frac{75^2}{515.250} + \frac{15^2}{185.250} + \frac{426^2}{488.1610} + \frac{364^2}{392.1610} + \frac{210^2}{280.1610} + \frac{440^2}{515.1610} + \frac{170^2}{185.1610} - 1 \right) = 50,83$$

Tra bảng của phân bố Khi bình phương ta được:

$$\chi_{s-1}^2(\alpha) = \chi_4^2(0,05) = 9,49$$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng tỷ lệ sâu non trong các thời kỳ là khác nhau, sự khác nhau này là có ý nghĩa.

b) Xác định thời kỳ có tỷ lệ sâu mới nở cao nhất. Ước lượng tỷ lệ này. Để ước lượng có sai số không vượt quá 0,02 cần bao nhiêu sâu? ( $\alpha = 0,05$ ):

Tỷ lệ mẫu về sâu mới nở trong mỗi thời kỳ là:

$$p_1^* = \frac{62}{488} = 0,127 \quad p_4^* = \frac{75}{515} = 0,1456$$

$$p_2^* = \frac{28}{392} = 0,0714 \quad p_5^* = \frac{15}{185} = 0,0811$$

$$p_3^* = \frac{70}{280} = 0,25$$

Ta thấy:  $p_3^* > p_4^* > p_1^* > p_5^* > p_2^*$ . Trước hết, so sánh tỷ lệ sâu non mới nở trong tháng 3 với tháng 4:

Bài toán:  $H: p_3 = p_4 \mid K: p_3 > p_4 \quad (\text{mức ý nghĩa } \alpha = 0,05)$

Miền bác bỏ:  $S = \{ u \geq u(\alpha) \}$

$$u = \frac{\frac{m_3}{n_3} - \frac{m_4}{n_4}}{\sqrt{\frac{m_3 + m_4}{n_3 + n_4} \left(1 - \frac{m_3 + m_4}{n_3 + n_4}\right) \frac{n_3 + n_4}{n_3 n_4}}} = \frac{\frac{70}{280} - \frac{75}{515}}{\sqrt{\frac{70 + 75}{795} \left(1 - \frac{70 + 75}{795}\right) \frac{795}{280.515}}} \\ = 3,64$$

Tra bảng ta được:  $u(\alpha) = u(0,05) = 1,645$ .

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy, với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng tỷ lệ sâu non ở tháng 3 lớn hơn ở tháng 4. Kết hợp với tỷ lệ mẫu, suy ra tháng 3 có tỷ lệ sâu non lớn nhất.

\* *Ước lượng tỷ lệ sâu non ở tháng 3,  $\alpha = 0,05$ :*

Khoảng ước lượng:

$$p_3 \in \left( p_3^* - u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p_3^*(1-p_3^*)}{n_3}}, \quad p_3^* + u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p_3^*(1-p_3^*)}{n_3}} \right)$$

$$p_3 \in \left( 0,25 - u(0,025) \sqrt{\frac{0,25 \cdot 0,75}{280}}, \quad 0,25 + u(0,025) \sqrt{\frac{0,25 \cdot 0,75}{280}} \right)$$

$$p_3 \in \left( 0,25 - 1,96 \sqrt{\frac{0,25 \cdot 0,75}{280}}, \quad 0,25 + 1,96 \sqrt{\frac{0,25 \cdot 0,75}{280}} \right)$$

$p_3 \in (0,1993; \quad 0,3007)$

Vậy, với mức ý nghĩa 0,05 thì tỷ lệ sâu non ở tháng 3 nằm trong khoảng (0,1993; 0,3007).

\* *Cần bắt bao nhiêu sâu để sai số không vượt quá 0,02:*

Sai số không vượt quá 0,02 suy ra:

$$u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p_3^*(1-p_3^*)}{n_3}} \leq 0,02 \Leftrightarrow 1,96 \sqrt{\frac{0,25 \cdot 0,75}{n_3}} \leq 0,02$$

$$\Leftrightarrow n_3 \geq \frac{1,96^2}{0,02^2} \cdot 0,25 \cdot 0,75 \Leftrightarrow n_3 \geq 1800,75$$

Vậy, để sai số không vượt quá 0,02 thì ở tháng 3 phải bắt ít nhất 1801 con.

**Bài 17/201:** Đối với người Việt Nam, lượng huyết sắc tố trung bình là 138,3 g/l. Khám cho 80 công nhân ở nhà máy có tiếp xúc hóa chất thấy huyết sắc tố trung bình là 120 g/l;  $s = 15$  g/l. Từ kết quả trên có thể kết luận lượng huyết sắc tố trung bình của công nhân nhà máy này thấp hơn mức chung hay không ( $\alpha = 0,05$ ).

*Hướng dẫn: Dạng bài kiểm định giả thiết cho giá trị trung bình, trường hợp chưa biết phân bố chuẩn, chưa biết phương sai lý thuyết nhưng cỡ mẫu lớn.*

Bài toán:

$$H: \mu = 138,3 \quad | \quad K: \mu < 138,3 \quad (\text{mức ý nghĩa } \alpha = 0,05)$$

Miền bác bỏ:

$$S = \left\{ \frac{\bar{X} - \mu_0}{s} \sqrt{n-1} \leq -u(\alpha) \right\}$$

Ta có:

$$\frac{\bar{X} - \mu_0}{s} \sqrt{n-1} = \frac{120 - 138,3}{15} \sqrt{79} = -10,84$$

Tra bảng của phân bố chuẩn tắc ta được:  $-u(\alpha) = -u(0,05) = -1,645$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng lượng huyết sắc tố trung bình của công nhân thấp hơn mức chung của toàn quốc.

**Bài 18/201: Đo huyết sắc tố cho 50 công nhân nông trường thấy có 60% ở mức dưới 110 g/l. Số liệu chung của khu vực này là 30% ở mức dưới 110 g/l. Với mức ý nghĩa  $\alpha = 0,1$  có thể kết luận công nhân nông trường có tỷ lệ huyết sắc tố dưới 110 g/l cao hơn mức chung hay không?**

*Hướng dẫn: Dạng bài toán kiểm định giả thiết cho tỷ lệ.*

Bài toán:

$$H: p = 0,3 \quad | \quad K: p > 0,3 \quad (\text{mức ý nghĩa } \alpha = 0,1)$$

Miền bác bỏ:

$$S = \left\{ \frac{p^* - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} \geq u(\alpha) \right\}$$

Tính:

$$\frac{p^* - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} = \frac{0,6 - 0,3}{\sqrt{0,3 \cdot 0,7}} \sqrt{50} = 4,629$$

Tra bảng của phân bố chuẩn tắc ta được:

$$u(\alpha) = u(0,1) = 1,282$$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy với mức ý nghĩa  $\alpha = 0,1$  ta tạm thời cho rằng tỷ lệ huyết sắc tố dưới 110 g/l của công nhân cao hơn mức chung.

**Bài 19/202: Hàm lượng đường trong máu của công nhân sau 3 giờ làm việc với máy siêu cao tần đã được đo ở hai thời điểm trước và sau 3 giờ làm việc. Ta có kết quả sau:**

Trước:  $n_1 = 50; \bar{X} = 60 \text{ mg\%}; s_1 = 7$

Sau:  $n_2 = 40; \bar{Y} = 52 \text{ mg\%}; s_2 = 9,2$

Với mức ý nghĩa  $\alpha = 0,05$  có thể khẳng định hàm lượng đường trong máu sau 3 giờ làm việc đã giảm đi hay không?

Hướng dẫn: Dạng bài toán so sánh hai giá trị trung bình, trường hợp chưa biết phương sai, chưa biết phân bố chuẩn nhưng có cỡ mẫu lớn.

Gọi  $\mu_1$  và  $\mu_2$  lần lượt là kỳ vọng lý thuyết thời điểm trước và sau 3 giờ làm việc.

Bài toán:

$$H: \mu_1 = \mu_2 \quad | \quad K: \mu_1 > \mu_2 \quad (\text{mức ý nghĩa } \alpha = 0,05)$$

Miền bác bỏ:

$$S = \{u \geq u(\alpha)\}$$

Trong đó:

$$u = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n_1-1} + \frac{s_2^2}{n_2-1}}} = \frac{60 - 52}{\sqrt{\frac{7^2}{49} + \frac{9,2^2}{39}}} = 4,493$$

Tra bảng của phân bố chuẩn tắc ta được:  $u(\alpha) = u(0,05) = 1,645$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng hàm lượng đường trong máu của công nhân sau 3 giờ làm việc đã giảm đi.

**Bài 20/202<sup>1</sup>:** Đánh giá tác dụng của một chế độ ăn bồi dưỡng mà dấu hiệu quan sát là số hồng cầu. Người ta đếm số hồng cầu của 20 người trước và sau khi ăn bồi dưỡng:

x <sub>i</sub>	32	40	38	42	41	35	36	47	50	30
y <sub>i</sub>	40	45	42	50	52	43	48	45	55	34
x <sub>i</sub>	38	45	43	36	50	38	42	41	45	44
y <sub>i</sub>	32	54	58	30	60	35	50	48	40	50

Với  $\alpha = 0,05$  có thể kết luận hai dãy số liệu trên là thuần nhất (tức là hai mẫu đều rút ra từ một đại lượng ngẫu nhiên) hay không?

Hướng dẫn: Dạng bài so sánh hai giá trị trung bình, trường hợp phi tham số (sinh viên ngoài khoa Toán không cần ôn tập dạng bài này). Vì mẫu được cho dưới dạng cặp số liệu (mẫu phụ thuộc) nên sử dụng tiêu chuẩn Wilcoxon.

Bài toán:

$$\begin{cases} H: \text{Hai mẫu là thuần nhất (EX = EY)} \\ K: \text{Hai mẫu không thuần nhất (EX} \neq \text{EY)} \\ \text{Mức ý nghĩa } \alpha = 0,05 \end{cases}$$

+ Tính  $d_i = (x_i - y_i)$  và lấy các giá trị  $d_i \neq 0$  sắp xếp theo chiều tăng dần, ta được:

Giá trị d <sub>i</sub>	-15	-12	-11	-10	-9	-8	-8	-8	-8	-7
STT	1	2	3	4	5	6	7	8	9	10
Giá trị d <sub>i</sub>	-6	-5	-5	-4	-4	2	3	5	6	6
STT	11	12	13	14	15	16	17	18	19	20

Gọi n là số  $d_i \neq 0$ . Suy ra: n = 20

<sup>1</sup> Dạng bài so sánh hai giá trị trung bình, tiêu chuẩn phi tham số. Sinh viên khoa ngoài không cần ôn tập bài này.

+ Tính hạng của  $|d_i|$ :

Giá trị $ d_i $	2	3	4	5	6	7	8	9	10	11	12	15
Rank	1	2	3,5	6	9	11	13,5	16	17	18	19	20

+ Tính T, ET, DT:

$$T = \sum_{d_i > 0} \text{Rank}(|d_i|) = \text{rank}(2) + \text{rank}(3) + \text{rank}(5) + 2 \cdot \text{rank}(6)$$

$$= 1 + 2 + 6 + 2 \cdot 9 = 27$$

$$ET = \frac{n(n+1)}{4} = \frac{20 \cdot 21}{4} = 105$$

$$DT = \frac{n(n+1)(2n+1)}{24} = \frac{20 \cdot 21 \cdot 41}{24} = 717,5$$

⇒ Miền bắc bỏ:

$$S = \left\{ \frac{|T - ET|}{\sqrt{DT}} \geq u(\alpha/2) \right\}$$

Tính:

$$\frac{|T - ET|}{\sqrt{DT}} = \frac{|27 - 105|}{\sqrt{717,5}} = 2,912$$

Tra bảng của phân bố chuẩn tắc ta được:  $u(\alpha/2) = u(0,025) = 1,96$

Do đó miền bắc bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng hai mẫu số liệu không thuận nhất (không được rút ra từ một đại lượng ngẫu nhiên). Ngoài ra, ta thấy  $\bar{X} < \bar{Y}$  nên có thể cho rằng chế độ ăn bồi dưỡng đã làm tăng số lượng hỏng cầu.

**Bài 21/202:** Gọi X là số người tới một trạm điện thoại trong thời gian 3 phút. Theo dõi 50 khoảng thời gian như vậy ta có các số liệu sau:

Số người đến (X)	0	1	2	3	4	5	6
Số khoảng xảy ra	8	15	12	9	4	1	1

Với mức ý nghĩa  $\alpha = 0,05$  có thể kết luận X tuân theo luật phân bố Poisson hay không?

*Hướng dẫn:* Dạng bài toán “Tiêu chuẩn phù hợp Khi bình phương”. Trong trường hợp này, phân bố Poisson phụ thuộc tham số  $\lambda$  và  $\lambda$  được tính xấp xỉ bằng ước lượng điểm.

Bài toán:

$$\begin{cases} H: X \text{ có phân bố Poisson} \\ K: X \text{ không có phân bố Poisson} \\ \text{Mức ý nghĩa } \alpha = 0,05 \end{cases}$$

Uớc lượng điểm cho tham ẩn  $\lambda$  của phân bố Poisson:

$$\lambda^* = \bar{X} = 1,86$$

Tiến hành gộp khoảng để thỏa mãn  $m_i \geq 5 \forall i$ . Ta được:

Số người đến (X)	0	1	2	3	4
Số khoảng xảy ra	8	15	12	9	6

Tính các xác suất  $p_i$  từ tham ẩn vừa tìm được:

$$P(X = 0) = e^{-1,86} \frac{1,86^0}{0!} = 0,1557$$

$$P(X = 1) = e^{-1,86} \frac{1,86^1}{1!} = 0,2896$$

$$P(X = 2) = e^{-1,86} \frac{1,86^2}{2!} = 0,2693$$

$$P(X = 3) = e^{-1,86} \frac{1,86^3}{3!} = 0,1670$$

$$P(X \geq 4) = 1 - P(X < 4) = 0,1184$$

$\Rightarrow$  Miền bắc bỏ của bài toán:

$$S = \left\{ \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n \geq \chi_{k-r-1}^2(\alpha) \right\}$$

(với  $k = 5$  là số khoảng và  $r = 1$  là số lượng tham ẩn)

Tính:

$$\frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n = \frac{1}{50} \left( \frac{8^2}{0,1557} + \frac{15^2}{0,2896} + \frac{12^2}{0,2693} + \frac{9^2}{0,1670} + \frac{6^2}{0,1184} \right) - 50 = 0,236$$

Tra bảng của phân bố Khi bình phương ta được:

$$\chi_{k-r-1}^2(\alpha) = \chi_{5-1-1}^2(0,05) = \chi_3^2(0,05) = 7,81$$

Do đó miền bắc bỏ không xảy ra, ta chấp nhận  $H$  và bác bỏ  $K$ . Vậy, với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng  $X$  tuân theo luật Poisson.

**Bài 22/203: Gieo đồng thời 2 đồng tiền 50 lần. Tần số xuất hiện số mặt sấp là:**

Số mặt sấp	0	1	2
Tần số xuất hiện	10	28	12

Với mức ý nghĩa  $\alpha = 0,05$  có thể kết luận 2 đồng tiền là cân đối và đồng chất hay không?

Hướng dẫn: Dạng bài toán “Tiêu chuẩn phù hợp Khi bình phương”. Nếu hai đồng tiền cân đối và đồng chất thì số lần xuất hiện mặt sấp (0, 1, 2) sẽ ứng với tỷ lệ 0,25:0,5:0,25

Gọi X là số lần xuất hiện mặt sấp. Bài toán:

$$\begin{cases} H: X \text{ có phân bố } 0,25 : 0,5 : 0,25 \text{ (hai đồng tiền cân đối và đồng chất)} \\ K: X \text{ không có phân bố như trên} \\ \text{Mức ý nghĩa } \alpha = 0,05 \end{cases}$$

Miền bác bỏ:

$$S = \left\{ \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n \geq \chi_{k-1}^2(\alpha) \right\} \quad (\text{với } k = 3 \text{ là số khoảng})$$

Ta có:

$$\frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n = \frac{1}{50} \left( \frac{10^2}{0,25} + \frac{28^2}{0,5} + \frac{12^2}{0,25} \right) - 50 = 0,88$$

Tra bảng của phân bố Khi bình phương ta được:

$$\chi_{k-1}^2(\alpha) = \chi_2^2(0,05) = 5,99$$

Do đó miền bác bỏ không xảy ra, ta chấp nhận H và bác bỏ K. Vậy, với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng hai đồng tiền là cân đối và đồng chất

**Bài 23/203:** Tiến hành 50 quan sát độc lập về thời gian ngồi uống bia của khách, ta được các số liệu sau:

Khoảng thời gian (T phút)	(0, 5)	[5, 10)	[10, 15)	[15, 20)	$\geq 20$
Số người	10	20	8	7	5

Với  $\alpha = 0,1$  thử xem T có tuân theo phân bố mũ hay không?

Hướng dẫn: Dạng bài toán “Tiêu chuẩn phù hợp Khi bình phương”. Phân bố mũ với tham số  $\lambda$  được xấp xỉ bằng ước lượng điểm.

Bài toán:

$$\begin{cases} H: T \text{ có phân bố mũ} \\ K: T \text{ không có phân bố mũ} \\ \text{Mức ý nghĩa } \alpha = 0,1 \end{cases}$$

Sử dụng máy tính ta có:

$$\bar{T} = 10,2$$

$\Rightarrow$  Ước lượng điểm cho tham số  $\lambda$  của phân bố mũ là:

$$\lambda^* = \frac{1}{\bar{T}} = \frac{1}{10,2} = 0,098$$

Hàm phân bố của phân bố mũ có dạng:  $F(t) = 1 - e^{-\lambda t}$ . Tính các xác suất  $p_i$  dựa vào tham số vừa tìm được, ứng với từng khoảng giá trị của  $t$  trong mẫu số liệu:

$$p_1 = F(5) - F(0) = F(5) = 1 - e^{-0,098 \cdot 5} = 0,3874$$

$$p_2 = F(10) - F(5) = e^{-0,098 \cdot 5} - e^{-0,098 \cdot 10} = 0,2373$$

$$p_3 = F(15) - F(10) = e^{-0,098 \cdot 10} - e^{-0,098 \cdot 15} = 0,1454$$

$$p_4 = F(20) - F(15) = e^{-0,098 \cdot 15} - e^{-0,098 \cdot 20} = 0,0890$$

$$p_5 = 1 - F(20) = e^{-0,098 \cdot 20} = 0,1409$$

⇒ Miền bác bỏ của bài toán:

$$S = \left\{ \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n \geq \chi_{k-r-1}^2(\alpha) \right\}$$

(với  $k = 5$  là số khoảng và  $r = 1$  là số lượng tham số)

Tính:

$$\frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n = \frac{1}{50} \left( \frac{10^2}{0,3874} + \frac{20^2}{0,2373} + \frac{8^2}{0,1454} + \frac{7^2}{0,0890} + \frac{5^2}{0,1409} \right) - 50 = 12,24$$

Tra bảng của phân bố Khi bình phương ta được:

$$\chi_{k-r-1}^2(\alpha) = \chi_{5-1-1}^2(\alpha) = \chi_3^2(0,1) = 6,25$$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy, với mức ý nghĩa  $\alpha = 0,1$  ta tạm thời cho rằng T không tuân theo phân bố mũ.

**Bài 24/203:** Trong đợt thi đua, phân xưởng I báo cáo chất lượng sản phẩm làm ra như sau: có 85% loại I, 10% loại II và 5% loại III. Ban thi đua đã lấy ngẫu nhiên từ lô sản phẩm chưa phân loại của phân xưởng I ra 100 sản phẩm, thấy có 80 loại I, 13 loại II và 7 loại III. Với mức ý nghĩa  $\alpha = 0,1$  có thể kết luận gì về báo cáo của phân xưởng I.

*Hướng dẫn: Dạng bài “Tiêu chuẩn phù hợp Khi bình phương”. Kiểm tra xem mẫu số liệu có phân bố  $0,85 : 0,10 : 0,05$  hay không.*

Bài toán:  $\begin{cases} H: X \text{ có phân bố } 0,85 : 0,10 : 0,05 \text{ (với } X \text{ là chất lượng sản phẩm)} \\ K: X \text{ không có phân bố như trên} \\ \text{Mức ý nghĩa } \alpha = 0,1 \end{cases}$

Miền bác bỏ:

$$S = \left\{ \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n \geq \chi_{k-1}^2(\alpha) \right\} \quad (\text{với } k = 3 \text{ là số loại})$$

Ta có:

$$\frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n = \frac{1}{100} \left( \frac{80^2}{0,85} + \frac{13^2}{0,1} + \frac{7^2}{0,05} \right) - 100 = 1,994$$

Tra bảng của phân bố Khi bình phương ta được:

$$\chi^2_{k-1}(\alpha) = \chi^2_2(0,1) = 4,61$$

Do đó miền bác bỏ không xảy ra, ta chấp nhận H và bác bỏ K. Vậy, với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng báo cáo của phân xưởng I là đúng.

**Bài 25/203:** Một nhà máy có 3 phân xưởng cùng sản xuất một loại sản phẩm. Chất lượng sản phẩm được chia thành 3 loại. Kiểm tra, phân loại ngẫu nhiên một số sản phẩm từ lô sản phẩm của 3 phân xưởng ta có số liệu sau:

Chất lượng \ Phân xưởng	Phân xưởng I	Phân xưởng II	Phân xưởng III
Loại I	70	80	60
Loại II	25	20	15
Loại III	5	10	5

Với  $\alpha = 0,05$  có thể kết luận chất lượng sản phẩm phụ thuộc vào nơi làm ra chúng hay không?

*Hướng dẫn: Dạng bài toán kiểm tra tính độc lập.*

Bài toán:

$$\begin{cases} H: \text{Chất lượng sản phẩm độc lập với nơi sản xuất} \\ K: \text{Chất lượng sản phẩm phụ thuộc vào nơi sản xuất} \\ \text{Mức ý nghĩa } \alpha = 0,05 \end{cases}$$

$\Rightarrow$  Miền bác bỏ:  $S = \left\{ \chi^2 \geq \chi^2_{(r-1),(s-1)}(\alpha) \right\}$

(với  $r = 3$  là số mức chất lượng,  $s = 3$  là số phân xưởng)

Ta có:

$$\begin{aligned} \chi^2 &= n \left( \sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_{i\cdot} \cdot n_{\cdot j}} - 1 \right) = 290 \left( \frac{70^2}{100 \cdot 210} + \frac{80^2}{110 \cdot 210} + \frac{60^2}{80 \cdot 210} + \frac{25^2}{100 \cdot 60} + \right. \\ &\quad \left. + \frac{20^2}{110 \cdot 60} + \frac{15^2}{80 \cdot 60} + \frac{5^2}{100 \cdot 20} + \frac{10^2}{110 \cdot 20} + \frac{5^2}{80 \cdot 20} - 1 \right) = 2,87 \end{aligned}$$

Tra bảng của phân bố Khi bình phương:  $\chi^2_{(r-1),(s-1)}(\alpha) = \chi^2_4(0,05) = 9,49$

Do đó miền bác bỏ không xảy ra, ta chấp nhận H và bác bỏ K. Vậy, với mức ý nghĩa  $\alpha = 0,05$  ta tạm thời cho rằng chất lượng sản phẩm độc lập với nơi sản xuất (không phụ thuộc vào nơi làm ra chúng).

## 5.2. Nhận xét bài tập chương 5

Chương 5 chứa nội dung chủ yếu của phần thống kê và **không thể không** xuất hiện trong đề thi cuối kỳ. Các câu hỏi đôi khi có phần lắt léo nên phải cẩn thận trong việc xác định dạng bài toán, một số câu hỏi yêu cầu hiểu được ý nghĩa của phép kiểm định mới xác định đúng dạng cần áp dụng.

+ Các dạng bài thường có câu hỏi rõ ràng và xác định được ngay: kiểm định giả thiết cho giá trị trung bình, kiểm định giả thiết cho phương sai, kiểm định giả thiết cho tỷ lệ, so sánh hai giá trị trung bình, so sánh hai phương sai, so sánh hai tỷ lệ, kiểm tra tính độc lập, so sánh nhiều tỷ lệ.

+ Dạng bài “Tiêu chuẩn phù hợp Khi bình phương” hầu như kỳ nào cũng xuất hiện và một số cách hỏi có thể gây nhầm lẫn hoặc khó xác định dạng.

- Kiểm tra xem các số liệu mẫu có tuân theo tỷ lệ  $p_1 : p_2 : \dots : p_n$  cho trước hay không. Nhiều trường hợp làm nhầm thành nhiều bài toán kiểm định giả thiết cho tỷ lệ.
- Kiểm tra xem các số liệu mẫu có tuân theo luật phân bố  $F(x)$  nào đó hay không. Phân bố  $F(x)$  chưa xác định và phụ thuộc tham số. Ước lượng điểm cho tham số và tính các xác suất  $p_i$ . Chú ý gộp khoảng sao cho  $m_i \geq 5 \forall i$  (nhiều sinh viên quên gộp khoảng)

+ Dạng bài kiểm định xem tỷ lệ nào lớn nhất. Từ số liệu đã cho, ta tính được các tỷ lệ mẫu:  $p_1^*, p_2^*, \dots, p_n^*$  nhưng không nhất thiết phải làm đến  $C_n^2$  bài toán kiểm định. So sánh các tỷ lệ mẫu với nhau, giả sử thấy  $p_1^* > p_2^* > \dots > p_n^*$ ; trước hết kiểm định bài toán  $p_1 = p_2 | p_1 > p_2$ . Nếu kết quả nhận được  $p_1 > p_2$  thì kết luận luôn  $p_1$  lớn nhất. Nếu kết quả nhận được  $p_1 = p_2$  thì kiểm định tiếp 2 bài toán: giữa  $p_1$  và  $p_3$ , giữa  $p_2$  và  $p_3$ .

# CHƯƠNG 6: BÀI TOÁN TƯƠNG QUAN VÀ HỒI QUY

## A. LÝ THUYẾT

### 6.1. Hệ số tương quan mẫu

Giả sử mẫu ngẫu nhiên có n về hai đại lượng ngẫu nhiên X, Y được cho theo từng cặp  $(x_i, y_i)$ . Rút gọn mẫu thành k khoảng và gọi  $m_i$  là số lần xuất hiện cặp  $(x_i, y_i)$  sao cho:  $\sum m_i = n$

$x_i$	$x_1$	$x_2$	...	$x_k$
$y_i$	$y_1$	$y_2$	...	$y_k$
$m_i$	$m_1$	$m_2$	...	$m_k$

Hệ số tương quan mẫu:

$$r = \frac{\frac{1}{n} \sum_{i=1}^k x_i y_i m_i - \bar{X} \cdot \bar{Y}}{s_x s_y}$$

Giá trị r đặc trưng cho mức độ phụ thuộc tuyến tính của mẫu về X và Y

- + Miền giá trị:  $|r| \leq 1$
- + Tương quan tuyến tính rất chặt:  $0,9 \leq |r| \leq 1$
- + Tương quan tuyến tính chặt:  $0,8 \leq |r| < 0,9$
- + Tương quan tuyến tính khá chặt:  $0,6 \leq |r| < 0,8$

### 6.2. Đường hồi quy tuyến tính thực nghiệm<sup>1</sup>

\* Phương trình đường hồi quy tuyến tính thực nghiệm của Y theo X:

$$y - \bar{Y} = r \frac{s_y}{s_x} (x - \bar{X})$$

Sai số bình phương trung bình khi xấp xỉ Y theo X:

$$s_{Y/X}^2 = s_y^2 (1 - r^2)$$

\* Phương trình đường hồi quy tuyến tính thực nghiệm của X theo Y:

$$x - \bar{X} = r \frac{s_x}{s_y} (y - \bar{Y})$$

Sai số bình phương trung bình khi xấp xỉ X theo Y:

$$s_{X/Y}^2 = s_x^2 (1 - r^2)$$

Ý nghĩa của phương trình đường hồi quy tuyến tính thực nghiệm: dùng để xấp xỉ Y theo X (hoặc xấp xỉ X theo Y).

<sup>1</sup> Tên đầy đủ là: "Đường hồi quy bình phương trung bình tuyến tính thực nghiệm"

## B. BÀI TẬP

### 6.1. Bài tập trong giáo trình 2 (G<sub>2</sub>)

(Xác suất thống kê, Đào Hữu Hò, trang 231)

**Bài 1/231:** Nghiên cứu mối liên hệ giữa X là số tiền đầu tư cho việc phòng bệnh tính trên đầu người và Y là tỷ lệ người mắc bệnh ở 50 địa phương thu được bảng tương quan thực nghiệm sau:

X (đồng) \ Y (%)	2	2,5	3	3,5	4
100				2	3
200			3	6	2
300		4	6	3	
400	1	6	4	1	
500	6	3			

- a) Tìm hệ số tương quan mẫu
- b) Xây dựng đường hồi quy tuyến tính thực nghiệm của Y theo X.
- c) Nếu năm sau đầu tư cho phòng bệnh là 600đ/người thì tỷ lệ mắc bệnh khoảng bao nhiêu %?
- d) Ước lượng phương sai phần dư.

a) Tìm hệ số tương quan mẫu:

$$r = \frac{\frac{1}{n} \sum_{i=1}^k x_i y_i m_i - \bar{X} \cdot \bar{Y}}{s_X s_Y}$$

Sử dụng máy tính ta có<sup>1</sup>:

$$\bar{X} = 318; s_X^2 = 15476$$

$$\bar{Y} = 2,95; s_Y^2 = 0,3625$$

Do đó:

$$r = \frac{\frac{1}{50} \left( 100 \cdot 3,5 \cdot 2 + 100 \cdot 4 \cdot 3 + 200 \cdot 3 \cdot 3 + 200 \cdot 3,5 \cdot 6 + 200 \cdot 4 \cdot 2 + 300 \cdot 2,5 \cdot 4 + 300 \cdot 3 \cdot 6 + 300 \cdot 3,5 \cdot 3 + 400 \cdot 2 \cdot 1 + 400 \cdot 2,5 \cdot 6 + 400 \cdot 3 \cdot 4 + 400 \cdot 3,5 \cdot 1 + 500 \cdot 2 \cdot 6 + 500 \cdot 2,5 \cdot 3 \right) - 318 \cdot 2,95}{\sqrt{15476 \cdot 0,3625}}$$

$$\Leftrightarrow r = \frac{\frac{43800}{50} - 318 \cdot 2,95}{\sqrt{15476 \cdot 0,3625}} \Leftrightarrow r = -0,8291$$

<sup>1</sup> Được phép sử dụng máy tính bỏ túi để tính toán giá trị trung bình mẫu, phương sai mẫu, độ lệch mẫu nhưng phải viết rõ công thức tính hệ số tương quan r.

b) Xây dựng đường hồi quy tuyến tính thực nghiệm của Y theo X:

Đường hồi quy tuyến tính thực nghiệm của Y theo X có dạng:

$$y - \bar{Y} = r \frac{s_Y}{s_X} (x - \bar{X}) \Leftrightarrow y - 2,95 = -0,8291 \cdot \sqrt{\frac{0,3625}{15476}} (x - 318)$$

$$\Leftrightarrow y = -0,004013x + 4,226$$

c) Tỷ lệ mắc bệnh là bao nhiêu năm sau đầu tư cho phòng bệnh là 600đ/người:

Sử dụng đường hồi quy tuyến tính thực nghiệm để xấp xỉ tỷ lệ người mắc bệnh theo số tiền đầu tư cho việc phòng bệnh.

⇒ Tỷ lệ người mắc bệnh của năm sau:

$$y = -0,004013 \cdot 600 + 4,226 = 1,8182 \quad (\%)$$

d) Ước lượng phuông sai phần dư:

Sinh viên ngoài khoa Toán không cần ôn tập dạng này!

**Bài 2/231:** Qua nhiều tác giả nghiên cứu cho thấy giữa lượng đạm (N) và carbon (C) trong mùn có liên hệ với nhau theo dạng tuyến tính. Hãy xác nhận lại nhận định trên qua ví dụ về đất lâm nghiệp ở Quảng Ninh sau đây: (số liệu trích từ bộ môn Đất, trường Đại học Lâm nghiệp)

Hàm lượng C	1,79	4,39	3,07	4,4	3,1	5,6	7,81	3,95	4,71
Hàm lượng N	0,06	0,42	0,18	0,3	0,22	0,38	0,46	0,23	0,42

Tìm hệ số tương quan mẫu và xây dựng đường hồi quy tuyến tính thực nghiệm của N theo C.

\* Hệ số tương quan mẫu:

$$r = \frac{\frac{1}{n} \sum_{i=1}^k x_i y_i m_i - \bar{X} \bar{Y}}{s_X s_Y}$$

Gọi X là hàm lượng carbon và Y là hàm lượng nitơ. Sử dụng máy tính ta có:

$$\bar{X} = 4,3133; s_X^2 = 2,6192$$

$$\bar{Y} = 0,2967; s_Y^2 = 0,016$$

Do đó:

$$r = \frac{\frac{1}{9} \left( 1,79 \cdot 0,06 + 4,39 \cdot 0,42 + 3,07 \cdot 0,18 + 4,4 \cdot 0,3 + 3,1 \cdot 0,22 + \right. \\ \left. + 5,6 \cdot 0,38 + 7,81 \cdot 0,46 + 3,95 \cdot 0,23 + 4,71 \cdot 0,42 \right) - 4,3133 \cdot 0,2967}{\sqrt{2,6192 \cdot 0,016}}$$

$$\Leftrightarrow r = \frac{\frac{13,1131}{9} - 4,3133 \cdot 0,2967}{\sqrt{2,6192 \cdot 0,016}} \Leftrightarrow r = 0,8659 \quad (\text{tương quan chặt})$$

\* Đường hồi quy tuyến tính thực nghiệm của N theo C:

Đường hồi quy tuyến tính thực nghiệm của N theo C có dạng:

$$y - \bar{Y} = r \frac{s_y}{s_x} (x - \bar{X}) \Leftrightarrow y - 0,2967 = 0,8659 \cdot \sqrt{\frac{0,016}{2,6192}} (x - 4,3133)$$

$$\Leftrightarrow y = 0,0677x - 0,0048$$

**Bài 3/232** và **Bài 4/232** không thuộc nội dung ôn tập cho sinh viên ngoài khoa Toán!

## 6.2. Nhận xét bài tập chương 6

Dạng bài tập chương 6 khá đơn giản và rõ ràng, chỉ cần tính toán cẩn thận:

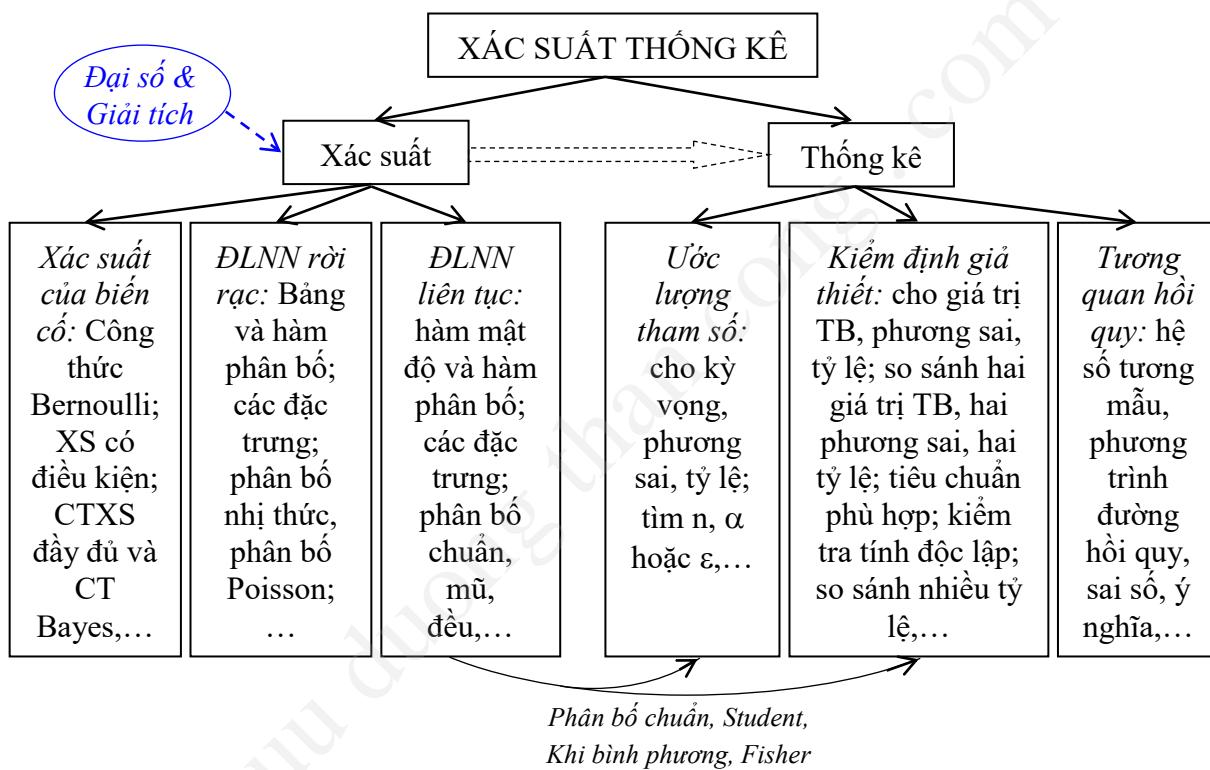
- + Tìm hệ số tương quan mẫu r: các giá trị trung gian như  $\bar{X}$ ,  $\bar{Y}$ ,  $s_x$ ,  $s_y$  có thể dùng bằng máy tính bỏ túi nhưng phải viết rõ công thức tính r.
- + Tìm đường hồi quy tuyến tính thực nghiệm: lưu ý sự khác nhau giữa phương trình hồi quy tuyến tính của Y theo X và của X theo Y, tránh nhầm lẫn.
- + Ước lượng sai số khi xấp xỉ Y theo X hoặc X theo Y.

# KẾT LUẬN

Trên đây là những kiến thức cơ bản về Xác suất Thống kê dành cho sinh viên ngoài khoa Toán. Do kiến thức khá nhiều so với thời lượng 3 tín chỉ nên các bạn cần ôn tập ngay từ đầu, cố gắng nắm vững lý thuyết trên lớp trước khi áp dụng vào làm bài tập cụ thể.

Các dạng bài tập cơ bản đã được tóm tắt trong phần nhận xét cuối mỗi chương, chỉ cần vận dụng chúng sao cho linh hoạt. Các nội dung phụ thuộc chặt chẽ với nhau và việc hiểu phần trước là cơ sở để hiểu phần sau, kiến thức xác suất là cơ sở cho các phép ước lượng, kiểm định giả thiết trong thống kê. Do đó, nhất thiết phải nắm được toàn bộ nội dung cơ bản như một thể thống nhất.

Có thể hình dung sơ đồ môn học như sau:



Đối với những bạn không có điều kiện ôn tập từ đầu, sắp đến ngày thi cuối kỳ mà chưa học được nhiều thì có thể xem một số dạng thường gặp dưới đây:

\* Phân xác suất:

- (1) Công thức xác suất đầy đủ và công thức Bayes (rất hay xuất hiện)
  - (2) Đại lượng ngẫu nhiên rời rạc: lập bảng phân bố xác suất; tìm kỳ vọng, phương sai; phân bố nhị thức, phân bố Poisson.
  - (3) Phân bố chuẩn, phân bố mũ: tìm kỳ vọng, phương sai, median, tính xác suất

### \* Phân thống kê:

- (4) Uớc lượng khoảng cho giá trị trung bình, tìm số quan sát cần thiết.
  - (5) Uớc lượng khoảng cho tỷ lệ.

- (6) Kiểm định giả thiết cho giá trị trung bình (hay xuất hiện)
- (7) Kiểm định giả thiết cho tỷ lệ.
- (8) So sánh hai giá trị trung bình (hay xuất hiện)
- (9) Tiêu chuẩn phù hợp Khi bình phương (rất hay xuất hiện)
- (10) Tìm hệ số tương quan mẫu, lập phương trình hồi quy tuyến tính và nêu ý nghĩa của phương trình, sai số bình phương trung bình (rất hay xuất hiện).

*Bảng thống kê một số dạng thường gặp trong các đề thi cuối kỳ (đề thi bắt đầu từ trang 118):*

	Dạng (1)	Dạng (2)	Dạng (3)	Dạng (4)	Dạng (5)	Dạng (6)	Dạng (7)	Dạng (8)	Dạng (9)	Dạng (10)
Đề 1		Câu 1	Câu 2	Câu 3a				Câu 3b	Câu 5	Câu 4
Đề 2	Câu 1		Câu 2		Câu 3c		Câu 3b		Câu 3a	Câu 4
Đề 3	Câu 1b		Câu 1a, 2		Câu 3b			Câu 3a	Câu 3c	Câu 4
Đề 4		Câu 1a	Câu 2	Câu 4a			Câu 4b		Câu 4c	Câu 3
Đề 5	Câu 2	Câu 1		Câu 4b				Câu 4a	Câu 3	Câu 5
Đề 6	Câu 1		Câu 2a	Câu 3a		Câu 3c			Câu 3d	Câu 4
Đề 7	Câu 1a	Câu 2			Câu 3c	Câu 3a				Câu 3d,e
Đề 8	Dành riêng cho sinh viên khoa Vật lý, hệ chuẩn									
Đề 9	Câu 1					Câu 3		Câu 4		Câu 5
Đề 10	Dành riêng cho sinh viên khoa Vật lý, hệ chuẩn									
Đề 11		Câu 1		Câu 3a				Câu 3b		Câu 4

# MỘT SỐ ĐỀ THI CUỐI KỲ

## 1. Đề thi cuối kỳ II năm học 2012 – 2013

ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

ĐỀ THI KẾT THÚC HỌC KỲ II  
NĂM HỌC 2012 – 2013

### Môn thi: Xác suất thống kê

Mã môn học: MAT 1101      Số tín chỉ: 03

Dành cho sinh viên các khoa: Hóa học, Sinh học, Địa lý, Địa chất, Môi trường, KTTV-HDH, Y-Dược

Dạng đề thi: **Được sử dụng tài liệu**

Thời gian làm bài: 90 phút (không kể thời gian phát đề)

**Câu 1.** Cho  $X$  là biến ngẫu nhiên rời rạc có thể nhận các giá trị  $x_i = \{1, 2, 4, 8, 16\}$ . Biết rằng  $P(X = x_i) = k/x_i$  với  $k > 0$ :

- a) Lập bảng phân phối của  $X$  và tính kỳ vọng  $E(X)$ , phương sai  $D(X)$ .
- b) Cho  $Y = (X - 4)^2$ . Tính  $E(Y), D(Y)$ .

**Câu 2.** Các nghiên cứu cho thấy tuổi bắt đầu biết đi của trẻ em tuân theo phân phối chuẩn với kỳ vọng  $\mu = 13$  (tháng) và độ lệch chuẩn  $\sigma = 1,5$  (tháng).

- a) Tính xác suất để một em bé bắt đầu biết đi trước 11 tháng tuổi? Sau 15 tháng tuổi?
- b) Xác suất để một em bé biết đi trong khoảng thời gian từ 11 tháng tuổi đến 15 tháng tuổi.
- c) Xác suất để một em bé bắt đầu biết đi vào đúng 13 tháng tuổi.
- d) Gia đình bé An nói rằng bé sẽ biết đi sớm nhất là 12 tháng và muộn nhất là khi được 15 tháng. Khẳng định này của gia đình bé An có thể sai với xác suất là bao nhiêu?

**Câu 3.** Cho  $X$  là biến ngẫu nhiên liên tục tuân theo phân phối chuẩn. Quan sát 13 giá trị của  $X$  trên nhóm đối tượng A và 15 giá trị của  $X$  trên nhóm đối tượng B, được các số liệu sau:

Nhóm A:  $n_A = 13$ ;  $\sum x_A = 1761$ ;  $\sum (x_A)^2 = 238787$

Nhóm B:  $n_B = 15$ ;  $\sum x_B = 2119$ ;  $\sum (x_B)^2 = 299819$

a) Với độ tin cậy 95%, tính ước lượng khoảng của  $\mu_A$  và  $\mu_B$  (lần lượt là kỳ vọng của  $X$  trên nhóm đối tượng A và B).

b) Hãy kiểm tra kết luận  $\mu_A < \mu_B$  với mức ý nghĩa  $\alpha = 5\%$ . Biết rằng hai phương sai  $\sigma_A^2$  và  $\sigma_B^2$  (lần lượt là phương sai của  $X$  trên nhóm A và B) là bằng nhau.

**Câu 4.** Đo lượng cholesterol trong máu Y (g/l) của 10 người ở các độ tuổi khác nhau X (năm) được bảng số liệu sau (giả thuyết rằng X, Y đều tuân theo phân phối chuẩn):

*Hoàng Văn Trọng – 0974.971.149*

x <sub>i</sub>	30	60	40	20	50	30	40	20	70	60
y <sub>i</sub>	1,6	2,5	2,2	1,4	2,7	1,8	2,1	1,5	2,8	2,6

- a) Tính hệ số tương quan mẫu
- b) Lập phương trình hồi quy tuyến tính thực nghiệm của Y theo X.
- c) Đường hồi quy tìm được có thể dùng làm gì? Vì sao?
- d) Uớc lượng sai số bình phương trung bình.

**Câu 5.** Quan sát các giá trị x<sub>i</sub> của biến ngẫu nhiên X thu được kết quả số lần xuất hiện m<sub>i</sub> như sau:

x <sub>i</sub>	1	2	3	4	5	6	7	> 7
m <sub>i</sub>	2	4	6	7	6	4	3	1

Với mức ý nghĩa 5%, hãy kiểm tra xem X có tuân theo luật phân bố Poát – xông hay không?

Lời giải:

**Câu 1:**

a) *Lập bảng phân phối của X và tính EX, DX:*

Theo đầu bài ta có: P(X = x<sub>i</sub>) = k/x<sub>i</sub>. Do đó:

$$P(X = 1) = k/1$$

$$P(X = 2) = k/2$$

$$P(X = 4) = k/4$$

$$P(X = 8) = k/8$$

$$P(X = 16) = k/16$$

Mà tổng các xác suất phải bằng 1 nên:

$$\frac{k}{1} + \frac{k}{2} + \frac{k}{4} + \frac{k}{8} + \frac{k}{16} = 1 \Leftrightarrow k\left(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16}\right) = 1$$

$$\Leftrightarrow k \cdot \frac{31}{16} = 1 \Leftrightarrow k = \frac{16}{31}$$

+ *Bảng phân phối (phân bố) xác suất của X:*

X	1	2	4	8	16
P(X = x <sub>i</sub> )	$\frac{16}{31}$	$\frac{8}{31}$	$\frac{4}{31}$	$\frac{2}{31}$	$\frac{1}{31}$

+ *Kỳ vọng của X là:*

$$EX = \sum_{i=1}^5 x_i p_i \quad \text{với } p_i = P(X = x_i)$$

$$EX = 1 \cdot \frac{16}{31} + 2 \cdot \frac{8}{31} + 4 \cdot \frac{4}{31} + 8 \cdot \frac{2}{31} + 16 \cdot \frac{1}{31} = \frac{5 \cdot 16}{31} = \boxed{\frac{80}{31}}$$

+ Phương sai của X:

$$\begin{aligned} DX &= EX^2 - (EX)^2 = \sum_{i=1}^5 x_i^2 p_i - (EX)^2 \\ &= 1^2 \cdot \frac{16}{31} + 2^2 \cdot \frac{8}{31} + 4^2 \cdot \frac{4}{31} + 8^2 \cdot \frac{2}{31} + 16^2 \cdot \frac{1}{31} - \left( \frac{80}{31} \right)^2 = \frac{496}{31} - \left( \frac{80}{31} \right)^2 = \frac{8976}{961} \\ &= \boxed{9,3403} \end{aligned}$$

b) Tính EY, DY với  $Y = (X - 4)^2$ :

Các giá trị có thể có của Y:

X	1	2	4	8	16
$Y = (X - 4)^2$	9	4	0	16	144

⇒ Bảng phân bố xác suất của Y:

Y	0	4	9	16	144
$P(Y = y_i)$	$\frac{4}{31}$	$\frac{8}{31}$	$\frac{16}{31}$	$\frac{2}{31}$	$\frac{1}{31}$

+ Kỷ vọng của Y:

$$EY = \sum_{i=1}^5 y_i p_i = 0 \cdot \frac{4}{31} + 4 \cdot \frac{8}{31} + 9 \cdot \frac{16}{31} + 16 \cdot \frac{2}{31} + 144 \cdot \frac{1}{31} = \boxed{\frac{352}{31}}$$

+ Phương sai của Y:

$$\begin{aligned} DY &= EY^2 - (EY)^2 = \sum_{i=1}^5 y_i^2 p_i - (EY)^2 \\ &= 0^2 \cdot \frac{4}{31} + 4^2 \cdot \frac{8}{31} + 9^2 \cdot \frac{16}{31} + 16^2 \cdot \frac{2}{31} + 144^2 \cdot \frac{1}{31} - \left( \frac{352}{31} \right)^2 = \frac{578928}{961} \\ &= \boxed{602,4225} \end{aligned}$$

## Câu 2:

Gọi X là tuổi bắt đầu biết đi của trẻ em. Ta có:

$$X \sim N(13; 1,5^2)$$

a) Xác suất để một em bé bắt đầu biết đi trước 11 tháng tuổi, sau 15 tháng tuổi:

$$P(X < 11) = P\left(\frac{X - 13}{1,5} < \frac{11 - 13}{1,5}\right) = P\left(Z < -\frac{2}{1,5}\right) \quad \text{với } Z \sim N(0, 1)$$

$$= P(Z < -1,33) = \Phi(-1,33) = 1 - \Phi(1,33) = 1 - 0,9082 = \boxed{0,0918}$$

(giá trị  $\Phi(x)$  được tra trong bảng phân bố chuẩn tắc)

Vì phân bố chuẩn có tính chất đối xứng nên:

$$P(X < \mu - \delta) = P(X > \mu + \delta) \quad (\mu \text{ là kỳ vọng và } \delta \text{ là số dương bất kỳ})$$

Thay  $\mu = 13$  và  $\delta = 2$  ta được:

$$P(X > 15) = P(X < 11) = \boxed{0,0918}$$

b) Xác suất để một em bé biết đi trong khoảng thời gian từ 11 đến 15 tháng tuổi:

$$\begin{aligned} P(11 < X < 15) &= 1 - P(X < 11) - P(X > 15) \\ &= 1 - 0,0918 - 0,0918 \\ &= \boxed{0,8164} \end{aligned}$$

c) Xác suất để một em bé bắt đầu biết đi vào đúng 13 tháng tuổi:

$$P(X = 13) = 0 \text{ vì } X \text{ là đại lượng ngẫu nhiên liên tục}^1$$

d) Xác suất sai của khẳng định của gia đình bé An:

Khẳng định của gia đình bé An là sai khi bé An biết đi trước 12 tháng hoặc sau 15 tháng. Xác suất sai là:

$$\begin{aligned} &P(X < 12) + P(X > 15) \\ &= \Phi\left(\frac{12-13}{1,5}\right) + 0,0918 = \Phi(-0,67) + 0,0918 = 1 - \Phi(0,67) + 0,0918 \\ &= 1 - 0,7486 + 0,0918 = \boxed{0,3432} \end{aligned}$$

### Câu 3:

a) Ước lượng khoảng cho kỳ vọng  $\mu_A$  và  $\mu_B$ , độ tin cậy 95%:

Dạng bài toán ước lượng khoảng cho kỳ vọng, trường hợp biết phân bố chuẩn nhưng chưa biết phương sai lý thuyết.

\* Ước lượng khoảng cho  $\mu_A$ , độ tin cậy 95%:

Ta có:

$$\bar{X}_A = \frac{\sum x_A}{n_A} = \frac{1761}{13} = 135,4615$$

$$s_A^2 = \frac{\sum x_A^2}{n_A} - (\bar{X}_A)^2 = \frac{238787}{13} - \left(\frac{1761}{13}\right)^2 = 18,4024$$

$\Rightarrow$  Ước lượng khoảng cho  $\mu_A$  với độ tin cậy 95%:

<sup>1</sup> Lưu ý: giá trị 13 ở đây là giá trị chính xác tuyệt đối, không được làm tròn.

$$\mu_A \in \left( \bar{X}_A - t_{n_A-1} \left( \frac{\alpha}{2} \right) \frac{s_A}{\sqrt{n_A-1}}; \quad \bar{X}_A + t_{n_A-1} \left( \frac{\alpha}{2} \right) \frac{s_A}{\sqrt{n_A-1}} \right)$$

$$\mu_A \in \left( 135,4615 - t_{12}(0,025) \sqrt{\frac{18,4024}{12}}; \quad 135,4615 + t_{12}(0,025) \sqrt{\frac{18,4024}{12}} \right)$$

$$\mu_A \in \left( 135,4615 - 2,179 \sqrt{\frac{18,4024}{12}}; \quad 135,4615 + 2,179 \sqrt{\frac{18,4024}{12}} \right)$$

$\mu_A \in (132,76; \quad 138,16)$

\* Uớc lượng khoảng cho  $\mu_A$ , độ tin cậy 95%:

Ta có:

$$\bar{X}_B = \frac{\sum x_B}{n_B} = \frac{2119}{15} = 141,2667$$

$$s_B^2 = \frac{\sum x_B^2}{n_B} - (\bar{X}_B)^2 = \frac{299819}{15} - \left( \frac{2119}{15} \right)^2 = 31,6622$$

$\Rightarrow$  Uớc lượng khoảng cho  $\mu_B$  với độ tin cậy 95%:

$$\mu_B \in \left( \bar{X}_B - t_{n_B-1} \left( \frac{\alpha}{2} \right) \frac{s_B}{\sqrt{n_B-1}}; \quad \bar{X}_B + t_{n_B-1} \left( \frac{\alpha}{2} \right) \frac{s_B}{\sqrt{n_B-1}} \right)$$

$$\mu_B \in \left( 141,2667 - t_{14}(0,025) \sqrt{\frac{31,6622}{14}}; \quad 141,2667 + t_{14}(0,025) \sqrt{\frac{31,6622}{14}} \right)$$

$$\mu_B \in \left( 141,2667 - 2,145 \sqrt{\frac{31,6622}{14}}; \quad 141,2667 + 2,145 \sqrt{\frac{31,6622}{14}} \right)$$

$\mu_B \in (138,04; \quad 144,49)$

Vậy với độ tin cậy 95% ta có thể cho rằng giá trị trung bình của X trên nhóm đối tượng A nằm trong khoảng (132,76; 138,16) và giá trị trung bình của X trên nhóm đối tượng B nằm trong khoảng (138,04; 144,49)

b) Kiểm tra kết luận  $\mu_A < \mu_B$  với mức ý nghĩa 5%

Thuộc dạng bài so sánh hai giá trị trung bình, trường hợp thứ chia biết hai phương sai nhưng hai phương sai bằng nhau và biết phân bố chuẩn.

Bài toán:  $\begin{cases} \text{Giả thiết H: } \mu_A = \mu_B \\ \text{Đối thiết K: } \mu_A < \mu_B \\ \text{Mức ý nghĩa: } \alpha = 0,05 \end{cases}$

⇒ Miền bác bỏ của bài toán:

$$S = \{ t \leq -t_{n_A+n_B-2}(\alpha) \}$$

Ta có:

$$\begin{aligned} t &= \frac{\bar{X}_A - \bar{X}_B}{\sqrt{n_A s_A^2 + n_B s_B^2}} \sqrt{\frac{n_A n_B (n_A + n_B - 2)}{n_A + n_B}} \\ &= \frac{135,4615 - 141,2667}{\sqrt{13.18,4024 + 15.31,6622}} \sqrt{\frac{13.15.(13+15-2)}{13+15}} = -2,923 \end{aligned}$$

Tra bảng phân bố Student ta được:

$$-t_{n_A+n_B-2}(\alpha) = -t_{26}(0,05) = -1,706$$

Do đó,  $t < -t_{n_A+n_B-2}(\alpha)$ . Miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K.

Vậy, với mức ý nghĩa 5% ta tạm thời cho rằng giá trị trung bình của X trên nhóm đối tượng A nhỏ hơn giá trị trung bình của X trên nhóm đối tượng B ( $\mu_A < \mu_B$ ), cho tới khi có thêm thông tin mới.

#### Câu 4:

X là độ tuổi, Y là lượng cholesterol trong máu.

$x_i$	30	60	40	20	50	30	40	20	70	60
$y_i$	1,6	2,5	2,2	1,4	2,7	1,8	2,1	1,5	2,8	2,6

a) Tìm hệ số tương quan mẫu:

$$r = \frac{\frac{1}{n} \sum x_i y_i m_i - \bar{X} \cdot \bar{Y}}{s_X s_Y}$$

Sử dụng máy tính bỏ túi ta tính được<sup>1</sup>:

$$\begin{aligned} + \frac{1}{n} \sum x_i y_i m_i &= 96,9 \\ + \bar{X} &= 42 & + s_X^2 &= 276 \\ + \bar{Y} &= 2,12 & + s_Y^2 &= 0,2456 \end{aligned}$$

⇒ Hệ số tương quan mẫu:

$$r = \frac{96,9 - 42.2,12}{\sqrt{276.0,2456}} = \boxed{0,9547}$$

Hệ số tương quan r rất lớn, do đó mối quan hệ tuyến tính giữa X và Y rất chặt.

<sup>1</sup> Xem cách tìm các chỉ số thống kê bằng máy tính bỏ túi tại *Phụ lục P.2.2*, trang 192.

b) *Lập phương trình đường hồi quy tuyến tính thực nghiệm của Y theo X:*

Phương trình đường hồi quy của Y theo X có dạng:

$$y - \bar{Y} = r \frac{s_Y}{s_X} (x - \bar{X}) \Leftrightarrow y = r \frac{s_Y}{s_X} (x - \bar{X}) + \bar{Y}$$

$$\Leftrightarrow y = 0,9547 \sqrt{\frac{0,2456}{276}} (x - 42) + 2,12$$

$$\Leftrightarrow \boxed{y = 0,0285x + 0,9239}$$

c) *Phương trình đường hồi quy tuyến tính thực nghiệm có thể dùng để dự báo (xấp xỉ) hàm lượng cholesterol trong máu theo độ tuổi nhất định.*

Ví dụ: khi độ tuổi là 80 thì hàm lượng cholesterol trong máu là:

$$0,0285 \cdot 80 + 0,9239 = 3,2 \text{ (g/l)}$$

d) *Uớc lượng sai số bình phương trung bình:*

Sai số bình phương trung bình khi xấp xỉ Y theo X:

$$s_{Y/X}^2 = s_Y^2 (1 - r^2) = 0,2456 \cdot (1 - 0,9547^2) = \boxed{0,0217}$$

### Câu 5:

Hướng dẫn: Dạng bài toán “Tiêu chuẩn phù hợp khi bình phương” vì cần kiểm tra xem các số liệu quan sát có tuân theo phân bố F nào đó hay không.

Bài toán:

- Giả thiết H: X có phân bố Poisson
- Đối thiết K: X không có phân bố Poisson
- Mức ý nghĩa  $\alpha = 0,05$

Uớc lượng điểm cho tham số  $\lambda$  của phân bố Poisson:

$$\lambda^* = \bar{X} = \frac{1.2 + 2.4 + 3.6 + 4.7 + 5.6 + 6.4 + 7.3 + 8.1}{33} = \frac{139}{33} = 4,2121$$

Tiến hành gộp khoảng để thỏa mãn  $m_i \geq 5 \forall i$ . Ta được bảng số liệu mới như sau:

$x_i$	< 3	3	4	5	> 5
$m_i$	6	6	7	6	8

Tính các xác suất  $p_i$  lý thuyết theo phân bố Poisson ( $\lambda^*$ ):

$$P(X < 3) = e^{-4,2121} \cdot \frac{4,2121^0}{0!} + e^{-4,2121} \cdot \frac{4,2121^1}{1!} + e^{-4,2121} \cdot \frac{4,2121^2}{2!} = 0,2086$$

$$P(X = 3) = e^{-4,2121} \cdot \frac{4,2121^3}{3!} = 0,1845$$

$$P(X = 4) = e^{-4,2121} \cdot \frac{4,2121^4}{4!} = 0,1943$$

$$P(X = 5) = e^{-4,2121} \cdot \frac{4,2121^5}{5!} = 0,1637$$

$$P(X > 5) = 1 - P(X \leq 5) = 1 - 0,7511 = 0,2489$$

$\Rightarrow$  Miền bác bỏ của bài toán:

$$S = \left\{ \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n \geq \chi_{k-r-1}^2 (\alpha) \right\}$$

(với  $k = 5$  là số khoảng và  $r = 1$  là số lượng tham ẩn)

Tính:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n &= \frac{1}{33} \left( \frac{6^2}{0,2086} + \frac{6^2}{0,1845} + \frac{7^2}{0,1943} + \frac{6^2}{0,1637} + \frac{8^2}{0,2489} \right) - 33 \\ &= 0,2404 \end{aligned}$$

Tra bảng của phân bố Khi bình phương ta được:

$$\chi_{k-r-1}^2 (\alpha) = \chi_3^2 (0,05) = 7,81$$

Do đó miền bác bỏ không xảy ra, ta chấp nhận  $H$  và bác bỏ  $K$ . Vậy, với mức ý nghĩa 5% ta tạm thời cho rằng  $X$  tuân theo luật Poisson.

## 2. Đề thi cuối kỳ I năm học 2013 – 2014

ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

ĐỀ THI KẾT THÚC HỌC KỲ I  
NĂM HỌC 2013 – 2014

### Môn thi: Xác suất thống kê

Mã môn học: MAT 1101 Số tín chỉ: 03

Dành cho sinh viên các khoa: Hóa học, Sinh học, Địa lý, Địa chất, Môi trường, KTTV-HDH, Y-Dược

Dạng đề thi: **Được sử dụng tài liệu**

Thời gian làm bài: 90 phút (không kể thời gian phát đề)

**Câu 1:** Người ta dùng một test để phát hiện một loại virus trong máu gây bệnh cho người (test dương tính là dấu hiệu để chẩn đoán có virus, test âm tính là dấu hiệu để chẩn đoán không có virus). Test trên có đặc tính sau:

- 95% số người mang virus khi test cho kết quả dương tính.
- 90% số người không mang virus khi test cho kết quả âm tính.

Người ta chọn ngẫu nhiên một người từ một vùng có tỉ lệ người nhiễm virus là 2/3.

- Tìm xác suất để người được chọn ra có kết quả dương tính khi thực hiện test.
- Tính xác suất để test cho chẩn đoán sai về tình trạng nhiễm virus của người được chọn ra ở trên.

**Câu 2:** Tuổi thọ của một loài côn trùng nào đó là một biến ngẫu nhiên X (đơn vị là tháng) với hàm mật độ như sau:

$$f(x) = \begin{cases} kx^2(2-x) & \text{khi } 0 \leq x \leq 2 \\ 0 & \text{khác} \end{cases}$$

- Hãy xác định hằng số k.
- Tìm hàm phân phối F(x)
- Tính xác suất để côn trùng chết trước khi nó được một tháng tuổi.
- Tìm EX, DX.

**Câu 3:** Một cửa hàng lớn chuyên buôn bán giày cũ nhân viên đến một nhà cung cấp mua một lô giày hợp đồng đang bán chạy để bán trong dịp đón năm mới tết đến với tỉ lệ các cỡ giày là: 10% cỡ to, 25% cỡ vừa, 50% cỡ trung, 15% cỡ nhỏ. Nhà cung cấp mới nhập về một lô giày thuộc loại trên bán với điều kiện phải mua cả lô, vì số lượng lớn nhân viên bên cửa hàng kiểm tra ngẫu nhiên 100 đôi thấy có 2 đôi cỡ to, 10 đôi cỡ vừa, 48 đôi cỡ trung, 40 đôi cỡ nhỏ.

- Với mức ý nghĩa  $\alpha = 0,05$  có thể chấp nhận lô hàng cung cấp phù hợp với các tỉ lệ về cỡ giày của cửa hàng không?
- Cửa hàng có thể chấp nhận tỉ lệ cỡ giày nào trong các tỉ lệ mà nhà cung cấp có không (với  $\alpha = 0,05$ )?

c) Với xác suất 0,95 tỉ lệ giày cỡ nhỏ của nhà cung cấp thấp nhất là bao nhiêu?

$$(Cho biết \chi^2_3(0,05) = 7,81; u(0,025) = 1,96)$$

**Câu 4:** Để nghiên cứu mối quan hệ giữa số người tiêm chích ma túy (X) với số người nhiễm HIV (Y) giả sử người ta thống kê ở địa phương (tỉnh, thành phố) A trọng điểm trong cả nước có các số liệu sau:

x <sub>i</sub>	100	150	200	250	300	300	350
y <sub>i</sub>	60	90	120	150	150	180	180
Số địa phương m <sub>i</sub>	1	1	1	2	1	3	1

- a) Tính hệ số tương quan mẫu r.
- b) Có nhận xét gì về sự phụ thuộc giữa số người nhiễm HIV và số người tiêm chích ma túy?
- c) Xây dựng đường hồi quy bình phương trung bình tuyến tính thực nghiệm của Y theo X.
- d) Đường hồi quy nhận được dùng để làm gì? Vì sao? Cho ví dụ.
- e) Ước lượng sai số bình phương trung bình mắc phải khi dùng đường hồi quy trên.

Lời giải:

### **Câu 1:**

Giả sử: A = “Người mang virus”

B = “Người không mang virus”

H = “Test cho kết quả dương tính”

a) Tìm xác suất để người được chọn ra có kết quả dương tính khi thực hiện test:

Theo giả thiết:

$$P(A) = 2/3; P(B) = 1/3; P(H|A) = 0,95; P(H|B) = 0,1$$

⇒ Xác suất để người được chọn ra cho kết quả dương tính khi được test:

$$P(H) = P(H | A). P(A) + P(H | B). P(B)$$

$$\Leftrightarrow P(H) = 0,95 \cdot \frac{2}{3} + 0,1 \cdot \frac{1}{3} = \boxed{\frac{2}{3}}$$

b) Xác suất để test cho chẩn đoán sai:

Test cho chẩn đoán sai khi người mang virus nhưng test lại cho kết quả âm tính hoặc người không mang virus nhưng test lại cho kết quả dương tính.

Gọi: K = “Test cho chẩn đoán sai”. Áp dụng công thức xác suất đầy đủ ta được:

$$P(K) = P(\bar{H}). P(A) + P(H). P(B) = 0,05 \cdot \frac{2}{3} + 0,1 \cdot \frac{1}{3} = \frac{2}{30} = \boxed{\frac{1}{15}}$$

**Câu 2:**

Ta có hàm mật độ của X:

$$f(x) = \begin{cases} kx^2(2-x) & \text{khi } 0 \leq x \leq 2 \\ 0 & \text{khác} \end{cases}$$

a) Xác định hằng số k:

Vì  $f(x)$  là hàm mật độ của X nên:

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x) dx = 1 &\Leftrightarrow \int_{-\infty}^0 f(x) dx + \int_0^2 f(x) dx + \int_2^{+\infty} f(x) dx = 1 \\ \Leftrightarrow \int_0^2 f(x) dx = 1 &\Leftrightarrow \int_0^2 kx^2(2-x) dx = 1 \Leftrightarrow k \left( \frac{2x^3}{3} - \frac{x^4}{4} \right) \Big|_0^2 = 1 \\ \Leftrightarrow k \left( \frac{16}{3} - \frac{16}{4} \right) = 1 &\Leftrightarrow k \frac{4}{3} = 1 \Leftrightarrow \boxed{k = \frac{3}{4}} \end{aligned}$$

b) Tìm hàm phân phối (phân bố)  $F(x)$  của X:

Hàm phân phối  $F(x)$  có dạng:

$$F(x) = P(X < x) = \int_{-\infty}^x f(t) dt$$

+ Nếu  $x \leq 0$  thì  $F(x) = 0$

+ Nếu  $0 < x \leq 2$  thì:

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t) dt = \int_0^x f(t) dt = \int_0^x \frac{3}{4} t^2 (2-t) dt \\ &= \frac{3}{4} \left( \frac{2}{3} t^3 - \frac{t^4}{4} \right) \Big|_0^x = \frac{3}{4} \left( \frac{2}{3} x^3 - \frac{x^4}{4} \right) = \frac{8x^3 - 3x^4}{16} \end{aligned}$$

+ Nếu  $x > 2$  thì  $F(x) = 1$

$\Rightarrow$  Hàm phân phối  $F(x)$ :

$$F(x) = \begin{cases} 0 & \text{khi } x \leq 0 \\ \frac{8x^3 - 3x^4}{16} & \text{khi } 0 < x \leq 2 \\ 1 & \text{khi } x > 2 \end{cases}$$

c) Xác suất để con trùng chét trước khi nó được một tháng tuổi:

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

$$= \frac{3}{4} \left( \frac{2}{3} x^3 - \frac{x^4}{4} \right) \Big|_0^1 = \frac{3}{4} \left( \frac{2}{3} - \frac{1}{4} \right) = \boxed{\frac{5}{16}}$$

d) Tìm kỳ vọng và phương sai của X:

+ Kỳ vọng của X:

$$\begin{aligned} EX &= \int_{-\infty}^{+\infty} x f(x) dx = \frac{3}{4} \int_0^2 x^3 (2-x) dx = \frac{3}{4} \left( \frac{x^4}{2} - \frac{x^5}{5} \right) \Big|_0^2 \\ &= \frac{3}{4} \left( 8 - \frac{32}{5} \right) = \frac{3}{4} \cdot \frac{8}{5} = \boxed{\frac{6}{5}} \end{aligned}$$

+ Phương sai của X:  $DX = EX^2 - (EX)^2$

$$\begin{aligned} EX^2 &= \int_{-\infty}^{+\infty} x^2 f(x) dx = \frac{3}{4} \int_0^2 x^4 (2-x) dx = \frac{3}{4} \left( \frac{2x^5}{5} - \frac{x^6}{6} \right) \Big|_0^2 \\ &= \frac{3}{4} \left( \frac{64}{5} - \frac{64}{6} \right) = \frac{3}{4} \cdot \frac{64}{30} = \frac{8}{5} \\ \Rightarrow DX &= EX^2 - (EX)^2 = \frac{8}{5} - \left( \frac{6}{5} \right)^2 = \frac{40 - 36}{25} = \boxed{\frac{4}{25}} \end{aligned}$$

### Câu 3:

a) Lô hàng có phù hợp với tỉ lệ cỡ giày của cửa hàng không, mức ý nghĩa 0,05:

*Hướng dẫn: Dạng bài toán “Tiêu chuẩn phù hợp Khi bình phương” vì kiểm tra xem mẫu có tuân theo phân bố nào đó hay không. Lưu ý phải gộp khoảng do số lượng giày cỡ to kiểm tra được bé hơn 5.*

Gọi X là kích cỡ đôi giày của nhà cung cấp. Bài toán:

$$\begin{cases} \text{Giả thiết H: X có tỉ lệ } 0,35 : 0,5 : 0,15 \\ \text{Đối thiết K: X không có tỉ lệ như trên} \\ \text{Mức ý nghĩa } \alpha = 0,05 \end{cases}$$

Miền bác bỏ:

$$S = \left\{ \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n \geq \chi_{k-1}^2(\alpha) \right\} \quad (\text{với } k = 3 \text{ là số loại kích cỡ giày})$$

Ta có:

$$\frac{1}{n} \sum_{i=1}^4 \frac{m_i^2}{p_i} - n = \frac{1}{100} \left( \frac{12^2}{0,35} + \frac{48^2}{0,5} + \frac{40^2}{0,15} \right) - 100 = 56,86$$

Tra bảng của phân bố Khi bình phương ta được:  $\chi_2^2(0,05) = 5,99$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy, với mức ý nghĩa 5% ta tạm thời cho rằng lô hàng không phù hợp với các tỷ lệ về cỡ giày của cửa hàng.

b) Cửa hàng có thể chấp nhận tỉ lệ cỡ giày nào trong các tỉ lệ mà nhà cung cấp có, mức ý nghĩa  $\alpha = 0,05$ :

Hướng dẫn: <sup>1</sup>Dạng bài toán kiểm định giả thiết cho tỷ lệ (làm 4 bài toán)

\* Đối với tỷ lệ giày cỡ to:

Bài toán:

$$H: p_1 = 0,1 \mid K: p_1 \neq 0,1 \quad (\text{mức ý nghĩa } \alpha = 0,05)$$

Tỷ lệ mẫu:

$$p_1^* = \frac{2}{100} = 0,02$$

Miền bác bỏ:

$$S = \left\{ \frac{|p_1^* - p_1|}{\sqrt{p_1(1-p_1)}} \sqrt{n} \geq u(\alpha/2) \right\}$$

Tính:

$$\frac{|p_1^* - p_1|}{\sqrt{p_1(1-p_1)}} \sqrt{n} = \frac{|0,02 - 0,1|}{\sqrt{0,1 \cdot 0,9}} \sqrt{100} = 2,667$$

Tra bảng của phân bố chuẩn tắc ta được:  $u(\alpha/2) = u(0,025) = 1,96$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy, với mức ý nghĩa  $\alpha = 0,05$  cửa hàng sẽ không chấp nhận tỷ lệ giày cỡ to.

\* Đối với tỷ lệ giày cỡ vừa:

Bài toán:

$$H: p_2 = 0,25 \mid K: p_2 \neq 0,25 \quad (\text{mức ý nghĩa } \alpha = 0,05)$$

Tỷ lệ mẫu:

$$p_2^* = \frac{10}{100} = 0,1$$

Miền bác bỏ:

$$S = \left\{ \frac{|p_2^* - p_2|}{\sqrt{p_2(1-p_2)}} \sqrt{n} \geq u(\alpha/2) \right\}$$

$$\text{Tính: } \frac{|p_2^* - p_2|}{\sqrt{p_2(1-p_2)}} \sqrt{n} = \frac{|0,1 - 0,25|}{\sqrt{0,25 \cdot 0,75}} \sqrt{100} = 3,464$$

<sup>1</sup> Dựa vào số liệu mẫu có thể đoán được chỉ có tỷ lệ giày cỡ trung là thỏa mãn.

Tra bảng của phân bố chuẩn tắc ta được:  $u(\alpha/2) = u(0,025) = 1,96$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy, với mức ý nghĩa  $\alpha = 0,05$  cửa hàng sẽ không chấp nhận tỷ lệ giày cỡ vừa.

\* Đối với tỷ lệ giày cỡ trung:

Bài toán:

$$H: p_3 = 0,5 \quad | \quad K: p_3 \neq 0,5 \quad (\text{mức ý nghĩa } \alpha = 0,05)$$

Tỷ lệ mẫu:

$$p_3^* = \frac{48}{100} = 0,48$$

Miền bác bỏ:

$$S = \left\{ \frac{|p_3^* - p_3|}{\sqrt{p_3(1-p_3)}} \sqrt{n} \geq u(\alpha/2) \right\}$$

Tính:

$$\frac{|p_3^* - p_3|}{\sqrt{p_3(1-p_3)}} \sqrt{n} = \frac{|0,48 - 0,5|}{\sqrt{0,5 \cdot 0,5}} \sqrt{100} = 0,4$$

Tra bảng của phân bố chuẩn tắc ta được:  $u(\alpha/2) = u(0,025) = 1,96$

Do đó miền bác bỏ không xảy ra, ta chấp nhận H và bác bỏ K. Vậy, với mức ý nghĩa  $\alpha = 0,05$  cửa hàng sẽ chấp nhận tỷ lệ giày cỡ trung.

\* Đối với tỷ lệ giày cỡ nhỏ:

Bài toán:

$$H: p_4 = 0,15 \quad | \quad K: p_4 \neq 0,15 \quad (\text{mức ý nghĩa } \alpha = 0,05)$$

Tỷ lệ mẫu:

$$p_4^* = \frac{40}{100} = 0,4$$

Miền bác bỏ:

$$S = \left\{ \frac{|p_4^* - p_4|}{\sqrt{p_4(1-p_4)}} \sqrt{n} \geq u(\alpha/2) \right\}$$

$$\text{Tính: } \frac{|p_4^* - p_4|}{\sqrt{p_4(1-p_4)}} \sqrt{n} = \frac{|0,4 - 0,15|}{\sqrt{0,15 \cdot 0,85}} \sqrt{100} = 7$$

Tra bảng của phân bố chuẩn tắc ta được:  $u(\alpha/2) = u(0,025) = 1,96$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy, với mức ý nghĩa  $\alpha = 0,05$  cửa hàng sẽ không chấp nhận tỷ lệ giày cỡ nhỏ.

c) Với xác suất 0,95 tỷ lệ giày cỡ nhỏ của nhà cung cấp thấp nhất là bao nhiêu:

Hướng dẫn: Dạng bài toán ước lượng khoảng cho tỷ lệ.

Tỷ lệ mẫu của giày cỡ nhỏ  $p_4^* = 0,4$ . Khoảng ước lượng là:

$$\begin{aligned} p_4 &\in \left( p_4^* - u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p_4^*(1-p_4^*)}{n}}; \quad p_4^* + u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p_4^*(1-p_4^*)}{n}} \right) \\ \Leftrightarrow p_4 &\in \left( 0,4 - u(0,025) \sqrt{\frac{0,4 \cdot 0,6}{100}}; \quad 0,4 + u(0,025) \sqrt{\frac{0,4 \cdot 0,6}{100}} \right) \\ \Leftrightarrow p_4 &\in (0,304; \quad 0,496) \end{aligned}$$

Vậy với xác suất 0,95 thì tỷ lệ giày cỡ nhỏ của nhà cung cấp thấp nhất là 30,4%

**Câu 4:** X là số người tiêm chích ma túy, Y là số người nhiễm HIV

a) Tính hệ số tương quan mẫu r:

$$r = \frac{\frac{1}{n} \sum x_i y_i m_i - \bar{X} \cdot \bar{Y}}{s_X s_Y}$$

Sử dụng máy tính bỏ túi ta tính được<sup>1</sup>:

$$\begin{aligned} + \frac{1}{n} \sum x_i y_i m_i &= 38850 \\ + \bar{X} &= 250 \quad + s_X^2 = 5500 \\ + \bar{Y} &= 144 \quad + s_Y^2 = 1584 \end{aligned}$$

$\Rightarrow$  Hệ số tương quan mẫu:

$$r = \frac{38850 - 250 \cdot 144}{\sqrt{5500 \cdot 1584}} = [0,9656]$$

b) Có nhận xét gì về sự phụ thuộc giữa số người nhiễm HIV và số người tiêm chích ma túy:

Hệ số tương quan mẫu r rất lớn, do đó mối quan hệ tuyến tính giữa X và Y là rất chặt. Số người nhiễm HIV phụ thuộc rất chặt chẽ vào số người tiêm chích ma túy.

c) Xây dựng đường hồi quy bình phương trung bình tuyến tính thực nghiệm của Y theo X:

Phương trình đường hồi quy của Y theo X có dạng:

$$y - \bar{Y} = r \frac{s_Y}{s_X} (x - \bar{X}) \Leftrightarrow y = r \frac{s_Y}{s_X} (x - \bar{X}) + \bar{Y}$$

<sup>1</sup> Xem cách tìm các chỉ số thống kê bằng máy tính bỏ túi tại Phụ lục P.2.2, trang 192.

$$\Leftrightarrow y = 0,9656\sqrt{\frac{1584}{5500}}(x - 250) + 144$$

$$\Leftrightarrow \boxed{y = 0,5182x + 14,4512}$$

d) Đường hồi quy nhận được dùng để làm gì? Vì sao? Cho ví dụ:

Đường hồi quy nhận được dùng để ước xấp xỉ Y theo X (dự đoán số người nhiễm HIV dựa vào số người tiêm chích ma túy) vì nó là hàm của Y phụ thuộc vào X.

Ví dụ: khi điều tra ở địa phương B nào đó thấy có 330 người tiêm chích ma túy, khi đó dự đoán số người nhiễm HIV ở địa phương B là:

$$0,5182 \cdot 330 + 14,4512 \approx 186 \text{ (người)}$$

e) Ước lượng sai số bình phương trung bình mắc phải khi dùng đường hồi quy:

Sai số bình phương trung bình khi xấp xỉ Y theo X:

$$s_{Y/X}^2 = s_Y^2(1 - r^2) = 1584(1 - 0,9656^2) = \boxed{107,1}$$

### 3. Đề thi cuối kỳ II năm học 2013 – 2014

ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

ĐỀ THI KẾT THÚC HỌC KỲ II  
NĂM HỌC 2013 – 2014

#### Môn thi: Xác suất thống kê

Mã môn học: MAT 1101

Số tín chỉ: 03

Dành cho sinh viên các khoa: Hóa học, Sinh học, Địa lý, Địa chất, Môi trường, KTTV-HDH, Y-Dược

Dạng đề thi: Được sử dụng tài liệu

Thời gian làm bài: 120 phút (không kể thời gian phát đề)

**Câu 1 (3đ):** Một nhà vườn ươm cây phi lao giống cho lâm trường, có 3 vườn ươm: A, B và C. Biết các cây giống trong các vườn ươm này có chiều cao tuân theo phân bố chuẩn với trung bình lần lượt là 0,9m; 1m; 1,1m và độ lệch tiêu chuẩn đều bằng 0,2m.

a) Chọn ngẫu nhiên 100 cây trong vườn ươm A thì số cây có chiều cao lớn hơn 1m là bao nhiêu (theo ước lượng của bạn)?

b) Một cán bộ lâm trường xuống nhà vườn kiểm tra xem cây giống có đem ra trồng được chưa. Cán bộ này chọn ngẫu nhiên một vườn ươm (trong 3 vườn ươm A, B, C với xác suất bằng nhau) sau đó chọn 5 cây bất kỳ từ vườn ươm đó và đo chiều cao, nếu thấy có từ 4 cây trở lên có chiều cao lớn hơn 1m thì sẽ cho phép đem ra trồng đại trà. Biết rằng người cán bộ này không cho phép đem ra trồng. Tính xác suất để 5 cây phi lao anh ta chọn đo chiều cao là của vườn ươm A.

**Câu 2 (2đ):** Thống kê số khách trên một ôtô buýt tại một tuyến giao thông thu được số liệu như sau:

Số khách trên một tuyến	20	25	30	35	40	45
Xác suất tương ứng	0,15	0,15	0,2	0,2	0,2	0,1

a) Tim kỳ vọng và phương sai của số khách đi mỗi chuyến và giải thích ý nghĩa của kết quả thu được.

b) Giá sử chi phí cho mỗi chuyến xe là 1 triệu VNĐ không phụ thuộc vào số khách đi trên xe. Để công ty xe buýt có thể thu lãi bình quân cho mỗi chuyến xe là 300 nghìn VNĐ thì phải qui định giá vé là bao nhiêu?

**Câu 3 (3đ):** Để đánh giá tác dụng của hai loại phân bón, người ta tổng kết năng suất lúa (tấn/ha) X và Y khi dùng lần lượt hai loại phân bón (PB) I và II kết quả như sau:

Khoảng giá trị năng suất (tấn/ha)	< 4,5	[4,5; 4,6)	[4,6; 4,7)	[4,7; 4,8)	[4,8; 4,9)	$\geq 4,9$
Số diện tích ruộng dùng loại PB I (ha)	5	10	15	10	7	4
Số diện tích ruộng dùng loại PB II (ha)	4	12	15	11	6	3

a) Với  $\alpha = 0,05$  hãy kiểm tra xem tác dụng của hai loại phân bón đến năng suất lúa trung bình có như nhau hay loại nào tốt hơn?

b) Nếu năng suất  $\geq 4,7$  được xem là cao. Hãy ước lượng cho số tỉ lệ diện tích ruộng có năng suất cao khi dùng loại phân bón II? (với xác suất 95%).

c) Một người khẳng định “loại phân bón I thực ra chính là loại mẩy năm nay đang dùng nhưng chỉ đổi mác thôi”. Biết số liệu thống kê đã có trong mẩy năm nay tương ứng với các khoảng giá trị sản lượng ở bảng trên là:  $p_1 = 9\%$ ;  $p_2 = 19\%$ ;  $p_3 = 21\%$ ;  $p_4 = 30\%$ ;  $p_5 = 13\%$  và  $p_6 = 8\%$ .

Theo bạn thì khẳng định của người này có đáng tin không? (với  $\alpha = 0,05$ )

**Câu 4 (2đ):** Người ta cho 6 bệnh nhân ở lứa tuổi khác nhau uống cùng một loại thuốc và theo dõi thời gian phân hủy loại thuốc đó trong cơ thể, được kết quả:

Tuổi X (năm)	25	30	35	40	50	65
Thời gian phân hủy Y (phút)	15	20	22	25	28	31

a) Hãy ước lượng hệ số tương quan và tìm đường hồi quy tuyến tính thực nghiệm của Y theo X?

b) Dự báo thời gian phân hủy thuốc của bệnh nhân 27 tuổi?

Lời giải:

### Câu 1:

Gọi X, Y, Z lần lượt là chiều cao cây giống trong các vườn ươm A, B, C. Theo đề bài ta có:

$$X \sim N(0,9; 0,2^2)$$

$$Y \sim N(1,0; 0,2^2)$$

$$Z \sim N(1,1; 0,2^2)$$

a) *Ước lượng số cây có chiều cao lớn hơn 1m ở vườn ươm A:*

Tỷ lệ cây có chiều cao lớn hơn 1m ở vườn A:

$$\begin{aligned} P(X > 1) &= 1 - P(X < 1) = 1 - P\left(\frac{X - 0,9}{0,2} < \frac{1 - 0,9}{0,2}\right) \\ &= 1 - \Phi(0,5) = 1 - 0,6915 = 0,3085 \end{aligned}$$

$\Rightarrow$  Ước lượng số cây có chiều cao lớn hơn 1m:

$$100 \cdot 0,3085 = 30,85 \text{ (cây)} \approx 31 \text{ (cây)}$$

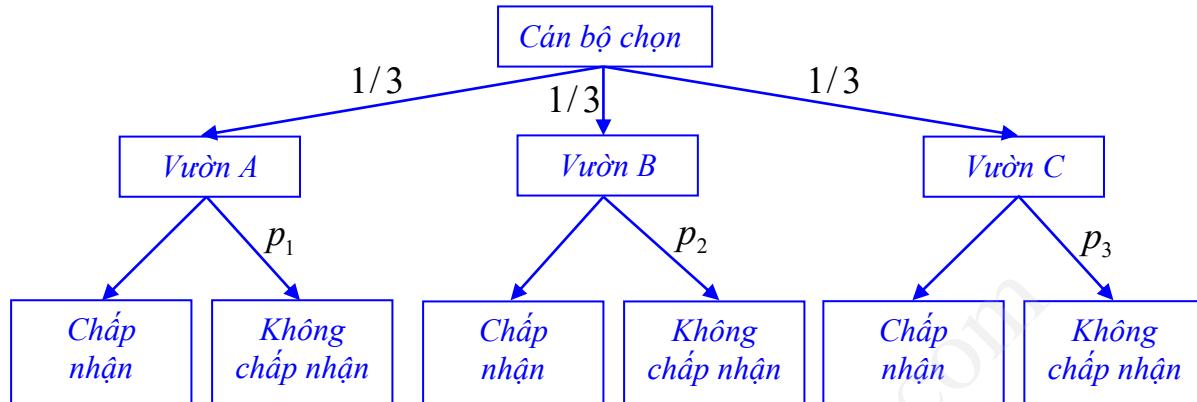
b) *Xác suất để 5 cây chọn ra là của vườn ươm A:*

*Hướng dẫn:* Người cán bộ không cho phép đem đi trồng đại trà tức là có ít hơn 4 cây có chiều cao lớn hơn 1m. Bài toán đã cho biết kết quả xảy ra, yêu cầu đi tìm xác suất của một nguyên nhân dẫn đến kết quả đó. Trong kết quả xảy ra lại có dạng phép

thử lặp Bernoulli (với  $n = 5$  và  $p$  tương ứng với từng vườn ươm,  $p$  được tính dựa vào phân bố chuẩn)

⇒ Dạng bài toán kết hợp cả việc tính xác suất của một biến ngẫu nhiên phân bố chuẩn, phép thử lặp Bernoulli, công thức xác suất đầy đủ và công thức Bayes.

Sơ đồ cây có dạng như sau:



Các giá trị  $p_1$ ,  $p_2$  và  $p_3$  được tính thông qua Phép thử lặp Bernoulli và phân bố chuẩn tương ứng với từng vườn ươm.

Gọi các biến cỏ:

A = “Chọn được vườn ươm A”.

B = “Chọn được vườn ươm B”.

C = “Chọn được vườn ươm C”.

H = “Không chấp nhận đem đi trồng đại trà”.

Theo bài ra ta có:  $P(A) = P(B) = P(C) = 1/3$

+ Nếu chọn được vườn A và sau đó chọn 1 cây giống bất kỳ thì xác suất để cây đó cao hơn 1m là: 0,3085 (theo kết quả câu a)

Khi chọn 5 cây trong vườn A tương đương với 5 phép thử độc lập, mỗi phép thử có xác suất chọn được cây cao hơn 1m là 0,3085. Do đó xác suất có ít hơn 4 cây cao hơn 1m (không được đem đi trồng đại trà) là:

$$\begin{aligned} p_1 &= P(H | A) = 1 - \left( C_5^4 \cdot 0,3085^4 \cdot 0,6915 + C_5^5 \cdot 0,3085^5 \right) \\ &= 1 - 0,0341 = 0,9659 \end{aligned}$$

+ Nếu chọn được vườn B và sau đó chọn 1 cây giống bất kỳ thì xác suất để cây đó cao hơn 1m là:

$$P(Y > 1) = 0,5 \text{ (vì Median } Y = 1)$$

Khi chọn 5 cây trong vườn B tương đương với 5 phép thử độc lập. Xác suất có ít hơn 4 cây cao hơn 1m là:

$$\begin{aligned} p_2 &= P(H | B) = 1 - \left( C_5^4 \cdot 0,5^4 + C_5^5 \cdot 0,5^5 \right) \\ &= 1 - 0,1875 = 0,8125 \end{aligned}$$

+ Nếu chọn được vườn C và sau đó chọn 1 cây giống bất kỳ thì xác suất để cây đó cao hơn 1m là:

$$\begin{aligned} P(Z > 1) &= 1 - P(Z < 1) = 1 - \Phi\left(\frac{1-1,1}{0,2}\right) = 1 - (1 - \Phi(0,5)) = \Phi(0,5) \\ &= 0,6915 \end{aligned}$$

Khi chọn 5 cây trong vườn C tương đương với 5 phép thử độc lập. Xác suất có ít hơn 4 cây cao hơn 1m là:

$$\begin{aligned} p_3 &= P(H | C) = 1 - (C_5^4 \cdot 0,6915^4 \cdot 0,3085 + C_5^5 \cdot 0,6915^5) \\ &= 1 - 0,5108 = 0,4892 \end{aligned}$$

$\Rightarrow$  Xác suất để 5 cây được chọn ra là của vườn ươm A: áp dụng công thức xác suất Bayes.

$$\begin{aligned} P(A | H) &= \frac{P(H | A) \cdot P(A)}{P(H | A) \cdot P(A) + P(H | B) \cdot P(B) + P(H | C) \cdot P(C)} \\ &= \frac{0,9659 \cdot \frac{1}{3}}{0,9659 \cdot \frac{1}{3} + 0,8125 \cdot \frac{1}{3} + 0,4892 \cdot \frac{1}{3}} = \frac{0,9659}{0,9659 + 0,8125 + 0,4892} \\ &= \boxed{0,426} \end{aligned}$$

## Câu 2:

Gọi X là số khách đi trên một chuyến xe buýt.

a) Tìm kỳ vọng và phương sai của số khách đi trên mỗi chuyến và giải thích ý nghĩa của kết quả thu được:

Số khách trên một chuyến	20	25	30	35	40	45
Xác suất tương ứng	0,15	0,15	0,2	0,2	0,2	0,1

+ Kỳ vọng của X:

$$\begin{aligned} EX &= \sum x_i p_i = 20 \cdot 0,15 + 25 \cdot 0,15 + 30 \cdot 0,2 + 35 \cdot 0,2 + 40 \cdot 0,2 + 45 \cdot 0,1 \\ &= \boxed{32,25} \end{aligned}$$

+ Phương sai của X:

$$\begin{aligned} DX &= EX^2 - (EX)^2 = \sum x_i^2 p_i - (EX)^2 \\ &= 20^2 \cdot 0,15 + 25^2 \cdot 0,15 + 30^2 \cdot 0,2 + 35^2 \cdot 0,2 + 40^2 \cdot 0,2 + 45^2 \cdot 0,1 - 32,25^2 \\ &= 1101,25 - 32,25^2 = \boxed{61,19} \end{aligned}$$

- Ý nghĩa:

Kỳ vọng ( $EX$ ) thể hiện số khách trung bình đi trên một chuyến xe, sau khi quan sát số lượng khách của nhiều chuyến. Phuông sai ( $DX$ ) thể hiện mức độ phân tán của số khách trong mỗi chuyến so với số khách trung bình của các chuyến,  $DX$  càng lớn thì mức phân tán càng lớn,  $DX$  càng nhỏ thì số khách trong các chuyến càng tập trung gần số khách trung bình.

b) Giá vé quy định là bao nhiêu:

Để mỗi chuyến lãi bình quân là 300.000 VNĐ thì doanh thu bình quân của mỗi chuyến là 1.300.000 VNĐ (vì phải trừ đi 1 triệu tiền chi phí). Mà số khách trung bình trên mỗi chuyến là 32,25 người do đó giá vé phải quy định là:

$$1300\,000 : 32,25 = 40\,310 \text{ (VNĐ)}$$

**Câu 3:**

Gọi X và Y là năng suất lúa khi dùng hai loại phân bón I và II.

a) Tác dụng của hai loại phân bón đến năng suất lúa trung bình có như nhau hay không:

Hướng dẫn: Dạng bài toán so sánh hai giá trị trung bình; trường hợp chưa biết phuông sai lý thuyết, chưa biết phân bố chuẩn nhưng hai cỡ mẫu lớn. Trước tiên, tính năng suất lúa trung bình của hai mẫu và so sánh chúng với nhau để xác định đối thiết của bài toán là  $EX < EY$  hay  $EX > EY$ .

Từ bảng dữ liệu ta tính được:

$$n_X = n_Y = 51; \bar{X} = 4,6814; \bar{Y} = 4,6735; s_X^2 = 0,0194; s_Y^2 = 0,0171$$

Ta thấy,  $\bar{X} > \bar{Y}$ . Do đó, tiến hành kiểm định bài toán sau:

$$\begin{cases} H: EX = EY \\ K: EX > EY \end{cases} \quad (\alpha = 0,05)$$

Miền bác bỏ của bài toán:

$$S = \{ U \geq u(\alpha) \}$$

$$\text{Tính: } U = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n_X - 1} + \frac{s_Y^2}{n_Y - 1}}} = \frac{4,6814 - 4,6735}{\sqrt{\frac{0,0194}{50} + \frac{0,0171}{50}}} = 0,2924$$

Tra bảng của phân bố chuẩn tắc ta được:  $u(\alpha) = u(0,05) = 1,645$

⇒ Ta thấy  $U < u(\alpha)$  do đó miền bác bỏ không đã xảy ra. Ta chấp nhận  $H$  và bác bỏ  $K$ . Vậy với mức ý nghĩa 0,05 ta tạm thời chấp nhận rằng hai loại phân bón cho năng suất lúa như nhau.

b) Ước lượng tỉ lệ diện tích ruộng cho năng suất cao ( $\geq 4,7$  tấn/ha) khi dùng phân bón loại II (với  $1 - \alpha = 0,95$ ):

Tỷ lệ mẫu của ruộng cho năng suất cao:

$$p^* = \frac{11+6+3}{51} = \frac{20}{51}$$

Với  $1 - \alpha = 0,95$  thì khoảng ước lượng như sau:

$$p \in \left( p^* - u \left( \frac{\alpha}{2} \right) \sqrt{\frac{p^*(1-p^*)}{n}}, \quad p^* + u \left( \frac{\alpha}{2} \right) \sqrt{\frac{p^*(1-p^*)}{n}} \right)$$

$$p \in \left( \frac{20}{51} - 1,96 \sqrt{\frac{20.31}{51^3}}, \quad \frac{20}{51} + 1,96 \sqrt{\frac{20.31}{51^3}} \right)$$

$$p \in (0,2582; \quad 0,5262)$$

Vậy với xác suất 95% thì tỷ lệ diện tích ruộng cho năng suất cao khi dùng phân bón loại II thuộc khoảng (0,2582; 0,5262).

c) *Tỷ lệ diện tích ruộng dùng phân bón loại I có như khẳng định không ( $\alpha = 0,05$ ):*

*Hướng dẫn: Dạng bài toán Tiêu chuẩn phù hợp Khi bình phương, lưu ý gộp khoảng để thỏa mãn  $m_i \geq 5 \forall i$*

Vì số quan sát trong khoảng cuối cùng  $< 5$  nên ta tiến hành gộp khoảng thành:

Khoảng giá trị năng suất (tấn/ha)	$< 4,5$	[4,5; 4,6]	[4,6; 4,7]	[4,7; 4,8]	$\geq 4,8$
Số diện tích ruộng dùng loại PB I (ha)	5	10	15	10	11

Bài toán:

$$\begin{cases} \text{Giả thiết H: Diện tích ruộng theo tỷ lệ } 0,09 : 0,19 : 0,21 : 0,3 : 0,21 \\ \text{Đối thiết K: Diện tích ruộng không theo tỷ lệ như trên} \\ \text{Mức ý nghĩa } 0,05 \end{cases}$$

Miền bác bỏ của bài toán:

$$S = \left\{ \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n \geq \chi_{k-1}^2(\alpha) \right\}$$

Trong đó: +  $m_i$  là số quan sát trong khoảng thứ  $i$ .

+  $n$  là tổng số quan sát,  $n = 51$

+  $p_i$  là xác suất tính theo lý thuyết của khoảng thứ  $i$ .

+  $k$  là số khoảng,  $k = 5$

Tính:

$$\frac{1}{n} \sum_{i=1}^5 \frac{m_i^2}{p_i} - n = \frac{1}{51} \left( \frac{5^2}{0,09} + \frac{10^2}{0,19} + \frac{15^2}{0,21} + \frac{10^2}{0,3} + \frac{11^2}{0,21} \right) - 51 = 3,61$$

Tra bảng phân bố  $\chi^2$  ta được:  $\chi_{k-1}^2(\alpha) = \chi_4^2(0,05) = 9,49$

Do đó miền bác bỏ không xảy ra, ta chấp nhận H và bác bỏ K.

Vậy, với mức ý nghĩa 0,05 ta tạm thời chấp nhận rằng “phân bón loại I chính là loại mẩy năm nay đang dùng nhưng chỉ đổi mác mà thôi”.

#### Câu 4:

a) *Hệ số tương quan mẫu r và phương trình đường hồi quy tuyến tính thực nghiệm của Y theo X:*

Ta có: X là tuổi bệnh nhân (năm) và Y là thời gian phân hủy của thuốc (phút).

\* *Hệ số tương quan mẫu:*

$$r = \frac{\frac{1}{n} \sum x_i y_i m_i - \bar{X} \cdot \bar{Y}}{s_x s_y}$$

Sử dụng máy tính bỏ túi ta tính được<sup>1</sup>:

$$\begin{aligned} + \frac{1}{n} \sum x_i y_i m_i &= \frac{6160}{6} \\ + \bar{X} &= \frac{245}{6} & + s_x^2 &= \frac{6425}{36} \\ + \bar{Y} &= \frac{141}{6} & + s_y^2 &= \frac{993}{36} \end{aligned}$$

⇒ Hệ số tương quan mẫu:

$$r = \frac{\frac{6160}{6} - \frac{245}{6} \cdot \frac{141}{6}}{\sqrt{\frac{6425 \cdot 993}{36^2}}} = \boxed{0,9561} \quad (\text{mối quan hệ tuyến tính rất chặt})$$

\* *Phương trình đường hồi quy tuyến tính thực nghiệm của Y theo X:*

$$y - \bar{Y} = r \frac{s_y}{s_x} (x - \bar{X}) \Leftrightarrow y = r \frac{s_y}{s_x} (x - \bar{X}) + \bar{Y}$$

$$y = 0,9561 \cdot \sqrt{\frac{993}{36} \frac{36}{6425}} \left( x - \frac{245}{6} \right) + \frac{141}{6}$$

$$\Leftrightarrow \boxed{y = 0,3759x + 8,1518}$$

b) *Dự báo thời gian phân hủy thuốc của bệnh nhân 27 tuổi:*

Phương trình đường hồi quy tuyến tính thực nghiệm của Y theo X có thể dùng để dự báo thời gian phân hủy thuốc của một bệnh nhân có tuổi nhất định.

Với bệnh nhân 27 tuổi ( $x = 27$ ) thì thời gian phân hủy thuốc là:

$$y = 0,3759 \cdot 27 + 8,1518 = 18,3 \text{ (phút)}$$

<sup>1</sup> Xem cách tìm các chỉ số thống kê bằng máy tính bỏ túi tại *Phụ lục P.2.2*, trang 192.

## 4. Đề thi cuối kỳ phụ – hè năm 2014

ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

ĐỀ THI KẾT THÚC HỌC KỲ PHỤ  
HÈ NĂM 2014

### Môn thi: Xác suất thống kê

Mã môn học: MAT 1101 Số tín chỉ: 03

Dành cho sinh viên các khoa: Hóa học, Sinh học, Địa lý, Địa chất, Môi trường, KT-TV-HDH, Y-Dược

Dạng đề thi: **Dược sử dụng tài liệu**

Thời gian làm bài: 120 phút (không kể thời gian phát đề)

**Câu 1 (2đ):** Chọn ngẫu nhiên 3 viên bi từ một túi có 6 bi đen và 4 bi trắng. Gọi X là số bi trắng được chọn. Gọi Y là số tiền nhận được, biết rằng nếu được mỗi bi trắng sẽ được 200 USD, mỗi bi đen được 300 USD.

- a) Tìm kỳ vọng, phương sai của Y. Nếu ý nghĩa các đại lượng.
- b) Tìm bảng phân bố đồng thời của X, Y. Tính cov(X, Y); kết luận gì về tính độc lập của X và Y.

**Câu 2 (3đ):** Trọng lượng sản phẩm X của một máy tự động sản xuất là một biến ngẫu nhiên có phân bố chuẩn với trung bình 100g, độ lệch tiêu chuẩn 1g. Sản phẩm được coi là đạt tiêu chuẩn kỹ thuật nếu trọng lượng của nó đạt từ 98g đến 102g.

- a) Tìm tỷ lệ sản phẩm đạt tiêu chuẩn kỹ thuật của nhà máy.
- b) Ước lượng số sản phẩm không đạt tiêu chuẩn từ lô hàng gồm 200 sản phẩm được sản xuất ra từ máy này.
- c) Tìm a để  $P(X > a) = 0,05$

**Câu 3 (2đ):** Sản lượng khai thác than ở một công ty than được ghi lại qua 9 năm như sau:

Năm (X)	1988	1989	1990	1991	1992	1993	1994	1995	1996
Sản lượng Y (vạn tấn)	60	61	64	65	66	66	69	70	72

a) Tìm ước lượng cho hệ số tương quan của X và Y. Kết luận gì về mối quan hệ giữa X và Y.

b) Tìm đường hồi quy tuyến tính thực nghiệm của X và Y. Hãy dự đoán về số than có thể khai thác được vào năm 2000?

**Câu 4 (3đ):** Số tai nạn giao thông xảy ra mỗi ngày ở thành phố quan sát được (trong 156 ngày quan sát):

Số tai nạn trong ngày (X)	0	1	2	3	4	5	6	7	8
Số ngày (n)	10	32	46	35	20	9	2	1	1

Hoàng Văn Trọng – 0974.971.149

a) Ước lượng cho số vụ tai nạn trung bình trong ngày ở thành phố với độ tin cậy 95%. Ý nghĩa của ước lượng này là gì?

b) Số vụ tai nạn trong ngày lớn hơn 4 thì ngày đó được xem là ngày **tử thần**, hãy kiểm tra nhận xét cho rằng tỷ lệ ngày tử thần ở thành phố này là lớn hơn 8%. Hãy nêu ý nghĩa của kết luận của bạn?

c) Có ý kiến cho rằng số tai nạn giao thông ở thành phố mỗi ngày có phân bố Poisson. Hãy kết luận về nhận xét này với mức ý nghĩa  $\alpha = 0,01$ .

Lời giải:

### Câu 1:

a) Tìm  $EY$ ,  $DY$ . Nêu ý nghĩa các đại lượng:

X là số bi trắng trong 3 viên bi được lấy ra. Ta có:  $X(\Omega) = \{0, 1, 2, 3\}$

Y là số tiền nhận được:  $Y = 200X + 300(3 - X) = 900 - 100X$

+ Tính các xác suất tương ứng:

$$P(X=0) = \frac{C_4^0 \cdot C_6^3}{C_{10}^3} = \frac{1}{6} \Rightarrow P(Y=900) = \frac{1}{6}$$

$$P(X=1) = \frac{C_4^1 \cdot C_6^2}{C_{10}^3} = \frac{1}{2} \Rightarrow P(Y=800) = \frac{1}{2}$$

$$P(X=2) = \frac{C_4^2 \cdot C_6^1}{C_{10}^3} = \frac{3}{10} \Rightarrow P(Y=700) = \frac{3}{10}$$

$$P(X=3) = \frac{C_4^3 \cdot C_6^0}{C_{10}^3} = \frac{1}{30} \Rightarrow P(Y=600) = \frac{1}{30}$$

$\Rightarrow$  Bảng phân bố xác suất của Y:

Y	600	700	800	900
$P(Y=y_i)$	$1/30$	$3/10$	$1/2$	$1/6$

+ Kỷ vọng của Y:

$$EY = \sum y_i p_i = 600 \cdot \frac{1}{30} + 700 \cdot \frac{3}{10} + 800 \cdot \frac{1}{2} + 900 \cdot \frac{1}{6} = \boxed{780} \text{ (USD)}$$

+ Phuơng sai của Y:

$$DY = EY^2 - (EY)^2 = 600^2 \cdot \frac{1}{30} + 700^2 \cdot \frac{3}{10} + 800^2 \cdot \frac{1}{2} + 900^2 \cdot \frac{1}{6} - 780^2 = \boxed{5600}$$

\* Ý nghĩa các đại lượng:

+ EY thể hiện giá trị trung bình của số tiền nhận được.

+ DY thể hiện mức độ phân tán (mức độ tản mát) của các số tiền nhận được xung quanh giá trị trung bình của chúng. DY càng lớn thì mức độ phân tán càng lớn, DY càng nhỏ thì các số tiền càng tập trung quanh giá trị trung bình.

b) *Bảng phân bố đồng thời của X và Y. Tính cov(X, Y). Kết luận về tính độc lập giữa X và Y:*

+ *Bảng phân bố đồng thời của X và Y:*

X \ Y	600	700	800	900
0	0	0	0	1/6
1	0	0	1/2	0
2	0	3/10	0	0
3	1/30	0	0	0

+ *Tính cov(X, Y):*

$$\text{cov}(X, Y) = EXY - EX.EY$$

Ta có:

$$EXY = \sum_{i=1}^4 \sum_{j=1}^4 x_i y_j p_{ij} = 600 \cdot 3 \cdot \frac{1}{30} + 700 \cdot 2 \cdot \frac{3}{10} + 800 \cdot 1 \cdot \frac{1}{2} + 900 \cdot 0 \cdot \frac{1}{6} = 880$$

$$EX = \sum x_i p_i = 0 \cdot \frac{1}{6} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{3}{10} + 3 \cdot \frac{1}{30} = 1,2$$

$$EY = 780 \quad (\text{theo câu a})$$

$$\text{Suy ra: } \text{cov}(X, Y) = 880 - 1,2 \cdot 780 = \boxed{-56}$$

+ *Kết luận về tính độc lập giữa X và Y:*

X và Y độc lập khi và chỉ khi:

$$P(X = x_i; Y = y_j) = P(X = x_i) \cdot P(Y = y_j) \quad \forall i, j.$$

Xét một trường hợp cụ thể ta thấy:

$$0 = P(X = 0; Y = 600) \neq P(X = 0) \cdot P(Y = 600) = \frac{1}{6} \cdot \frac{1}{30} = \frac{1}{180}$$

Vậy X và Y không độc lập (hay X và Y phụ thuộc với nhau).

## Câu 2:

X là trọng lượng sản phẩm do máy tự động sản xuất:  $X \sim N(100, 1^2)$

a) *Tìm tỷ lệ sản phẩm đạt tiêu chuẩn kỹ thuật của nhà máy:*

Sản phẩm đạt tiêu chuẩn là những sản phẩm có trọng lượng từ 98 đến 102g. Xác suất cần tìm:

$$P(98 \leq X \leq 102) = P\left(\frac{98-100}{1} \leq \frac{X-100}{1} \leq \frac{102-100}{1}\right) = P(-2 < Z < 2)$$

(với  $Z = (X - 100)/1$  là ĐLNN có phân bố chuẩn tắc)

$$= \Phi(2) - \Phi(-2) = 2 \cdot \Phi(2) - 1 = 2 \cdot 0,9772 - 1 = \boxed{0,9544}$$

b) *Uớc lượng số sản phẩm không đạt tiêu chuẩn từ lô hàng gồm 200 sản phẩm:*

Khi lấy ngẫu nhiên một sản phẩm, xác suất để sản phẩm này không đạt tiêu chuẩn là:  $1 - 0,9544 = 0,0456$

Nếu gọi  $Y$  là số sản phẩm không đạt tiêu chuẩn trong 200 sản phẩm được lấy ra thì:  $Y \sim B(200; 0,0456)$ . Uớc lượng cho số sản phẩm không đạt tiêu chuẩn là tần suất vọng của  $Y$  (uớc lượng điểm).

Số sản phẩm không đạt tiêu chuẩn theo uớc lượng là:

$$200 \cdot 0,0456 = 9,12 \approx \boxed{10} \text{ (sản phẩm)}$$

c) *Tìm  $a$  để  $P(X > a) = 0,05$ :*

$$P(X > a) = 0,05 \Leftrightarrow P(X < a) = 0,95 \Leftrightarrow \Phi\left(\frac{a - 100}{1}\right) = 0,95$$

$$\Leftrightarrow a - 100 = 1,645 \Leftrightarrow \boxed{a = 101,645}$$

### Câu 3:

X: năm

Y: sản lượng (vạn tấn)

a) *Uớc lượng cho hệ số tương quan mầu của  $X$  và  $Y$ :*

\* Hệ số tương quan mầu:

$$r = \frac{\frac{1}{n} \sum x_i y_i m_i - \bar{X} \bar{Y}}{s_X s_Y}$$

Sử dụng máy tính bỏ túi ta tính được<sup>1</sup>:

$$\begin{aligned} + \frac{1}{n} \sum x_i y_i m_i &= \frac{1181342}{9} \\ + \bar{X} &= 1992 & + s_X^2 &= \frac{20}{3} \\ + \bar{Y} &= \frac{593}{9} & + s_Y^2 &= 14,0988 \end{aligned}$$

$\Rightarrow$  Hệ số tương quan mầu:

$$r = \frac{\frac{1181342}{9} - 1992 \cdot \frac{593}{9}}{\sqrt{\frac{20}{3} \cdot 14,0988}} = \boxed{0,9856}$$

\* Mỗi quan hệ giữa  $X$  và  $Y$  là mối quan hệ tuyến tính thuận, rất chặt.

<sup>1</sup> Xem cách tìm các chỉ số thống kê bằng máy tính bỏ túi tại *Phụ lục P.2.2*, trang 192.

b) Đường hồi quy tuyến tính thực nghiệm của X và Y; dự đoán số than có thể khai thác được trong năm 2000:

+ Đường hồi quy tuyến tính thực nghiệm của X theo Y:

$$\begin{aligned} x - \bar{X} = r \frac{s_x}{s_y} (y - \bar{Y}) &\Leftrightarrow x - 1992 = 0,9856 \cdot \sqrt{\frac{20}{3.14,0988}} \left( y - \frac{593}{9} \right) \\ &\Leftrightarrow \boxed{x = 0,6777 y + 1947,3444} \end{aligned}$$

+ Đường hồi quy tuyến tính thực nghiệm của Y theo X:

$$\begin{aligned} y - \bar{Y} = r \frac{s_y}{s_x} (x - \bar{X}) &\Leftrightarrow y - \frac{593}{9} = 0,9856 \cdot \sqrt{\frac{3.14,0988}{20}} (x - 1992) \\ &\Leftrightarrow \boxed{y = 1,4333 x - 2789,247} \end{aligned}$$

Dự đoán số than có thể khai thác được vào năm 2000: dựa vào phương trình đường hồi quy tuyến tính thực nghiệm của Y theo X ta có:

$$y_{2000} = 1,4333 \cdot 2000 - 2789,247 = 77,353 \quad (\text{vạn tấn})$$

#### Câu 4:

a) Ước lượng số vụ tai nạn trung bình trong ngày, độ tin cậy 95%. Ý nghĩa của ước lượng:

*Hướng dẫn: Dạng bài toán ước lượng khoảng cho giá trị trung bình, trường hợp chưa biết phương sai, chưa biết phân bố chuẩn nhưng cỡ mẫu lớn.*

Gọi X là số tai nạn giao thông trong ngày. Dựa vào số liệu quan sát ta có:

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum x_i m_i = \frac{0.10 + 1.32 + 2.46 + 3.35 + 4.20 + 5.9 + 6.2 + 7.1 + 8.1}{156} \\ &= \frac{381}{156} = 2,4423 \\ s^2 &= \frac{0^2 \cdot 10 + 1^2 \cdot 32 + 2^2 \cdot 46 + 3^2 \cdot 35 + 4^2 \cdot 20 + 5^2 \cdot 9 + 6^2 \cdot 2 + 7^2 \cdot 1 + 8^2 \cdot 1}{156} - \left( \frac{381}{156} \right)^2 \\ &= 2,1185 \end{aligned}$$

$\Rightarrow$  Khoảng ước lượng:

$$\begin{aligned} EX &\in \left( \bar{X} - u\left(\frac{\alpha}{2}\right) \frac{s}{\sqrt{n-1}}, \bar{X} + u\left(\frac{\alpha}{2}\right) \frac{s}{\sqrt{n-1}} \right) \\ EX &\in \left( 2,4423 - u(0,025) \sqrt{\frac{2,1185}{155}}, 2,4423 + u(0,025) \sqrt{\frac{2,1185}{155}} \right) \\ EX &\in \left( 2,4423 - 1,96 \sqrt{\frac{2,1185}{155}}, 2,4423 + 1,96 \sqrt{\frac{2,1185}{155}} \right) \\ &\boxed{EX \in (2,2132; 2,6714)} \end{aligned}$$

Vậy, với độ tin cậy 95% thì số tai nạn giao thông trung bình/ngày nằm trong khoảng (2,2132; 2,6714).

\* Ý nghĩa của ước lượng:

Số tai nạn giao thông trung bình/ngày năm trong khoảng (2,2132; 2,6714) với xác suất 95% và nằm ngoài khoảng trên với xác suất 5%.

b) Kiểm tra nhận xét: Tỷ lệ ngày tử thán (số tai nạn > 4) lớn hơn 8%. Nếu ý nghĩa kết luận:

Hướng dẫn: Dạng bài kiểm định giả thiết cho tỷ lệ.

Tỷ lệ ngày tử thán dựa trên mẫu quan sát là:

$$p^* = \frac{9+2+1+1}{156} = \frac{13}{156}$$

Bài toán:

$$\begin{cases} \text{Giả thiết } H: p = 0,08 \\ \text{Đối thiết } K: p > 0,08 \\ \text{Mức ý nghĩa } 5\%. \end{cases}$$

Miền bác bỏ:

$$S = \left\{ \frac{p^* - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} \geq u(\alpha) \right\}$$

$$\text{Ta có: } \frac{p^* - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} = \frac{13/156 - 0,08}{\sqrt{0,08 \cdot 0,92}} \sqrt{156} = 0,153$$

Tra bảng ta được:  $u(\alpha) = u(0,05) = 1,645$

Do đó, miền bác bỏ không xảy ra, ta chấp nhận H và bác bỏ K. Vậy với mức ý nghĩa 5% ta tạm thời **không** chấp nhận lời nhận xét cho rằng tỷ lệ ngày tử thán ở thành phố trên là lớn hơn 8%, cho tới khi có thêm thông tin mới.

\* Ý nghĩa của kết luận trên:

Kết luận đưa ra ở trên có xác suất đúng là 95%, xác suất sai là 5%.

c) Kiểm tra nhận xét: Số tai nạn giao thông ở thành phố này có phân bố Poisson, mức ý nghĩa 0,01:

Để cho số quan sát tối thiểu bằng 5, ta tiến hành gộp khoảng thành như sau:

Số tai nạn (X)	0	1	2	3	4	$\geq 5$
Số ngày (n)	10	32	46	35	20	13

Ước lượng điểm cho tham số  $\lambda$  của phân bố Poisson:

$$\lambda^* = \bar{X} = 2,4423 \quad (\text{theo câu a})$$

+ *Tính các xác suất lý thuyết theo phân bố Poisson ( $\lambda^*$ ):*

$$P(X = 0) = e^{-2,4423} \frac{2,4423^0}{0!} = 0,0870$$

$$P(X=1) = e^{-2,4423} \frac{2,4423^1}{1!} = 0,2124$$

$$P(X=2) = e^{-2,4423} \frac{2,4423^2}{2!} = 0,2594$$

$$P(X=3) = e^{-2,4423} \frac{2,4423^3}{3!} = 0,2111$$

$$P(X=4) = e^{-2,4423} \frac{2,4423^4}{4!} = 0,1289$$

$$P(X \geq 5) = 1 - P(X < 5) = 0,1012$$

Bài toán:

$$\begin{cases} H: X \text{ có phân bố Poisson} \\ K: X \text{ không có phân bố Poisson} \\ \text{Mức ý nghĩa } \alpha = 0,01 \end{cases}$$

Miền bác bỏ:

$$S = \left\{ \chi^2 \geq \chi_{k-r-1}^2(\alpha) \right\}$$

(với  $k = 6$  là số khoảng và  $r = 1$  là số lượng tham ẩn)

$$\text{Tính: } \chi^2 = \frac{1}{n} \sum_{i=1}^k \frac{m_i^2}{p_i} - n = \frac{1}{156} \left( \frac{10^2}{0,0870} + \frac{32^2}{0,2124} + \frac{46^2}{0,2594} + \frac{35^2}{0,2111} + \frac{20^2}{0,1289} + \frac{13^2}{0,1012} \right) - 156 = 2,358$$

Tra bảng của phân bố Khi bình phương ta được:  $\chi_{k-r-1}^2(\alpha) = \chi_4^2(0,01) = 13,28$

Do đó, miền bác bỏ không xảy ra, ta chấp nhận  $H$  và bác bỏ  $K$ . Vậy với mức ý nghĩa 0,01 ta tạm thời cho rằng số tai nạn trong một ngày có phân bố Poisson cho tới khi có thêm thông tin mới.

## 5. Đề thi cuối kỳ I năm học 2014 – 2015

ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

ĐỀ THI KẾT THÚC HỌC KỲ I  
NĂM HỌC 2014 – 2015

### Môn thi: Xác suất thống kê

Mã môn học: MAT 1101 Số tín chỉ: 03

Dành cho sinh viên các khoa: *Hóa học, Sinh học, Địa lý, Địa chất, Môi trường, KTTV-HDH, Y-Dược*

Dạng đề thi: ***Được sử dụng tài liệu***

Thời gian làm bài: 90 phút (không kể thời gian phát đề)

**Câu 1:** Có hai xạ thủ, mỗi người bắn một viên đạn vào một cái đích. Xác suất bắn trúng của hai xạ thủ lần lượt là 0,7; 0,8. Gọi X là số viên đạn trúng đích, tìm phân phối xác suất của X và tìm EX.

**Câu 2:** Một thùng chứa rất nhiều cam với tỷ lệ cam tốt là 80%. Một người chọn ngẫu nhiên 2 quả từ thùng cam bỏ vào một rổ cam trong đó đã có sẵn 3 quả cam tốt và 1 quả cam xấu. Sau đó người này lại chọn ngẫu nhiên 1 quả từ rổ.

a) Tính xác suất để quả cam lấy ra từ rổ là cam tốt.

b) Giả sử quả cam lấy ra từ rổ là cam tốt. Tính xác suất để trong 2 quả lấy ra từ thùng cam có ít nhất 1 quả cam tốt.

**Câu 3:** Tung con xúc xắc 100 lần ta thu được kết quả như sau:

Mặt	1	2	3	4	5	6
Số lần xuất hiện	16	15	17	18	9	25

Với mức ý nghĩa 5% có thể kết luận con xúc xắc này không cân đối đồng chất hay không?

**Câu 4:** Người ta cho 6 bệnh nhân sử dụng một loại thuốc A và sau đó tiến hành đo thời gian (tính bằng mili giây) phản ứng với thuốc của 6 bệnh nhân này thì thu được kết quả: 91, 87, 99, 77, 88, 91.

Sau đó người ta cho 6 bệnh nhân khác sử dụng loại thuốc B và cũng tiến hành đo thời gian (tính bằng mili giây) phản ứng với thuốc của 6 bệnh nhân này thì thu được kết quả: 101, 110, 103, 93, 99, 85. Giả thiết rằng thời gian phản ứng với các loại thuốc trên xấp xỉ phân bố chuẩn và độ lệch tiêu chuẩn của chúng là như nhau.

a) Với mức ý nghĩa 5% hãy kiểm tra xem thời gian phản ứng trung bình với thuốc B lớn hơn so với thuốc A hay không?

b) Với độ tin cậy 95% hãy ước lượng khoảng cho thời gian phản ứng trung bình với thuốc A và với thuốc B.

**Câu 5:** Doanh thu từng năm (tỷ đồng) của một công ty tính từ năm thành lập được cho trong bảng sau:

Năm thứ (X)	1	2	3	4	5
Doanh thu (Y)	12	19	29	37	45

- a) Lập phương trình hồi quy tuyến tính  $y = ax + b$ .  
 b) Sử dụng đường hồi quy trên để dự báo doanh thu của công ty trong năm thứ 8.

Lời giải:

### Câu 1:

Gọi X là tổng số viên đạn bắn trúng đích của hai xạ thủ. Mỗi xạ thủ bắn 1 viên nên X có thể nhận các giá trị là:  $X(\Omega) = \{0, 1, 2\}$

Tính các xác suất tương ứng:

$$\begin{aligned} P(X=0) &= 0,3 \cdot 0,2 = 0,06 && (\text{cả hai xạ thủ đều bắn trượt}) \\ P(X=1) &= 0,7 \cdot 0,2 + 0,3 \cdot 0,8 = 0,38 && (1 \text{ xạ thủ bắn trúng}, 1 \text{ xạ thủ bắn trượt}) \\ P(X=2) &= 0,7 \cdot 0,8 = 0,56 && (\text{cả hai xạ thủ đều bắn trúng}) \end{aligned}$$

⇒ Bảng phân bố xác suất của X:

X	0	1	2
P	0,06	0,38	0,56

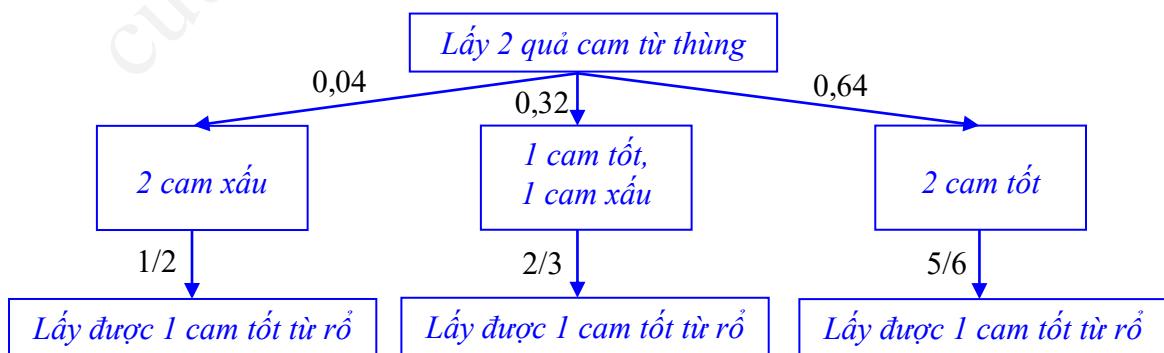
⇒ Số viên trúng đích trung bình:

$$EX = \sum_{i=1}^3 x_i p_i = 0 \cdot 0,06 + 1 \cdot 0,38 + 2 \cdot 0,56 = 1,5 \text{ (viên)}$$

### Câu 2:

Hướng dẫn: Dạng bài toán sử dụng công thức xác suất đầy đủ (toute phần) và công thức Bayes nhưng cách hồi dẽ gây hiểu nhầm lẫn. Có thể hình dung thành hai hành động: hành động I lấy ngẫu nhiên 2 quả từ thùng cam và hành động II lấy ngẫu nhiên 1 quả từ rổ cam, trong đó xác suất lấy được cam tốt từ rổ cam còn phụ thuộc vào việc lấy 2 quả từ thùng cam trước đó.

Có thể hình dung sơ đồ cây như sau:



- a) Tính xác suất để quả cam lấy ra từ rổ là cam tốt:

Gọi: X là số cam tốt lấy ra từ thùng cam (trong lần lấy đầu tiên)

H là biến cố lấy được cam tốt từ rổ cam (trong lần lấy thứ hai)

Tính các xác suất tương ứng:

$$P(X = 0) = C_2^0 \cdot 0,8^0 \cdot 0,2^2 = 0,04$$

$$P(X = 1) = C_2^1 \cdot 0,8^1 \cdot 0,2^1 = 0,32$$

$$P(X = 2) = C_2^2 \cdot 0,8^2 \cdot 0,2^0 = 0,64$$

+ Nếu không lấy được quả cam tốt nào từ lần lấy đầu tiên thì xác suất lấy được cam tốt ở lần thứ hai là:

$$P(H | X = 0) = \frac{3}{6} = \frac{1}{2}$$

+ Nếu lấy được một quả cam tốt từ lần lấy đầu tiên thì xác suất lấy được cam tốt ở lần thứ hai là:

$$P(H | X = 1) = \frac{4}{6} = \frac{2}{3}$$

+ Nếu lấy được cả hai quả cam tốt từ lần lấy đầu tiên thì xác suất lấy được cam tốt ở lần thứ hai là:

$$P(H | X = 2) = \frac{5}{6}$$

$\Rightarrow$  Xác suất lấy được cam tốt từ rổ cam: áp dụng CTXS đầy đủ

$$\begin{aligned} P(H) &= P(H | X = 0) \cdot P(X = 0) + P(H | X = 1) \cdot P(X = 1) + P(H | X = 2) \cdot P(X = 2) \\ &= 0,04 \cdot \frac{1}{2} + 0,32 \cdot \frac{2}{3} + 0,64 \cdot \frac{5}{6} = \boxed{\frac{23}{30}} \end{aligned}$$

b) Cam lấy ra từ rổ là cam tốt. Tính xác suất để trong hai quả lấy ra từ thùng cam có ít nhất một cam tốt.

*Hướng dẫn:* Bài toán cho biết trước kết quả xảy ra và có 3 nguyên nhân dẫn đến kết quả này: lần đầu không lấy được cam tốt, lấy được 1 cam tốt, lấy được 2 cam tốt. Yêu cầu tính xác suất để lần đầu lấy được 1 cam tốt hoặc 2 cam tốt (áp dụng công thức Bayes)

Áp dụng công thức Bayes ta được:

$$\begin{aligned} P(X \geq 1 | H) &= \frac{P(H | X = 1) \cdot P(X = 1) + P(H | X = 2) \cdot P(X = 2)}{P(H)} \\ &= \frac{0,32 \cdot \frac{2}{3} + 0,64 \cdot \frac{5}{6}}{\frac{23}{30}} = \frac{56}{75} : \frac{23}{30} = \boxed{\frac{112}{115}} \end{aligned}$$

### Câu 3:

*Hướng dẫn:* Con xúc xắc cân đối và đồng chất khi xác suất xuất hiện các mặt là như nhau. Bài toán dạng “Tiêu chuẩn phù hợp Khi bình phương”.

Bài toán:

$$\begin{cases} H: \text{con xúc xắc cân đối và đồng chất} \\ K: \text{con xúc xắc không cân đối và đồng chất} \\ \text{Mức ý nghĩa: } \alpha = 0,05 \end{cases}$$

Miền bác bỏ của bài toán:

$$S = \left\{ \chi^2 \geq \chi_{k-l}^2(\alpha) \right\} \quad (\text{với } k=6 \text{ là số khoảng chia})$$

$$\begin{aligned} \text{Tính: } \chi^2 &= \frac{1}{n} \sum_{i=1}^6 \frac{m_i^2}{p_i} - n = \frac{1}{100} \left( \frac{16^2}{1/6} + \frac{15^2}{1/6} + \frac{17^2}{1/6} + \frac{18^2}{1/6} + \frac{9^2}{1/6} + \frac{25^2}{1/6} \right) - 100 \\ &= 108 - 100 = 8 \end{aligned}$$

$$\text{Tra bảng ta được: } \chi_{k-l}^2(\alpha) = \chi_5^2(0,05) = 11,07$$

Do đó, miền bác bỏ không xảy ra, ta chấp nhận H và bác bỏ K. Vậy với mức ý nghĩa 5% ta chưa thể cho rằng con xúc xắc này **không** cân đối và đồng chất.

#### Câu 4:

a) VỚI MỨC Ý NGHĨA 5% HÃY KIỂM TRA XEM THỜI GIAN PHẢN ỨNG TRUNG BÌNH VỚI THUỐC B LỚN HƠN SO VỚI THUỐC A HAY KHÔNG:

*Hướng dẫn: Dạng bài toán so sánh hai giá trị trung bình, trường hợp thứ hai: biệt phân bố chuẩn nhưng chưa biết phương sai.*

Gọi X là thời gian phản ứng với thuốc A, Y là thời gian phản ứng với thuốc B.

Bài toán:

$$\begin{cases} H: EX = EY \\ K: EX < EY \quad (\alpha = 0,05) \end{cases}$$

Miền bác bỏ của bài toán:

$$\left\{ t \leq -t_{n_A+n_B-2}(\alpha) \right\}$$

Tính:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{n_A s_X^2 + n_B s_Y^2}} \sqrt{\frac{n_A n_B (n_A + n_B - 2)}{n_A + n_B}}$$

Sử dụng máy tính ta có:

$$\bar{X} = \frac{533}{6}; \bar{Y} = \frac{591}{6}$$

$$s_X^2 = 42,8056; s_Y^2 = 61,9167$$

$$\text{Suy ra: } t = \frac{(533 - 591)/6}{\sqrt{\sqrt{6.42,8056 + 6.61,9167}}} \sqrt{\frac{36.10}{12}} = -2,112$$

Tra bảng của phân bố Student ta được:  $-t_{n_A+n_B-2}(\alpha) = -t_{10}(0,05) = -1,812$

Do đó, miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy với mức ý nghĩa 5%, ta có thể kết luận rằng phản ứng trung bình của người bệnh với thuốc B lớn hơn với thuốc A.

b) *Uớc lượng khoảng cho thời gian phản ứng thuốc trung bình với thuốc A và thuốc B:*

*Hướng dẫn: Dạng bài toán ước lượng khoảng cho giá trị trung bình, trường hợp biết phân bố chuẩn nhưng chưa biết phương sai.*

\* Đối với thuốc A, ước lượng khoảng cho EX:

$$s_x^2 = 42,8056$$

⇒ Uớc lượng khoảng cho EX với độ tin cậy 95%:

$$\begin{aligned} EX &\in \left( \bar{X} - t_{n_A-1}\left(\frac{\alpha}{2}\right) \frac{s_x}{\sqrt{n_A-1}}; \quad \bar{X} + t_{n_A-1}\left(\frac{\alpha}{2}\right) \frac{s_x}{\sqrt{n_A-1}} \right) \\ EX &\in \left( \frac{533}{6} - t_5(0,025) \sqrt{\frac{42,8056}{5}}; \quad \frac{533}{6} + t_5(0,025) \sqrt{\frac{42,8056}{5}} \right) \\ EX &\in \left( \frac{533}{6} - 2,571 \sqrt{\frac{42,8056}{5}}; \quad \frac{533}{6} + 2,571 \sqrt{\frac{42,8056}{5}} \right) \end{aligned}$$

$$EX \in (81,31; 96,36)$$

Vậy, với độ tin cậy 95% thì thời gian phản ứng trung bình với thuốc A nằm trong khoảng (81,31; 96,36) mili giây.

\* Đối với thuốc B, ước lượng khoảng cho EY:

$$s_y^2 = 61,9167$$

⇒ Uớc lượng khoảng cho EY với độ tin cậy 95%:

$$\begin{aligned} EY &\in \left( \bar{Y} - t_{n_B-1}\left(\frac{\alpha}{2}\right) \frac{s_y}{\sqrt{n_B-1}}; \quad \bar{Y} + t_{n_B-1}\left(\frac{\alpha}{2}\right) \frac{s_y}{\sqrt{n_B-1}} \right) \\ EY &\in \left( \frac{591}{6} - t_5(0,025) \sqrt{\frac{61,9167}{5}}; \quad \frac{591}{6} + t_5(0,025) \sqrt{\frac{61,9167}{5}} \right) \\ EY &\in \left( \frac{591}{6} - 2,571 \sqrt{\frac{61,9167}{5}}; \quad \frac{591}{6} + 2,571 \sqrt{\frac{61,9167}{5}} \right) \end{aligned}$$

$$EY \in (89,45; 107,55)$$

Vậy, với độ tin cậy 95% thì thời gian phản ứng trung bình với thuốc B nằm trong khoảng (89,45; 107,55) mili giây.

**Câu 5:**

Gọi: X là thứ tự của năm

Y là doanh thu của năm tương ứng.

a) *Lập phương trình đường hồi quy tuyến tính thực nghiệm  $y = ax + b$ :*

\* *Hệ số tương quan mẫu:*

$$r = \frac{\frac{1}{n} \sum x_i y_i m_i - \bar{X} \cdot \bar{Y}}{s_x s_y}$$

Sử dụng máy tính bỏ túi ta tính được<sup>1</sup>:

$$\begin{aligned} &+ \frac{1}{n} \sum x_i y_i m_i = 102 \\ &+ \bar{X} = 3 \quad + s_x^2 = 2 \\ &+ \bar{Y} = 28,4 \quad + s_y^2 = 141,44 \end{aligned}$$

$\Rightarrow$  Hệ số tương quan mẫu:

$$r = \frac{102 - 3 \cdot 28,4}{\sqrt{2 \cdot 141,44}} = 0,9989 \quad (\text{mối quan hệ tuyến tính rất chặt})$$

\* *Phương trình đường hồi quy tuyến tính thực nghiệm của Y theo X:*

$$y - \bar{Y} = r \frac{s_y}{s_x} (x - \bar{X}) \Leftrightarrow y = r \frac{s_y}{s_x} (x - \bar{X}) + \bar{Y}$$

$$y = 0,9989 \sqrt{\frac{141,44}{2}} (x - 3) + 28,4$$

$$\Leftrightarrow \boxed{y = 8,4x + 3,1992}$$

b) *Dự báo doanh thu của công ty trong năm thứ 8:*

Phương trình đường hồi quy tuyến tính thực nghiệm của Y theo X có thể dùng để dự báo doanh thu của công ty tại một năm nhất định.

$$\text{Doanh thu trong năm thứ 8 là: } 8,4 \cdot 8 + 3,1992 = \boxed{70,4} \quad (\text{đơn vị tiền tệ})$$

<sup>1</sup> Xem cách tìm các chỉ số thống kê bằng máy tính bỏ túi tại *Phụ lục P.2.2*, trang 192.

## 6. Đề thi cuối kỳ II năm học 2014 – 2015

ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

ĐỀ THI KẾT THÚC HỌC KỲ II  
NĂM HỌC 2014 – 2015

### Môn thi: Xác suất thống kê

Mã môn học: MAT 1101 Số tín chỉ: 03

Dành cho sinh viên các khoa: Hóa học, Sinh học, Địa lý, Địa chất, Môi trường, KT-TV-HDH, Y-Dược

Dạng đề thi: **Được sử dụng tài liệu**

Thời gian làm bài: 90 phút (không kể thời gian phát đề)

**Câu 1.** Một công ty bảo hiểm phân chia khách hàng thành 3 loại: rủi ro nhất, rủi ro trung bình và ít rủi ro với tỷ lệ 2:5:3. Một báo cáo đưa ra khả năng một người rủi ro nhất, rủi ro trung bình và ít rủi ro gặp tai nạn trong khoảng thời gian 1 năm tương ứng là 0,05; 0,15 và 0,3.

a) Tính xác suất để một khách hàng bất kỳ gặp tai nạn trong một năm?

b) Nếu khách hàng A không có tai nạn nào trong năm 1987, tính xác suất để A thuộc lớp khách hàng rủi ro nhất?

**Câu 2.** Cho biến ngẫu nhiên X có phân bố với hàm mật độ xác suất

$$f(x) = \begin{cases} kx & \text{khi } 0 \leq x \leq 4 \\ 0 & \text{khác} \end{cases}$$

a) Tìm xác suất  $P(1 < X < 2)$

b) Quan sát X 10 lần và gọi Y là số lần X rơi vào khoảng (1, 2). Tìm phân bố xác suất của Y,  $EY$ ,  $DY$ .

**Câu 3.** Một bài viết trên *Tạp chí Khoa học Nông nghiệp* [“Sử dụng Phần dư hợp lý cực đại cho đặc điểm chất lượng mầm hạt của lúa mì với nhiều ảnh hưởng, khí hậu và phân bón nito” (1997, Vol.128, tr 135-142)] đã nghiên cứu kỳ vọng của hàm lượng protein thô lúa mì hạt (CP) và Hagberg rơi (HFN) được khảo sát ở Vương quốc Anh. Các phân tích sử dụng các áp dụng phân bón (kg N/ha), nhiệt độ ( $^{\circ}\text{C}$ ) và tổng lượng mưa tháng (mm). Số liệu dưới đây mô tả nhiệt độ cho lúa mì sinh trưởng tại Đại học Nông nghiệp Harper Adams từ 1981 đến 2014. Nhiệt độ đo được trong các tháng 6 như sau:

15,2; 14,2; 14,0; 12,2; 14,4; 12,5; 14,3; 14,2; 13,5; 11,8; 15,2; 15,4; 14,2; 14,0; 12,2; 14,4; 12,5; 14,3; 14,2; 13,5; 11,8; 15,2; 15,3; 14,2; 14,0; 12,2; 14,4; 12,5; 14,3; 14,2; 13,5; 11,8; 15,2

Giả sử nhiệt độ có phân bố chuẩn với độ lệch tiêu chuẩn  $\sigma = 0,5$ .

a) Xây dựng khoảng tin cậy 99% cho kỳ vọng nhiệt độ. Nếu sai số trong khoảng ước lượng nhỏ hơn  $2^{\circ}\text{C}$  thì kích cỡ mẫu sẽ là bao nhiêu?

b) Nếu độ rộng của khoảng tin cậy cho nhiệt độ trung bình là  $1,5^{\circ}\text{C}$  nhận được từ số liệu mẫu trên thì độ tin cậy bao nhiêu?

c) Kiểm tra nhiệt độ trung bình trong các tháng 6 có cao hơn  $13^{\circ}\text{C}$  hay không với mức ý nghĩa 5%?

d) Từ mẫu số liệu nhận được ta có thể khẳng định được có 30% các tháng 6 có nhiệt độ dưới 13; 40% các tháng 6 có nhiệt độ từ 13 đến 15 và 30% các tháng 6 có nhiệt độ từ 15 trở lên hay không với mức ý nghĩa 5%.

**Câu 4.** Một bài báo trong *Technometrics* của Narula & JFWellington [“Dự báo, hồi quy tuyến tính và tổng tối thiểu của sai số tương đối” (1977, Vol. 19)] trình bày dữ liệu về giá bán và thuế hàng năm cho 21 ngôi nhà.

Giá bán (X)	25,9	29,5	27,9	25,9	29,9	29,9	30,9	28,9	35,9	31,5
Thuế (Y)	4,918	5,021	4,543	4,557	5,060	3,891	5,898	5,604	5,828	5,300
Giá bán (X)	40,5	43,9	37,5	37,9	44,5	37,9	38,9	36,9	45,8	31,0
Thuế (Y)	7,784	9,038	5,989	7,542	8,795	6,083	8,361	8,140	9,142	6,271
										5,959

a) Xây dựng đường hồi quy tuyến tính thực nghiệm dự báo giá bán nhà theo thuế và ước lượng bình phương sai số khi dùng đường hồi quy trên.

b) Dự báo giá bán nhà nếu thuế là 7,50

Lời giải:

### **Câu 1:**

*Hướng dẫn: Dạng bài toán sử dụng công thức xác suất đầy đủ và công thức Bayes.*

a) Xác suất để một khách hàng bất kỳ gặp tai nạn trong năm:

Gọi các biến cố:

A = “Khách hàng rủi ro nhất”

B = “Khách hàng rủi ro trung bình”

C = “Khách hàng ít rủi ro”

H = “Khách hàng gặp tai nạn trong năm”

Theo bài ra ta có:

$$P(A) = 0,2; \quad P(B) = 0,5; \quad P(C) = 0,3$$

$$P(H | A) = 0,05; \quad P(H | B) = 0,15; \quad P(H | C) = 0,3$$

⇒ Xác suất một khách hàng bất kỳ gặp tai nạn trong năm:

$$\begin{aligned} P(H) &= P(HA) + P(HB) + P(HC) = P(H | A).P(A) + P(H | B).P(B) + P(H | C).P(C) \\ &= 0,05 \cdot 0,2 + 0,15 \cdot 0,5 + 0,3 \cdot 0,3 = \boxed{0,175} \end{aligned}$$

b) Người A không có tai nạn nào trong năm. Xác suất để A thuộc lớp khách hàng rủi ro nhất:

+ Xác suất để người A không bị tai nạn:

$$P(\bar{H}) = 1 - P(H) = 1 - 0,175 = 0,825$$

+ Xác suất để người A vừa không bị tai nạn vừa thuộc nhóm người rủi ro nhất:

$$P(\bar{H}A) = P(\bar{H} | A) \cdot P(A) = 0,95 \cdot 0,2 = 0,19$$

$\Rightarrow$  Xác suất để A thuộc nhóm người rủi ro nhất khi biết A là khách hàng không bị tai nạn nào trong năm:

$$P(A | \bar{H}) = \frac{P(\bar{H}A)}{P(\bar{H})} = \frac{0,19}{0,825} = \boxed{0,2303}$$

## Câu 2:

a) Tìm xác suất  $P(1 < X < 2)$ :

Vì  $f(x)$  là hàm mật độ của  $X$  nên suy ra:

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \Leftrightarrow \int_0^4 kx dx = 1 \Leftrightarrow k \frac{x^2}{2} \Big|_0^4 = 1 \Leftrightarrow 8k = 1 \Leftrightarrow k = \frac{1}{8}$$

$\Rightarrow$  Xác suất cần tìm:

$$P(1 < X < 2) = \int_1^2 \frac{1}{8} x dx = \frac{1}{8} \frac{x^2}{2} \Big|_1^2 = \frac{3}{16} = \boxed{0,1875}$$

b) Tìm phân bố xác suất của  $Y$  và  $EY, DY$ :

$Y$  là số lần quan sát thấy  $X$  rơi vào khoảng  $(1, 2)$  trong tổng số 10 lần quan sát nên  $Y$  có phân bố nhị thức:

$$Y \sim B(10; 0,1875)$$

+ Xác suất để có đúng  $k$  lần xuất hiện  $X$  trong 10 lần quan sát là:

$$P(Y = k) = C_{10}^k \cdot 0,1875^k \cdot 0,8125^{10-k}$$

$\Rightarrow$  Bảng phân bố xác suất của  $Y$ :

Y	0	1	2	3	4	5	6	7	8	9	10
P	0,1254	0,2893	0,3005	0,1849	0,0747	0,0207	0,0040	0,0005	5.10 <sup>-5</sup>	2.10 <sup>-6</sup>	5.10 <sup>-8</sup>

+ Kỳ vọng của  $Y$ :

$$EY = np = 10 \cdot 0,1875 = \boxed{1,875} (\text{lần})$$

+ Phuơng sai của  $Y$ :

$$DY = np(1-p) = 10 \cdot 0,1875 \cdot 0,8125 = \boxed{1,5234}$$

**Câu 3:**

a) Xây dựng khoảng tin cậy 99% cho kỳ vọng nhiệt độ. Tìm kích thước mẫu biết sai số nhỏ hơn  $2^{\circ}\text{C}$ .

Hướng dẫn: Dạng bài toán ước lượng khoảng cho giá trị trung bình, trường hợp biết phân bố chuẩn và biết phương sai lý thuyết.

Gọi  $X$  là nhiệt độ cho lúa mì. Nhiệt độ trung bình đo được từ mẫu là:

$$\bar{X} = \frac{\sum_{i=1}^{33} x_i}{33} = 13,78$$

$\Rightarrow$  Khoảng ước lượng cho kỳ vọng nhiệt độ:

$$EX \in \left( \bar{X} - u\left(\frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}}; \bar{X} + u\left(\frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}} \right)$$

$$EX \in \left( 13,78 - 2,576 \frac{0,5}{\sqrt{33}}; 13,78 + 2,576 \frac{0,5}{\sqrt{33}} \right)$$

$$EX \in (13,56; 14)$$

Vậy, với độ tin cậy 99% thì nhiệt độ trung bình nằm trong khoảng  $(13,56; 14)^{\circ}\text{C}$

+ Sai số của ước lượng nhỏ hơn 2 suy ra:

$$u\left(\frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}} < 2 \Leftrightarrow \sqrt{n} > \frac{u(\alpha/2).\sigma}{2} \Leftrightarrow n > \left( \frac{u(\alpha/2).\sigma}{2} \right)^2$$

$$\Leftrightarrow n > \left( \frac{2,576 \cdot 0,5}{2} \right)^2 = 0,415$$

$\Rightarrow$  Kích thước mẫu nhỏ nhất bằng 1.

b) Tìm độ tin cậy của ước lượng biết độ rộng của khoảng tin cậy là  $1,5^{\circ}\text{C}$ :

Độ rộng của khoảng tin cậy là  $1,5^{\circ}\text{C}$  suy ra:

$$2\varepsilon = 1,5 \Leftrightarrow 2u\left(\frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}} = 1,5 \Leftrightarrow u\left(\frac{\alpha}{2}\right) = \frac{1,5\sqrt{n}}{2\sigma} = \frac{1,5\sqrt{33}}{2 \cdot 0,5} = 8,62$$

$$\Rightarrow \Phi(u(\alpha/2)) = \Phi(8,62) \approx 1 \Rightarrow \alpha/2 \approx 0 \Rightarrow \alpha \approx 0$$

Do đó, độ tin cậy của ước lượng:  $1 - \alpha \approx 1$ . Độ tin cậy xấp xỉ 100%

c) Kiểm tra nhiệt độ trung bình của các tháng 6 có cao hơn  $13^{\circ}\text{C}$  hay không với mức ý nghĩa  $\alpha = 0,05$ .

Hướng dẫn: Dạng bài kiểm định giả thiết cho giá trị trung bình, trường hợp biết phân bố chuẩn và biết phương sai lý thuyết.

Bài toán:

$$\begin{cases} \text{Giả thiết H: } EX = 13 \\ \text{Đối thiết K: } EX > 13 \\ \text{Mức ý nghĩa: } \alpha = 0,05 \end{cases}$$

Miền bác bỏ của bài toán:

$$S = \left\{ \frac{\bar{X} - 13}{\sigma} \sqrt{n} \geq u(\alpha) \right\}$$

Ta có:

$$\frac{\bar{X} - 13}{\sigma} \sqrt{n} = \frac{13,78 - 13}{0,5} \sqrt{33} = 8,962$$

$$u(\alpha) = u(0,05) = 1,645$$

Do đó, miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy với mức ý nghĩa 5% ta tạm thời cho rằng nhiệt độ trung bình của các tháng 6 là cao hơn  $13^{\circ}\text{C}$ .

d) Các khoảng nhiệt độ ( $< 13^{\circ}\text{C}$ ); ( $13, 15^{\circ}\text{C}$ ); ( $15^{\circ}\text{C} <$ ) có theo tỷ lệ  $0,3 : 0,4 : 0,3$  hay không:

*Hướng dẫn: Dạng bài toán “Tiêu chuẩn phù hợp Khi bình phương” vì kiểm tra xem các khoảng nhiệt độ có tuân theo phân bố F nào đó hay không.*

Bài toán:

$$\begin{cases} H: \text{các khoảng nhiệt độ tuân theo tỷ lệ } 0,3 : 0,4 : 0,3 \\ K: \text{các khoảng nhiệt độ không tuân theo tỷ lệ trên} \\ \text{Mức ý nghĩa: } \alpha = 0,05 \end{cases}$$

Bảng thống kê được rút gọn thành:

Nhiệt độ X	$< 13$	( $13, 15$ )	$> 15$
$m_i$	9	18	6

Miền bác bỏ của bài toán:

$$S = \left\{ \chi^2 \geq \chi_{k-1}^2(\alpha) \right\} \quad (\text{với } k = 3 \text{ là số khoảng chia})$$

$$\text{Ta có: } \chi^2 = \frac{1}{n} \sum_{i=1}^3 \frac{m_i^2}{p_i} - n = \frac{1}{33} \left( \frac{9^2}{0,3} + \frac{18^2}{0,4} + \frac{6^2}{0,3} \right) - 33 = 36,36 - 33 = 3,36$$

Tra bảng của phân bố Khi bình phương ta được:  $\chi_{k-1}^2(\alpha) = \chi_2^2(0,05) = 5,99$

Do đó, miền bác bỏ không xảy ra, ta chấp nhận H và bác bỏ K. Vậy với mức ý nghĩa 5% ta tạm thời cho rằng các khoảng nhiệt độ tuân theo tỷ lệ  $0,3 : 0,4 : 0,3$ , cho tới khi có thêm thông tin mới.

**Câu 4:**

a) Xây dựng phương trình đường hồi quy tuyến tính thực nghiệm dự báo giá bán nhà theo thuế và ước lượng bình phương sai số khi dùng đường hồi quy:

+ Hệ số tương quan mẫu:

$$r = \frac{\frac{1}{n} \sum x_i y_i m_i - \bar{X} \cdot \bar{Y}}{s_X s_Y}$$

Sử dụng máy tính bỏ túi ta tính được<sup>1</sup>:

$$+ \frac{1}{n} \sum x_i y_i m_i = 227,405$$

$$+ \bar{X} = 34,3762 \quad + s_X^2 = 35,6313$$

$$+ \bar{Y} = 6,3678 \quad + s_Y^2 = 2,4867$$

⇒ Hệ số tương quan mẫu:

$$r = \frac{227,405 - 34,3762 \cdot 6,3678}{\sqrt{35,6313 \cdot 2,4867}} = 0,9035 \text{ (quan hệ tuyến tính rất chặt)}$$

\* Phương trình đường hồi quy tuyến tính thực nghiệm của giá bán nhà theo thuế (của X theo Y):

$$x - \bar{X} = r \frac{s_X}{s_Y} (y - \bar{Y}) \Leftrightarrow x = r \frac{s_X}{s_Y} (y - \bar{Y}) + \bar{X}$$

$$\Leftrightarrow x = 0,9035 \sqrt{\frac{35,6313}{2,4867}} (y - 6,3678) + 34,3762$$

$$\Leftrightarrow \boxed{x = 3,42 y + 12,598}$$

⇒ Ước lượng sai số bình phương trung bình khi dùng đường hồi quy trên:

$$s_{X/Y}^2 = s_X^2 (1 - r^2) = 35,6313 \cdot (1 - 0,9035^2) = \boxed{6,545}$$

b) Dự báo giá bán nhà nếu thuế là 7,5:

Phương trình đường hồi quy tuyến tính thực nghiệm có thể dùng để dự báo giá bán nhà theo thuế nhất định nào đó.

Nếu thuế là 7,5 thì dự báo giá bán nhà:

$$3,42 \cdot 7,5 + 12,598 \approx \boxed{38,25} \text{ (đơn vị tiền tệ)}$$

<sup>1</sup> Xem cách tìm các chỉ số thống kê bằng máy tính bỏ túi tại *Phụ lục P.2.2*, trang 192.

## 7. Đề thi cuối kỳ I năm học 2015 – 2016

ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

ĐỀ THI KẾT THÚC HỌC KỲ I  
NĂM HỌC 2015 – 2016

### Môn thi: Xác suất thống kê

Mã môn học: MAT 1101 Số tín chỉ: 03

Dành cho sinh viên các khoa: Hóa học, Sinh học, Địa lý, Địa chất, Môi trường, KT-TV-HDH, Y-Dược

Dạng đề thi: **Được sử dụng tài liệu**

Thời gian làm bài: 90 phút (không kể thời gian phát đề)

**Câu 1:** Một bệnh nhân bị nghi mắc một trong ba bệnh A, B, C với xác suất tương ứng là 0,3; 0,4; 0,3. Để biết chính xác bị bệnh gì người đó đi xét nghiệm. Biết rằng xác suất để xét nghiệm cho kết quả dương tính khi người đó mắc bệnh A, B, C lần lượt là 0,2; 0,5; 0,7.

- a) Tính xác suất xét nghiệm cho kết quả dương tính.
- b) Nếu bệnh nhân đó làm xét nghiệm 10 lần thì số lần cho kết quả âm tính tin chắc nhất là bao nhiêu?
- c) Giả sử xét nghiệm cho kết quả âm tính. Tính xác suất bệnh nhân mắc bệnh A.

**Câu 2:** Gieo một con xúc xắc cho đến khi được số nốt lớn hơn 4 thì dừng lại. Gọi X là số lần gieo cần thiết.

- a) Lập bảng phân bố xác suất của X.
- b) Vẽ trung bình phải gieo bao nhiêu lần mới được số nốt lớn hơn 4?

**Câu 3:** X(%) và Y (g/mm) là hai chỉ tiêu của một sản phẩm. Kiểm tra một số sản phẩm ta có:

X Y	0 - 5	5 - 10	10 - 15	15 - 20	20 - 25
115 - 125	7				
125 - 135	12	8	10		
135 - 145		20	15	2	
145 - 155		19	16	9	5
155 - 165				8	3

a) Giả sử trung bình tiêu chuẩn của Y là 120g/mm. Với mức ý nghĩa 1% hỏi tình hình sản xuất có thấp hơn tiêu chuẩn đề ra không?

b) Để ước lượng trung bình chỉ tiêu Y với độ chính xác 0,6 g/mm thì đảm bảo độ tin cậy là bao nhiêu?

c) Sản phẩm có chỉ tiêu  $X \geq 15\%$  là sản phẩm loại A. Tìm khoảng tin cậy cho tỷ lệ sản phẩm loại A với độ tin cậy 99%.

- d) Tính hệ số tương quan của X và Y  
e) Lập phương trình hồi quy tuyến tính của X theo Y. Dự đoán X biết Y = 145g/mm.

Lời giải:

### Câu 1:

Hướng dẫn: Câu a sử dụng công thức xác suất đầy đủ (toàn phần). Câu b là giá trị Mode của phân bố nhị thức tương ứng. Câu c cho biết kết quả xét nghiệm và yêu cầu tìm xác suất mắc bệnh A, áp dụng công thức Bayes.

a) Xác suất xét nghiệm cho kết quả dương tính:

Gọi các biến cố:

A = “Bệnh nhân mắc bệnh A”

B = “Bệnh nhân mắc bệnh B”

C = “Bệnh nhân mắc bệnh C”

H = “Xét nghiệm cho kết quả dương tính”

Theo bài ra ta có:  $P(A) = 0,3$ ;  $P(B) = 0,4$ ;  $P(C) = 0,3$

$P(H|A) = 0,2$ ;  $P(H|B) = 0,5$ ;  $P(H|C) = 0,7$

$\Rightarrow$  Xác suất bệnh nhân đi xét nghiệm cho kết quả dương tính:

$$\begin{aligned} P(H) &= P(HA) + P(HB) + P(HC) \\ &= P(H | A).P(A) + P(H | B).P(B) + P(H | C).P(C) \\ &= 0,2 \cdot 0,3 + 0,5 \cdot 0,4 + 0,7 \cdot 0,3 \Leftrightarrow P(H) = 0,47 \end{aligned}$$

b) Đi xét nghiệm 10 lần, số lần cho kết quả âm tính tin chắc nhất là bao nhiêu:

Xác suất cho kết quả âm tính trong 1 lần đi xét nghiệm là:

$$P(\bar{H}) = 1 - P(H) = 1 - 0,47 = 0,53$$

Gọi X là số lần xảy ra kết quả âm tính trong 10 lần đi xét nghiệm thì X có phân bố nhị thức:  $X \sim B(10; 0,53)$ . Số lần cho kết quả âm tính tin chắc nhất là:

$$\begin{aligned} \text{Mod } X &= [(n+1)p] \quad (\text{lấy giá trị phần nguyên}) \\ &= [11 \cdot 0,53] = 5 \end{aligned}$$

Vậy, số lần xảy ra kết quả âm tính có khả năng nhất (tin chắc nhất) là 5.

c) Xét nghiệm cho kết quả âm tính, tính xác suất bệnh nhân mắc bệnh A:

$$P(A | \bar{H}) = \frac{P(A\bar{H})}{P(\bar{H})} = \frac{P(\bar{H} | A).P(A)}{P(\bar{H})} = \frac{0,8 \cdot 0,3}{0,53}$$

$$\Leftrightarrow P(A | \bar{H}) = 0,4528$$

**Câu 2:** X là số lần gieo xúc xắc cho đến khi xuất hiện mặt có số nốt lớn hơn 4.

a) *Lập bảng phân bố xác suất của X:*

Trong mỗi lần gieo, xác suất ra số nốt lớn hơn 4 là: 1/3

+ Xác suất gieo được số nốt lớn hơn 4 trong lần gieo đầu tiên:  $P(X = 1) = 1/3$

+ Xác suất phải gieo đến lần thứ hai:

$P(X = 2) = 2/3 \cdot 1/3$  (lần đầu gieo hỏng và lần thứ hai gieo thành công)

+ Xác suất phải gieo đến lần thứ ba:

$P(X = 3) = (2/3)^2 \cdot 1/3$  (hai lần đầu gieo hỏng và lần thứ ba thành công)

Do đó, xác suất phải gieo đến lần thứ n:

$$P(X = n) = \left(\frac{2}{3}\right)^{n-1} \cdot \frac{1}{3} \quad (\text{với } n - 1 \text{ lần đầu gieo hỏng và lần thứ } n \text{ thành công})$$

=> Bảng phân bố xác suất của X:

X	1	2	3	...	n	...
P	$\frac{1}{3}$	$\frac{2}{3} \cdot \frac{1}{3}$	$\left(\frac{2}{3}\right)^2 \cdot \frac{1}{3}$	...	$\left(\frac{2}{3}\right)^{n-1} \cdot \frac{1}{3}$	...

b) Trung bình phải gieo bao nhiêu lần mới được số nốt lớn hơn 4:

$$\begin{aligned} EX &= \sum_{i=1}^n x_i p_i \quad (\text{với } p_i = P(X = x_i)) \\ &= 1 \cdot \frac{1}{3} + 2 \cdot \frac{2}{3} \cdot \frac{1}{3} + 3 \cdot \left(\frac{2}{3}\right)^2 \cdot \frac{1}{3} + \dots + n \cdot \left(\frac{2}{3}\right)^{n-1} \cdot \frac{1}{3} \\ &= \frac{1}{3} \left[ 1 + 2 \cdot \frac{2}{3} + 3 \cdot \left(\frac{2}{3}\right)^2 + \dots + n \cdot \left(\frac{2}{3}\right)^{n-1} \right] = \frac{1}{3} \sum_{n=1}^{\infty} n \left(\frac{2}{3}\right)^{n-1} \quad (1) \end{aligned}$$

Xét trường hợp tính tổng của chuỗi số có dạng  $\sum_{n=1}^{\infty} n x^{n-1}$ . Ta có:

$$\sum_{n=1}^{\infty} n x^{n-1} = \sum_{n=1}^{\infty} (x^n)'$$

Các  $x^n$  và tổng của chúng đều là hàm khả vi (đều lấy được đạo hàm) nên tổng các đạo hàm bằng đạo hàm của tổng:

$$\sum_{n=1}^{\infty} (x^n)' = \left( \sum_{n=1}^{\infty} x^n \right)' = (x + x^2 + x^3 + \dots x^n)'$$

Trong ngoặc đơn là tổng của cấp số nhân lùi vô hạn với công bội  $x < 1$ , suy ra:

$$(x + x^2 + x^3 + \dots x^n)' = \left( x \frac{1-x^n}{1-x} \right)'$$

Vì  $x < 1$  nên  $x^n \approx 0$  khi  $n \rightarrow \infty$ . Do đó:

$$\left( x \frac{1-x^n}{1-x} \right)' = \left( \frac{x}{1-x} \right)' = \frac{1}{(1-x)^2}$$

Kết hợp với (1) và thay  $x = 2/3$  ta được kết quả cần tìm:

$$EX = \frac{1}{3} \cdot \frac{1}{(1/3)^2} \Leftrightarrow [EX = 3]$$

### Câu 3: X(%) và Y (g/mm) là hai chỉ tiêu của một sản phẩm

Từ bảng tần suất đồng thời của X và Y, suy ra các bảng tần suất thành phần:

X	0 - 5	5 - 10	10 - 15	15 - 20	20 - 25
$m_x$	19	47	41	19	8

Y	115 - 125	125 - 135	135 - 145	145 - 155	155 - 165
$m_y$	7	30	37	49	11

a) Với mức ý nghĩa 1% hỏi tình hình sản xuất có thấp hơn tiêu chuẩn đề ra hay không (mức tiêu chuẩn về giá trị trung bình của Y là 120 g/mm):

Hướng dẫn: Dạng bài toán kiểm định giả thiết cho giá trị trung bình, trường hợp chưa biết phân bố chuẩn, chưa biết phương sai lý thuyết nhưng cỡ mẫu lớn.

Bài toán:

$$\begin{cases} \text{Giả thiết H: } EY = 120 \\ \text{Đối thiết K: } EY < 120 \\ \text{Mức ý nghĩa: } \alpha = 0,01 \end{cases}$$

Miền bác bỏ của bài toán:

$$S = \left\{ \frac{\bar{Y} - \mu_0}{s_Y} \sqrt{n_Y - 1} \leq -u(\alpha) \right\}$$

Sử dụng máy tính ta có:

$$\bar{Y} = 142$$

$$s_Y^2 = 112,8657$$

$$\Rightarrow \frac{\bar{Y} - \mu_0}{s_Y} \sqrt{n_Y - 1} = \frac{142 - 120}{\sqrt{112,8657}} \sqrt{133} = 23,88$$

Tra bảng của phân bố chuẩn tắc ta được:  $-u(\alpha) = -u(0,01) = -2,33$

Do đó miền bác bỏ không xảy ra, ta chấp nhận H và bác bỏ K. Vậy với mức ý nghĩa 1% thì tình hình sản xuất **không** thấp hơn tiêu chuẩn đề ra<sup>(1)</sup>.

b) Độ tin cậy là bao nhiêu để ước lượng cho giá trị trung bình của Y có độ chính xác là 0,6:

Độ chính xác khi ước lượng khoảng cho giá trị trung bình của Y:

$$\varepsilon = u\left(\frac{\alpha}{2}\right) \frac{s_Y}{\sqrt{n_Y - 1}} = 0,6 \Leftrightarrow u\left(\frac{\alpha}{2}\right) = 0,6 \frac{\sqrt{n_Y - 1}}{s_Y} = 0,6 \sqrt{\frac{133}{112,8657}}$$

$$\Leftrightarrow u\left(\frac{\alpha}{2}\right) = 0,65 \Leftrightarrow 1 - \frac{\alpha}{2} = 0,7422 \Leftrightarrow \frac{\alpha}{2} = 0,2578 \Leftrightarrow \alpha = 0,5156$$

Vậy, để độ chính xác của ước lượng là 0,6 thì độ tin cậy là:  $1 - \alpha = 0,4844$

c) Tìm khoảng tin cậy cho tỷ lệ sản phẩm loại A (sản phẩm có chỉ tiêu X ≥ 15%) với độ tin cậy 99%:

$$\text{Tỷ lệ mẫu: } p^* = \frac{19+8}{134} = \frac{27}{134}$$

⇒ Khoảng ước lượng cho tỷ lệ sản phẩm loại A:

$$p \in \left( p^* - u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p^*(1-p^*)}{n}}, \quad p^* + u\left(\frac{\alpha}{2}\right) \sqrt{\frac{p^*(1-p^*)}{n}} \right)$$

$$p \in \left( \frac{27}{134} - u(0,005) \sqrt{\frac{27 \cdot 107}{134^3}}, \quad \frac{27}{134} + u(0,005) \sqrt{\frac{27 \cdot 107}{134^3}} \right)$$

$$p \in \left( \frac{27}{134} - 2,58 \sqrt{\frac{27 \cdot 107}{134^3}}, \quad \frac{27}{134} + 2,58 \sqrt{\frac{27 \cdot 107}{134^3}} \right)$$

$$p \in (0,1121; \quad 0,2909)$$

Vậy, khoảng tin cậy 99% cho tỷ lệ sản phẩm loại A là (0,1121; 0,2909).

d) Tính hệ số tương quan của X và Y:

Hệ số tương quan mẫu:

$$r = \frac{\frac{1}{n} \sum_{i=1}^k x_i y_i m_i - \bar{X} \bar{Y}}{s_X s_Y}$$

<sup>(1)</sup> Cách khác ngắn gọn hơn: Từ trung bình mẫu  $\bar{Y} = 142$  ta có thể kết luận miền bác bỏ không xảy ra, không cần tính  $\frac{\bar{Y} - \mu_0}{s_Y} \sqrt{n_Y - 1}$  cũng như tra giá trị  $u(\alpha)$ . Lý do là vì  $\bar{Y} > \mu_0$  nên vế phải dương trong khi vế trái âm, do đó miền bác bỏ không xảy ra.

Sử dụng máy tính ta có:

$$\frac{1}{n} \sum_{i=1}^k x_i y_i m_i = \frac{207275}{134}$$

$$\bar{X} = 10,63; s_x^2 = 29,07$$

$$\bar{Y} = 142; s_y^2 = 112,8657 \text{ (theo câu a)}$$

Do đó:

$$r = \frac{\frac{207275}{134} - 10,63 \cdot 142}{\sqrt{29,07 \cdot 112,8657}} = \boxed{0,6524} \text{ (mức tương quan tuyến tính khá chặt)}$$

e) Lập phương trình hồi quy tuyến tính của X theo Y, dự đoán X biết Y = 145g/mm:

+ Phương trình đường hồi quy tuyến tính thực nghiệm của X theo Y:

$$x - \bar{X} = r \frac{s_x}{s_y} (y - \bar{Y}) \Leftrightarrow x = r \frac{s_x}{s_y} (y - \bar{Y}) + \bar{X}$$

$$x = 0,6524 \sqrt{\frac{29,07}{112,8657}} (y - 142) + 10,63$$

$$\Leftrightarrow \boxed{x = 0,3311y - 36,3858}$$

+ Dự đoán X biết Y = 145g/mm:

Phương trình đường hồi quy tuyến tính thực nghiệm của X theo Y dùng để dự đoán X khi biết Y, với Y = 145 thì dự đoán X là:

$$x = 0,3311 \cdot 145 - 36,3858 = 11,62 \text{ (%)}$$

## 8. Đề thi cuối kỳ I năm học 2015 – 2016 (dành cho sinh viên khoa Vật lý, hệ chuẩn)

ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

ĐỀ THI KẾT THÚC HỌC KỲ I  
NĂM HỌC 2015 – 2016

### Môn thi: Xác suất thống kê

Mã môn học: MAT 1101 Số tín chỉ: 03

Dành cho sinh viên: Khoa Vật lý (K59 Sư phạm Vật lý, K59 Khoa học vật liệu và K59 Công nghệ hạt nhân)

Dạng đề thi: ***Không được sử dụng tài liệu***

Thời gian làm bài: 90 phút (không kể thời gian phát đề)

**Câu 1:** Cho hàm mật độ phân phối xác suất  $f(x)$  của đại lượng ngẫu nhiên liên tục  $X$  có dạng:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(với  $\lambda$  là hằng số,  $\lambda > 0$ ). Hãy tìm kỳ vọng toán và phương sai của đại lượng ngẫu nhiên liên tục  $X$ .

**Câu 2:** Hai cầu thủ bóng rổ, mỗi người được ném bóng 3 lần. Xác suất trúng rổ của người thứ nhất trong mỗi lần ném bằng 0,8 và của người thứ hai trong mỗi lần ném bằng 0,6. Hãy tính xác suất:

- a) Sao cho số lần trúng rổ của họ như nhau.
- b) Sao cho số lần trúng rổ của người thứ nhất lớn hơn của người thứ hai.

**Câu 3:** Đại lượng ngẫu nhiên  $X$  có hàm mật độ phân phối xác suất:

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{a^2} e^{-\frac{x^2}{2a^2}} & x \geq 0 \end{cases}$$

(với  $a$  là hằng số,  $a > 0$ ). Hãy xác định mốc và trung vị của đại lượng ngẫu nhiên  $X$ .

**Câu 4:** Cho hàm mật độ phân phối xác suất của đại lượng ngẫu nhiên liên tục  $f(x)$  dạng:

$$f(x) = \begin{cases} 0 & x < 0 \\ ax^2 & 0 \leq x < 1 \\ a(2-x)^2 & 1 \leq x < 2 \\ 0 & x \geq 2 \end{cases}$$

Hãy xác định hệ số  $a$ , sau đó tính mômen gốc và mômen trung tâm tới bậc 4

**Câu 5:** Một xí nghiệp có hai phân xưởng với các tỷ lệ phế phẩm tương ứng là 3% và 5%. Biết rằng phân xưởng 1 sản xuất 60% sản phẩm còn phân xưởng 2 sản xuất 40% sản phẩm của xí nghiệp:

- Tìm xác suất để từ kho chung của xí nghiệp chọn ngẫu nhiên được 1 phế phẩm.
- Giả sử lấy được một phế phẩm, tìm xác suất để nó do phân xưởng 2 sản xuất ra.

Lời giải:

### Câu 1:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

a) *Kỳ vọng toán của X:*

$$EX = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^{+\infty} \lambda x e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^{+\infty} t e^{-t} dt \quad (\text{với } t = \lambda x)$$

Sử dụng phương pháp tích phân từng phần<sup>(1)</sup>: Đặt  $u = t$  và  $dv = e^{-t} dt$

$$EX = \frac{1}{\lambda} \left( -te^{-t} \Big|_0^{+\infty} - \int_0^{+\infty} -e^{-t} dt \right) = \frac{1}{\lambda} \left( 0 + \int_0^{+\infty} e^{-t} dt \right) = \frac{1}{\lambda} \left( -e^{-t} \Big|_0^{+\infty} \right)$$

$$\Leftrightarrow EX = \boxed{\frac{1}{\lambda}}$$

b) *Phương sai của X:*

$$DX = \int_{-\infty}^{+\infty} x^2 f(x) dx - (EX)^2$$

Ta có:

$$\int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx$$

Sử dụng phương pháp tích phân từng phần 2 lần:

+ Đặt  $u = \lambda x^2$  và  $dv = e^{-\lambda x} dx$  ta được:

$$\int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx = -x^2 e^{-\lambda x} \Big|_0^{+\infty} - \int_0^{+\infty} -2xe^{-\lambda x} dx = 0 + 2 \int_0^{+\infty} xe^{-\lambda x} dx$$

+ Đặt  $u = x$  và  $dv = e^{-\lambda x} dx$  ta được:

<sup>(1)</sup> Công thức tích phân từng phần:  $\int u dv = uv - \int v du$

$$2 \int_0^{+\infty} xe^{-\lambda x} dx = -\frac{2}{\lambda} xe^{-\lambda x} \Big|_0^{+\infty} - 2 \int_0^{+\infty} -\frac{1}{\lambda} e^{-\lambda x} dx = \frac{2}{\lambda} \int_0^{+\infty} e^{-\lambda x} dx = \frac{2}{\lambda} \left( \frac{-1}{\lambda} e^{-\lambda x} \right) \Big|_0^{+\infty} = \frac{2}{\lambda^2}$$

Do đó, phương sai cần tìm là:

$$DX = \int_{-\infty}^{+\infty} x^2 f(x) dx - (EX)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \Leftrightarrow DX = \boxed{\frac{1}{\lambda^2}}$$

### Câu 2:

Gọi X là số lần trúng rõ của người thứ nhất và Y là số lần trúng rõ của người thứ hai. Do đó, X và Y có thể nhận các giá trị 0, 1, 2, 3.

+ Các xác suất tương ứng của cầu thủ thứ nhất.

$$P(X=0) = C_3^0 \cdot 0,8^0 \cdot 0,2^3 = 0,008$$

$$P(X=1) = C_3^1 \cdot 0,8^1 \cdot 0,2^2 = 0,096$$

$$P(X=2) = C_3^2 \cdot 0,8^2 \cdot 0,2^1 = 0,384$$

$$P(X=3) = C_3^3 \cdot 0,8^3 \cdot 0,2^0 = 0,512$$

+ Các xác suất tương ứng của cầu thủ thứ hai.

$$P(Y=0) = C_3^0 \cdot 0,6^0 \cdot 0,4^3 = 0,064$$

$$P(Y=1) = C_3^1 \cdot 0,6^1 \cdot 0,4^2 = 0,288$$

$$P(Y=2) = C_3^2 \cdot 0,6^2 \cdot 0,4^1 = 0,432$$

$$P(Y=3) = C_3^3 \cdot 0,6^3 \cdot 0,4^0 = 0,216$$

a) Tính xác suất số lần trúng rõ của hai người là nhau nhau:

Vì hai người ném độc lập với nhau nên:  $P(X=x_i; Y=y_j) = P(X=x_i) \cdot P(Y=y_j) \quad \forall i, j$

Xác suất cần tìm là:

$$\begin{aligned} P(X=0; Y=0) + P(X=1; Y=1) + P(X=2; Y=2) + P(X=3; Y=3) \\ = 0,008 \cdot 0,064 + 0,096 \cdot 0,288 + 0,384 \cdot 0,432 + 0,512 \cdot 0,216 = \boxed{0,3046} \end{aligned}$$

b) Số lần trúng rõ của người thứ nhất lớn hơn của người thứ hai:

$$\begin{aligned} P(X > Y) &= P(X=1; Y=0) + P(X=2; Y=0) + P(X=3; Y=0) + \\ &\quad + P(X=2; Y=1) + P(X=3; Y=1) + P(X=3; Y=2) \\ &= 0,096 \cdot 0,064 + 0,384 \cdot 0,064 + 0,512 \cdot 0,064 + 0,384 \cdot 0,288 + 0,512 \cdot 0,288 \\ &\quad + 0,512 \cdot 0,432 = \boxed{0,5427} \end{aligned}$$

**Câu 3:** Đại lượng ngẫu nhiên X có hàm mật độ:

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{a^2} e^{-\frac{x^2}{2a^2}} & x \geq 0 \end{cases}$$

a) Tìm mode X:

Mode X là giá trị  $x_0$  mà tại đó hàm mật độ  $f(x)$  đạt giá trị lớn nhất.

Khảo sát hàm  $f(x)$  trên  $[0, +\infty)$ . Ta có:

$$f'(x) = \frac{1}{a^2} e^{-\frac{x^2}{2a^2}} - \frac{x}{a^2} \frac{x}{a^2} e^{-\frac{x^2}{2a^2}} = e^{-\frac{x^2}{2a^2}} \left( \frac{1}{a^2} - \frac{x^2}{a^4} \right)$$

Cho:  $f'(x) = 0$

$$\Leftrightarrow e^{-\frac{x^2}{2a^2}} \left( \frac{1}{a^2} - \frac{x^2}{a^4} \right) = 0 \Leftrightarrow \frac{1}{a^2} = \frac{x^2}{a^4} \Leftrightarrow x^2 = a^2 \Leftrightarrow x = a \quad (\text{vì } x > 0)$$

Lập bảng xét dấu  $f'(x)$ :

x	0	a
	$\frac{1}{a^2}$	+
	0	-


CD

Vậy: Mod X = a

b) Tìm trung vị của X:

Trung vị của X là giá trị m sao cho  $P(X < m) = P(X > m) = 1/2$

Do đó:

$$\int_{-\infty}^m f(x) dx = \frac{1}{2} \Leftrightarrow \int_0^m \frac{x}{a^2} e^{-\frac{x^2}{2a^2}} dx = \frac{1}{2}$$

Đặt  $t = x^2 \Rightarrow dt = 2x dx$ , ta được:

$$\begin{aligned} \int_0^{m^2} \frac{1}{2a^2} e^{-\frac{t}{2a^2}} dt = \frac{1}{2} &\Leftrightarrow \frac{1}{2a^2} (-2a^2) e^{-\frac{t}{2a^2}} \Big|_0^{m^2} = \frac{1}{2} \Leftrightarrow -e^{-\frac{t}{2a^2}} \Big|_0^{m^2} = \frac{1}{2} \\ &\Leftrightarrow -\left( e^{-\frac{m^2}{2a^2}} - 1 \right) = \frac{1}{2} \Leftrightarrow e^{-\frac{m^2}{2a^2}} = \frac{1}{2} \Leftrightarrow -\frac{m^2}{2a^2} = \ln \frac{1}{2} \Leftrightarrow \boxed{m = \sqrt{2a^2 \ln 2}} \end{aligned}$$

**Câu 4:** Đại lượng ngẫu nhiên X có hàm mật độ:

$$f(x) = \begin{cases} 0 & x < 0 \\ ax^2 & 0 \leq x < 1 \\ a(2-x)^2 & 1 \leq x < 2 \\ 0 & x \geq 2 \end{cases}$$

\* Xác định hệ số a:

Vì  $f(x)$  là hàm mật độ của X nên:

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \Leftrightarrow \int_0^1 ax^2 dx + \int_1^2 a(2-x)^2 dx = 1$$

$$\Leftrightarrow \frac{ax^3}{3} \Big|_0^1 + a \left( 4x - 2x^2 + \frac{x^3}{3} \right) \Big|_1^2 = 1 \Leftrightarrow \frac{a}{3} + \frac{a}{3} = 1 \Leftrightarrow \boxed{a = \frac{3}{2}}$$

\* Tính mômen gốc bậc 4:

$$v_4 = E(X^4) = \int_{-\infty}^{+\infty} x^4 f(x) dx \Leftrightarrow \frac{3}{2} \int_0^1 x^6 dx + \frac{3}{2} \int_1^2 x^4 (2-x)^2 dx$$

$$= \frac{3}{2} \frac{x^7}{7} \Big|_0^1 + \frac{3}{2} \left( \frac{4x^5}{5} - \frac{4x^6}{6} + \frac{x^7}{7} \right) \Big|_1^2 = \frac{3}{14} + \frac{3}{2} \left( \frac{128}{5} - \frac{256}{6} + \frac{128}{7} - \frac{4}{5} + \frac{4}{6} - \frac{1}{7} \right)$$

$$= \frac{3}{14} + \frac{3}{2} \cdot \frac{33}{35} \Leftrightarrow \boxed{v_4 = \frac{57}{35}}$$

\* Tính mômen trung tâm bậc 4:

$$\mu_4 = E[(X - EX)^4]$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx \Leftrightarrow \frac{3}{2} \int_0^1 x^3 dx + \frac{3}{2} \int_1^2 x(2-x)^2 dx = \frac{3x^4}{8} \Big|_0^1 + \frac{3}{2} \left( 2x^2 - \frac{4}{3}x^3 + \frac{x^4}{4} \right) \Big|_1^2$$

$$= \frac{3}{8} + \frac{3}{2} \cdot \frac{5}{12} = 1$$

$\Rightarrow$  Mômen trung tâm bậc 4:

$$\mu_4 = E[(X-1)^4] = \int_{-\infty}^{+\infty} (x-1)^4 f(x) dx = \frac{3}{2} \int_0^1 (x-1)^4 x^2 dx + \frac{3}{2} \int_1^2 (x-1)^4 (2-x)^2 dx$$

$$= \frac{3}{2} \int_0^1 (x^4 - 4x^3 + 6x^2 - 4x + 1) x^2 dx + \frac{3}{2} \int_1^2 (x^4 - 4x^3 + 6x^2 - 4x + 1) (2-x)^2 dx$$

$$\begin{aligned}
&= \frac{3}{2} \int_0^1 (x^6 - 4x^5 + 6x^4 - 4x^3 + x^2) dx + \\
&\quad + \frac{3}{2} \int_1^2 (x^6 - 8x^5 + 26x^4 - 44x^3 + 41x^2 - 20x + 4) dx \\
&= \frac{3}{2} \left( \frac{x^7}{7} - \frac{2x^6}{3} + \frac{6x^5}{5} - x^4 + \frac{x^3}{3} \right) \Big|_0^1 + \frac{3}{2} \left( \frac{x^7}{7} - \frac{4x^6}{3} + \frac{26x^5}{5} - 11x^4 + \frac{41x^3}{3} - 10x^2 + 4x \right) \Big|_1^2 \\
&= \frac{3}{2} \cdot \frac{1}{105} + \frac{3}{2} \left( \frac{24}{35} - \frac{71}{105} \right) = \frac{3}{2} \cdot \frac{1}{105} + \frac{3}{2} \cdot \frac{1}{105} \Leftrightarrow \boxed{\mu_4 = \frac{3}{105}}
\end{aligned}$$

**Câu 5:**

Gọi: A = “Sản phẩm do phân xưởng 1 sản xuất”

B = “Sản phẩm do phân xưởng 2 sản xuất”

H = “Lấy được phé phẩm”

Theo bài ra ta có:  $P(A) = 0,6$ ;  $P(B) = 0,4$ ;  $P(H|A) = 0,03$ ;  $P(H|B) = 0,05$

a) *Tìm xác suất để từ kho chung của xí nghiệp chọn ngẫu nhiên được 1 phé phẩm:*

Áp dụng công thức xác suất đầy đủ:

$$\begin{aligned}
P(H) &= P(H | A).P(A) + P(H | B).P(B) \\
&= 0,03 \cdot 0,6 + 0,05 \cdot 0,4 = \boxed{0,038}
\end{aligned}$$

b) *Giả sử lấy được phé phẩm, tìm xác suất để nó do phân xưởng 2 sản xuất:*

Áp dụng công thức Bayes:

$$P(B | H) = \frac{P(HB)}{P(H)} = \frac{P(H | B).P(B)}{P(H)} = \frac{0,05 \cdot 0,4}{0,38} = \boxed{0,5263}$$

## 9. Đề thi cuối kỳ II năm học 2015 – 2016

ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

ĐỀ THI KẾT THÚC HỌC KỲ II  
NĂM HỌC 2015 – 2016

### Môn thi: Xác suất thống kê

Mã môn học: MAT 1101 Số tín chỉ: 03

Dành cho sinh viên các khoa: Hóa học, Sinh học, Địa lý, Địa chất, Môi trường, KTTV-HDH, Y-Dược

Dạng đề thi: *Không được sử dụng tài liệu*

Thời gian làm bài: 90 phút (không kể thời gian phát đề)

**Câu 1:** Thống kê tình hình mắc bệnh A trong nhiều năm cho thấy tỷ lệ mắc bệnh A là 10%. Một người ốm muốn kiểm tra xem mình có mắc bệnh A không và bệnh viện thực hiện hai xét nghiệm T<sub>1</sub> và T<sub>2</sub> một cách độc lập cho người này. Biết rằng nếu có bệnh A thì xét nghiệm T<sub>1</sub> cho dương tính với xác suất 0,95; còn xét nghiệm T<sub>2</sub> cho dương tính với xác suất 0,85. Trong trường hợp không mắc bệnh A thì xét nghiệm T<sub>1</sub> cho dương tính với xác suất 0,2 còn xét nghiệm T<sub>2</sub> cho dương tính với xác suất 0,1.

- a. Nếu người đó mắc bệnh A thì xác suất để ít nhất 1 trong hai xét nghiệm T<sub>1</sub> và T<sub>2</sub> cho kết quả dương tính là bao nhiêu?
- b. Nếu cả hai xét nghiệm T<sub>1</sub> và T<sub>2</sub> cho kết quả đều dương tính thì xác suất mắc bệnh A của người đó là bao nhiêu?

**Câu 2:** Giả sử tuổi thọ của một loại thiết bị A (tính theo năm) là một biến ngẫu nhiên có hàm mật độ như sau:

$$f(x) = \begin{cases} C(5-x) & \text{khi } 0 < x < 5 \\ 0 & \text{khác} \end{cases}$$

Trong đó, C là một hằng số nào đó.

- a. Tìm C và tính tuổi thọ trung bình của thiết bị A
- b. Một người có 10 thiết bị A. Tính xác suất để có ít nhất 4 thiết bị mà tuổi thọ của chúng ít nhất là 3 năm.

**Câu 3:** Tiến hành đo hàm lượng chất Salmonella trong 9 lô kem được chọn ngẫu nhiên từ một nhà máy, ta thu được hàm lượng như sau (đơn vị MPN/g):

0,593 0,142 0,329 0,691 0,231 0,793 0,519 0,392 0,418

Với mức ý nghĩa 0,05 có thể kết luận hàm lượng trung bình của Salmonella trong kem có lớn hơn 0,3 MPN/g hay không?

**Câu 4:** Cho biết nồng độ alkalinity của các con sông ở khu vực thượng nguồn có phân bố chuẩn với độ lệch tiêu chuẩn là 10mg/l. Trong khi đó nồng độ alkalinity của các con sông ở khu vực hạ nguồn có phân bố chuẩn với độ lệch tiêu chuẩn là 25mg/l.

Đo nồng độ alkalinity của một con sông (đo nhiều lần) ở khu vực thượng nguồn và hạ nguồn ta thu được kết quả như sau:

Kết quả đo thượng nguồn	91	75	91	88	94	63	86	77	71	69
Kết quả đo hạ nguồn	86	95	135	121	68	64	113	108	79	
	62	143	108	121	85	97				

Với mức ý nghĩa 0,01 liệu có thể kết luận nồng độ alkalinity trung bình ở khu vực hạ nguồn cao hơn ở khu vực thượng nguồn không?

**Câu 5:** Điểm môn Xác suất và môn Thống kê của 5 sinh viên được cho dưới đây:

Sinh viên	1	2	3	4	5
Xác suất (x)	95	85	80	70	60
Thống kê (y)	85	95	70	65	70

a. Lập phương trình hồi quy tuyến tính  $y = ax + b$

b. Nếu một sinh viên có điểm Xác suất là 80, hãy ước lượng điểm môn Thống kê của sinh viên này.

Cho biết:

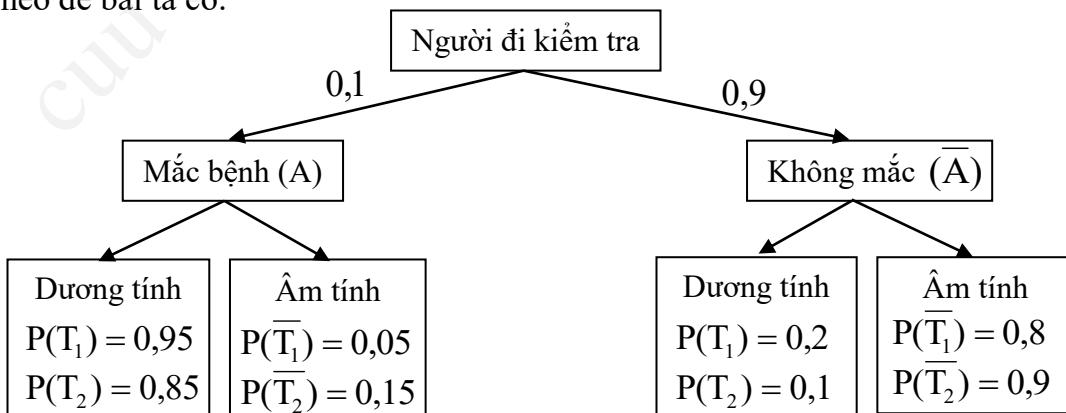
$$t_8(0,05) = 1,860; t_8(0,025) = 2,306; t_{23}(0,01) = 2,5; t_{23}(0,005) = 2,807$$

$$u(0,05) = 1,65; u(0,025) = 1,96; u(0,01) = 2,327; u(0,005) = 2,576$$

Lời giải:

**Câu 1:**

Theo đề bài ta có:



Giả sử các biến cố:

A = “Mắc bệnh A”

T<sub>1</sub> = “Xét nghiệm T<sub>1</sub> cho kết quả dương tính”

T<sub>2</sub> = “Xét nghiệm T<sub>2</sub> cho kết quả dương tính”

a) Xác suất để ít nhất 1 trong hai xét nghiệm cho kết quả dương tính, nếu người đó mắc bệnh A:

Nếu người đó mắc bệnh A thì xác suất để cả hai xét nghiệm cùng cho kết quả **âm tính** là:

$$P(\overline{T_1} \cdot \overline{T_2} | A) = 0,05 \cdot 0,15 = 0,0075$$

=> Xác suất để ít nhất 1 trong hai xét nghiệm cho kết quả dương tính:

$$1 - P(\overline{T_1} \cdot \overline{T_2} | A) = 1 - 0,0075 = \boxed{0,9925}$$

b) Nếu cả hai xét nghiệm  $T_1$  và  $T_2$  cho kết quả đều dương tính thì xác suất mắc bệnh A của người đó là bao nhiêu:

+ Nếu người đó mắc bệnh A thì xác suất để cả hai xét nghiệm cùng dương tính là:

$$P(T_1 T_2 | A) = 0,95 \cdot 0,85 = 0,8075$$

+ Nếu người đó không mắc bệnh A thì xác suất để cả hai xét nghiệm cùng dương tính là:

$$P(T_1 T_2 | \overline{A}) = 0,2 \cdot 0,1 = 0,02$$

=> Xác suất cần tìm: (áp dụng công thức Bayes)

$$\begin{aligned} P(A | T_1 T_2) &= \frac{P(T_1 T_2 | A) \cdot P(A)}{P(T_1 T_2 | A) \cdot P(A) + P(T_1 T_2 | \overline{A}) \cdot P(\overline{A})} = \frac{0,8075 \cdot 0,1}{0,8075 \cdot 0,1 + 0,02 \cdot 0,9} \\ &= \boxed{0,8177} \end{aligned}$$

## Câu 2:

Giả sử, tuổi thọ của thiết bị A là một biến ngẫu nhiên X có hàm mật độ:

$$f(x) = \begin{cases} C(5-x) & \text{khi } 0 < x < 5 \\ 0 & \text{khác} \end{cases}$$

a) Tìm C và tính tuổi thọ trung bình của thiết bị A:

Vì  $f(x)$  là hàm mật độ của X nên suy ra:

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x) dx = 1 &\Leftrightarrow \int_0^5 C(5-x) dx = 1 \Leftrightarrow C \left( 5x - \frac{x^2}{2} \right) \Big|_0^5 = 1 \Leftrightarrow C \frac{25}{2} = 1 \\ &\Leftrightarrow \boxed{C = \frac{2}{25}} \end{aligned}$$

b) Tính xác suất để có ít nhất 4 trong 10 thiết bị mà tuổi thọ của chúng ít nhất là 3 năm:

Lấy ngẫu nhiên 1 thiết bị A, xác suất để thiết bị này có tuổi thọ ít nhất là 3 năm:

$$\begin{aligned} P(X \geq 3) &= \int_{3}^{5} \frac{2}{25} (5-x) dx = \frac{2}{25} \left( 5x - \frac{x^2}{2} \right) \Big|_3^5 = \frac{2}{25} \left( 25 - \frac{25}{2} - 15 + \frac{9}{2} \right) \\ &= \frac{4}{25} \end{aligned}$$

Gọi Y là số thiết bị có tuổi thọ ít nhất là 3 năm. Ta có  $Y \sim B(10; 4/25)$   
 $\Rightarrow$  Xác suất có ít nhất 4 thiết bị mà tuổi thọ ít nhất 3 năm là:

$$P(Y \geq 4) = 1 - [P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3)]$$

Ta có:

$$P(Y = 0) = C_{10}^0 \left( \frac{4}{25} \right)^0 \left( \frac{21}{25} \right)^{10} = 0,1749$$

$$P(Y = 1) = C_{10}^1 \left( \frac{4}{25} \right)^1 \left( \frac{21}{25} \right)^9 = 0,3331$$

$$P(Y = 2) = C_{10}^2 \left( \frac{4}{25} \right)^2 \left( \frac{21}{25} \right)^8 = 0,2856$$

$$P(Y = 3) = C_{10}^3 \left( \frac{4}{25} \right)^3 \left( \frac{21}{25} \right)^7 = 0,145$$

Do đó, xác suất cần tìm:

$$P(Y \geq 4) = 1 - (0,1749 + 0,3331 + 0,2856 + 0,1450)$$

$$= [0,0614]$$

**Câu 3:** Có thể kết luận hàm lượng trung bình của Salmonella trong kem có lớn hơn 0,3 MPN/g hay không?

Gọi X là đại lượng ngẫu nhiên chỉ hàm lượng Salmonella trong kem. Dạng bài toán kiểm định giả thiết cho giá trị trung bình, trường hợp chưa biết phương sai lý thuyết và đại lượng ngẫu nhiên chuẩn<sup>1</sup>.

Bài toán:

$$\begin{cases} \text{Giả thiết } H: EX = 0,3 \\ \text{Đối thiết } K: EX > 0,3 \\ \text{Mức ý nghĩa: } \alpha = 0,05 \end{cases}$$

$$\text{Miền bác bỏ của bài toán: } S = \left\{ \frac{\bar{X} - 0,3}{s} \sqrt{n-1} \geq t_{n-1}(\alpha) \right\}$$

<sup>1</sup> Trong đề bài Câu 3, lẽ ra phải cho “Hàm lượng chất Salmonella là biến ngẫu nhiên chuẩn” thì mới đủ dữ kiện để giải bài toán.

Sử dụng máy tính ta có:

$$\bar{X} = 0,4564; \quad s = 0,0403$$

$$\Rightarrow \frac{\bar{X} - 0,3}{s} \sqrt{n-1} = \frac{0,4564 - 0,3}{0,0403} \sqrt{8} = 10,98$$

Tra bảng của phân bố Student ta được:  $t_{n-1}(\alpha) = t_8(0,05) = 1,86$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy, với mức ý nghĩa 0,05 ta tạm thời cho rằng hàm lượng Salmonella trung bình lớn hơn 0,3MPN/g; cho đến khi có thêm thông tin mới.

**Câu 4:** Có thể kết luận nồng độ độ alkalinity trung bình ở khu vực hạ nguồn sông cao hơn ở khu vực thượng nguồn không?

Gọi X là đại lượng ngẫu nhiên thể hiện nồng độ alkalinity ở khu vực thượng nguồn và Y ứng với khu vực hạ nguồn. Dạng bài toán so sánh 2 giá trị trung bình, trường hợp đã biết 2 phương sai lý thuyết và 2 đại lượng ngẫu nhiên có phân bố chuẩn.

Bài toán:

$$H : EX = EY \quad | \quad K : EX < EY \quad \text{mức ý nghĩa } \alpha = 0,01$$

Miền bác bỏ:

$$S = \left\{ \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq -u(\alpha) \right\}$$

(trong đó  $n_1, n_2$ , lần lượt là cỡ mẫu của X và Y)

Sử dụng máy tính ta có:  $\bar{X} = 80,5; \bar{Y} = 99$

$$\Rightarrow \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{80,5 - 99}{\sqrt{\frac{10^2}{10} + \frac{25^2}{15}}} = -2,57$$

Tra bảng của phân bố chuẩn ta được:  $-u(\alpha) = -u(0,01) = -2,327$

Do đó miền bác bỏ đã xảy ra, ta bác bỏ H và chấp nhận K. Vậy với mức ý nghĩa 0,01 ta tạm thời cho rằng nồng độ alkalinity ở khu vực hạ nguồn cao hơn ở khu vực thượng nguồn, cho đến khi có thêm thông tin mới.

**Câu 5:**

a) Lập phương trình hồi quy tuyến tính  $y = ax + b$ :

Hệ số tương quan mẫu:

$$r = \frac{\frac{1}{n} \sum x_i y_i m_i - \bar{x} \bar{y}}{s_x s_y}$$

Sử dụng máy tính ta có:

$$+ \frac{1}{n} \sum x_i y_i m_i = 6100$$

$$+ \bar{x} = 78 \quad + s_x = 12,083$$

$$+ \bar{y} = 77 \quad + s_y = 11,225$$

$$\Rightarrow r = \frac{6100 - 78.77}{12,083.11,225} = 0,693 \text{ (tương quan tuyến tính khá chặt)}$$

$\Rightarrow$  Phương trình đường hồi quy tuyến tính của y theo x:

$$y - \bar{y} = r \frac{s_y}{s_x} (x - \bar{x}) \Leftrightarrow y = r \frac{s_y}{s_x} (x - \bar{x}) + \bar{y}$$

$$\Leftrightarrow y = 0,693 \frac{11,225}{12,083} (x - 78) + 77$$

$$\Leftrightarrow \boxed{y = 0,6438x + 26,78}$$

b) Uớc lượng cho điểm môn Thống kê của một sinh viên nếu sinh viên đó có điểm môn Xác suất là 80:

Phương trình đường hồi quy tuyến tính thực nghiệm của y theo x có thể dùng để dự báo (xấp xỉ) điểm môn Thống kê theo điểm môn Xác suất.

Nếu sinh viên có điểm môn Xác suất là 80 thì ước lượng điểm môn Thống kê là:

$$0,6438.80 + 26,78 \approx \boxed{78,3} \text{ (điểm)}$$

## 10. Đề thi cuối kỳ II năm học 2015 – 2016 (dành cho sinh viên khoa Vật lý, hệ chuẩn)

ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

ĐỀ THI KẾT THÚC HỌC KỲ II  
NĂM HỌC 2015 – 2016

### Môn thi: Xác suất thống kê

Mã môn học: Số tín chỉ: 03

Dành cho sinh viên: Khoa Vật lý, K60

Dạng đề thi: Không được sử dụng tài liệu

Thời gian làm bài: 90 phút (không kể thời gian phát đề)

**Câu 1** (2 điểm): Phát biểu và chứng minh định lý về xác suất của tổng hai biến có không xung khắc.

**Câu 2** (2 điểm): Một rổ cam gồm 3 loại: loại 1 có 20%, loại 2 có 30%, loại 3 có 50%; với các tỷ lệ cam hỏng tương ứng là 2%, 3% và 1%.

- Người ta lấy hú họa ra 1 quả cam. Tính xác suất để quả cam đó là quả hỏng.
- Giả sử lấy được 1 quả hỏng, tính xác suất để nó là quả cam loại 1.

**Câu 3** (2 điểm): X là đại lượng ngẫu nhiên tuân theo phân phối Poisson, tức là có bảng phân phối xác suất dạng:

X	0	1	2	...	k	...
P	$e^{-\lambda}$	$\frac{\lambda e^{-\lambda}}{1!}$	$\frac{\lambda^2 e^{-\lambda}}{2!}$	...	$\frac{\lambda^k e^{-\lambda}}{k!}$	...

Tìm kỳ vọng toán và phương sai của X.

**Câu 4** (3 điểm): Cho đại lượng ngẫu nhiên X có hàm mật độ:

$$f(x) = \begin{cases} \frac{k}{x^3} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

- Tìm k, tìm hàm phân phối xác suất của X
- Tìm  $P(1 < X < 3)$ . Tìm trung vị của X

**Câu 5** (1 điểm): Cho 1 bảng phân phối thực nghiệm như sau

X	9,8	9,9	10	10,1	10,2
W (tần suất)	0,05	0,25	0,4	0,2	0,1

Tìm hàm phân phối mẫu.

Lời giải:

**Câu 1:** Định lý về xác suất của tổng hai biến có không xung khắc:

\* Nội dung định lý:

Hoàng Văn Trọng – 0974.971.149

Xác suất của tổng hai biến cố không xung khắc bằng tổng các xác suất của chúng trừ đi xác suất tích của chúng.

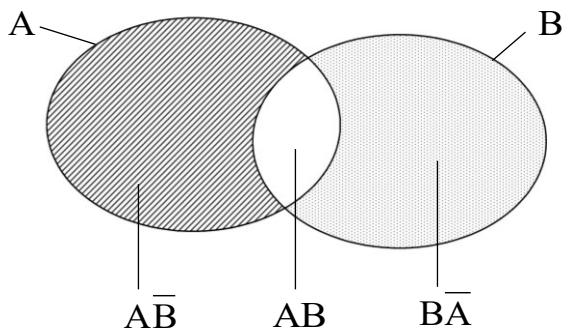
$$\text{Biểu thức: } P(A + B) = P(A) + P(B) - P(AB)$$

\* *Chứng minh:*

Biểu diễn theo quan hệ của các biến cố, ta có:

$$A + B = A\bar{B} + AB + B\bar{A}$$

Vì các biến cố  $A\bar{B}, AB, B\bar{A}$  là xung khắc với nhau nên suy ra:



$$P(A + B) = P(A\bar{B}) + P(AB) + P(B\bar{A}) \quad (1)$$

$$\text{Mà: } \begin{cases} P(A) = P(A\bar{B}) + P(AB) \\ P(B) = P(B\bar{A}) + P(AB) \end{cases} \Rightarrow \begin{cases} P(A\bar{B}) = P(A) - P(AB) \\ P(B\bar{A}) = P(B) - P(AB) \end{cases} \quad (2)$$

Thay (2) vào (1) ta được:

$$\begin{aligned} P(A + B) &= P(A\bar{B}) + P(AB) + P(B\bar{A}) \\ &= P(A) - P(AB) + P(AB) + P(B) - P(AB) \\ \Leftrightarrow [P(A + B) &= P(A) + P(B) - P(AB)] \text{ (điều phải chứng minh)} \end{aligned}$$

## Câu 2:

Gọi các biến cố:

$A$  = “Cam loại 1”

$B$  = “Cam loại 2”

$C$  = “Cam loại 3”

$H$  = “Cam lấy ra là cam hỏng”

Theo đề bài ta có:  $P(A) = 0,2; P(B) = 0,3; P(C) = 0,5;$

$$P(H|A) = 0,02; P(H|B) = 0,03; P(H|C) = 0,01$$

a) *Tính xác suất để quả cam lấy ra là quả hỏng:*

Áp dụng công thức xác suất đầy đủ:

$$\begin{aligned} P(H) &= P(HA) + P(HB) + P(HC) \\ &= P(H | A).P(A) + P(H | B).P(B) + P(H | C).P(C) \\ &= 0,02 \cdot 0,2 + 0,03 \cdot 0,3 + 0,01 \cdot 0,5 \\ \Leftrightarrow [P(H) &= 0,018] \end{aligned}$$

b) Giả sử lấy được 1 quả hồng, tính xác suất để nó là cam loại 1:

Áp dụng công thức Bayes:

$$P(A | H) = \frac{P(HA)}{P(HA) + P(HB) + P(HC)} = \frac{P(H | A).P(A)}{P(H)} = \frac{0,02.0,2}{0,018}$$

$$\Leftrightarrow P(A | H) = 0,2222$$

**Câu 3:** X là đại lượng ngẫu nhiên tuân theo phân phối Poisson

\* Kỳ vọng toán của X:

$$EX = \sum_{k=0}^{\infty} x_k p_k = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} \text{ (vì số hạng đầu tiên bằng 0)}$$

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \text{ (cho k chạy từ 0 thì thay k - 1 bằng k)}$$

$$\text{Mà: } \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots + \frac{\lambda^k}{k!} = e^{\lambda} \text{ (công thức khai triển Taylor của } e^{\lambda})$$

Do đó:

$$EX = \lambda e^{-\lambda} e^{\lambda} \Leftrightarrow EX = \lambda$$

\* Phương sai của X:

$$DX = EX^2 - (EX)^2$$

Ta có:

$$EX^2 = \sum_{k=0}^{\infty} x_k^2 p_k = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} k^2 \frac{\lambda^k e^{-\lambda}}{k!} \text{ (vì số hạng đầu tiên bằng 0)}$$

$$= \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} = \lambda \sum_{k=0}^{\infty} (k+1) \frac{\lambda^k e^{-\lambda}}{k!} \text{ (cho k chạy từ 0 và thay k - 1 bằng k)}$$

$$= \lambda \left[ \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} + \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \right] = \lambda [E(X) + 1] = \lambda [\lambda + 1] = \lambda^2 + \lambda$$

Do đó phương sai cần tìm là:

$$DX = EX^2 - (EX)^2 = \lambda^2 + \lambda - \lambda^2 \Leftrightarrow DX = \lambda$$

**Câu 4:** Đại lượng ngẫu nhiên X có hàm mật độ:

$$f(x) = \begin{cases} \frac{k}{x^3} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

a) Tìm k và hàm phân phối xác suất của X:

\* Tìm k:

Vì f(x) là hàm mật độ của X nên suy ra:

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x)dx = 1 &\Leftrightarrow \int_1^{+\infty} \frac{k}{x^3} dx = 1 \Leftrightarrow k \int_1^{+\infty} \frac{1}{x^3} dx = 1 \\ &\Leftrightarrow k \left( -\frac{1}{2x^2} \right) \Big|_1^{+\infty} = 1 \quad (\text{theo công thức nguyên hàm: } \int \frac{1}{x^n} dx = \frac{-1}{(n-1)x^{n-1}}) \\ &\Leftrightarrow k \frac{1}{2} = 1 \Leftrightarrow \boxed{k = 2} \end{aligned}$$

\* Tìm hàm phân phối xác suất của X:

Hàm phân phối của X có dạng:  $F(x) = P(X < x)$

+ Nếu  $x \leq 1$  thì  $F(x) = 0$

+ Nếu  $x > 1$  thì:

$$F(x) = \int_1^x \frac{2}{t^3} dt = 2 \left( -\frac{1}{2t^2} \right) \Big|_1^x = 2 \left( -\frac{1}{2x^2} + \frac{1}{2} \right) = 1 - \frac{1}{x^2}$$

$\Rightarrow$  Hàm phân phối xác suất của X là:

$$F(x) = \begin{cases} 1 - \frac{1}{x^2} & x > 1 \\ 0 & x \leq 1 \end{cases}$$

b) Tìm  $P(1 < X < 3)$  và trung vị của X:

+ Xác suất  $P(1 < X < 3)$ :

$$P(1 < X < 3) = F(3) = 1 - \frac{1}{3^2} = \boxed{\frac{8}{9}}$$

+ Trung vị của X là giá trị m sao cho  $F(m) = 1/2$

$$F(m) = \frac{1}{2} \Leftrightarrow \int_1^m \frac{2}{x^3} dx = \frac{1}{2} \Leftrightarrow 2 \left( -\frac{1}{2x^2} \right) \Big|_1^m = \frac{1}{2} \Leftrightarrow 2 \left( -\frac{1}{2m^2} + \frac{1}{2} \right) = \frac{1}{2}$$

$$\Leftrightarrow -\frac{1}{m^2} + 1 = \frac{1}{2} \Leftrightarrow \frac{1}{m^2} = \frac{1}{2} \Leftrightarrow \boxed{m = \sqrt{2}}$$

**Câu 5:**

Hàm phân phối mẫu có dạng:  $G(x) = \frac{m}{n}$

(trong đó n là tổng số phần tử của mẫu, m là số phần tử của mẫu ứng với giá trị quan sát nhỏ hơn x. Ta có  $n = \sum m$ )

⇒ Hàm phân phối mẫu cần tìm:

$$G(x) = \begin{cases} 0 & x \leq 9,8 \\ 0,05 & x \leq 9,9 \\ 0,05 + 0,25 & x \leq 10 \\ 0,05 + 0,25 + 0,4 & x \leq 10,1 \\ 0,05 + 0,25 + 0,4 + 0,2 & x \leq 10,2 \\ 0,05 + 0,25 + 0,4 + 0,2 + 0,1 & x > 10,2 \end{cases}$$

$$G(x) = \begin{cases} 0 & x \leq 9,8 \\ 0,05 & x \leq 9,9 \\ 0,3 & x \leq 10 \\ 0,7 & x \leq 10,1 \\ 0,9 & x \leq 10,2 \\ 1 & x > 10,2 \end{cases}$$

## 11. Đề thi cuối kỳ phụ - hè năm 2016

ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

ĐỀ THI KẾT THÚC HỌC KỲ PHỤ  
HÈ NĂM 2016

### Môn thi: Xác suất thống kê

Mã môn học: MAT 1101 Số tín chỉ: 03

Dành cho sinh viên các khoa: Hóa học, Sinh học, Địa lý, Địa chất, Môi trường, KTTV-HDH, Y-Dược

Dạng đề thi: **Không được sử dụng tài liệu**

Thời gian làm bài: 90 phút (không kể thời gian phát đề)

**Câu 1:** Một sân bóng được chia làm 4 sân nhỏ để cho thuê. Biết số đội đến thuê sân vào buổi sáng thứ 7 là một đại lượng X có phân bố Poisson với tham số  $\lambda = 3$ .

- a. Tìm xác suất để không phải tất cả 4 sân đều được thuê.
- b. Trung bình bao nhiêu sân được thuê trong một buổi sáng thứ 7?

**Câu 2:** Diện tích của một chiếc lá của một loài cây là đại lượng ngẫu nhiên liên tục X (đơn vị  $\text{cm}^2$ ) có hàm mật độ:

$$f(x) = Cx^2(2-x) \text{ nếu } 0 \leq x \leq 2 \text{ và bằng } 0 \text{ nếu ngược lại}$$

- a. Tìm hằng số C và tính kỳ vọng của X?

b. Lấy ngẫu nhiên 5 chiếc lá. Tính xác suất để trong 5 chiếc lá đó có 3 chiếc lá có diện tích lớn hơn  $1\text{cm}^2$ .

**Câu 3:** Quan sát thời gian sản xuất của một sản phẩm bằng 2 phương pháp ta có số liệu sau:

Thời gian (X)	$\leq 10$	(10 - 12]	(12 - 14]	(14 - 16]	(16 - 18]	(18 - 20]	$> 20$
Phương pháp I (Số sản phẩm)	3	8	9	11	7	6	2
Phương pháp II (Số sản phẩm)	1	1	4	8	12	8	5

a. Hãy ước lượng thời gian sản xuất trung bình của một sản phẩm bằng phương pháp I (với độ tin cậy 95%). Nếu muốn độ chính xác của ước lượng bằng 0,4 thì phải cần thêm bao nhiêu giá trị mẫu nữa?

b. Với mức ý nghĩa  $\alpha = 0,025$ ; có thể kết luận thời gian sản xuất trung bình một sản phẩm bằng phương pháp I là nhanh hơn phương pháp II không? Ý nghĩa của kết luận là gì?

**Câu 4:** Để nghiên cứu mối quan hệ giữa số người tiêm chích ma túy (X) và số người nhiễm HIV (Y), người ta thống kê ở 8 địa phương thu được số liệu sau:

X (người)	100	150	200	250	300	300	350	360
Y (người)	60	90	120	150	160	180	200	210

a. Hãy ước lượng hệ số tương quan và tìm đường hồi quy tuyến tính thực nghiệm của Y theo X?

b. Dự báo số người bị nhiễm HIV ở một địa phương có 220 người tiêm chích ma túy?

Lời giải:

### Câu 1:

X là số đội đến thuê sân. Theo bài ra ta có:  $X \sim \text{Poisson}(3)$

a) Xác suất để không phải tất cả 4 sân đều được thuê:

Không phải tất cả 4 sân đều được thuê tương đương với 4 trường hợp: không có sân nào được thuê, hoặc 1 sân được thuê, hoặc 2 sân được thuê, hoặc 3 sân được thuê. Xác suất cần tìm là:

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

$$= e^{-3} \frac{3^0}{0!} + e^{-3} \frac{3^1}{1!} + e^{-3} \frac{3^2}{2!} + e^{-3} \frac{3^3}{3!} = e^{-3} \left( 1 + 3 + \frac{9}{2} + \frac{27}{6} \right) = \boxed{13e^{-3}}$$

b) Trung bình bao nhiêu sân được thuê trong một buổi sáng thứ 7:

Gọi Y là số sân được thuê trong một buổi sáng thứ 7. Các giá trị có thể có của Y là:  $Y(\Omega) = \{0, 1, 2, 3, 4\}$ . Tính các xác suất tương ứng:

$$+ P(Y = 0) = P(X = 0) = e^{-3} \frac{3^0}{0!} = e^{-3}$$

$$+ P(Y = 1) = P(X = 1) = e^{-3} \frac{3^1}{1!} = 3e^{-3}$$

$$+ P(Y = 2) = P(X = 2) = e^{-3} \frac{3^2}{2!} = \frac{9}{2} e^{-3}$$

$$+ P(Y = 3) = P(X = 3) = e^{-3} \frac{3^3}{3!} = \frac{27}{6} e^{-3}$$

$$+ P(Y = 4) = P(X \geq 4) = 1 - P(X \leq 3) = 1 - 13e^{-3}$$

$\Rightarrow$  Bảng phân bố xác suất của Y:

Y	0	1	2	3	4
P	$e^{-3}$	$3e^{-3}$	$\frac{9}{2} e^{-3}$	$\frac{27}{6} e^{-3}$	$1 - 13e^{-3}$

Do đó, số sân trung bình được thuê trong một buổi sáng thứ 7 là:

$$EY = \sum_{i=0}^4 y_i p_i = 0 \cdot e^{-3} + 1 \cdot 3e^{-3} + 2 \cdot \frac{9}{2} e^{-3} + 3 \cdot \frac{27}{6} e^{-3} + 4 \cdot (1 - 13e^{-3}) \Leftrightarrow \boxed{EY = 2,68}$$

**Câu 2:**

Diện tích của một chiếc lá là đại lượng ngẫu nhiên liên tục X có hàm mật độ:

$$f(x) = Cx^2(2-x) \text{ nếu } 0 \leq x \leq 2 \text{ và bằng } 0 \text{ nếu ngược lại}$$

a) *Tìm hằng số C và tính kỳ vọng của X:*

Vì  $f(x)$  là hàm mật độ của X nên suy ra:

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \Leftrightarrow \int_0^2 Cx^2(2-x) dx = 1 \Leftrightarrow C \left( 2 \frac{x^3}{3} - \frac{x^4}{4} \right) \Big|_0^2 = 1$$

$$\Leftrightarrow C \left( \frac{16}{3} - 4 \right) = 1 \Leftrightarrow C \frac{4}{3} = 1 \Leftrightarrow \boxed{C = \frac{3}{4}}$$

Kỳ vọng của X:

$$EX = \int_{-\infty}^{+\infty} xf(x) dx = \frac{3}{4} \int_0^2 x^3(2-x) dx = \frac{3}{4} \left( 2 \frac{x^4}{4} - \frac{x^5}{5} \right) \Big|_0^2 = \frac{3}{4} \left( 8 - \frac{32}{5} \right)$$

$$\Leftrightarrow EX = \boxed{\frac{6}{5}}$$

b) *Lấy ngẫu nhiên 5 chiếc lá. Tính xác suất để trong 5 chiếc lá đó có 3 chiếc lá có diện tích lớn hơn  $1\text{cm}^2$ :*

+ Nếu lấy 1 chiếc lá thì xác suất để chiếc lá đó có diện tích lớn hơn  $1\text{cm}^2$  là:

$$\begin{aligned} P(X > 1) &= \int_1^{+\infty} f(x) dx = \frac{3}{4} \int_1^2 x^2(2-x) dx = \frac{3}{4} \left( 2 \frac{x^3}{3} - \frac{x^4}{4} \right) \Big|_1^2 \\ &= \frac{3}{4} \left( 2 \frac{8}{3} - 4 - \frac{2}{3} + \frac{1}{4} \right) = \boxed{\frac{11}{16}} \end{aligned}$$

+ Gọi Y là số chiếc lá có diện tích lớn hơn  $1\text{cm}^2$  trong số 5 chiếc được lấy ngẫu nhiên. Ta có:  $Y \sim B(5; 11/16)$ . Xác suất cần tìm là:

$$P(Y = 3) = C_5^3 \left( \frac{11}{16} \right)^3 \left( \frac{5}{16} \right)^2 = \boxed{0,3173}$$

**Câu 3:**

Gọi X là thời gian sản xuất trung bình một sản phẩm ứng với phương pháp I và Y là thời gian sản xuất trung bình ứng với phương pháp II.

a) *Ước lượng thời gian sản xuất trung bình của một sản phẩm bằng phương pháp I (độ tin cậy 95%). Nếu muốn độ chính xác của ước lượng bằng 0,4 thì phải cần thêm bao nhiêu giá trị mẫu nữa:*

+ *Ước lượng thời gian sản xuất trung bình của một sản phẩm bằng phương pháp I:*

Hoàng Văn Trọng – 0974.971.149

(dạng bài ước lượng khoảng cho kỳ vọng; trường hợp chưa biết phương sai lý thuyết, chưa biết phân bố chuẩn nhưng cỡ mẫu lớn)

Sử dụng máy tính ta có:

$$+ Kỳ vọng mẫu: \bar{X} = 14,61$$

$$+ Phương sai mẫu: s_x = 3,1656$$

$\Rightarrow$  Ước lượng khoảng cho kỳ vọng, ứng với phương pháp I:

$$EX \in \left( \bar{X} - u\left(\frac{\alpha}{2}\right) \frac{s_x}{\sqrt{n_x - 1}}; \bar{X} + u\left(\frac{\alpha}{2}\right) \frac{s_x}{\sqrt{n_x - 1}} \right)$$

$$\Leftrightarrow EX \in \left( 14,61 - u(0,025) \frac{3,1656}{\sqrt{45}}; 14,61 + u(0,025) \frac{3,1656}{\sqrt{45}} \right)$$

$$\Leftrightarrow EX \in \left( 14,61 - 1,96 \cdot \frac{3,1656}{\sqrt{45}}; 14,61 + 1,96 \cdot \frac{3,1656}{\sqrt{45}} \right)$$

$$\Leftrightarrow EX \in (13,69; 15,53)$$

Độ chính xác của ước lượng bằng 0,4 suy ra:

$$u\left(\frac{\alpha}{2}\right) \frac{s_x}{\sqrt{n_x - 1}} = 0,4 \Leftrightarrow 1,96 \frac{3,1656}{\sqrt{n_x - 1}} = 0,4 \Leftrightarrow \sqrt{n_x - 1} = \frac{1,96 \cdot 3,1656}{0,4}$$

$$n_x = \left( \frac{1,96 \cdot 3,1656}{0,4} \right)^2 + 1 \approx 242$$

$\Rightarrow$  Nếu muốn độ chính xác là 0,4 thì phải cần thêm  $(242 - 46 = 196)$  mẫu nữa.

b) Có thể kết luận thời gian sản xuất trung bình một sản phẩm bằng phương pháp I là nhanh hơn phương pháp II không? ( $\alpha = 0,025$ ). Ý nghĩa của kết luận?

(dạng bài toán so sánh hai giá trị trung bình, trường hợp chưa biết hai phương sai lý thuyết, chưa biết tính chuẩn của hai đại lượng ngẫu nhiên nhưng cỡ mẫu lớn)

Bài toán kiểm định:

$$H: EX = EY \quad | \quad K: EX < EY, \text{ mức ý nghĩa } \alpha = 0,025;$$

Miền tiêu chuẩn (miền bác bỏ) của bài toán:

$$S = \left\{ \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{n_x - 1} + \frac{s_y^2}{n_y - 1}}} \leq -u(\alpha) \right\}$$

Sử dụng máy tính ta có:  $\bar{Y} = 16,74$

$$s_x^2 = 10,02; s_y^2 = 7,83$$

Tính:

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{n_x - 1} + \frac{s_y^2}{n_y - 1}}} = \frac{14,61 - 16,74}{\sqrt{\frac{10,02}{45} + \frac{7,83}{38}}} = -3,25$$

Tra bảng phân bố chuẩn ta được:  $-u(\alpha) = -u(0,025) = -1,96$

Do đó, miền tiêu chuẩn đã xảy ra. Vậy, với mức ý nghĩa  $\alpha = 0,025$  ta tạm thời cho rằng thời gian sản xuất trung bình một sản phẩm theo phương pháp I là nhanh hơn so với phương pháp II, cho đến khi có thêm thông tin mới.

\* Ý nghĩa của kết luận:

Kết luận trên có nghĩa là: Thời gian sản xuất trung bình một sản phẩm theo phương pháp I nhanh hơn theo phương pháp II, với xác suất 97,5%.

#### Câu 4:

a) *Uớc lượng hệ số tương quan và tìm đường hồi quy tuyến tính thực nghiệm của Y theo X:*

Hệ số tương quan mău giữa Y và X:

$$r = \frac{\frac{1}{n} \sum_{i=1}^k x_i y_i m_i - \bar{X} \bar{Y}}{s_x s_y} \quad (\text{trong đó } n \text{ là số cặp } (x, y))$$

Sử dụng máy tính ta có:

$$\frac{1}{n} \sum_{i=1}^k x_i y_i m_i = \frac{328600}{8} = 41075$$

$$\bar{X} = 251,25; \bar{Y} = 146,25; s_x = 88,0962; s_y = 49,4817$$

$\Rightarrow$  Hệ số tương quan mău:

$$r = \frac{41075 - 251,25 \cdot 146,25}{88,0962 \cdot 49,4817} = \boxed{0,9932} \quad (\text{mối tương quan rất chặt})$$

$\Rightarrow$  Đường hồi quy tuyến tính thực nghiệm của Y theo X:

$$y - \bar{Y} = r \frac{s_y}{s_x} (x - \bar{X}) \Leftrightarrow y - 146,25 = 0,9932 \cdot \frac{49,4817}{88,0962} (x - 251,25)$$

$$\Leftrightarrow \boxed{y = 0,5576x + 6,088}$$

b) Dự báo số người bị nhiễm HIV ở một địa phương có 220 người tiêm chích ma túy:

Đường hồi quy trung bình tuyến tính thực nghiệm có thể được dùng để dự báo Y theo X. Nếu một địa phương có 220 người tiêm chích ma túy thì dự báo số người bị nhiễm HIV là:

$$y = 0,5576x + 6,088$$

$$\Leftrightarrow y = 0,5576 \cdot 220 + 6,088 = 128,76$$

$$\Leftrightarrow \boxed{y \approx 129} \quad (\text{người})$$

## PHỤ LỤC

### P.1. Kiến thức chuẩn bị

#### a. Hoán vị

Cho tập hợp  $A$  gồm  $n$  phần tử ( $n \geq 1$ ). Mỗi kết quả của sự sắp xếp thứ tự  $n$  phần tử của tập hợp  $A$  được gọi là một hoán vị của  $n$  phần tử đó.

Gọi  $P_n$  là số hoán vị của  $n$  phần tử. Ta có:

$$P_n = n(n-1)(n-2)\dots2.1$$

#### b. Chính hợp

Cho tập hợp  $A$  gồm  $n$  phần tử ( $n \geq 1$ ). Kết quả của việc lấy  $k$  phần tử ( $k \leq n$ ) khác nhau từ  $n$  phần tử của tập hợp  $A$  và sắp xếp chúng theo một thứ tự nào đó được gọi là một chính hợp chập  $k$  của  $n$  phần tử đã cho.

Gọi  $A_n^k$  là số chính hợp chập  $k$  của  $n$  phần tử. Ta có:

$$A_n^k = n(n-1)(n-2)\dots(n-k+1)$$

+ Tính chất:

- $A_n^n = P_n$
- $A_n^k = \frac{n!}{(n-k)!}$  (với  $0! = 1$  và  $n! = P_n$ )

Trong một chính hợp, mỗi phần tử của tập  $A$  xuất hiện không quá 1 lần.

#### c. Chính hợp lặp

Cho tập hợp  $A$  gồm  $n$  phần tử ( $n \geq 1$ ). Chính hợp lặp của  $k$  phần tử ( $k$  có thể lớn hơn  $n$ ) là một nhóm có thứ tự gồm  $k$  phần tử trong đó mỗi phần tử có thể xuất hiện từ 1 đến  $k$  lần.

Gọi  $\tilde{A}_n^k$  là số chính hợp lặp chập  $k$  của  $n$  phần tử. Ta có:

$$\tilde{A}_n^k = n^k$$

#### d. Tổ hợp

Cho tập hợp  $A$  gồm  $n$  phần tử ( $n \geq 1$ ). Mỗi tập con gồm  $k$  phần tử của  $A$  được gọi là một tổ hợp chập  $k$  của  $n$  phần tử đã cho ( $k$  phần tử này không cần xếp theo thứ tự).

Gọi  $C_n^k$  là số tổ hợp chập  $k$  của  $n$  phần tử. Ta có:  $C_n^k = \frac{n!}{k!(n-k)!}$

+ Tính chất:

- $C_n^k = C_n^{n-k}$
- $C_n^k = C_{n-1}^k + C_{n-1}^{k-1}$

### e. Tích phân và tích phân suy rộng

+)*Định nghĩa tích phân hàm một biến:*

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{b-a}{n} f(x_i) \quad (\text{nếu tồn tại giới hạn})$$

Công thức Newton – Leibnitz:

$$\int_a^b f(x) dx = F(x)|_a^b = F(b) - F(a)$$

(với  $F(x)$  là một nguyên hàm của hàm số  $f(x)$  trên đoạn  $[a, b]$ )

+)*Nguyên hàm của một số hàm số sơ cấp cơ bản (với  $C$  là số thực bất kỳ):*

$$\int dx = x + C$$

$$\int x^\alpha dx = \frac{x^{\alpha+1}}{\alpha+1} + C$$

$$\int e^x dx = e^x + C$$

$$\int \frac{dx}{x} = \ln|x| + C$$

$$\int \sin x dx = -\cos x + C$$

$$\int \cos x dx = \sin x + C$$

$$\int \frac{dx}{\sin^2 x} = -\cot x + C$$

$$\int \frac{dx}{\cos^2 x} = \tan x + C$$

$$\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C$$

$$\int \frac{dx}{1+x^2} = \arctan x + C$$

+)*Tích phân suy rộng của hàm một biến:*

$$\int_{-\infty}^{+\infty} f(x) dx = F(x)|_{-\infty}^{+\infty} = \lim_{x \rightarrow +\infty} F(x) - \lim_{x \rightarrow -\infty} F(x)$$

(với  $F(x)$  là một nguyên hàm của  $f(x)$  trên toàn tập xác định)

## P.2. Tính toán chỉ số thống kê bằng máy tính bỏ túi

Phần này giới thiệu cách tính các chỉ số thống kê của mẫu số liệu về **một** đại lượng ngẫu nhiên (trung bình mẫu, phương sai và phương sai hiệu chỉnh mẫu, độ lệch tiêu chuẩn và độ lệch tiêu chuẩn hiệu chỉnh mẫu) và tương quan giữa **hai** đại lượng ngẫu nhiên (hệ số tương quan mẫu, giá trị a và b trong phương trình đường hồi quy tuyến tính thực nghiệm:  $y = ax + b$ ) bằng các loại máy tính bỏ túi: Casio fx-500MS (máy fx-570MS làm tương tự), Casio fx-570ES PLUS (máy fx-570ES và fx-570VN PLUS làm tương tự).

### P.2.1. Tính toán các chỉ số thống kê mẫu của một đại lượng ngẫu nhiên

Cho một mẫu số liệu thống kê dưới dạng bảng tần suất như sau:

$x_i$	$x_1$	$x_2$	$x_3$	....	$x_n$
$m_i$	$m_1$	$m_2$	$m_3$	.....	$m_n$

a) Đối với máy Casio fx-500MS:

(máy fx-570MS thì làm tương tự, chỉ khác ở bước 2 khi chuyển về chế độ thống kê mẫu một biến thì nhấn **Mode** **Mode** **1**)

+ Bước 1: Reset toàn bộ máy, nhấn liên tiếp 4 phím:

**Shift** **Mode** **3** **=**

+ Bước 2: Chuyển về chế độ thống kê và nhập số liệu đầu vào:

- Chuyển về chế độ thống kê mẫu một biến: **Mode** **2**
- Nhập số liệu vào (các giá trị  $x_1, m_1, \dots$  là các số cụ thể trong bảng):

**x<sub>1</sub>** **Shift** **,** **m<sub>1</sub>** **M+**  
**x<sub>2</sub>** **Shift** **,** **m<sub>2</sub>** **M+**  
.....  
**x<sub>n</sub>** **Shift** **,** **m<sub>n</sub>** **M+**

Sau khi nhập xong, kiểm tra dữ liệu nhập vào (nếu muốn có thể bỏ qua) bằng cách nhấn liên tiếp phím mũi tên xuống **▼**

Nhấn **AC** để thoát chế độ nhập dữ liệu (nếu muốn có thể bỏ qua).

+ Bước 3: Hiển thị kết quả

Nhấn **Shift** **2** để vào giao diện hiển thị kết quả. Có 3 tùy chọn:

- Nhấn **1** **=** để hiển thị trung bình mẫu, hoặc:  
Nhấn **2** **=** để hiển thị độ lệch tiêu chuẩn mẫu, hoặc:  
Nhấn **3** **=** để hiển thị độ lệch tiêu chuẩn hiệu chỉnh của mẫu

(khi muốn xem thêm chỉ số thống kê khác thì nhấn tiếp **Shift** **2**)

b) Đối với máy Casio fx-570ES PLUS:

(máy fx-570ES hoặc fx-570VN PLUS thì làm tương tự)

+ Bước 1: Reset toàn bộ máy, nhấn liên tiếp 5 phím:

**Shift** **9** **3** **=** **AC**

+ Bước 2: Chuyển về chế độ thống kê và nhập số liệu đầu vào

- Chuyển về chế độ thống kê mẫu một biến:

**Shift** **Mode** **▼** **4** **1**  
**Mode** **3** **1**

- Nhập số liệu vào (các giá trị  $x_1, x_2, \dots$  là các số cụ thể trong bảng):

Chuột đang ở vị trí dòng 1 cột X, nhấn **x<sub>1</sub>** **=**

Chuột chuyển xuống dòng 2 cột X, nhấn **x<sub>2</sub>** **=**

Khi kết thúc cột X, nhấn **▼** **▷** để nhập các giá trị  $m_i$  ở cột FREQ

- Nhấn **AC** để thoát chế độ nhập dữ liệu.

+ Bước 3: Hiển thị kết quả

Nhấn **Shift** **1** **4** để vào giao diện hiển thị kết quả. Có 4 tùy chọn:

Nhấn **1** **=** để hiển thị kích thước mẫu (cỡ mẫu), hoặc:

Nhấn **2** **=** để hiển thị trung bình mẫu, hoặc:

Nhấn **3** **=** để hiển thị độ lệch tiêu chuẩn mẫu, hoặc:

Nhấn **4** **=** để hiển thị độ lệch tiêu chuẩn hiệu chỉnh của mẫu.

(khi muốn xem thêm chỉ số thống kê khác thì nhấn tiếp **Shift** **1** **4**)

### P.2.2. Tính toán tương quan giữa hai đại lượng ngẫu nhiên

Cho một mẫu số liệu thống kê dưới dạng bảng tần suất như sau:

X <sub>i</sub>	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	....	X <sub>n</sub>
y <sub>i</sub>	y <sub>1</sub>	y <sub>2</sub>	y <sub>3</sub>	....	y <sub>n</sub>
m <sub>i</sub>	m <sub>1</sub>	m <sub>2</sub>	m <sub>3</sub>	.....	m <sub>n</sub>

a) Đối với máy Casio fx-500MS:

(máy fx-570MS thì làm tương tự, chỉ khác ở bước 2 khi chuyển về chế độ thống kê mẫu hai biến thì nhấn **Mode** **Mode** **2** **1**)

+ Bước 1: Reset toàn bộ máy, nhấn liên tiếp 4 phím:

**Shift** **Mode** **3** **=**

+ Bước 2: Chuyển về chế độ thống kê và nhập số liệu đầu vào:

- Chuyển về chế độ thống kê mẫu hai biến: Mode 3 1
- Nhập số liệu vào máy:

$x_1$  ,  $y_1$  Shift ,  $m_1$  M+

$x_2$  ,  $y_2$  Shift ,  $m_2$  M+

.....

$x_n$  ,  $y_n$  Shift ,  $m_n$  M+

Sau khi nhập xong, kiểm tra dữ liệu nhập vào (nếu muốn có thể bỏ qua) bằng cách nhấn liên tiếp phím mũi tên xuống  $\nabla$

Nhấn AC để thoát chế độ nhập dữ liệu (nếu muốn có thể bỏ qua).

+ Bước 3: Hiển thị kết quả

- Nhấn Shift 2 màn hình hiển thị ba tùy chọn của đại lượng ngẫu nhiên X.
- Nhấn phím sang phải  $\rightarrow$  màn hình sẽ hiển thị ba tùy chọn của đại lượng ngẫu nhiên Y.
- Nhấn tiếp phím sang phải  $\rightarrow$  để hiển thị ba tùy chọn: hệ số tương quan mẫu r, hai giá trị A và B trong phương trình đường hồi quy tuyến tính thực nghiệm (lưu ý: các máy tính Casio đã mặc định phương trình đường hồi quy có dạng:  $y = A + Bx$ )

(khi muốn xem thêm chỉ số thống kê khác thì nhấn tiếp Shift 2)

b) Đối với máy Casio fx-570ES PLUS:

(máy fx-570ES hoặc fx-570VN PLUS thì làm tương tự)

+ Bước 1: Reset toàn bộ máy, nhấn liên tiếp 5 phím:

Shift 9 3 = AC

+ Bước 2: Chuyển về chế độ thống kê và nhập số liệu đầu vào

- Chuyển về chế độ thống kê mẫu hai biến:

Shift Mode  $\nabla$  4 1

Mode 3 2

- Nhập số liệu vào máy (các giá trị  $x_1, x_2, \dots$  là các số cụ thể trong bảng):

Chuột đang ở vị trí dòng 1 cột X, nhấn  $x_1$  =

Chuột chuyển xuống dòng 2 cột X, nhấn  $x_2$  =

Khi kết thúc cột X, nhấn  $\nabla$   $\rightarrow$  để nhập các giá trị  $y_i$  ở cột Y

Khi kết thúc cột Y, nhấn  $\nabla$   $\rightarrow$  để nhập các giá trị  $m_i$  ở cột FREQ

- Nhấn **AC** để thoát chế độ nhập dữ liệu.
- + Bước 3: Hiển thị kết quả
- Nhấn **Shift** **1** **4** để hiển thị cỡ mẫu, trung bình, phương sai, phương sai hiệu chỉnh của hai đại lượng ngẫu nhiên, hoặc:
- Nhấn **Shift** **1** **5** để hiển thị hệ số tương quan mẫu r, hai giá trị A và B trong phương trình đường hồi quy tuyến tính thực nghiệm, lưu ý các máy tính Casio đã mặc định phương trình có dạng:  $y = A + Bx$
- (khi muốn xem thêm chỉ số thống kê khác thì nhấn tiếp **Shift** **1** **4** hoặc **Shift** **1** **5**)

### P.3. Tính toán xác suất thống kê bằng hàm trong Excel

Sử dụng hàm trong Excel giúp tính toán nhanh hơn, chính xác hơn. Đôi khi gặp phải trường hợp muốn tra 1 bảng phân bố xác suất nào đó nhưng lại không có giáo trình bên cạnh thì việc sử dụng Excel là một cách giải quyết hiệu quả.

#### P.3.1. Hàm SUMPRODUCT

= SUMPRODUCT (array1, [array2], [array3], ...)

*Trả về tổng của tích các nhóm phần tử tương ứng trong các mảng dữ liệu.*

+ Điều kiện: các đối số của hàm là các mảng có cùng kích thước (cùng số lượng phần tử)

Ví dụ: Tính kỳ vọng của đại lượng ngẫu nhiên X có bảng phân bố xác suất:

A3	f(x)	=SUMPRODUCT(B1:E1,B2:E2)
	A	B
1	X	1
2	P	0.2
3		2.4

#### P.3.2. Hàm BINOMDIST

= BINOMDIST (number\_s, trials, probability\_s, cumulative)

*Trả về giá trị của phân bố nhị thức (theo công thức Bernoulli).*

Nếu X là đại lượng ngẫu nhiên rời rạc có phân bố nhị thức  $X \sim B(n, p)$  thì:

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

Các đối số trong hàm:

+ number\_s: là giá trị k trong công thức trên

+ trials: là giá trị n

+ probability\_s: giá trị xác suất p

+ cumulative: giá trị tích lũy. Nếu cho cumulative = 0 thì ứng với  $P(X = k)$ , nếu cho cumulative = 1 thì ứng với  $P(X \leq k)$ .

Ví dụ 1: Lập bảng phân bố xác suất của ĐLNN  $X \sim B(4; 0,2)$

B2	f(x)	=BINOMDIST(B1,4,0.2,0)
	A	B
1	X	0
2	P	0.4096

Ví dụ 2: Cho  $X \sim B(6; 0,25)$ . Tìm  $P(X \leq 3)$

B1	f(x)	=BINOMDIST(3,6,0.25,1)
	A	B
1		0.9624

### P.3.3. Hàm POISSON

= POISSON (x, mean, cumulative)

Trả về giá trị của phân bố Poisson.

Nếu X là đại lượng ngẫu nhiên rời rạc có phân bố Poisson với tham số λ thì:

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Các đối số trong hàm:

+ x: là giá trị k trong công thức trên

+ mean: là giá trị λ (giá trị trung bình của phân bố Poisson)

+ cumulative: giá trị tích lũy. Nếu cho cumulative = 0 thì ứng với  $P(X = k)$ , nếu cho cumulative = 1 thì ứng với  $P(X \leq k)$ .

Ví dụ 1: Lập bảng phân bố xác suất của ĐLNN  $X \sim \text{Poisson}(4)$

B2	f(x)	=POISSON(B1,4,0)	A	B	C	D	E	F	G	H
1	X	0	1	2	3	4	5	...		
2	P	0.0183	0.0733	0.1465	0.1954	0.1954	0.1563	...		

Ví dụ 2: Cho  $X \sim \text{Poisson}(6)$ . Tìm  $P(X \leq 4)$

B1	f(x)	=POISSON(4,6,1)	A	B
1		0.2851		

### P.3.4. Hàm NORMDIST

= NORMDIST (x, mean, standard\_dev, cumulative)

Trả về giá trị hàm phân bố  $F(x)$  của phân bố chuẩn.

Nếu  $X \sim N(\mu, \sigma^2)$  thì hàm phân bố  $F(x)$  được tính bằng công thức sau:

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

Các đối số trong hàm:

+ x: là giá trị x trong công thức trên

+ mean: là giá trị μ (giá trị trung bình của phân bố chuẩn)

+ standard\_dev: là giá trị σ (độ lệch tiêu chuẩn)

+ cumulative: giá trị tích lũy. Cho cumulative = 1 để trả về F(x)

Ví dụ 1: Cho  $X \sim N(12; 0,5^2)$ . Tìm  $P(X < 13)$

B1	f <sub>x</sub>	=NORMDIST(13,12,0.5,1)
A	B	0.9772

Ví dụ 2: Lập hàm phân bố chuẩn hóa  $Z \sim N(0, 1)$  với các giá trị x từ 0 → 0,29

B2	f <sub>x</sub>	=NORMDIST(\$A\$2+B1,0,1,1)
A	B	C
1	0.00	0.01
2	0.0	0.5000
3	0.1	0.5398
4	0.2	0.5793
		0.02
		0.03
		0.04
		0.05
		0.06
		0.07
		0.08
		0.09
		0.5040
		0.5080
		0.5120
		0.5160
		0.5199
		0.5239
		0.5279
		0.5319
		0.5359
		0.5438
		0.5478
		0.5517
		0.5557
		0.5596
		0.5636
		0.5675
		0.5714
		0.5753
		0.5832
		0.5871
		0.5910
		0.5948
		0.5987
		0.6026
		0.6064
		0.6103
		0.6141

### P.3.5. Hàm NORMINV

= NORMINV (probability, mean, standard\_dev)

Trả về giá trị nghịch đảo (giá trị x) của hàm phân bố chuẩn, biết  $F(x) = p$ .

Các đối số trong hàm:

+ probability: giá trị xác suất p sao cho hàm phân bố  $F(x) = p$

+ mean: trung bình của phân bố chuẩn

+ standard\_dev: độ lệch tiêu chuẩn.

Ví dụ 1: Cho  $X \sim N(3; 0,4^2)$ . Tìm giá trị a sao cho  $P(X < a) = 0,2$

B1	f <sub>x</sub>	=NORMINV(0.2,3,0.4)
A	B	2.66

Ví dụ 2: Tìm  $u(0,025)$  của phân bố chuẩn hóa  $N(0, 1)$

B1	f <sub>x</sub>	=NORMINV(1-0.025,0,1)
A	B	1.96

### P.3.6. Hàm TINV

= TINV (probability, deg\_freedom)

Trả về giá trị nghịch đảo  $t_k(\alpha)$  của phân bố Student.

Các đối số trong hàm:

+ probability: mức ý nghĩa  $\alpha$

+ deg\_freedom: số bậc tự do k

Lưu ý: giá trị trả về của hàm TINV chỉ đúng cho trường hợp mức ý nghĩa được chia đều sang hai bên của phân bố (trường hợp hai phía).

Ví dụ 1: Tìm giá trị  $t_{30}(0,05)$  trong trường hợp hai phía:

B1	$f_x$	=TINV(0.05,30)
A	B	
1		2.042

Tuy nhiên, giá trị  $t_k(\alpha)$  trong các công thức được học trên lớp đều thuộc trường hợp mức ý nghĩa nằm hết về một phía (bên phải của phân bố). Do đó khi sử dụng hàm TINV thì phải cho đối số “probability” tăng lên hai lần.

Ví dụ 2: Khi cần tìm  $t_{30}(0,05)$  thì viết hàm như sau:

B1	$f_x$	=TINV(0.05*2,30)
A	B	
1		1.697

(mặc dù tìm  $t_{30}(0,05)$  nhưng đối số thứ nhất phải tăng lên hai lần)

### P.3.7. Hàm CHIINV

= CHIINV (probability, deg\_freedom)

Trả về giá trị nghịch đảo  $\chi^2_k(\alpha)$  của phân bố Khi bình phương.

Các đối số trong hàm:

+ probability: mức ý nghĩa  $\alpha$

+ deg\_freedom: số bậc tự do k

Ví dụ: Tìm giá trị  $\chi^2_4(0,05)$

B1	$f_x$	=CHIINV(0.05,4)
A	B	
1		9.488

### P.3.8. Hàm FINV

= FINV (probability, deg\_freedom1, deg\_freedom2)

Trả về giá trị nghịch đảo  $F_{m,n}(\alpha)$  của phân bố Fisher.

Các đối số trong hàm:

+ probability: mức ý nghĩa  $\alpha$

+ deg\_freedom1: số bậc tự do m ứng với dãy ĐLNN X.

+ deg\_freedom2: số bậc tự do n ứng với dãy ĐLNN Y.

Ví dụ: Tìm giá trị  $F_{4,5}(0,05)$

B1	$f_x$	=FINV(0.05,4,5)
A	B	
1		5.192

## P.4. Bảng tra cứu một số phân bố thường gấp

### P.4.1. Bảng phân bố nhị thức

Giả sử X là đại lượng ngẫu nhiên rời rạc có phân bố nhị thức với tham số n và p:

$$X \sim B(n, p)$$

Ta có:

$$P(X = k) = C_n^k p^k q^{n-k} \quad (\text{với } q = 1 - p)$$

Dưới đây là bảng phân bố xác suất của ĐLNN có phân bố nhị thức ứng với một số giá trị n và p thường gấp ( $n = 2, 5$  và  $p = 0.1, 0.9$ )

Cách tra bảng: giả sử đề bài yêu cầu lập bảng phân bố xác suất của  $X \sim B(4; 0,5)$  thì chọn bảng ứng với  $n = 4$  sau đó lấy các giá trị của dòng  $p = 0,5$ . Kết quả ta được bảng phân bố xác suất của X như sau:

X	0	1	2	3	4
P	0.0625	0.2500	0.3750	0.2500	0.0625

a) Với  $n = 2$ :

$\begin{matrix} k \\ p \end{matrix}$	0	1	2
0.1	0.81	0.18	0.01
0.2	0.64	0.32	0.04
0.3	0.49	0.42	0.09
0.4	0.36	0.48	0.16
0.5	0.25	0.5	0.25
0.6	0.16	0.48	0.36
0.7	0.09	0.42	0.49
0.8	0.04	0.32	0.64
0.9	0.01	0.18	0.81

b) Với  $n = 3$ :

$\begin{matrix} k \\ p \end{matrix}$	0	1	2	3
0.1	0.729	0.243	0.027	0.001
0.2	0.512	0.384	0.096	0.008
0.3	0.343	0.441	0.189	0.027
0.4	0.216	0.432	0.288	0.064
0.5	0.125	0.375	0.375	0.125
0.6	0.064	0.288	0.432	0.216
0.7	0.027	0.189	0.441	0.343
0.8	0.008	0.096	0.384	0.512
0.9	0.001	0.027	0.243	0.729

c) Với  $n = 4$ :

$\begin{array}{c} k \\ \diagdown \\ p \end{array}$	0	1	2	3	4
0.1	0.6561	0.2916	0.0486	0.0036	0.0001
0.2	0.4096	0.4096	0.1536	0.0256	0.0016
0.3	0.2401	0.4116	0.2646	0.0756	0.0081
0.4	0.1296	0.3456	0.3456	0.1536	0.0256
0.5	0.0625	0.2500	0.3750	0.2500	0.0625
0.6	0.0256	0.1536	0.3456	0.3456	0.1296
0.7	0.0081	0.0756	0.2646	0.4116	0.2401
0.8	0.0016	0.0256	0.1536	0.4096	0.4096
0.9	0.0001	0.0036	0.0486	0.2916	0.6561

d) Với  $n = 5$ :

$\begin{array}{c} k \\ \diagdown \\ p \end{array}$	0	1	2	3	4	5
0.1	0.5905	0.3281	0.0729	0.0081	0.0005	0.0000
0.2	0.3277	0.4096	0.2048	0.0512	0.0064	0.0003
0.3	0.1681	0.3602	0.3087	0.1323	0.0284	0.0024
0.4	0.0778	0.2592	0.3456	0.2304	0.0768	0.0102
0.5	0.0313	0.1563	0.3125	0.3125	0.1563	0.0313
0.6	0.0102	0.0768	0.2304	0.3456	0.2592	0.0778
0.7	0.0024	0.0284	0.1323	0.3087	0.3602	0.1681
0.8	0.0003	0.0064	0.0512	0.2048	0.4096	0.3277
0.9	0.0000	0.0005	0.0081	0.0729	0.3281	0.5905

e) Với  $n = 6$ :

$\begin{array}{c} k \\ \diagdown \\ p \end{array}$	0	1	2	3	4	5	6
0.1	0.5314	0.3543	0.0984	0.0146	0.0012	0.0001	0.0000
0.2	0.2621	0.3932	0.2458	0.0819	0.0154	0.0015	0.0001
0.3	0.1176	0.3025	0.3241	0.1852	0.0595	0.0102	0.0007
0.4	0.0467	0.1866	0.3110	0.2765	0.1382	0.0369	0.0041
0.5	0.0156	0.0938	0.2344	0.3125	0.2344	0.0938	0.0156
0.6	0.0041	0.0369	0.1382	0.2765	0.3110	0.1866	0.0467
0.7	0.0007	0.0102	0.0595	0.1852	0.3241	0.3025	0.1176
0.8	0.0001	0.0015	0.0154	0.0819	0.2458	0.3932	0.2621
0.9	0.0000	0.0001	0.0012	0.0146	0.0984	0.3543	0.5314

#### P.4.2. Bảng phân bố Poisson

Giả sử X là đại lượng ngẫu nhiên rời rạc có phân bố Poisson với tham số  $\lambda$  thì:

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (\text{với } 0 \leq k < +\infty)$$

Bảng phân bố Poisson dưới đây tính cho các trường hợp tham số  $\lambda$  từ 1 đến 8.

Cách tra bảng: giả sử đề bài cho  $X \sim \text{Poisson}(4)$ , yêu cầu lập bảng phân bố xác suất của X và tính  $P(X < 3)$ . Chọn các giá trị trên dòng ứng với  $\lambda = 4$  ta được:

X	0	1	2	3	4	5	....
P	0.0183	0.0733	0.1465	0.1954	0.1954	0.1563	....

$$\begin{aligned} \text{Từ đó suy ra: } P(X < 3) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= 0,0183 + 0,0733 + 0,1465 = 0,2381 \end{aligned}$$

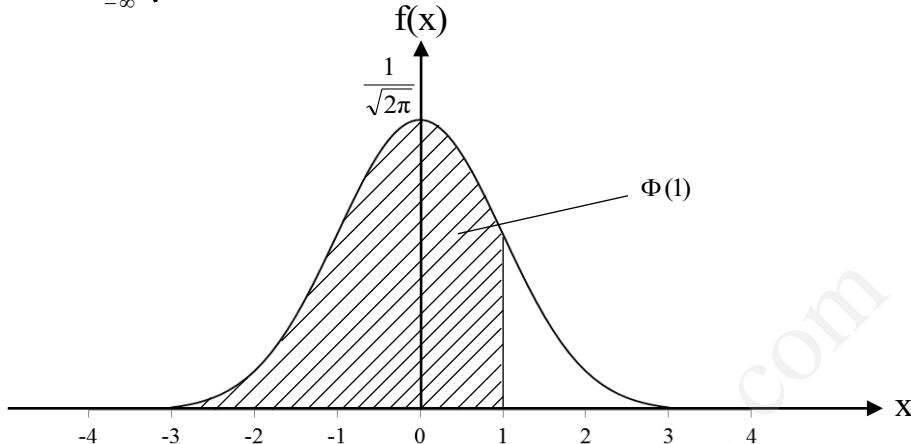
$\lambda \backslash k$	0	1	2	3	4	5	6	7	8	9	10
1	0.3679	0.3679	0.1839	0.0613	0.0153	0.0031	0.0005	0.0001	0.0000	0.0000	.....
2	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.0120	0.0034	0.0009	0.0002	.....
3	0.0498	0.1494	0.2240	0.2240	0.1680	0.1008	0.0504	0.0216	0.0081	0.0027	.....
4	0.0183	0.0733	0.1465	0.1954	0.1954	0.1563	0.1042	0.0595	0.0298	0.0132	.....
5	0.0067	0.0337	0.0842	0.1404	0.1755	0.1755	0.1462	0.1044	0.0653	0.0363	.....
6	0.0025	0.0149	0.0446	0.0892	0.1339	0.1606	0.1606	0.1377	0.1033	0.0688	.....
7	0.0009	0.0064	0.0223	0.0521	0.0912	0.1277	0.1490	0.1490	0.1304	0.1014	.....
8	0.0003	0.0027	0.0107	0.0286	0.0573	0.0916	0.1221	0.1396	0.1396	0.1241	.....

### P.4.3. Hàm phân bố chuẩn tắc

Giả sử  $Z$  là đại lượng ngẫu nhiên liên tục có phân bố chuẩn tắc:  $Z \sim N(0, 1)$

⇒ Hàm phân bố xác suất của  $Z$ :

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$



Giá trị hàm phân bố chuẩn tắc dưới đây tính cho các trường hợp  $x$  từ 0 đến 3,99.

Cách tra bảng:

- Tra xuôi: ví dụ muốn tra  $\Phi(0,82)$  thì đóng theo hàng 0,8 và cột 0,02 ta được giá trị cần tìm là:  $\Phi(0,82) = 0,7939$ 
  - + Khi muốn tìm giá trị  $\Phi(-x)$  thì tra thông qua  $\Phi(x)$  với:  $\Phi(-x) = 1 - \Phi(x)$
  - + Xem bảng tra xuôi ở trang tiếp theo...
- Tra ngược: ví dụ muốn tra  $u(0,025)$ . Thực chất  $u(0,025)$  là một giá trị  $x$  nào đó sao cho  $P(Z \geq x) = 0,025$  hay  $\Phi(x) = 0,975$ . Tra bảng theo 2 bước sau:
  - + Lấy 1 trừ cho 0,025 bằng 0,975.
  - + Dò trong bảng phân bố chuẩn tắc, xem có giá trị nào bằng hoặc gần với 0,975 nhất. Dóng ngược lại theo hàng và cột để được giá trị cần tìm. Kết quả là:  $u(0,025) = 1,96$

Một số giá trị  $u(\alpha)$  trong bảng tra ngược:

Mức ý nghĩa $\alpha$	0.005	0.01	0.015	0.02	0.025
$u(\alpha)$	2.58	2.33	2.17	2.05	1.96

Mức ý nghĩa $\alpha$	0.03	0.04	0.05	0.1	0.2
$u(\alpha)$	1.88	1.75	1.64	1.28	0.84

Giá trị hàm phân bố  $\Phi(x)$  của phân bố chuẩn tắc ( $x = hàng + cột$ )

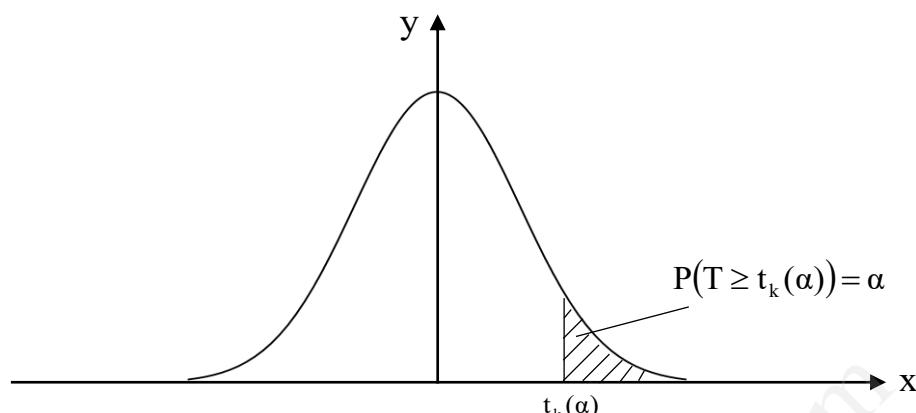
(x)	<b>0</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>
<b>0</b>	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
<b>0.1</b>	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
<b>0.2</b>	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
<b>0.3</b>	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
<b>0.4</b>	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
<b>0.5</b>	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
<b>0.6</b>	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
<b>0.7</b>	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
<b>0.8</b>	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
<b>0.9</b>	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
<b>1</b>	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
<b>1.1</b>	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
<b>1.2</b>	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
<b>1.3</b>	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
<b>1.4</b>	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
<b>1.5</b>	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
<b>1.6</b>	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
<b>1.7</b>	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
<b>1.8</b>	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
<b>1.9</b>	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
<b>2</b>	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
<b>2.1</b>	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
<b>2.2</b>	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
<b>2.3</b>	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
<b>2.4</b>	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
<b>2.5</b>	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
<b>2.6</b>	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
<b>2.7</b>	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
<b>2.8</b>	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
<b>2.9</b>	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
<b>3</b>	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
<b>3.1</b>	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
<b>3.2</b>	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
<b>3.3</b>	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
<b>3.4</b>	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
<b>3.5</b>	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
<b>3.6</b>	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
<b>3.7</b>	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
<b>3.8</b>	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
<b>3.9</b>	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Giá trị  $\Phi(3,84)$

#### P.4.4. Hàm ngược của hàm phân bố Student

Bảng dưới đây dùng để tra giá trị  $t_k(\alpha)$  sao cho:  $P(T \geq t_k(\alpha)) = \alpha$

(với T có phân bố Student)



Chú ý: những công thức trong phần ước lượng tham số và kiểm định giả thiết liên quan đến việc tra cứu  $t_k(\alpha)$  đều áp dụng cho trường hợp một phía. Vì vậy, các giá trị trong bảng dưới đây ứng với trường hợp một phía.

Cách tra bảng: giá trị  $t_k(\alpha)$  là giao của hàng k và cột  $\alpha$ . Nếu bậc tự do k quá lớn và không có trong bảng thì lấy theo hàng có k gần nhất. Ví dụ:  $t_{99}(\alpha) \approx t_{100}(\alpha)$

Giá trị  $t_k(\alpha)$  của phân bố Student:

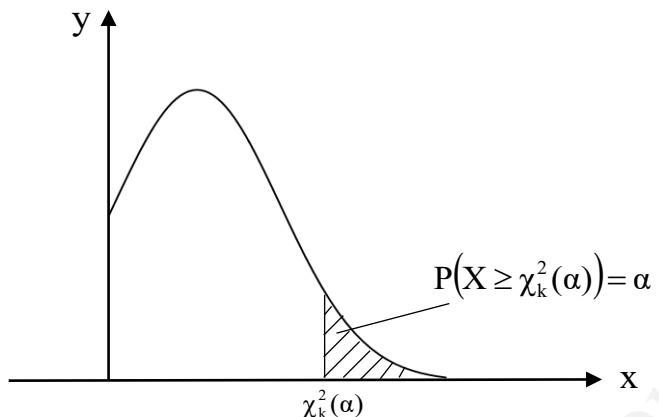
$\frac{\alpha}{k}$	0.005	0.01	0.015	0.02	0.025	0.03	0.04	0.05	0.1	0.2
1	63.657	31.821	21.205	15.895	12.706	10.579	7.916	6.314	3.078	1.376
2	9.925	6.965	5.643	4.849	4.303	3.896	3.320	2.920	1.886	1.061
3	5.841	4.541	3.896	3.482	3.182	2.951	2.605	2.353	1.638	0.978
4	4.604	3.747	3.298	2.999	2.776	2.601	2.333	2.132	1.533	0.941
5	4.032	3.365	3.003	2.757	2.571	2.422	2.191	2.015	1.476	0.920
6	3.707	3.143	2.829	2.612	2.447	2.313	2.104	1.943	1.440	0.906
7	3.499	2.998	2.715	2.517	2.365	2.241	2.046	1.895	1.415	0.896
8	3.355	2.896	2.634	2.449	2.306	2.189	2.004	1.860	1.397	0.889
9	3.250	2.821	2.574	2.398	2.262	2.150	1.973	1.833	1.383	0.883
10	3.169	2.764	2.527	2.359	2.228	2.120	1.948	1.812	1.372	0.879
11	3.106	2.718	2.491	2.328	2.201	2.096	1.928	1.796	1.363	0.876
12	3.055	2.681	2.461	2.303	2.179	2.076	1.912	1.782	1.356	0.873
13	3.012	2.650	2.436	2.282	2.160	2.060	1.899	1.771	1.350	0.870
14	2.977	2.624	2.415	2.264	2.145	2.046	1.887	1.761	1.345	0.868
15	2.947	2.602	2.397	2.249	2.131	2.034	1.878	1.753	1.341	0.866
16	2.921	2.583	2.382	2.235	2.120	2.024	1.869	1.746	1.337	0.865
17	2.898	2.567	2.368	2.224	2.110	2.015	1.862	1.740	1.333	0.863
18	2.878	2.552	2.356	2.214	2.101	2.007	1.855	1.734	1.330	0.862
19	2.861	2.539	2.346	2.205	2.093	2.000	1.850	1.729	1.328	0.861
20	2.845	2.528	2.336	2.197	2.086	1.994	1.844	1.725	1.325	0.860

(xem tiếp trang bên)

$k \backslash \alpha$	<b>0.005</b>	<b>0.01</b>	<b>0.015</b>	<b>0.02</b>	<b>0.025</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.1</b>	<b>0.2</b>
<b>21</b>	2.831	2.518	2.328	2.189	2.080	1.988	1.840	1.721	1.323	0.859
<b>22</b>	2.819	2.508	2.320	2.183	2.074	1.983	1.835	1.717	1.321	0.858
<b>23</b>	2.807	2.500	2.313	2.177	2.069	1.978	1.832	1.714	1.319	0.858
<b>24</b>	2.797	2.492	2.307	2.172	2.064	1.974	1.828	1.711	1.318	0.857
<b>25</b>	2.787	2.485	2.301	2.167	2.060	1.970	1.825	1.708	1.316	0.856
<b>26</b>	2.779	2.479	2.296	2.162	2.056	1.967	1.822	1.706	1.315	0.856
<b>27</b>	2.771	2.473	2.291	2.158	2.052	1.963	1.819	1.703	1.314	0.855
<b>28</b>	2.763	2.467	2.286	2.154	2.048	1.960	1.817	1.701	1.313	0.855
<b>29</b>	2.756	2.462	2.282	2.150	2.045	1.957	1.814	1.699	1.311	0.854
<b>30</b>	2.750	2.457	2.278	2.147	2.042	1.955	1.812	1.697	1.310	0.854
<b>31</b>	2.744	2.453	2.275	2.144	2.040	1.952	1.810	1.696	1.309	0.853
<b>32</b>	2.738	2.449	2.271	2.141	2.037	1.950	1.808	1.694	1.309	0.853
<b>33</b>	2.733	2.445	2.268	2.138	2.035	1.948	1.806	1.692	1.308	0.853
<b>34</b>	2.728	2.441	2.265	2.136	2.032	1.946	1.805	1.691	1.307	0.852
<b>35</b>	2.724	2.438	2.262	2.133	2.030	1.944	1.803	1.690	1.306	0.852
<b>36</b>	2.719	2.434	2.260	2.131	2.028	1.942	1.802	1.688	1.306	0.852
<b>37</b>	2.715	2.431	2.257	2.129	2.026	1.940	1.800	1.687	1.305	0.851
<b>38</b>	2.712	2.429	2.255	2.127	2.024	1.939	1.799	1.686	1.304	0.851
<b>39</b>	2.708	2.426	2.252	2.125	2.023	1.937	1.798	1.685	1.304	0.851
<b>40</b>	2.704	2.423	2.250	2.123	2.021	1.936	1.796	1.684	1.303	0.851
<b>41</b>	2.701	2.421	2.248	2.121	2.020	1.934	1.795	1.683	1.303	0.850
<b>42</b>	2.698	2.418	2.246	2.120	2.018	1.933	1.794	1.682	1.302	0.850
<b>43</b>	2.695	2.416	2.244	2.118	2.017	1.932	1.793	1.681	1.302	0.850
<b>44</b>	2.692	2.414	2.243	2.116	2.015	1.931	1.792	1.680	1.301	0.850
<b>45</b>	2.690	2.412	2.241	2.115	2.014	1.929	1.791	1.679	1.301	0.850
<b>46</b>	2.687	2.410	2.239	2.114	2.013	1.928	1.790	1.679	1.300	0.850
<b>47</b>	2.685	2.408	2.238	2.112	2.012	1.927	1.789	1.678	1.300	0.849
<b>48</b>	2.682	2.407	2.237	2.111	2.011	1.926	1.789	1.677	1.299	0.849
<b>49</b>	2.680	2.405	2.235	2.110	2.010	1.925	1.788	1.677	1.299	0.849
<b>50</b>	2.678	2.403	2.234	2.109	2.009	1.924	1.787	1.676	1.299	0.849
<b>55</b>	2.668	2.396	2.228	2.104	2.004	1.920	1.784	1.673	1.297	0.848
<b>60</b>	2.660	2.390	2.223	2.099	2.000	1.917	1.781	1.671	1.296	0.848
<b>65</b>	2.654	2.385	2.219	2.096	1.997	1.914	1.778	1.669	1.295	0.847
<b>70</b>	2.648	2.381	2.215	2.093	1.994	1.912	1.776	1.667	1.294	0.847
<b>75</b>	2.643	2.377	2.212	2.090	1.992	1.910	1.775	1.665	1.293	0.846
<b>80</b>	2.639	2.374	2.209	2.088	1.990	1.908	1.773	1.664	1.292	0.846
<b>85</b>	2.635	2.371	2.207	2.086	1.988	1.906	1.772	1.663	1.292	0.846
<b>90</b>	2.632	2.368	2.205	2.084	1.987	1.905	1.771	1.662	1.291	0.846
<b>95</b>	2.629	2.366	2.203	2.082	1.985	1.904	1.770	1.661	1.291	0.845
<b>100</b>	2.626	2.364	2.201	2.081	1.984	1.902	1.769	1.660	1.290	0.845
<b>105</b>	2.623	2.362	2.200	2.080	1.983	1.901	1.768	1.659	1.290	0.845
<b>110</b>	2.621	2.361	2.199	2.078	1.982	1.900	1.767	1.659	1.289	0.845
<b>115</b>	2.619	2.359	2.197	2.077	1.981	1.900	1.766	1.658	1.289	0.845
<b>120</b>	2.617	2.358	2.196	2.076	1.980	1.899	1.766	1.658	1.289	0.845
<b>+∞</b>	2.576	2.326	2.170	2.054	1.960	1.881	1.751	1.645	1.282	0.842

#### P.4.5. Hàm ngược của hàm phân bố Khi bình phương

Bảng dưới đây dùng để tra giá trị  $\chi_k^2(\alpha)$  sao cho:  $P(X \geq \chi_k^2(\alpha)) = \alpha$   
 (với X có phân bố Khi bình phương)



Cách tra bảng: giá trị  $\chi_k^2(\alpha)$  là giao của hàng k và cột  $\alpha$ .

- + Nếu hàng k không có trong bảng thì lấy theo hàng gần nhất với hàng k.
- + Xem bảng tra ở trang bên...

Giá trị  $\chi_k^2(\alpha)$  của phân bố  $\chi^2$  (với  $\alpha$  từ 0,005 đến 0,2)

$\frac{\alpha}{k}$	<b>0.005</b>	<b>0.01</b>	<b>0.015</b>	<b>0.02</b>	<b>0.025</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.1</b>	<b>0.2</b>
<b>1</b>	7.88	6.63	5.92	5.41	5.02	4.71	4.22	3.84	2.71	1.64
<b>2</b>	10.60	9.21	8.40	7.82	7.38	7.01	6.44	5.99	4.61	3.22
<b>3</b>	12.84	11.34	10.47	9.84	9.35	8.95	8.31	7.81	6.25	4.64
<b>4</b>	14.86	13.28	12.34	11.67	11.14	10.71	10.03	9.49	7.78	5.99
<b>5</b>	16.75	15.09	14.10	13.39	12.83	12.37	11.64	11.07	9.24	7.29
<b>6</b>	18.55	16.81	15.78	15.03	14.45	13.97	13.20	12.59	10.64	8.56
<b>7</b>	20.28	18.48	17.40	16.62	16.01	15.51	14.70	14.07	12.02	9.80
<b>8</b>	21.95	20.09	18.97	18.17	17.53	17.01	16.17	15.51	13.36	11.03
<b>9</b>	23.59	21.67	20.51	19.68	19.02	18.48	17.61	16.92	14.68	12.24
<b>10</b>	25.19	23.21	22.02	21.16	20.48	19.92	19.02	18.31	15.99	13.44
<b>11</b>	26.76	24.72	23.50	22.62	21.92	21.34	20.41	19.68	17.28	14.63
<b>12</b>	28.30	26.22	24.96	24.05	23.34	22.74	21.79	21.03	18.55	15.81
<b>13</b>	29.82	27.69	26.40	25.47	24.74	24.12	23.14	22.36	19.81	16.98
<b>14</b>	31.32	29.14	27.83	26.87	26.12	25.49	24.49	23.68	21.06	18.15
<b>15</b>	32.80	30.58	29.23	28.26	27.49	26.85	25.82	25.00	22.31	19.31
<b>16</b>	34.27	32.00	30.63	29.63	28.85	28.19	27.14	26.30	23.54	20.47
<b>17</b>	35.72	33.41	32.01	31.00	30.19	29.52	28.44	27.59	24.77	21.61
<b>18</b>	37.16	34.81	33.38	32.35	31.53	30.84	29.75	28.87	25.99	22.76
<b>19</b>	38.58	36.19	34.74	33.69	32.85	32.16	31.04	30.14	27.20	23.90
<b>20</b>	40.00	37.57	36.09	35.02	34.17	33.46	32.32	31.41	28.41	25.04
<b>21</b>	41.40	38.93	37.43	36.34	35.48	34.76	33.60	32.67	29.62	26.17
<b>22</b>	42.80	40.29	38.77	37.66	36.78	36.05	34.87	33.92	30.81	27.30
<b>23</b>	44.18	41.64	40.09	38.97	38.08	37.33	36.13	35.17	32.01	28.43
<b>24</b>	45.56	42.98	41.41	40.27	39.36	38.61	37.39	36.42	33.20	29.55
<b>25</b>	46.93	44.31	42.73	41.57	40.65	39.88	38.64	37.65	34.38	30.68
<b>26</b>	48.29	45.64	44.03	42.86	41.92	41.15	39.89	38.89	35.56	31.79
<b>27</b>	49.64	46.96	45.33	44.14	43.19	42.41	41.13	40.11	36.74	32.91
<b>28</b>	50.99	48.28	46.63	45.42	44.46	43.66	42.37	41.34	37.92	34.03
<b>29</b>	52.34	49.59	47.91	46.69	45.72	44.91	43.60	42.56	39.09	35.14
<b>30</b>	53.67	50.89	49.20	47.96	46.98	46.16	44.83	43.77	40.26	36.25
<b>31</b>	55.00	52.19	50.48	49.23	48.23	47.40	46.06	44.99	41.42	37.36
<b>32</b>	56.33	53.49	51.75	50.49	49.48	48.64	47.28	46.19	42.58	38.47
<b>33</b>	57.65	54.78	53.02	51.74	50.73	49.88	48.50	47.40	43.75	39.57
<b>34</b>	58.96	56.06	54.29	53.00	51.97	51.11	49.72	48.60	44.90	40.68
<b>35</b>	60.27	57.34	55.55	54.24	53.20	52.34	50.93	49.80	46.06	41.78
<b>36</b>	61.58	58.62	56.81	55.49	54.44	53.56	52.14	51.00	47.21	42.88
<b>37</b>	62.88	59.89	58.07	56.73	55.67	54.78	53.34	52.19	48.36	43.98
<b>38</b>	64.18	61.16	59.32	57.97	56.90	56.00	54.55	53.38	49.51	45.08
<b>39</b>	65.48	62.43	60.57	59.20	58.12	57.22	55.75	54.57	50.66	46.17
<b>40</b>	66.77	63.69	61.81	60.44	59.34	58.43	56.95	55.76	51.81	47.27
<b>45</b>	73.17	69.96	67.99	66.56	65.41	64.45	62.90	61.66	57.51	52.73
<b>50</b>	79.49	76.15	74.11	72.61	71.42	70.42	68.80	67.50	63.17	58.16
<b>55</b>	85.75	82.29	80.17	78.62	77.38	76.34	74.66	73.31	68.80	63.58
<b>60</b>	91.95	88.38	86.19	84.58	83.30	82.23	80.48	79.08	74.40	68.97
<b>65</b>	98.11	94.42	92.16	90.50	89.18	88.07	86.27	84.82	79.97	74.35
<b>70</b>	104.21	100.43	98.10	96.39	95.02	93.88	92.02	90.53	85.53	79.71
<b>75</b>	110.29	106.39	104.00	102.24	100.84	99.66	97.75	96.22	91.06	85.07
<b>80</b>	116.32	112.33	109.87	108.07	106.63	105.42	103.46	101.88	96.58	90.41
<b>85</b>	122.32	118.24	115.72	113.87	112.39	111.16	109.14	107.52	102.08	95.73
<b>90</b>	128.30	124.12	121.54	119.65	118.14	116.87	114.81	113.15	107.57	101.05
<b>95</b>	134.25	129.97	127.34	125.40	123.86	122.56	120.45	118.75	113.04	106.36
<b>100</b>	140.17	135.81	133.12	131.14	129.56	128.24	126.08	124.34	118.50	111.67

Hoàng Văn Trọng – 0974.971.149

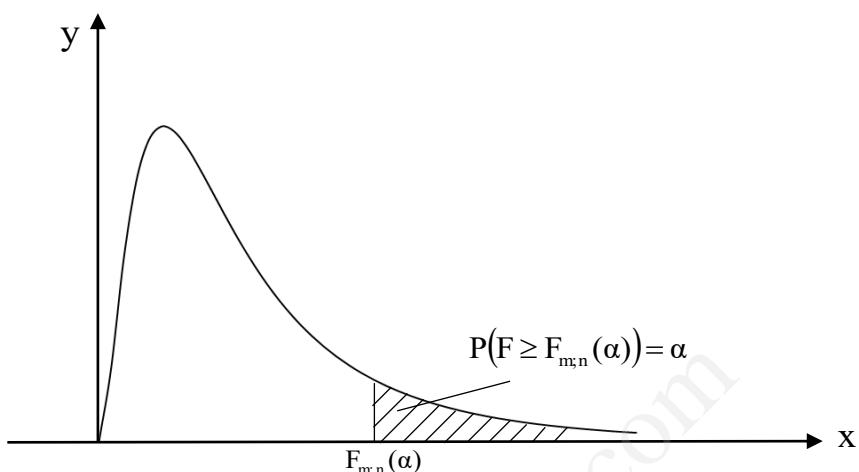
Giá trị  $\chi_k^2(\alpha)$  của phân bố  $\chi^2$  (với  $\alpha$  từ 0,8 đến 0,995)

$\frac{\alpha}{k}$	<b>0.8</b>	<b>0.9</b>	<b>0.95</b>	<b>0.96</b>	<b>0.97</b>	<b>0.975</b>	<b>0.98</b>	<b>0.985</b>	<b>0.99</b>	<b>0.995</b>
<b>1</b>	0.0642	0.0158	0.0039	0.0025	0.0014	0.00098	0.00063	0.000353	0.000157	0.000039
<b>2</b>	0.446	0.211	0.103	0.0816	0.0609	0.0506	0.0404	0.0302	0.02010	0.01003
<b>3</b>	1.01	0.584	0.352	0.300	0.245	0.216	0.185	0.152	0.115	0.0717
<b>4</b>	1.65	1.06	0.711	0.627	0.535	0.484	0.429	0.368	0.297	0.207
<b>5</b>	2.34	1.61	1.15	1.03	0.903	0.831	0.752	0.662	0.554	0.412
<b>6</b>	3.07	2.20	1.64	1.49	1.33	1.24	1.13	1.02	0.872	0.676
<b>7</b>	3.82	2.83	2.17	2.00	1.80	1.69	1.56	1.42	1.24	0.989
<b>8</b>	4.59	3.49	2.73	2.54	2.31	2.18	2.03	1.86	1.65	1.34
<b>9</b>	5.38	4.17	3.33	3.10	2.85	2.70	2.53	2.33	2.09	1.73
<b>10</b>	6.18	4.87	3.94	3.70	3.41	3.25	3.06	2.84	2.56	2.16
<b>11</b>	6.99	5.58	4.57	4.31	4.00	3.82	3.61	3.36	3.05	2.60
<b>12</b>	7.81	6.30	5.23	4.94	4.60	4.40	4.18	3.91	3.57	3.07
<b>13</b>	8.63	7.04	5.89	5.58	5.22	5.01	4.77	4.48	4.11	3.57
<b>14</b>	9.47	7.79	6.57	6.24	5.86	5.63	5.37	5.06	4.66	4.07
<b>15</b>	10.31	8.55	7.26	6.91	6.50	6.26	5.98	5.65	5.23	4.60
<b>16</b>	11.15	9.31	7.96	7.60	7.16	6.91	6.61	6.26	5.81	5.14
<b>17</b>	12.00	10.09	8.67	8.29	7.83	7.56	7.26	6.88	6.41	5.70
<b>18</b>	12.86	10.86	9.39	8.99	8.51	8.23	7.91	7.52	7.01	6.26
<b>19</b>	13.72	11.65	10.12	9.70	9.20	8.91	8.57	8.16	7.63	6.84
<b>20</b>	14.58	12.44	10.85	10.42	9.90	9.59	9.24	8.81	8.26	7.43
<b>21</b>	15.44	13.24	11.59	11.14	10.60	10.28	9.91	9.47	8.90	8.03
<b>22</b>	16.31	14.04	12.34	11.87	11.31	10.98	10.60	10.14	9.54	8.64
<b>23</b>	17.19	14.85	13.09	12.61	12.03	11.69	11.29	10.81	10.20	9.26
<b>24</b>	18.06	15.66	13.85	13.35	12.75	12.40	11.99	11.50	10.86	9.89
<b>25</b>	18.94	16.47	14.61	14.10	13.48	13.12	12.70	12.19	11.52	10.52
<b>26</b>	19.82	17.29	15.38	14.85	14.22	13.84	13.41	12.88	12.20	11.16
<b>27</b>	20.70	18.11	16.15	15.61	14.96	14.57	14.13	13.58	12.88	11.81
<b>28</b>	21.59	18.94	16.93	16.37	15.70	15.31	14.85	14.29	13.56	12.46
<b>29</b>	22.48	19.77	17.71	17.14	16.45	16.05	15.57	15.00	14.26	13.12
<b>30</b>	23.36	20.60	18.49	17.91	17.21	16.79	16.31	15.72	14.95	13.79
<b>31</b>	24.26	21.43	19.28	18.68	17.97	17.54	17.04	16.44	15.66	14.46
<b>32</b>	25.15	22.27	20.07	19.46	18.73	18.29	17.78	17.17	16.36	15.13
<b>33</b>	26.04	23.11	20.87	20.24	19.49	19.05	18.53	17.90	17.07	15.82
<b>34</b>	26.94	23.95	21.66	21.03	20.26	19.81	19.28	18.63	17.79	16.50
<b>35</b>	27.84	24.80	22.47	21.82	21.03	20.57	20.03	19.37	18.51	17.19
<b>36</b>	28.73	25.64	23.27	22.61	21.81	21.34	20.78	20.11	19.23	17.89
<b>37</b>	29.64	26.49	24.07	23.40	22.59	22.11	21.54	20.86	19.96	18.59
<b>38</b>	30.54	27.34	24.88	24.20	23.37	22.88	22.30	21.61	20.69	19.29
<b>39</b>	31.44	28.20	25.70	25.00	24.16	23.65	23.07	22.36	21.43	20.00
<b>40</b>	32.34	29.05	26.51	25.80	24.94	24.43	23.84	23.11	22.16	20.71
<b>45</b>	36.88	33.35	30.61	29.84	28.92	28.37	27.72	26.93	25.90	24.31
<b>50</b>	41.45	37.69	34.76	33.94	32.95	32.36	31.66	30.82	29.71	27.99
<b>55</b>	46.04	42.06	38.96	38.08	37.03	36.40	35.66	34.76	33.57	31.73
<b>60</b>	50.64	46.46	43.19	42.27	41.15	40.48	39.70	38.74	37.48	35.53
<b>65</b>	55.26	50.88	47.45	46.48	45.31	44.60	43.78	42.77	41.44	39.38
<b>70</b>	59.90	55.33	51.74	50.72	49.50	48.76	47.89	46.84	45.44	43.28
<b>75</b>	64.55	59.79	56.05	55.00	53.71	52.94	52.04	50.93	49.48	47.21
<b>80</b>	69.21	64.28	60.39	59.29	57.96	57.15	56.21	55.06	53.54	51.17
<b>85</b>	73.88	68.78	64.75	63.61	62.22	61.39	60.41	59.22	57.63	55.17
<b>90</b>	78.56	73.29	69.13	67.94	66.51	65.65	64.63	63.39	61.75	59.20
<b>95</b>	83.25	77.82	73.52	72.30	70.82	69.92	68.88	67.60	65.90	63.25
<b>100</b>	87.95	82.36	77.93	76.67	75.14	74.22	73.14	71.82	70.06	67.33

Hoàng Văn Trọng – 0974.971.149

#### P.4.6. Hàm ngược của hàm phân bố Fisher

Bảng dưới đây dùng để tra giá trị  $F_{m; n}(\alpha)$  sao cho:  $P(F \geq F_{m; n}(\alpha)) = \alpha$   
 (với  $F$  có phân bố Fisher)



Cách tra bảng: Mỗi bảng ứng với một giá trị  $\alpha$  nhất định (chỉ lập cho 2 giá trị thường dùng của  $\alpha$  là 0,01 và 0,05); giá trị  $F_{m; n}(\alpha)$  là giao của **cột m** và **hàng n**.

- + Xem bảng tra ở trang bên...
  - + Nếu cột m và hàng n không có trong bảng thì lấy theo cột và hàng gần nhất.
- Trường hợp muốn tra chính xác thì gõ công thức sau vào Excel:

$$= FINV(\alpha, m, n)$$

(với  $\alpha, m, n$  là các giá trị cụ thể)

Giá trị  $F_{m,n}(\alpha)$  của phân bố Fisher (với  $\alpha = 0,01$ )

$n \backslash m$	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	150	200	250	300	500	+∞	
<b>30</b>	2.39	2.34	2.30	2.27	2.25	2.22	2.21	2.19	2.18	2.17	2.16	2.15	2.14	2.14	2.13	2.09	2.07	2.06	2.05	2.03	2.01	
<b>33</b>	2.32	2.27	2.23	2.20	2.18	2.16	2.14	2.12	2.11	2.10	2.09	2.08	2.07	2.07	2.06	2.02	2.00	1.99	1.98	1.96	1.93	
<b>36</b>	2.26	2.21	2.18	2.14	2.12	2.10	2.08	2.07	2.05	2.04	2.03	2.02	2.02	2.01	2.00	1.96	1.94	1.93	1.92	1.90	1.87	
<b>39</b>	2.22	2.17	2.13	2.10	2.07	2.05	2.03	2.02	2.01	1.99	1.98	1.97	1.97	1.96	1.95	1.91	1.89	1.88	1.87	1.85	1.82	
<b>42</b>	2.18	2.13	2.09	2.06	2.03	2.01	1.99	1.98	1.96	1.95	1.94	1.93	1.93	1.92	1.91	1.87	1.85	1.83	1.82	1.80	1.78	
<b>45</b>	2.14	2.09	2.05	2.02	2.00	1.98	1.96	1.94	1.93	1.92	1.91	1.90	1.89	1.88	1.88	1.83	1.81	1.79	1.77	1.77	1.74	
<b>48</b>	2.12	2.06	2.02	1.99	1.97	1.95	1.93	1.91	1.90	1.89	1.88	1.87	1.86	1.85	1.84	1.80	1.78	1.76	1.75	1.73	1.70	
<b>51</b>	2.09	2.04	2.00	1.97	1.94	1.92	1.90	1.88	1.87	1.86	1.85	1.84	1.83	1.82	1.82	1.77	1.75	1.73	1.72	1.70	1.67	
<b>54</b>	2.07	2.02	1.98	1.94	1.92	1.90	1.90	1.88	1.86	1.85	1.83	1.82	1.81	1.81	1.80	1.79	1.75	1.72	1.71	1.70	1.68	1.65
<b>57</b>	2.05	1.99	1.95	1.92	1.90	1.87	1.86	1.84	1.83	1.81	1.80	1.79	1.78	1.78	1.77	1.72	1.70	1.68	1.67	1.65	1.62	
<b>60</b>	2.03	1.98	1.94	1.90	1.88	1.86	1.84	1.82	1.81	1.79	1.78	1.77	1.76	1.76	1.75	1.70	1.68	1.66	1.65	1.63	1.60	
<b>63</b>	2.01	1.96	1.92	1.89	1.86	1.84	1.82	1.80	1.79	1.78	1.77	1.76	1.75	1.74	1.73	1.68	1.66	1.64	1.63	1.61	1.58	
<b>66</b>	2.00	1.94	1.90	1.87	1.84	1.82	1.80	1.79	1.77	1.76	1.75	1.74	1.73	1.72	1.72	1.67	1.64	1.63	1.62	1.60	1.56	
<b>69</b>	1.98	1.93	1.89	1.86	1.83	1.81	1.79	1.77	1.76	1.75	1.73	1.72	1.72	1.71	1.70	1.65	1.63	1.61	1.60	1.58	1.55	
<b>72</b>	1.97	1.92	1.88	1.84	1.82	1.80	1.78	1.76	1.74	1.73	1.72	1.71	1.70	1.69	1.69	1.64	1.61	1.60	1.59	1.56	1.53	
<b>75</b>	1.96	1.91	1.87	1.83	1.81	1.78	1.76	1.75	1.73	1.72	1.71	1.70	1.69	1.68	1.67	1.62	1.60	1.58	1.57	1.55	1.52	
<b>78</b>	1.95	1.90	1.86	1.82	1.80	1.77	1.75	1.74	1.72	1.71	1.70	1.69	1.68	1.67	1.66	1.61	1.59	1.57	1.56	1.54	1.50	
<b>81</b>	1.94	1.89	1.85	1.81	1.79	1.76	1.74	1.73	1.71	1.70	1.69	1.68	1.67	1.66	1.65	1.60	1.58	1.56	1.55	1.53	1.49	
<b>84</b>	1.93	1.88	1.84	1.80	1.78	1.75	1.73	1.72	1.70	1.69	1.68	1.67	1.66	1.65	1.64	1.59	1.56	1.55	1.54	1.51	1.48	
<b>87</b>	1.92	1.87	1.83	1.79	1.77	1.74	1.72	1.71	1.69	1.68	1.67	1.66	1.65	1.64	1.63	1.58	1.55	1.54	1.53	1.50	1.47	
<b>90</b>	1.92	1.86	1.82	1.79	1.76	1.74	1.72	1.70	1.68	1.67	1.66	1.65	1.64	1.63	1.62	1.57	1.55	1.53	1.52	1.49	1.46	
<b>95</b>	1.90	1.85	1.81	1.77	1.75	1.72	1.70	1.69	1.67	1.66	1.65	1.64	1.63	1.62	1.61	1.56	1.53	1.51	1.50	1.48	1.44	
<b>100</b>	1.89	1.84	1.80	1.76	1.74	1.71	1.69	1.67	1.66	1.65	1.63	1.62	1.61	1.61	1.60	1.55	1.52	1.50	1.49	1.47	1.43	
<b>150</b>	1.83	1.77	1.73	1.69	1.66	1.64	1.62	1.60	1.59	1.57	1.56	1.55	1.54	1.53	1.52	1.46	1.43	1.42	1.40	1.38	1.33	
<b>200</b>	1.79	1.74	1.69	1.66	1.63	1.60	1.58	1.56	1.55	1.53	1.52	1.51	1.50	1.49	1.48	1.42	1.39	1.37	1.36	1.33	1.28	
<b>250</b>	1.77	1.72	1.67	1.64	1.61	1.58	1.56	1.54	1.53	1.51	1.50	1.49	1.48	1.47	1.46	1.40	1.36	1.34	1.33	1.30	1.24	
<b>300</b>	1.76	1.70	1.66	1.62	1.59	1.57	1.55	1.53	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.41	1.38	1.35	1.32	1.31	1.28	
<b>500</b>	1.74	1.68	1.63	1.60	1.57	1.54	1.52	1.50	1.48	1.47	1.45	1.44	1.43	1.42	1.41	1.34	1.31	1.28	1.27	1.23	1.16	
<b>+∞</b>	1.70	1.64	1.59	1.55	1.52	1.50	1.47	1.45	1.43	1.42	1.40	1.39	1.38	1.37	1.36	1.29	1.25	1.22	1.20	1.15	1.00	

Hoàng Văn Trọng – 0974.971.149

Giá trị  $F_{m;n}(\alpha)$  của phân bố Fisher (với  $\alpha = 0,05$ )

$n \backslash m$	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	150	200	250	300	500	+∞
<b>30</b>	1.84	1.81	1.79	1.77	1.76	1.75	1.74	1.73	1.72	1.72	1.71	1.71	1.70	1.70	1.70	1.67	1.66	1.65	1.65	1.64	1.62
<b>33</b>	1.81	1.78	1.76	1.74	1.72	1.71	1.70	1.69	1.69	1.68	1.67	1.67	1.66	1.66	1.66	1.63	1.62	1.61	1.61	1.60	1.58
<b>36</b>	1.78	1.75	1.73	1.71	1.69	1.68	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63	1.62	1.60	1.59	1.58	1.57	1.56	1.55
<b>39</b>	1.75	1.72	1.70	1.68	1.67	1.66	1.65	1.64	1.63	1.62	1.62	1.61	1.61	1.60	1.60	1.57	1.56	1.55	1.55	1.53	1.52
<b>42</b>	1.73	1.70	1.68	1.66	1.65	1.63	1.62	1.61	1.61	1.60	1.59	1.59	1.58	1.58	1.57	1.55	1.53	1.53	1.52	1.51	1.49
<b>45</b>	1.71	1.68	1.66	1.64	1.63	1.61	1.60	1.59	1.59	1.58	1.57	1.57	1.56	1.56	1.55	1.53	1.51	1.51	1.50	1.49	1.47
<b>48</b>	1.70	1.67	1.64	1.62	1.61	1.60	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.54	1.54	1.51	1.49	1.49	1.48	1.47	1.45
<b>51</b>	1.68	1.65	1.63	1.61	1.59	1.58	1.57	1.56	1.55	1.55	1.54	1.53	1.53	1.52	1.52	1.49	1.48	1.47	1.46	1.45	1.43
<b>54</b>	1.67	1.64	1.62	1.60	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.51	1.51	1.51	1.48	1.46	1.45	1.45	1.44	1.42
<b>57</b>	1.66	1.63	1.60	1.59	1.57	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.50	1.49	1.46	1.45	1.44	1.43	1.42	1.40
<b>60</b>	1.65	1.62	1.59	1.57	1.56	1.55	1.53	1.52	1.52	1.51	1.50	1.50	1.49	1.49	1.48	1.45	1.44	1.43	1.42	1.41	1.39
<b>63</b>	1.64	1.61	1.58	1.57	1.55	1.54	1.52	1.51	1.51	1.50	1.49	1.49	1.48	1.48	1.47	1.44	1.43	1.42	1.41	1.40	1.38
<b>66</b>	1.63	1.60	1.58	1.56	1.54	1.53	1.52	1.51	1.50	1.49	1.48	1.48	1.47	1.47	1.46	1.43	1.42	1.41	1.40	1.39	1.37
<b>69</b>	1.62	1.59	1.57	1.55	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.47	1.46	1.46	1.45	1.42	1.41	1.40	1.39	1.38	1.36
<b>72</b>	1.62	1.59	1.56	1.54	1.53	1.51	1.50	1.49	1.48	1.47	1.47	1.46	1.45	1.45	1.44	1.41	1.40	1.39	1.38	1.37	1.35
<b>75</b>	1.61	1.58	1.55	1.53	1.52	1.50	1.49	1.48	1.47	1.47	1.46	1.45	1.45	1.44	1.44	1.41	1.39	1.38	1.37	1.36	1.34
<b>78</b>	1.61	1.57	1.55	1.53	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.43	1.40	1.38	1.37	1.37	1.35	1.33
<b>81</b>	1.60	1.57	1.54	1.52	1.51	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.43	1.42	1.39	1.38	1.37	1.36	1.34	1.32
<b>84</b>	1.59	1.56	1.54	1.52	1.50	1.49	1.47	1.46	1.46	1.45	1.44	1.43	1.43	1.42	1.42	1.39	1.37	1.36	1.35	1.34	1.32
<b>87</b>	1.59	1.56	1.53	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.43	1.43	1.42	1.42	1.41	1.38	1.36	1.35	1.35	1.33	1.31
<b>90</b>	1.59	1.55	1.53	1.51	1.49	1.48	1.46	1.45	1.44	1.44	1.43	1.42	1.42	1.41	1.41	1.38	1.36	1.35	1.34	1.33	1.30
<b>95</b>	1.58	1.55	1.52	1.50	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40	1.40	1.37	1.35	1.34	1.33	1.32	1.29
<b>100</b>	1.57	1.54	1.52	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.41	1.41	1.40	1.40	1.39	1.36	1.34	1.33	1.32	1.31	1.28
<b>150</b>	1.54	1.50	1.48	1.45	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.36	1.35	1.34	1.31	1.29	1.28	1.27	1.25	1.22
<b>200</b>	1.52	1.48	1.46	1.43	1.41	1.40	1.39	1.37	1.36	1.35	1.35	1.34	1.33	1.33	1.32	1.28	1.26	1.25	1.24	1.22	1.19
<b>250</b>	1.50	1.47	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.32	1.31	1.31	1.27	1.25	1.23	1.22	1.20	1.17
<b>300</b>	1.50	1.46	1.43	1.41	1.39	1.38	1.36	1.35	1.34	1.33	1.32	1.31	1.31	1.30	1.30	1.26	1.23	1.22	1.21	1.19	1.15
<b>500</b>	1.48	1.45	1.42	1.40	1.38	1.36	1.35	1.33	1.32	1.31	1.30	1.30	1.29	1.28	1.28	1.23	1.21	1.19	1.18	1.16	1.11
<b>+∞</b>	1.46	1.42	1.39	1.37	1.35	1.33	1.32	1.30	1.29	1.28	1.27	1.26	1.26	1.25	1.24	1.20	1.17	1.15	1.14	1.11	1.00

Hoàng Văn Trọng – 0974.971.149

## TÀI LIỆU THAM KHẢO<sup>1</sup>

1. **Đặng Hùng Thắng**, *Mở đầu về lý thuyết xác suất và các ứng dụng*, NXB Giáo dục (Giáo trình 1 – G<sub>1</sub>).
2. **Đào Hữu Hồ**, *Xác suất thống kê*, NXB Đại học Quốc gia Hà Nội (G<sub>2</sub>).
3. **Đào Hữu Hồ**, *Hướng dẫn giải các bài toán Xác suất – Thống kê*, NXB Đại học Quốc gia Hà Nội (G<sub>3</sub>).
4. **Nguyễn Quang Báu**, *Lý thuyết xác suất và thống kê toán học*, NXB Đại học Quốc gia Hà Nội (G<sub>4</sub>).
5. **Đặng Hùng Thắng**, *Bài tập xác suất*, NXB Giáo dục (G<sub>5</sub>).

Các môn học: (dạng file do cá nhân tổng hợp)

### Hoàng Văn Trọng

1. *Microsoft Excel 2010* (sử dụng cho môn THCS 1)
2. *Những nguyên lý cơ bản của chủ nghĩa Mác – Lênin* (phân 2)  
(dành cho sinh viên toàn trường)
3. *Cơ – Nhiệt* (dành cho sinh viên ngoài khoa Vật lý)
4. *Điện – Quang* (dành cho sinh viên ngoài khoa Vật lý)
5. *Giải tích 2* (dành cho sinh viên ngoài khoa Toán)
6. *Xác suất – Thống kê* (dành cho sinh viên ngoài khoa Toán)

Link download<sup>2</sup>:

[facebook.com/hoangtrongdl/photos\\_albums](https://facebook.com/hoangtrongdl/photos_albums)

<sup>1</sup> Nên tham khảo tối thiểu 2 giáo trình: G<sub>1</sub> và G<sub>2</sub>

<sup>2</sup> Copy link bên dưới và dán vào trình duyệt web. Sau đó tìm tới ảnh của môn học cần tham khảo và download file môn học đó theo link ở phần mô tả của ảnh.