GIÁO TRÌNH MÔN HỌC
CHƯƠNG TRÌNH KHÔNG CHUYÊN
NGÀNH: KẾ TOÁN, QUẢN TRỊ KINH DOANH

STT	MÔN HỌC	GHI CHÚ
1	*Xác suất thống kê*	
2		
3		
4		
5		

TÊN MÔN HỌC	***XÁC SUẤT THỐNG KÊ***
MÃ SỐ	
THỜI LƯỢNG CHƯƠNG TRÌNH	Số tín chỉ: 04 (01 tín chỉ ứng với 15 tiết) Lý thuyết: 60 tiết Thực hành: 0 tiết Tổng cộng: 60 tiết
ĐIỀU KIỆN TIÊN QUYẾT	Đã được trang bị kiến thức Toán cao cấp
MÔ TẢ MÔN HỌC	• Cung cấp các khái niêm cơ bản về lý thuyết xác suất và thống kê toán học. • Trong phần xác suất, các khái niệm về biến cố, xác suất của biến cố. Biến cố ngẫu nhiên, phân phối xác suất được đề cập và nêu lên các đặc trưng. • Trong phần thống kê toán học, sinh viên sẽ học các khái niệm liên quan đến tập mẫu thống kê, lý thuyết ước lượng, kiểm định giả thuyết. • Sinh viên tiếp cận những kiến thức trên thông qua việc kết hợp bài giảng trên lớp, tự học và tìm hiểu thêm trong các tài liệu. • Trang bị kiến thức xác suất, thống kê bước đầu giúp sinh viên làm quen với một vài ứng dụng toán học trong cuộc sống.
ĐIỂM ĐẠT	- Hiện diện trên lớp: 10 % điểm (Danh sách các buổi thảo luận và bài tập nhóm). Vắng 12 tiết không được cộng điểm này. - Kiểm tra KQHT: 20 % điểm (2 bài kiểm tra giữa và cuối môn học: Có ba thang điểm: 2.0 (hai chẵn); 1.0 (một tròn); 0,0: (không chẵn). - Kiểm tra hết môn: 70% điểm (Bài thi hết môn)

Lưu ý: Danh sách các buổi thảo luận và các bài kiểm tra được hủy khi danh sách bảng điểm thi hết môn được công bố.

CẤU TRÚC MÔN HỌC	**KQHT 1: Khái quát những kiến thức cơ bản về lý thuyết xác suất** **KQHT 2: Giải các bài toán liên quan đến đại lượng ngẫu nhiên và Ứng dụng một số quy luật phân phối thông dụng.** **KQHT 3: Xác định tổng thể và mẫu.** **KQHT 4: Ước lượng các tham số đặc trưng của tổng thể.** **KQHT 5: Kiểm định giả thiết các tham số thống kê.** **KQHT 6: Xác định hàm hồi qui và tương quan.** ** Thực hành: Làm bài tập trên lớp+ Hoạt đông theo nhóm+ Thảo luận*

KẾT QUẢ VÀ CÁC BƯỚC HỌC TẬP

Kết quả học tập/ hình thức đánh giá	***Các bước học tập***	***Phương tiện, tài liệu, nơi học và cách đánh giá cho từng bước học***
1. Khái quát những kiến thức cơ bản về lý thuyết Xác suất. **Đánh giá:** Bài tập + **Đạt** : Trình bày được chính xác ít nhất một trong ba định nghĩa về xác suất và giải được các bài tập về: * Giải tích tổ hợp; * Biết cách biểu diễn một biến cố phức hợp thành tổng và tích của các biến cố đơn giản hơn. * Định nghĩa xác suất: Tính được các xác suất của một biến cố ở dạng đơn giản; * Áp dụng các công thức cộng, nhân, đầy đủ, tính được các xác suất.	1. Bổ sung về giải tích tổ hợp. 1.1 Nhắc lại Quy tắc đếm 1.2 Nhắc lại Chỉnh hợp (không lặp) 1.3 Nhắc lại Chỉnh hợp lặp 1.4 Nhắc lại Tổ hợp 1.5 Nhắc lại Hoán vị 2. Liệt kê các biến cố và quan hệ giữa các loại biến cố. 3. Định nghĩa xác suất. 3.1 Định nghĩa xác suất theo cổ điển. 3.2 Định nghĩa xác suất theo thống kê. 3.3 Định nghĩa xác suất theo hình học. 4. Đưa ra một số công thức tính xác suất. 4.1 Các định nghĩa 4.2 Công thức cộng 4.3 Công thức nhân xác suất 4.3.1 Xác suất có điều kiện 4.3.2 Công thức nhân xác suất 5. Công thức xác suất đầy đủ và công thức Bayer 5.1 Công thức xác suất đầy đủ 5.2 Công thức Bayes. 5.3 Công thức Bernoulli. 5.4 Công Thức Bernoulli Mở Rộng 5.4.1 Lược đồ Bernoulli mở rộng. 5.4.2 Công thức Bernoulli mở rộng.	+ Bảng đen + Kiến thức cơ bản về Giải tích tổ hợp. * Tài liệu chính: "**Xác suất thống kê"** * Các tài liệu tham khảo: + Đặng Hấn 1996 - Xác suất thống kê – NXB Thống kê. + Nguyễn Hữu Khánh – Bài giảng Xác suất thống kê – ĐH Cần Thơ. + Giải tích 12 (PTTH). + Học trong phòng. + Trả lời câu hỏi và bài tập nhóm, bài tập về nhà. + Bài tập về nhà.

<table>
<tr>
<td rowspan="4">2. Giải các bài toán liên quan đến đại lượng ngẫu nhiên và Ứng dụng một số quy luật phân phối thông dụng.
Đánh giá:
+ Đạt: Hoàn thành được các yêu cầu sau:
* Hiểu rõ các khái niệm: Đại lượng ngẫu nhiên và phân biệt được đại lượng ngẫu nhiên và biến cố ngẫu nhiên, đại lượng ngẫu nhiên liên tục và rời rạc.
* Viết đúng các công thức tính tham số của đại lượng ngẫu nhiên rời rạc và liên tục.
* Vận dụng công thức, giải các bài tập liên quan như kỳ vọng, phương sai,...
* Nhận biết đại lượng ngẫu nhiên có phân phối xác suất nào đó.
* Biết cách sử dụng các công thức gần đúng để tính xác suất và điều kiện để sử dụng các công thức đó.
* Hiểu rõ các khái niệm đại lượng ngẫu nhiên hai chiều, cách lập bảng phân phối xác suất của đại lượng ngẫu nhiên rời rạc.</td>
<td>1. Khái niệm đại lượng ngẫu nhiên
1.1 Khái niệm đại lượng ngẫu nhiên.
1.2 Liệt kê các đại lượng ngẫu nhiên.</td>
<td rowspan="4">+ Bảng, phấn.
+ Kiến thức Toán cao cấp, toán THPT.
* Tài liệu chính: “Xác suất thống kê”
* Các tài liệu tham khảo
+ Đặng Hấn, 1996 - Xác suất thống kê – NXB Thống kê.
+ Nguyễn Hữu Khánh – Bài giảng Xác suất thống kê – ĐH Cần Thơ.
+ Đinh Văn Gắng – Xác suất và Thống kê toán – NXB Thống kê

+ Học trong phòng.
+ Trả lời câu hỏi và bài tập nhỏ để nắm vững định nghĩa, tính chất, cách tính, bản chất và ý nghĩa của kỳ vọng toán, phương sai, độ lệch chuẩn và giá trị tin chắc nhất.
+ Các câu hỏi ngắn về xác định luật phân phối, về đại lượng ngẫu nhiên 2 chiều, luật số lớn.
+ Bài tập về nhà.</td>
</tr>
<tr>
<td>2. Đưa ra một số qui luật phân phối xác suất của đại lượng ngẫu nhiên.
2.1 Mô tả Bảng phân phối xác suất.
2.2 Khái niệm Hàm mật độ xác suất.
2.3 Khái niệm Hàm phân phối xác suất.
2.4 Khái niệm phân vị mức xác suất α</td>
</tr>
<tr>
<td>3. Liệt kê một số tham số đặc trưng của đại lượng ngẫu nhiên
3.1 Khái niệm Kỳ vọng
3.2 Khái niệm Phương sai.
3.3 Khái niệm Độ lệch tiêu chuẩn
3.4 Khái niệm Moment
3.5 Khái niệm Mode
3.6 Trung vị</td>
</tr>
<tr>
<td>4. Sử dụng một số qui luật phân phối xác suất thông dụng.
4.1 Phân phối nhị thức
4.2 Phân phối Poison
4.3 Phân phối siêu bội
4.4 Phân phối chuẩn
4.5 Phân phối mũ
4.6 Phân phối χ^2
4.7 Phân phối Student
4.8 Phân phối đều.</td>
</tr>
</table>

* Từ bảng phân phối xác suất của đại lượng ngẫu nhiên 2 chiều, có thể tính được kỳ vọng toán và phương sai của các đại lượng ngẫu nhiên thành phần. Tính được hiệp phương sai của đại lượng ngẫu nhiên 2 chiều. * Hiểu được ý nghĩa các định lý của luật số lớn.	5. Đại lượng ngẫu nhiên hai chiều. 5.1. Định nghĩa đại lượng ngẫu nhiên hai chiều. 5.2. Giới thiệu một số phân phối xác suất của đại lượng ngẫu nhiên hai chiều. 5.2.1 Bảng phân phối xác suất. 5.2.2 Hàm phân phối xác suất. 5.2.3 Hàm mật độ xác suất. 5.3 Các tham số đặc trưng của hàm một biến ngẫu nhiên. 5.3.1 Trường hợp (X,Y) rời rạc. 5.3.2 Trường hợp (X,Y) liên tục. 6. Luật số lớn. 6.1 Bất đẳng thức Markov 6.2 Bất đẳng thức Tchebyshev 6.3 Định lý Tchebyshev 6.4 Định lý Bernoulli	
3. Xác định Tổng thể và mẫu. **Đánh giá:** Câu hỏi ngắn Bài tập. **Đạt:** * Hiểu rõ các khái niệm: Tổng thể, mẫu, trung bình tổng thể, phương sai tổng thể, tỉ lệ tổng thể. * Thấy rõ sự khác nhau giữa mẫu ngẫu nhiên và mẫu cụ thể. * .Biết tính các tham số đặc trưng của mẫu. * Thực hành tính được các yếu tố $\overline{x}$, s’	1. Khái niệm Tổng thể và mẫu 1.1 Khái niệm Tổng thể 1.2 Khái niệm Mẫu 1.3 Đưa ra mô hình xác suất của tổng thể và mẫu 2. Tìm hiểu về Thống kê mẫu ngẫu nhiên. 2.1 Nêu Trung bình của mẫu ngẫu nhiên 2.2 Khái niệm Phương sai và phương sai điều chỉnh của mẫu ngẫu nhiên 2.3 Đưa ra công thức Độ lệch tiêu chuẩn và độ lệch tiêu chuẩn hiệu chỉnh. 3. Thu thập số liệu và sắp xếp số liệu. 3.1 Thu thập số liệu 3.2 Sắp xếp số liệu. 3.3 Thực hành tính các giá trị $\overline{x}$, s’	+ Bảng, phấn. + Kiến thức Toán cao cấp, toán THPT. * Tài liệu chính: **“Xác suất thống kê”** * Các tài liệu tham khảo + Đặng Hấn, 1996 - Xác suất thống kê – NXB Thống kê. + Nguyễn Hữu Khánh – Bài giảng Xác suất thống kê – ĐH Cần Thơ. + Đinh Văn Gắng – Xác suất và Thống kê toán – NXB Thống kê + Học trong phòng. + Trả lời câu hỏi và bài tập nhỏ để nắm vững các khái niệm và công thức. + Bài tập về nhà.

4. Ước lượng tham số của đại lượng ngẫu nhiên. **Đánh giá :** Câu hỏi ngắn Bài tập giải theo nhóm. **Đạt:** Đáp ứng được các yêu cầu sau đây: * Hiểu rõ các khái niệm ước lượng điểm, ước lượng khoảng, độ tin cậy, độ chính xác. * Biết tìm khoảng tin cậy của các tham số của tổng thể. * Biết tìm kích thước mẫu, độ tin cậy khi ước lượng trung bình và tỉ lệ của tổng thể.	1. Giới thiệu các phương pháp ước lượng 1.1 Mô tả phương pháp. 1.2 Đưa ra các phương pháp ước lượng điểm. 2. Ước lượng các tham số 2.1 Mô tả phương pháp 2.2 Ước lượng tham số trung bình 2.3 Ước lượng tham số tỉ lệ 2.4 Ước lượng tham số phương sai.	+ Bảng, phấn. + Kiến thức Toán cao cấp. * Tài liệu chính: **“Xác suất thống kê”** * Các tài liệu tham khảo + Đặng Hấn, 1996 - Xác suất thống kê – NXB Thống kê. + Nguyễn Hữu Khánh – Bài giảng Xác suất thống kê – ĐH Cần Thơ. + Đinh Văn Gắng – Xác suất và Thống kê toán – NXB Thống kê + Học trong phòng. + Trả lời câu hỏi và bài tập nhỏ. + Bài tập về nhà.
5. Kiểm định giả thuyết tham số thống kê. **Đánh giá :** Câu hỏi ngắn Bài tập thực hành theo nhóm. **Đạt:** * Hiểu rõ các khái niệm: Giả thiết thống kê, kiểm định giả thiết, giả thiết cần kiểm định, giả thiết đối, mức ý nghĩa, miền bác bỏ, các sai lầm và biết cách đặt giả thiết. * Làm được các bài tập vận dụng công thức để kiểm định các tham số.	1. Nêu các khái niệm về kiểm định 1.1 Nêu các khái niệm về kiểm định 1.2 Mô tả phương pháp kiểm định giả thiết thống kê. 2. Kiểm định các giả thuyết thống kê. 2.1 Kiểm định tham số trung bình 2.2 Kiểm định tham số tỷ lệ 2.3 Kiểm định giả thuyết về phương sai 2.4 Kiểm định giả thuyết về sự bằng nhau của hai trung bình 2.5 Kiểm định giả thuyết về sự bằng nhau của hai tỉ lệ 2.6 Kiểm định giả thuyết về sự bằng nhau của hai phương sai	+ Bảng, phấn. + Kiến thức Toán cao cấp. * Tài liệu chính: **“Xác suất thống kê”** * Các tài liệu tham khảo + Đặng Hấn, 1996 - Xác suất thống kê – NXB Thống kê. + Nguyễn Hữu Khánh – Bài giảng Xác suất thống kê – ĐH Cần Thơ. + Đinh Văn Gắng – Xác suất và Thống kê toán – NXB Thống kê + Học trong phòng. + Trả lời câu hỏi và bài tập nhỏ. + Bài tập về nhà.

6. Xác định hồi qui và tương quan tuyến tính. **Đánh giá:** Câu hỏi ngắn Bài tập **thực hành** **Đạt:** Đáp ứng được các yêu cầu sau: * Nắm được mối quan hệ giữa hai đại lượng ngẫu nhiên. * Vận dụng công thức để tìm được phương trình hồi qui và mối tương quan giữa chúng.	1. Nêu mối quan hệ giữa các đại lượng ngẫu nhiên. 2. Khái niệm hệ số tương quan. 2.1 Khái niệm Moment tương quan. 2.2 Khái niệm hệ số tương quan. 2.3 Ước lượng hệ số tương quan. 3. Xác định hồi qui. 3.1 Khái niệm kỳ vọng có điều kiện. 3.2 Khái niệm hàm hồi qui 3.3 Xác định hàm hồi qui	+ Bảng, phấn. + Kiến thức Toán cao cấp. * Tài liệu chính: **“Xác suất thống kê”** * Các tài liệu tham khảo + Đặng Hấn, 1996 - Xác suất thống kê – NXB Thống kê. + Nguyễn Hữu Khánh – Bài giảng Xác suất thống kê – ĐH Cần Thơ. + Đinh Văn Gắng – Xác suất và Thống kê toán – NXB Thống kê + Học trong phòng. + Trả lời câu hỏi và bài tập nhỏ. + Bài tập về nhà.

KẾ HOẠCH ĐÁNH GIÁ MÔN HỌC

Kết quả học tập	*Thời lượng giảng dạy*	*Mức độ yêu cầu đạt được*	*Hình thức đánh giá*					
			Viết	**Thao tác**	**Bài tập về nhà**	**Thực tập thực tế**	**Đề tài**	**Tự học**
1.	12,0	Giải được bài tập	X					
2.	14,0	Giải được bài tập	X		X			
3.	06,0	Giải được bài tập	X					
4.	09,0	Giải được bài tập	X		X			
5.	12,0	Giải được bài tập	X		X			
6.	07,0	Giải được bài tập	X					

ĐÁNH GIÁ CUỐI MÔN HỌC

HÌNH THỨC **Thi (tự luận)** .

THỜI GIAN **90 - 120 phút.**

NỘI DUNG ĐÁNH GIÁ

Trọng tâm:

- Các bài toán tính xác suất dạng cổ điển, các công thức cộng, nhân, đầy đủ, Bernuolli.
- Các bài toán về tính toán các tham số như kỳ vọng, phương sai, độ lệch tiêu chuẩn của đại lượng ngẫu nhiên.
- Sử dụng tính phân phối của đại lượng ngẫu nhiên để giải các bài tập như phân phối nhi thức, Poison, Chuẩn, mũ, đều,…
- Các bài tập về ước lượng tham số của đại lượng ngẫu nhiên.
- Các bài toán về kiểm định các tham số của đại lượng ngẫu nhiên.
- Tìm hàm hồi qui tuyến tính.

NỘI DUNG CHI TIẾT MÔN HỌC

KQHT 1: KHÁI QUÁT NHỮNG KIẾN THỨC CƠ BẢN VỀ LÝ THUYẾT XÁC SUẤT

Bước học 1. BỔ SUNG VỀ GIẢI TÍCH TỔ HỢP

1.1 Quy tắc đếm (quy tắc nhân):

Định nghĩa: *Giả sử một công việc phải trải qua k giai đoạn. Giai đoạn 1 có n_1 cách thực hiện, giai đoạn 2 có n_2 cách thực hiện,..., giai đoạn k có n_k cách thực hiện.*

Khi đó, để hoàn thành cả công việc thì ta có $\mathbf{n = n_1\, n_2\, n_3 \,..n_k}$ cách thực hiện.

Ví dụ 1: *Có 4 quyển sách toán, 2 quyển sách lý, 3 quyển sách văn. Hỏi có bao nhiêu cách để lấy ra mỗi loại một quyển sách?*

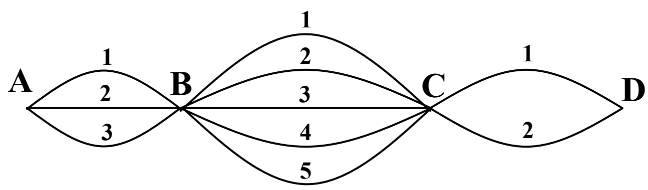
Có 3 giai đoạn: Giai đoạn 1, lấy 1 quyển toán $\rightarrow$ có 4 cách lấy.

Giai đoạn 2, lấy 1 quyển lý $\rightarrow$ có 2 cách lấy.

Giai đoạn 3, lấy 1 quyển văn $\rightarrow$ có 3 cách lấy.

$\Rightarrow$ Số cách lấy là n = 4.2.3 = 24 cách

Ví dụ 2: *Có 3 cách đi từ thành phố A đến thành phố B, có 5 cách đi từ thành phố B đến thành phố C và có 2 cách đi từ thành phố C đến thành phố D. Hỏi có bao nhiêu cách đi từ thành phố A đến thành phố D ?*



10

Số cách đi từ thành phố A đến thành phố D là : n = 3.5.2 = 30 (cách)

Ví dụ 3: *Các nhóm I , II , III , IV lần lượt có 8 ,10 ,12 , 9 sinh viên. Cần chọn 4 sinh viên, mỗi nhóm 1 sinh viên. Hỏi có bao nhiêu cách chọn như vậy?*

Việc chọn 4 sinh viên xem như được chia làm 4 giai đoạn:

Giai đoạn 1: Chọn 1 sinh viên của nhóm I : 8 cách.

Giai đoạn 2: Chọn 1 sinh viên của nhóm II : 10 cách.

Giai đoạn 3: Chọn 1 sinh viên của nhóm III : 12 cách.

Giai đoạn 4: Chọn 1 sinh viên của nhóm IV : 9 cách.

$\Rightarrow$ Số cách chọn: 8.10.12.9 = 8640 cách.

1.2 Chỉnh hợp (không lặp):

Định nghĩa: *Chỉnh hợp chập k của n phần tử ($k \leq n$) là một bộ (nhóm) có thứ tự gồm k phần tử khác nhau được chọn từ n phần tử đã cho. Chỉnh hợp chập k của n phần tử kí hiệu là:* A_n^k

♦ Vấn đề đặt ra là: Có n phần tử thì có thể lập được bao nhiêu chỉnh hợp chập k khác nhau?

Công thức: $$A_n^k = \frac{n!}{(n-k)!}$$

Chú ý: + n!: n giai thừa. n! = n.(n-1)……3.2.1

+ Qui ước: 0! = 1

Ví dụ 4: *Trong buổi hợp gồm 12 người. Hỏi có bao nhiêu cách chọn một chủ tọa và một thư ký?*

Số cách chọn là chỉnh hợp chập 2 của 12 $\Rightarrow$ có n = $A_{12}^2 = \frac{12!}{(12-2)!}$ = 12.11 =132 cách.

Ví dụ 5: *Cho một tập hợp gồm các số 0,1,2,3,4,5. Hỏi có bao nhiêu số tự nhiên gồm 4 chữ số khác nhau?*

Ta có các số 0123,0134,… không phải là số tự nhiên có 4 chữ số nên ta chia công việc ra làm hai giai đoạn.

Giai đoạn 1: Chọn chữ số đầu tiên phải khác 0. Vì còn lại 5 số nên có 5 cách chọn.

Giai đoạn 2: Chọn 3 số còn lại từ 5 số còn lại. Do có kể thứ tự, không trùng nhau nên số cách chọn là số chỉnh hợp chập 3 của 5: A_5^3 = 3.4.5= 60.

$\Rightarrow$ Số cách hoàn thành công việc là n = 5.60 = 300 cách.

Ví dụ 6: *Cho E = {1, 2, 3, 4}. Có bao nhiêu số tự nhiên bao gồm hai chữ số phân biệt được thành lập từ E.*

Mỗi số tự nhiên bao gồm hai chữ số phân biệt được thành lập từ E là một chỉnh hợp *(không lặp)* chập 2 của 4. Nên số các số tự nhiên cần tìm là:

$$\mathrm{A}_4^2 = \frac{4!}{2!} = \frac{4.3.2.1}{2.1} = 6$$

Ví dụ 7: *Một lớp có 8 môn học, mỗi ngày học 2 môn. Hỏi có bao nhiêu cách xếp thời khoá biểu trong một ngày?*

Số cách xếp thời khoá biểu trong một ngày chính là việc lấy 2 phần tử khác nhau từ tập hợp gồm 8 phần tử. Vì việc lấy gắn liền với việc xếp thời khoá biểu nên thứ tự là quan trọng.

Vậy số cách xếp thời khoá biểu cho một ngày là số chỉnh hợp chập 2 của 8 phần tử:

$$A_8^2 = \frac{8!}{(8-2)!} = \frac{8!}{6!} = 7.8 = 56$$ *(cách)*

1.3 Chỉnh hợp lặp:

Định nghĩa: *Chỉnh hợp lặp chập k của n phần tử là một bộ (nhóm) có thứ tự gồm k phần tử được chọn từ n phần tử đã cho, trong đó các phần tử trong nhóm có thể lặp lại 2,3,4,.., k lần.*

Gọi số chỉnh hợp lặp chập k của n phần tử là B_n^k, khi đó: $B_n^k = n^k$

Ví dụ 8: *Xếp ngẫu nhiên 5 quyển sách vào 3 ngăn kéo. Hỏi có bao nhiêu cách xếp?*

Mỗi cách xếp 5 quyển sách vào 3 ngăn kéo xem như một chỉnh hợp lặp chập 3 của 5 (mỗi lần xếp một quyển sách vào một ngăn, ta có thể xem như chọn một trong 3 ngăn $\Rightarrow$ Có 3 cách chọn. Do có 5 quyển sách nên số cách chọn là n = 3^5 = 243 cách.

Ví dụ 9: *Có bao nhiêu số tự nhiên gồm 4 chữ số từ các số: 1,2,3,4,5?*

Có $B_5^4 = 5^4 = 625$ số.

Ví dụ 10: *Có bao nhiêu cách sắp xếp 10 người lên một tàu hỏa có 3 toa?*

Số cách sắp xếp 10 người lên 3 toa tàu là số các chỉnh hợp lặp chập 10 của 3 phần tử. Số cách sắp xếp: $B_3^{10} = 3^{10}$

Ví dụ 11: *Mỗi vé số của mỗi tỉnh gồm có 6 chữ số. Hỏi mỗi tỉnh khi phát hành mỗi đợt sẽ phát hành được bao nhiêu vé số khác nhau?*

Ta có mỗi vé số gồm có 6 chữ số, nên ta có thể xem việc phát hành ra một vé số là việc chọn ra 6 số bất kỳ có thứ tự có thể trùng nhau từ 10 số từ 0 đến 9. Do đó mỗi vé số được phát hành có thể được xem là một chỉnh hợp lặp chập 6 của 10.

Vậy số vé số có thể phát hành mỗi đợt của mỗi tỉnh là số chỉnh hợp lặp chập 6 của 10:

$$B_{10}^6 = 10^6 = 1000000 \text{ (vé số)}$$

__Lưu ý:__ Trong chỉnh hợp không lặp thì $k \le n$ còn trong chỉnh hợp lặp thì có thể có k > n.

1.4 Hoán vị:

Định nghĩa: *Hoán vị của n phần tử là một bộ có thứ tự gồm đủ mặt n phần tử đã cho.*

Gọi số hoán vị của n phần tử là P_n, ta có công thức: $\mathbf{P_n = n!}$

Hai hoán vị khác nhau khi nào?

Do mỗi hoán vị đều có đủ mặt các phần tử, nên hai hoán vị khác nhau khi có ít nhất một thứ tự sắp xếp nào đó khác nhau. Chẳng hạn: 312 khác 321.

Ví dụ 12: *Hỏi có bao nhiêu cách xếp 4 học sinh vào một bàn có 4 chỗ ngồi?*

Số cách xếp là: n = P_4 = 4! = 24 cách.

Ví dụ 13: *Có 3 cuốn sách Toán, 2 cuốn sách Lý và 5 cuốn sách XSTK (các cuốn sách này khác nhau) được xếp vào 1 cái kệ. Hỏi có bao nhiêu cách sắp xếp sao cho các cuốn sách cùng loại đứng gần nhau?*

Để thỏa bài toán, ta chia công việc ra các giai đoạn sau :

Giai đoạn 1: Phân kệ thành 3 phần để xếp 3 loại sách: Có 3! cách sắp xếp.

Giai đoạn 2: Xếp 3 cuốn Toán $\rightarrow$ phần dành cho Toán: Có 3! cách sắp xếp.

Giai đoạn 3: Xếp 2 cuốn Lý $\rightarrow$ phần dành cho Lý: Có 2! cách sắp xếp.

Giai đoạn 4: Xếp 5 cuốn XSTK $\rightarrow$ phần dành cho XSTK: Có 5! cách sắp xếp.

$\Rightarrow$ Số cách sắp xếp cho cả bài toán: 3!.3!.2!.5! = 8640 (cách)

1.5 Tổ hợp:

Định nghĩa: *Tổ hợp chập k của n phần tử (k $\le$ n) là một bộ (nhóm) không kể thứ tự gồm k phần tử khác nhau được chọn từ n phần tử đã cho. Gọi số tổ hợp chập k của n phần tử là: C_n^k, có:*

$$C_n^k = \frac{n!}{k!(n-k)!}$$

Chú ý: $C_n^k = C_n^{n-k} \Rightarrow C_n^0 = C_n^n = 1$

Ví dụ 14: *Mỗi đề thi gồm có 3 câu hỏi khác nhau chọn từ 25 câu hỏi đã cho. Hỏi có thể thành lập được bao nhiêu đề thi khác nhau?*

Mỗi đề thi sẽ chọn 3 câu từ 25 câu đã cho. Do chọn không kể thứ tự, không trùng nhau nên số cách chọn là tổ hợp chập 3 của 25

$$\Rightarrow C_{25}^3 = \frac{25!}{3!(25-3)!} = \frac{25.24.23}{6} = 2300 \text{ cách.}$$

Ví dụ 15: *Trong một giải bóng chuyền chào mừng ngày Học sinh – Sinh viên của Trường. Có 12 đội bóng tham gia thi đấu vòng tròn một lượt. Hỏi có bao nhiêu trận đấu được tiến hành?*

Mỗi trận đấu có hai đội tham gia từ 12 đội, nên số trận đấu cần tiến hành là:

$$C_{12}^2 = \frac{12!}{2!10!} = \frac{11.12.10!}{2.10!} = 11.6 = 66$$

Ví dụ 16: *Từ lô hàng có 10 sản phẩm, ta rút ngẫu nhiên (đồng thời) 3 sản phẩm để kiểm tra. Tính số khả năng có thể xảy ra?*

Số khả năng có thể xảy ra là số tổ hợp chập 3 của 10 phần tử:

$$C_{10}^3 = \frac{10!}{3!(10-3)!} = 120$$

Ví dụ 17: *Nhóm A có 10 sinh viên và nhóm B có 12 sinh viên. Ta chọn ngẫu nhiên 9 sinh viên trong đó có 4 sinh viên nhóm A và 5 sinh viên nhóm B. Tính số khả năng có thể xảy ra?*

Chọn 4 sinh viên từ nhóm A có 10 sinh viên: Có $C_{10}^4 = \dfrac{10!}{4!(10-4)!} = 210$ cách.

Chọn 5 sinh viên từ nhóm B có 12 sinh viên: Có $C_{12}^5 = \dfrac{12!}{5!(12-5)!} = 792$ cách.

Áp dụng quy tắc nhân suy ra số khả năng có thể là: 210.792 = 166320

Lưu ý:

♦ Hai tổ hợp khác nhau khi nào?

♦ Chỉnh hợp khác tổ hợp khi nào?

BÀI TẬP

1. Một buổi liên hoan có 6 người trong đó có 2 người là vợ chồng

a. Nếu 6 người này ngồi quanh một cái bàn tròn có 6 cái ghế được đánh số. Hỏi có bao nhiêu cách sắp xếp sao cho 2 vợ chồng luôn ngồi cạnh nhau.

b. Nếu họ được xếp vào một cái bàn dài có 6 ghế, thì có bao nhiêu cách xếp để 2 vợ chồng luôn ngồi cạnh nhau.

2. Một nhóm gồm 5 vợ chồng đứng xếp hàng. Hỏi có bao nhiêu cách sắp xếp trong các trường hợp sau:

a. Nam và nữ đứng thành 2 nhóm riêng biệt.

b. Hai vợ chồng luôn đứng kế nhau.

c. Nếu mỗi người bắt tay một lần với người khác. Hỏi tất cả có bao nhiêu cái bắt tay.

d. Nếu trong nhóm có 3 người không bắt tay với nhau. Hỏi có bao nhiêu cái bắt tay trong trường hợp này.

3. Một lô hàng gồm có 6 sản phẩm được đánh các số thứ tự từ 1 đến 6, trong đó có 2 phế phẩm. Người ta lấy từ lô hàng lần lượt từng sản phẩm cho đến hết.

a. Có bao nhiêu trường hợp có thể xảy ra.

b. Có bao nhiêu trường hợp 2 phế phẩm được lấy sau cùng.

4. Một nhân viên bưu điện đưa ngẫu nhiên 3 lá thư cho 3 người khác nhau. Hỏi:

a. Có bao nhiêu trường hợp có thể xảy ra.

b. Có bao nhiêu trường hợp có ít nhất một người nhận đúng thư của mình.

5. Từ các số 1, 2, 3, 4, 5 ta có thể thành lập được bao nhiêu số trong các trường hợp sau:

a. Số có 3 chữ số.

b. Số chẵn có 3 chữ số khác nhau.

c. Số chia hết cho 5 có 3 chữ số khác nhau.

d. Số có 3 chữ số trong đó có số 1.

e. Số có 3 chữ số khác nhau gồm toàn số lẻ.

6. Từ các số 0, 1, 2, 3, 4, 5 ta có thể thành lập được bao nhiêu số trong các trường hợp sau:

a. Số có 3 chữ số.

b. Số chẵn có 3 chữ số khác nhau.

c. Số chia hết cho 5 có 3 chữ số khác nhau.

d. Số có 3 chữ số trong đó có số 1.

e. Số có 3 chữ số khác nhau gồm toàn số lẻ.

7. Giải bóng đá hạng nhất quốc gia gồm có 12 đội.

a. Nếu các đội thi đấu vòng tròn một lượt với nhau. Hỏi có bao nhiêu trận đấu đã xảy ra.

b. Nếu các đội được chia làm 3 bảng đều nhau, và mỗi đội trong bảng thi đấu vòng tròn một lượt với nhau thì có bao nhiêu trận đấu đã xảy ra.

8. Một lớp có 8 môn để học, mỗi ngày học 2 môn (sáng, chiều). Hỏi có bao nhiêu cách sắp xếp thời khoá biểu cho một ngày của lớp đó.

9. Một tổ gồm có 10 người, người ta muốn thành lập một tiểu ban gồm có 3 người.

a. Nếu 3 người này cùng làm một công việc thì có bao nhiêu cách chọn.

b. Nếu 3 người này được chọn làm 3 công việc khác nhau thì có bao nhiêu cách chọn.

10. Mỗi vé số của mỗi tỉnh khi phát hành có 6 chữ số.

a. Hỏi có bao nhiêu vé số khác nhau có thể phát hành mỗi đợt của mỗi tỉnh.

b. Nếu bạn trúng 2 số cuối cùng so với số số của giải này bạn sẽ được thưởng 20.000 đồng. Hỏi mỗi đợt phát hành có bao nhiêu vé số trúng 20.000 đồng.

11. Có n điểm khác nhau nằm trên một đường tròn.

a. Có bao nhiêu dây cung được tạo nên từ n điểm đó.

b. Có bao nhiêu đường chéo của đa giác tạo nên từ n điểm đó.

c. Đa giác nào có số đường chéo bằng số cạnh.

12. Có 6 dôi giày. Chọn ngẫu nhiên 4 chiếc giày. Hỏi có bao nhiêu cách chọn trong các trường hợp sau:

a. Chọn được 2 đôi giày.

b. Chọn được chỉ một đôi giày.

c. Không chọn được đôi giày nào cả.

13. Gieo một con xúc xắc liên tiếp 3 lần, có phân biệt thứ tự các lần gieo.

a. Có bao nhiêu kết quả khác nhau có thể xảy ra.

b. Có bao nhiêu kết quả xảy ra trong đó mặt mang số 6 không xuất hiện lần nào.

c. Có bao nhiêu kết quả xảy ra trong đó mặt mang số 6 xuất hiện ít nhất một lần.

14. Một khách sạn có 6 phòng đơn. Có 10 người khách đến thuê phòng, trong đó có 6 nam và 4 nữ. Người quản lý chọn ngẫu nhiên 4 người. Có bao nhiêu cách chọn trong các trường hợp sau:

a. Cả 6 người đều là nam.

b. Có 4 nam và 2 nữ.

c. Có ít nhất 2 nữ.

15. Một khoá số có 3 vòng, mỗi vòng được đánh số từ 0 đến 9 và chỉ có một khả năng để mở khoá. Một khả năng mở khoá là cách chọn đúng số theo thứ tự của 3 vòng. Một người muốn thử các trường hợp mở khoá. Hỏi người này mở tối đa bao nhiêu lần để chắc chắn sẽ chọn đúng số mở.

Bước học 2: LIỆT KÊ CÁC BIẾN CỐ VÀ QUAN HỆ GIỮA CÁC LOẠI BIẾN CỐ

1. Phép thử và biến cố:

Việc thực hiện một nhóm điều kiện xác định để quan sát một hiện tượng nào đó được gọi là một phép thử. Kết quả của phép thử được gọi là biến cố.

Ví dụ 1: *Khi một sinh viên đi thi môn Xác suất thống kê: thực hiện phép thử. Kết quả của phép thử là sinh viên thi đậu hoặc rớt. Đậu hoặc rớt là những sự kiện ngẫu nhiên.*

Tung một đồng xu là một phép thử, đồng xu xuất hiện mặt xấp hay ngữa là các biến cố.

Tung một con xúc xắc là một phép thử, xúc xắc xuất hiện mặt 1,..,6 là các biến cố.

Bắn một viên đạn đến một mục tiêu để xem viên đạn trúng hay trật.

♦ Điều kiện xác định của các hiện tượng ngẫu nhiên là gì?

♦ Hãy phân tích các yếu tố: Nhóm điều kiện, hiện tượng, kết quả của các phép thử trên. Cho các ví dụ khác và phân tích các yếu tố.

2. Các loại biến cố:

2.1. Biến cố chắc chắn:

Là biến cố chắc chắn xảy ra trong một phép thử, và người ta kí hiệu là: **W**

Ví dụ 2: *Tung một con xúc xắc. Gọi A là biến cố xúc xắc xuất hiện mặt có số chấm nhỏ hơn hoặc bằng 6. Khi đó ta nói A là biến cố chắc chắn, A = W.*

2.2. Biến cố không thể:

Là biến cố không thể xảy ra trong một phép thử, và người ta kí hiệu là: $\varnothing$

Ví dụ 3: *Tung một con xúc xắc. Gọi B là biến cố xúc xắc xuất hiện mặt 7 chấm. Khi đó ta nói A là biến cố không thể, A =* $\varnothing$.

2.3. Biến cố ngẫu nhiên:

Là biến cố có thể xảy ra cũng không thể xảy ra trong một phép thử. Ta thường dùng các chữ cái A, B, C,.. để kí hiệu cho biến cố ngẫu nhiên.

Ví dụ 4: *Một xạ thủ bắn vào một tấm bia, gọi A là biến cố xạ thủ bắn trúng bia, A là biến cố ngẫu nhiên.*

2.4. Biến cố thuận lợi (Biến cố kéo theo)

Biến cố A được gọi là thuận lợi cho biến cố B nếu A xảy ra thì B cũng xảy ra. Kí hiệu: $A \subset B$.

Ví dụ 5: *Tung một con xúc xắc. Gọi A là biến cố xúc xắc xuất hiện mặt 2 chấm và B là biến cố xuất hiện mặt chẵn. Khi đó ta nói $A \subset B$.*

Đặc biệt: Nếu $A \subset B$ và $B \subset A$ thì A và B là hai biến cố tương đương.

Kí hiệu A = B.

Ví dụ 6: *Mỗi số chấm trên mặt xúc xắc tương ứng 5 điểm. Gọi A là biến cố xúc xắc xuất hiện mặt 6 chấm, B là biến cố được 30 điểm. Khi đó A = B.*

2.5. Biến cố sơ cấp:

Biến cố A được gọi là biến cố sơ cấp nếu nó không có biến cố cố nào thuận lợi cho nó (trừ chính nó), tức là không thể phân tích được nữa.

Ví dụ 7: *Gọi A_i là biến cố xúc xắc xuất hiện mặt i chấm (i=1,..,6) thì $A_1, A_2, .. , A_6$ là các biến cố sơ cấp.*

Gọi B là biến cố thu được mặt có số chấm chẵn.

$\Rightarrow B = A_2 + A_4 + A_6 \Rightarrow$ B không phải là biến cố sơ cấp.

Tập hợp tất cả các biến cố sơ cấp của một phép thử được gọi là không gian các biến cố sơ cấp và kí hiệu: W

Ví dụ 8: $W = \{ A_1, A_2, A_3, A_4, A_5, A_6\}$.

2.6. Biến cố hiệu:

Hiệu của hai biến cố A và B, kí hiệu A-B (hay A\B) là một biến cố xảy ra $\Leftrightarrow$ A xảy ra nhưng B không xảy ra.

Ví dụ 9: *Tung một con xúc xắc.*

Gọi A là biến cố xúc xắc xuất hiện mặt có số chấm là số lẻ.

B là biến cố xúc xắc xuất hiện mặt có số chấm là số nguyên tố nhỏ hơn 5.

C là biến cố xúc xắc xuất hiện mặt có 5 chấm.

Ta có: $C = A \backslash B$

2.7. Biến cố tổng:

Tổng của hai biến cố A và B, kí hiệu A + B hay $A \cup B$ là một biến cố xảy ra $\Leftrightarrow$ ít nhất một trong hai biến cố A và B xảy ra.

Ví dụ 10: *Hai xạ thủ cùng bắn vào một con thú. Gọi A là biến cố xạ thủ thứ nhất bắn trúng, B là biến cố xạ thủ thứ hai bắn trúng. Khi đó biến cố thú bị trúng đạn là C = A + B.*

Ví dụ 11: *Có 2 xạ thủ, mỗi người bắn 1 viên đến 1 mục tiêu.*

Gọi A_i là biến cố xạ thủ thứ i bắn trúng mục tiêu (i = 1, 2).

Gọi $\overline{A_i}$ là biến cố xạ thủ thứ i không bắn trúng mục tiêu (i =1, 2).

Gọi B_i là biến cố mục tiêu bị bắn trúng i viên đạn.

Ta có: $B_0 = \overline{A_1}.\overline{A_2}$

$B_1 = A_1.\overline{A_2} + \overline{A_1}.A_2$

$B_2 = A_1.A_2$

Tổng quát: Tổng của n biến cố A_1, A_2, .., A_n là một biến cố xảy ra $\Leftrightarrow$ ít nhất một trong các biến cố A_i xảy ra (i = 1,..,n).

Kí hiệu: $A_1 + A_2 + .. + A_n$ hay $A_1 \cup A_2 \cup .. \cup A_n$

Chú ý: Biến cố chắc chắn W là tổng của mọi biến cố sơ cấp có thể, nghĩa là mọi biến cố sơ cấp đều thuận lợi cho W. Do đó, W còn được gọi là không gian các biến cố sơ cấp.

2.8. Biến cố tích:

Tích của hai biến cố A và B, kí hiệu: AB hay $A \cap B$ là một biến cố xảy ra $\Leftrightarrow$ cả hai biến cố A và B đồng thời xảy ra.

Ví dụ 12: *Hai xạ thủ cùng bắn vào một con thú. Gọi A là biến cố xạ thủ thứ nhất bắn trật, B là biến cố xạ thủ thứ hai bắn trật. Khi đó biến cố thú bị không bị trúng đạn là C = AB.*

Tổng quát: Tích của n biến cố A_1, A_2, .., A_n là một biến cố xảy ra $\Leftrightarrow$ tất cả các biến cố A_i đều xảy ra. Kí hiệu: $A_1 A_2 \ldots A_n$ hay $A_1 \cap A_2 \cap .. \cap A_n$

2.9. Biến cố xung khắc:

Hai biến cố A và B được gọi là xung khắc nếu chúng không đồng thời xảy ra trong một phép thử.

Ví dụ 13: *Tung một con xúc xắc, gọi A là biến cố xúc xắc xuất hiện mặt chẵn, B là biến cố xúc xắc xuất hiện mặt 3 chấm $\Rightarrow$ A, B xung khắc.*

2.10. Biến cố đối lập:

Biến cố không xảy ra biến cố A được gọi là biến cố đối lập của A. Kí hiệu: $\overline{A}$

A và $\overline{A}$ đối lập $\Leftrightarrow$ $A\overline{A} = \varnothing$ và $A \cup \overline{A}$ phải là biến cố chắc chắn, tức là trong phép thử có một và chỉ được một A hoặc $\overline{A}$ xảy ra.

Chú ý: Hai biến cố đối lập thì xung khắc nhưng ngược lại 2 biến cố xung khắc thì chưa chắc đối lập.

2.11. Biến cố đồng khả năng:

Các biến cố A, B, C,.. được gọi là đồng khả năng nếu chúng có cùng một khả năng xuất hiện như nhau trong một phép thử.

Ví dụ 14: *Tung một đồng xu, gọi S là biến cố đồng xu xuất hiện mặt xấp, N là biến cố xuất hiện mặt ngữa $\Rightarrow$ S, N là hai biến cố đồng khả năng.*

Tóm lại, qua các khái niệm trên, ta thấy các biến cố tổng, hiệu, tích, đối lập tương ứng với tập hợp, giao, hiệu, phần bù của lý thuyết tập hợp. Do đó có thể sử dụng các phép toán trên các tập hợp cho các phép toán trên các biến cố.

3. Các tính chất:

1. A + (B + C) = (A + B) + C ; A.(B.C) = (A.B).C
2. A + B = B + A ; A.B = B.A
3. A(B + C) = A.B + B.C

4. $A + A = A$; $A.A = A$
5. $A + W = W$; $A.W = A$
6. $A + \varnothing = A$; $A.\varnothing = \varnothing$
7. $B = \overline{A} \Rightarrow A = \overline{B}$ hay $\overline{(\overline{A})} = A$
8. $\overline{A+B} = \overline{A}.\overline{B}$; $\overline{A.B} = \overline{A} + \overline{B}$

Ví dụ 15: $\overline{B}(AB\overline{C} + A\overline{B}C + \overline{A}BC) = \overline{B}AB\overline{C} + \overline{B}A\overline{B}C + \overline{B}.\overline{A}BC$

$= A(\overline{B}B)\overline{C} + A(\overline{B}\,\overline{B})C + \overline{A}(\overline{B}B)C = A\phi\overline{C} + A\overline{B}C + \overline{A}\phi C$

$= \phi + A\overline{B}C + \phi = A\overline{B}C$.

BÀI TẬP

1. Có 3 xạ thủ, mỗi người độc lập bắn một viên vào một mục tiêu. Gọi Ai là biến cố xạ thủ thứ i bắn trúng mục tiêu.

a. Hãy mô tả các biến cố sau: $A_1A_2A_3$; $A_1 + A_2 + A_3$; $\overline{A_1 A_2}A_3$.

b. Xét các biến cố sau:

A: Có ít nhất 2 xạ thủ bắn trúng.

B: Có nhiều nhất một xạ thủ bắn trúng.

C: Chỉ có một xạ thủ bắn trúng.

D: Chỉ có một xạ thủ thứ 3 bắn trúng.

Hãy biểu diễn các biến cố A, B, C, D theo các biến cố A_i.

2. Cho 3 biến cố A, B, C. Hãy mô tả dưới dạng tập hợp các biến cố sau:

a. A, B, C đều xảy ra.

b. A, B xảy ra nhưng C không xảy ra.

c. Chỉ có một trong biến cố xảy ra.

d. Có ít nhất một biến cố xảy ra.

3. Một hộp có 5 sản phẩm, trong đó có 2 phế phẩm. Người ta lấy lần lượt từ hộp ra 2 sản phẩm cho đến khi phát hiện hết 2 phế phẩm thì dừng lại. Gọi A_i biến cố chọn được sản phẩm tốt lần thứ i.

a. Các biến cố A_i có độc lập toàn phần với nhau không? Tại sao?

b. Hãy biến diễn các biến cố sau theo các biến cố A_i

A: Việc kiểm tra dừng lại ở lần thứ 4.

B: Việc kiểm tra dừng lại ở lần lấy sau cùng.

4. Một đồng xu được tung 3 lần. Gọi S là biến cố đồng xu xuất hiện mặt sấp mỗi lần, N là biến cố đồng xu xuất hiện mặt ngữa mỗi lần.

a. S, N là có phải là các biến cố sơ cấp, đối lập nhau không?

b. Hãy tìm không gian các biến cố sơ cấp trong phép thử trên.

c. Hãy biểu diễn biến cố A: Có 2 lần đồng xu xuất hiện mặt ngữa.

5. Một hộp có 4 bi đỏ và 6 bi trắng

a. Chọn ngẫu nhiên từ hộp ra 5 bi. Gọi:

A là biến cố chọn được cả 5 bi đỏ.

B là biến cố chọn được ít nhất một bi trắng.

Xác định loại của biến cố A và biến cố B.

b. Chọn ngẫu nhiên từ hộp ra 4 bi. Gọi:

A_i là biến cố chọn được i bi trắng.

A là biến cố chọn được số bi trắng bằng số bi đỏ.

B là biến cố chọn được số bi trắng lớn hơn số bi đỏ.

C là biến cố có ít nhất một bi trắng.

i/. $\{A_i\}$, i = 0, .., 4 có phải là nhóm biến cố đầy đủ và xung khắc.

ii/. Xác định biến cố đối lặp của biến cố C.

iii/. Biểu diễn biến cố A, B qua các biến cố A_i.

Bước học 3: ĐỊNH NGHĨA XÁC SUẤT

3.1. Định nghĩa xác suất theo lối cổ điển:

Giả sử một phép thử có n biến cố sơ cấp đồng khả năng có thể xảy ra, trong đó có m biến cố sơ cấp đồng khả năng thuận lợi cho biến cố A. Khi đó xác suất của biến cố A (kí hiệu P(A)) được định nghĩa bởi công thức sau:

$P(A) = \dfrac{m}{n}$, *trong đó m là số biến cố thuận lợi cho A, n là biến cố đồng khả năng có thể xảy ra.*

Ví dụ 1: *Tung một con xúc xắc. Tính xác suất để xúc xắc xuất hiện mặt chẵn.*

Gọi A_i là biến cố xuất hiện mặt i chấm.

Gọi A là biến cố xuất hiện mặt chẵn, có $A = A_2 \cup A_4 \cup A_6$

Khi tung con xúc xắc có 6 biến cố đồng khả năng có thể xảy ra trong đó có 3 biến cố thuận lợi cho A. Khi đó: $P(A) = \dfrac{m}{n} = \dfrac{3}{6} = 0{,}5$

Ví dụ 2*: Tung ngẫu nhiên 1 con xúc xắc. Tính xác suất để con xúc xắc xuất hiện mặt có số chấm là số lẻ.*

Gọi A là biến cố xúc xắc xuất hiện mặt có chấm là số lẻ.

A_i là biến cố xúc xắc xuất hiện mặt có i chấm $(i = \overline{1,6})$.

Khi tung 1 con xúc xắc thì có 6 khả năng xảy ra tương ứng con xúc xắc có thể xuất hiện các mặt mang số chấm từ 1 đến 6. Ta có không gian các biến cố sơ cấp là:

$$W = \{A_1, A_2, A_3, A_4, A_5, A_6\}$$

Số trường hợp có thể của phép thử: 6.

Ta có các biến cố sơ cấp thuận lợi cho biến cố A: A_1, A_3, A_5.

Suy ra số trường hợp thuận lợi cho biến cố A: 3

Do đó: $p(A) = \dfrac{3}{6} = \dfrac{1}{2}$

Ví dụ 3: *Tung đồng thời 2 con xúc xắc. Tính xác suất để tổng số chấm xuất hiện trên 2 con xúc xắc là 7.*

Gọi A là biến cố tổng số chấm xuất hiện trên 2 con xúc xắc là 7.

A_i là biến cố xúc xắc thứ nhất xuất hiện mặt có i chấm $(i = \overline{1,6})$.

B_i là biến cố xúc xắc thứ hai xuất hiện mặt có i chấm $(i = \overline{1,6})$.

Ta thấy: Tương tự như ví dụ trên, khi ta tung 1 con xúc xắc thì có 6 khả năng. Do đó khi ta tung 2 con xúc xắc cùng lúc thì có thể có 6.6 = 36 khả năng xảy ra. Ta có không gian các biến cố sơ cấp là:

$$\begin{aligned} W = \{&(A_1,B_1);\ (A_1,B_2);\ ...;\ (A_1,B_6)\\ &(A_2,B_1);\ (A_2,B_2);\ ...;\ (A_2,B_6)\\ &...\quad ...\quad ...\quad ...\\ &(A_6,B_1);\ (A_6,B_2);\ ...;\ (A_6,B_6)\} \end{aligned}$$

Vậy số trường hợp có thể của phép thử là: 36

Ta có các biến cố thuận lợi cho biến cố A:

$$(A_1,B_6);\ (A_2,B_5);\ (A_3,B_4);\ (A_4,B_3);\ (A_5,B_2);\ (A_6,B_1)$$

Suy ra số trường hợp thuận lợi cho biến cố A là: 6.

$$\Rightarrow P(A) = \frac{6}{36} = \frac{1}{6}$$

Ví dụ 4: *Một người gọi điện thoại nhưng lại quên hai số cuối của số điện thoại, chỉ biết rằng hai số đó là khác nhau. Tính xác suất để người đó chỉ quay một lần đúng số cần gọi.*

Gọi B là biến cố người đó chỉ quay một lần đúng số cần gọi.

Số biến cố thuận lợi cho B là: m = 1

Số biến cố đồng khả năng có thể xảy ra là: n = A_{10}^2 = 90

$$\Rightarrow \text{P(A)} = \frac{1}{90}$$

Ví dụ 5: *Một hộp gồm 6 bi trắng và 4 bi đen, lấy ngẫu nhiên 2 bi từ hộp. Tính xác suất để : a) Có 1 bi trắng.*

b) Có 2 bi trắng.

Gọi A là biến cố có 1 bi đen trong 2 bi lấy ra.

Gọi B là biến cố có 2 bi trắng trong 2 bi lấy ra.

Có: $$\text{P(A)} = \frac{m}{n} = \frac{C_6^1 C_4^1}{C_{10}^2} = \frac{4.6}{45} = \frac{8}{15}$$

$$\text{P(B)} = \frac{m}{n} = \frac{C_6^2}{C_{10}^2} = \frac{1}{3}$$

Ví dụ 6: *Trong một hộp đựng 20 quả cầu trong đó có 14 quả cầu đỏ và 06 quả cầu trắng. Lấy ngẫu nhiên (không hoàn lại) 5 quả cầu từ trong hộp. Tính xác suất để trong 5 quả cầu lấy ra có 3 quả cầu đỏ. Biết rằng các quả cầu là cân đối và giống nhau.*

Vì các quả cầu là cân đối và giống nhau. Nên ta có: n = C_{20}^5

Gọi A là biến cố trong 5 quả cầu lấy ra có 3 quả cầu đỏ và 2 quả cầu trắng.

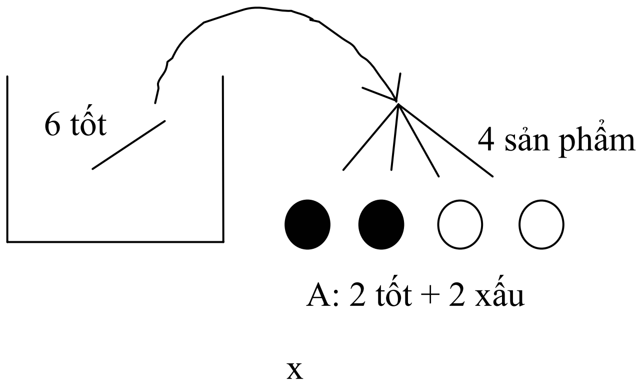
+ Số cách lấy 3 quả cầu đỏ: C_{14}^3

+ Số cách lấy 2 quả cầu trắng: C_6^2

$\Rightarrow$ m = $C_{14}^3 . C_6^2$

$$\Rightarrow P(A) = \frac{m}{n} = \frac{C_6^2.C_{14}^3}{C_{20}^5}$$

Ví dụ 7: *Hộp có 10 sản phẩm trong đó có 4 sản phẩm tốt còn lại là sản phẩm xấu. Chọn ngẫu nhiên từ hộp ra 4 sản phẩm. Tính xác suất để trong 4 sản phẩm rút ra có 2 sản phẩm tốt.*



Gọi A là biến cố có 2 sản phẩm tốt trong 4 sản phẩm được rút ra.

Ta có:

- Số trường hợp có thể xảy ra: n = C_{10}^4
- Số trường hợp thuận lợi:
  - ✓ Số trường hợp rút được 2 sản phẩm tốt trong 4 sản phẩm tốt: C_4^2
  - ✓ Số trường hợp rút được 2 sản phẩm xấu trong 6 sản phẩm xấu: C_6^2

  $\Rightarrow$ Số trường hợp thuận lợi của biến cố A: $C_4^2 . C_6^2$

$\Rightarrow$ Xác suất của A: $P(A) = \dfrac{C_4^2 C_6^2}{C_{10}^4} = \dfrac{24}{56} = 0.4286$

* Từ ví dụ trên ta có thể tổng quát thành bài toán lược đồ hộp kín:

Cho một hộp đựng N quả cầu cân đối và giống nhau trong đó có M quả cầu đỏ (M< N) và (N – M) quả cầu trắng.

Lấy ngẫu nhiên (không hoàn lại) p quả cầu (p ≤N) từ trong hộp.

Tính xác suất để trong p quả cầu lấy ra có q (q ≤p) quả cầu đỏ.

Gọi A là biến cố trong p quả cầu lấy ra có q quả cầu đỏ.

$\Rightarrow$ n = C_N^p.

* Số cách lấy q quả cầu đỏ: C_M^q

* Số cách lấy (p – q) quả cầu trắng: C_{N-M}^{p-q}

$\Rightarrow$ m = $C_N^p . C_{N-M}^{p-q} \Rightarrow P(A) = \dfrac{m}{n} = \dfrac{C_M^q.C_{N-M}^{p-q}}{C_N^p}$

Ví dụ 8: *Một nhóm gồm n người. Tính xác suất để có ít nhất hai người có cùng ngày sinh (cùng ngày cùng tháng).*

Gọi S là tập hợp các danh sách ngày sinh có thể của n người và E là biến cố có ít nhất hai người trong nhóm cùng ngày sinh trong năm.

Ta có $\overline{E}$ là biến cố không có hai người bất kỳ trong nhóm có cùng ngày sinh.

Số các trường hợp của S là: n(S) = $\underbrace{365.365.365...365}_{n}$ = 365^n

Số các trường hợp thuận lợi cho $\overline{E}$ là: n($\overline{E}$) = 365.364. . . [365 – (n – 1)]

$$= \frac{[365.364.363...(365-n+1)](365-n)!}{(365-n)!} = \frac{365!}{(365-n)!}$$

Vì các biến cố đồng khả năng nên: P($\overline{E}$) = $\dfrac{n(\overline{E})}{n(S)} = \dfrac{\frac{365!}{(365-n)!}}{365^n} = \dfrac{365!}{(365-n)!365^n}$

Do đó, xác suất để ít nhất hai người có cùng ngày sinh là:

P(E) = 1 - P($\overline{E}$) = 1 - $\dfrac{365!}{(365-n)!365^n}$

Chú ý: *Khi tính xác suất của các biến cố, ta không cần phải chỉ ra các biến cố sơ cấp có thể xảy ra và các biến cố sơ cấp thuận lợi mà chỉ cần chỉ ra số các biến cố sơ cấp có thể xảy ra, số các biến cố sơ cấp thuận lợi cho các biến cố đó.*

Nhận xét: Định nghĩa xác suất theo lối cổ điển có một vài hạn chế như sau:

- Chỉ xét cho hệ hữu hạn các biến cố sơ cấp.
- Không phải lúc nào cũng phân tích được thành tích các biến cố đồng khả năng.

3.2 Định nghĩa xác suất theo lối thống kê: (Bằng tần suất)

Định nghĩa: *Giả sử thực hiện 1 phép thử nào đó n lần độc lập (kết quả của phép thử sau không phụ thuộc vào kết quả của phép thử trước), trong đó biến cố A xảy ra m lần.*

Khi đó: m gọi là tần số xuất hiện của biến cố A.

$f = \dfrac{m}{n}$ *gọi là tần xuất của biến cố A.*

Khi n $\rightarrow \infty$, tần xuất f đạt giá trị ổn định và giá trị đó được xem là xác suất của biến cố A.

Ta có: $P(A) = \lim_{n\to\infty} f = \lim_{n\to\infty} \dfrac{m}{n}$

Ghi chú: Trong thực tế khi số phép thử đủ lớn thì P(A) = f.

Ví dụ 9: *Tiến hành sản xuất thử trên một hệ thống máy thu được kết quả như sau:*

Số sản phẩm n	*100*	*150*	*200*	*250*	*300*	...
Số phế phẩm m	*14*	*12*	*22*	*24*	*32*	...
Tần xuất f	*0,14*	*0,08*	*0,11*	*0,096*	*0,106*	...

Sản xuất một sản phẩm là thực hiện một phép thử. Biến cố A chúng ta quan tâm là sản phẩm trở thành phế phẩm. Như vậy, số sản phẩm sản xuất ra n là số phép thử độc lập, số phế phẩm thu được m là tần số của biến cố A.

Kết quả trên cho thấy khi n tăng dần, tần xuất f thay đổi và đạt tới giá trị ổn định 0,1. Có thể cho rằng, xác suất của biến cố A hay tỉ lệ phế phẩm của hệ thống là 0,1.

Chú ý: *Phương pháp định nghĩa xác suất theo lối thống kê được sử dụng trong thực tế khi liên quan đến số lượng lớn như xác định tỉ lệ phế phẩm của nhà máy, tỉ lệ bắn trúng bia của xạ thủ, tỉ lệ nam (nữ) trong khu vực dân cư lớn.*

Ví dụ 10: *Tung ngẫu nhiên một con xúc xắc .*

Gọi A là biến cố xúc xắc xuất hiện số chấm lẻ.

Goi B là biến cố xúc xắc xuất hiện số chấm: 5, 6.

Khi đó: $P(A) = \frac{3}{6} > P(B) = \frac{2}{6}$

Do đó, biến cố A dễ xảy ra hơn biến cố B. Tuy nhiên cần lưu ý rằng vẫn có trường hợp sự kiện B xảy ra nhưng sự kiện A không xảy ra, đó là trường hợp xúc xắc xuất hiện mặt 6 chấm.

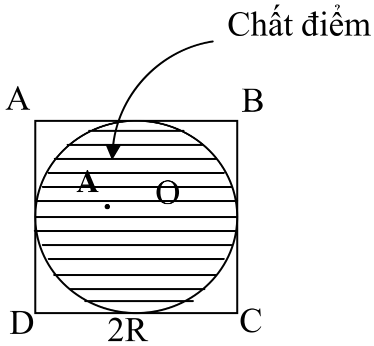
Ví dụ 11: *Một xạ thủ bắn 1000 viên đạn vào bia, trong đó có xấp xỉ 50 viên trúng bia.*

Gọi A là biến cố xạ thủ bắn trúng bia thì xác suất của A là $P(A) = \frac{50}{1000} = 0{,}05$.

3.3 Định nghĩa xác suất theo hình học:

Xét một phép thử có không gian các biến cố sơ cấp là miền hình học W (đoạn thẳng, hình phẳng, khối không gian,…) có số đo (độ dài, diện tích, thể tích…) hữu hạn, khác không. Giả sử xét một điểm rơi ngẫu nhiên vào miền W. Xét miền con A của W. Khi đó xác suất để điểm rơi vào miền A là:

$$P(A) = \frac{\text{Số đo miền A}}{\text{Số đo miền W}}$$



Ví dụ 12: *Ném 1 chất điểm vào trong hình vuông có cạnh dài 2R. Tính xác suất để chất điểm đó rơi vào hình tròn nội tiếp hình vuông.*

Gọi A là biến cố chất điểm rơi vào hình tròn nội tiếp hình vuông .

Trường hợp có thể của phép thử được biểu diễn bằng hình vuông ABCD.

Trường hợp thuận lợi của biến cố A được biểu diễn bằng hình tròn (O,3).

Suy ra : $P(A) = \frac{S_{(O,R)}}{S_{(ABCD)}} = \frac{S_{(O,R)}}{S_{(ABCD)}} = \frac{R^2\pi}{4R^2} = \frac{\pi}{4}$

Ví dụ 13: (Bài toán hai người gặp nhau)

Hai người hẹn gặp nhau ở một địa điểm xác định vào khoảng từ 7 giờ đến 8 giờ. Mỗi người đến (chắc chắn sẽ đến) điểm hẹn trong khoảng thời gian trên một cách độc lập với nhau, chờ trong 20 phút, nếu không thấy người kia sẽ bỏ đi. Tìm xác suất để hai người gặp nhau.

Gọi A là biến cố 2 người gặp nhau trong cuộc hẹn.

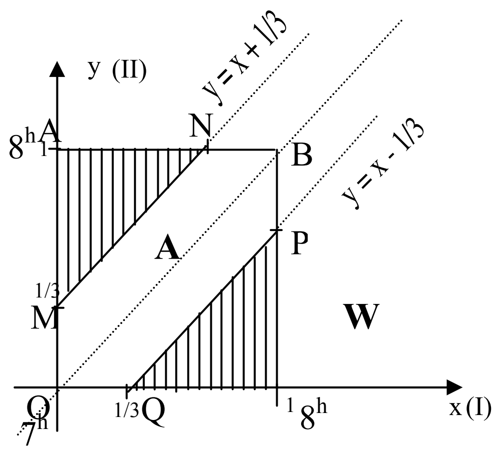
x, y lần lượt là thời gian đến điểm hẹn của người thứ 1 và người thứ 2.

Biểu diễn x, y lên hệ trục tọa độ Descartes. Chọn gốc tọa độ là lúc 7^h.

Trường hợp có thể của phép thử:

$W = \{(x, y): 0 \le x, y \le 1\}$ được biểu diễn bằng hình vuông OABC.

Trường hợp thuận lợi cho biến cố A:



Hình 4

$$|x-y| \le \frac{1}{3} \Leftrightarrow \begin{cases} x - y \le \dfrac{1}{3} \\ x - y \ge -\dfrac{1}{3} \end{cases} \Leftrightarrow \begin{cases} y \ge x - \dfrac{1}{3} \\ y \le x + \dfrac{1}{3} \end{cases}$$

được biểu diễn bằng miền gạch chéo trên hình vẽ: đa giác OMNBPQ.

Suy ra xác suất của A là:

$$P(A) = \frac{S_{(OMNBPQ)}}{S_{(OABC)}} = 1 - 2.\frac{S_{\Delta AMN}}{S_{\Delta ABC}} = 1 - 2.\frac{\frac{1}{2}\frac{2}{3}\frac{2}{3}}{1} = \frac{5}{9}$$

Ghi chú: *Định nghĩa xác suất theo hình học được xem như là sự mở rộng của định nghĩa xác suất theo lối cổ điển trong trường hợp số khả năng có thể xảy ra là vô hạn.*

♥ *Các tính chất của xác suất:*

i) $\forall A \in W: 0 \le P(A) \le 1$

ii) $P(\overline{A}) = 1 - P(A)$

iii) P(∅) = 0, với ∅ là biến cố rỗng.

iv) P(W) = 1, với W là biến cố chắc chắn.

v) Nếu A⊂ B thì P(A) ≤ P(B).

Ý nghĩa: *Xác suất của một biến cố là con số đặt trưng cho khả năng xảy ra ít hay nhiều của biến cố đó. Biến cố có xác suất càng lớn thì càng dễ xảy ra và ngược lại biến cố có xác suất càng nhỏ càng khó xảy ra.*

BÀI TẬP

Xác Suất Theo Lối Cổ Điển

1. Bảng số xe gắn máy gồm có phần chữ và phần số. Phần chữ gồm có 2 chữ được lấy từ 25 chữ La Tinh, phần số gồm có 4 số được lấy từ các số 0, 1, 2, … , 9. Tính xác suất trong các trường hợp sau:

a. Được bảng số xe có phần chữ và phần số khác nhau.

b. Được bảng số xe có chữ A và duy nhất số 5.

c. Có phần chữ giống nhau và phần số giống nhau.

2. Số điện thoại trước đây của mỗi tỉnh (không kể mã số tỉnh) gồm 5 chữ số. Để gia tăng số điện thoại, bưu điện gia tăng mỗi số điện thoại thêm một chữ số.

a. Tính số điện thoại thêm có thể cho việc gia tăng này.

b. Giả sử thành phố có 5 triệu dân, và mỗi người cần một số điện thoại khác nhau. Tính số chữ số tối thiểu cần phải có cho mỗi số điện thoại.

c. Giả sử bạn cần gọi một số điện thoại gồm 6 chữ số khác nhau. Bạn chỉ biết nó có các chữ số 3, 5, 7 nhưng bạn không biết vị trí của nó. Ba chữ số còn lại thì bạn không biết. Tính xác suất để bạn chọn đúng số điện thoại cần gọi.

d. Nếu ở câu (c) bạn biết rõ vị trí của 3 số 3, 5, 7 trong số điện thoại. Tính xác suất để bạn chọn đúng số điện thoại này.

3.

a. Có 10 cuốn sách khác nhau trong đó có 2 cuốn sách Xác Suất, 3 cuốn sách Vật Lý và 5 cuốn sách Toán được xếp vào một kệ sách. Có bao nhiêu cách xếp các cuốn sách đó sao cho các cuốn sách cùng loại thuộc cùng một nhóm.

b. Nếu 10 cuốn sách được xếp ngẫu nhiên vào 5 ngăn. Tính xác suất sao cho:

i/. 10 cuốn sách ở cùng một ngăn.

ii/. 2 cuốn sách Xác Suất ở 2 ngăn khác nhau.

iii/. Chỉ có 2 cuốn sách Xác Suất ở cùng một ngăn.

iv/. Chỉ có 2 cuốn sách Xác Suất ở 2 ngăn khác nhau.

4. Giải vòng loại cúp thế giới khu vực Đông Á gồm 12 đội, trong đó có VIỆT NAM và THÁI LAN được chia làm 3 bảng. Nếu việc chia bảng được thực hiện như sau: Chọn ngẫu nhiên 4 đội xếp vào một bảng nào đó. Sau đó tiếp tục chọn 4 đội xếp vào 1 trong 2 bảng còn lại, 4 đội cuối cùng được xếp vào bảng cuối cùng. Tính xác suất để VIỆT NAM và THÁI LAN chung một bảng.

5. Tung đồng thời 2 con xúc xắc. Tính xác suất trong các trường hợp sau:

a. Tổng số chấm 2 mặt xúc xắc là 9.

b. Trị tuyệt đối hiệu số chấm 2 mặt xúc xắc là 2.

6. Có 12 lọ thuốc trừ sâu được chia làm 6 nhóm (mỗi nhóm 2 lọ). Một nông dân chọn ngẫu nhiên 4 lọ để phun thuốc.

a. Tính xác suất để 4 lọ thuốc đó thuộc 2 nhóm.

b. Tính xác suất để trong 4 lọ thuốc đó chỉ có 2 lọ thuộc một nhóm.

7. Một tổ gồm 8 người tổ chức một buổi tiệc trong đó có 2 người là vợ chồng được xếp ngồi một cách ngẫu nhiên vào 8 cái ghế.

a. Nếu tất cả họ ngồi quanh một chiếc bàn tròn. Tìm xác suất để 2 người là vợ chồng không ngồi gần nhau.

b. Nếu 8 người đó ngồi trên một hàng ghế dài, thì xác suất để 2 vợ chồng ngồi cách nhau một ghế là bao nhiêu?

8. Câu lạc bộ nữ sinh tổ chức 3 hoạt động nhân ngày 8/3: cắm hoa, nấu nướng và may thêu. Một phòng có 10 nữ sinh (trong đó có A và B) đều ghi tên tham gia một hoạt động, ghi một cách ngẫu nhiên (khả năng chọn 3 hoạt động như nhau) và độc lập. Tính xác suất:

a. Cả 10 người ghi tên cắm hoa.

b. Cả 10 người ghi tên một hoạt động.

c. Có 5 người cắm hoa, 3 người nấu nướng và 2 người may thêu.

d. Hai bạn A và B cùng tham gia một hoạt động.

9. Mỗi vé số gồm có 5 chữ số (không kể số thứ tự lô). Khi mua một vé số, nếu bạn trúng 2 số cuối cùng bạn sẽ được thưởng 5 chục ngàn đồng, nếu bạn trúng cả 5 chữ số bạn sẽ được giải đặc biệt, nếu sai chỉ một số nào trong giải đặc biệt bạn sẽ được thưởng an ủi 5 chục ngàn đồng. Khi mua ngẫu nhiên một vé số, tính xác suất để:

a. Bạn trúng giải đặc biệt.

b. Bạn được thưởng 5 chục ngàn đồng.

10. Giả sử một kỹ thuật viên xét nghiệm máu để 10 mẫu máu của 10 người khác nhau trên một cái kệ. Giả sử người đó đưa ngẫu nhiên 10 mẫu máu cho 10 người. Tính xác suất trong các trường hợp sau:

a. Cả 10 mẫu máu đến đúng người nhận.

b. Người thứ nhất nhận đúng mẫu máu của mình.

c. 5 người đầu tiên nhận đúng mẫu máu của mình.

11. Xếp 10 người lên 7 toa tàu một cách ngẫu nhiên. Tìm xác suất để:

a. 10 người cùng lên toa đầu.

b. 10 người cung lên một toa.

c. 5 người đầu mỗi người một toa.

d. Có 2 người A và B lên cùng một toa.

e. Hai người A và B lên cùng một toa ngoài ra không có ai khác trên toa này.

12. Một bộ bài có 52 cây được chia làm 4 loại đều nhau, mỗi loại có một cây At. Chọn ngẫu nhiên 4 cây bài từ bộ bài. Tính xác suất trong các trường hợp sau:

a. 4 cây thuộc 4 loại khác nhau.

b. Tất cả đều là cây At.

c. Có ít nhất một cây At.

Xác Suất Hình Học

13. Một loài thực vật có hoa đực và hoa cái. Người ta nghiên cứu thấy rằng hoa đực và hoa cái nở ngẫu nhiên trong khoảng thời gian từ 1h – 2h. Tuy nhiên chúng chỉ kết hợp tạo thành trái nếu hai loại hoa nở cách nhau không quá 30 phút. Tính xác suất tạo thành trái của loại hoa trên.

14. Gieo ngẫu nhiên một điểm trong vòng tròn bán kính R. Tính xác suất để điểm đó rơi vào:

a. Hình vuông nội tiếp hình tròn.

b. Tam giác đều nội tiếp hình tròn.

15. Một đoạn thẳng có độ dài l được chia làm 3 đoạn bởi 2 điểm chia ngẫu nhiên. Tính xác suất để 3 đoạn đó tạo thành một tam giác.

Bước học 4: ĐƯA RA MỘT SỐ CÔNG THỨC TÍNH XÁC SUẤT

4.1 Các định nghĩa:

Định nghĩa 1*: Các biến cố A_1, A_2, ..., A_n được gọi là biến cố đầy đủ, xung khắc từng đôi nếu chúng xung khắc từng đôi và tổng của chúng là biến cố chắc chắn.*

Có: $A_i A_j = \varnothing$ và $A_1 \cup A_2 \cup . . \cup A_n = W$.

Định nghĩa 2: *Hai biến cố A và B được gọi là độc lập nếu việc xảy ra hay không xảy ra biến cố này không làm ảnh hưởng đến việc xảy ra hay không xảy ra biến cố kia và ngược lại.*

Định nghĩa 3: *Các biến cố A_1, A_2,..., A_n được gọi là độc lập toàn phần nếu mỗi biến cố trong chúng độc lập với tích của một tổ hợp bất kỳ các biến cố còn lại.*

4.2 Công thức cộng*:*

$$P(A+B) = P(A) + P(B) - P(AB),$$ với A và B là hai biến cố bất kỳ.

Tổng quát:

$$P(A_1+A_2+ \ldots +A_n) = \sum_{i=1}^{n} P(A_i) -$$

$$\sum_{i<j} P(A_i)P(A_j) + \sum_{i<j<k} P(A_j)P(A_j)P(A_k) + \ldots + (-1)^{n-1}P(A_1A_2 \ldots A_n)$$

Cụ thể khi n = 3, có:

$$P(A_1+A_2+A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) - P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3)$$

<u>Hệ quả</u>: i) Nếu A, B là hai biến cố xung khắc thì: $P(A+B) = P(A) + P(B)$

ii) Nếu A_1, A_2 , …, A_n là các biến cố xung khắc từng đôi thì:

$$P(A_1+A_2+ .. +A_n) = P(A_1) + P(A_2) + . . +P(A_n)$$

iii) Nếu A_1, A_2 , …, A_n là các biến cố độc lập toàn phần thì:

$$P(A_1+A_2+ . . +A_n) = 1 - P(\overline{A_1}).P(\overline{A_2})...P(\overline{A_n})$$

iv) Nếu A_1, A_2 ,…, A_n là nhóm các biến cố đầy đủ, xung khắc từng đôi thì: $\sum_{1}^{n} P(A_i) = 1$

Ví dụ 1: *Một lô hàng có 10 sản phẩm, trong đó có 2 phế phẩm. Lấy ngẫu nhiên không hoàn lại từ lô hàng ra 6 sản phẩm. Tìm xác suất để có không quá 1 phế phẩm trong 6 sản phẩm được lấy ra.*

Gọi A là biến cố không có phế phẩm trong 6 sản phẩm lấy ra

B là biến cố có đúng một phế phẩm.

C là biến cố có không quá một phế phẩm.

Khi đó A và B là hai biến cố xung khắc và C = A + B

Ta có $$P(A) = \frac{C_8^6}{C_{10}^6} = \frac{28}{210} = \frac{2}{15}$$

$$P(B) = \frac{C_2^1.C_8^5}{C_{10}^6} = \frac{112}{210} = \frac{8}{15}$$

Do đó: $P(C) = P(A) + P(B) = \frac{2}{15} + \frac{8}{15} = \frac{2}{3}$

Ví dụ 2: *Một lớp có 100 sinh viên, trong đó có 40 sinh viên giỏi ngoại ngữ, 30 sinh viên giỏi tin học, 20 sinh viên giỏi cả ngoại ngữ lẫn tin học. Sinh viên nào giỏi ít nhất một trong hai môn sẽ được thêm điểm trong kết quả học tập của học kỳ. Chọn ngẫu nhiên một sinh viên trong lớp. Tìm xác suất để sinh viên đó được thêm điểm.*

Gọi A là biến cố gọi được sinh viên được tăng điểm.

B là biến cố gọi được sinh viên giỏi ngoại ngữ.

C là biến cố gọi được sinh viên giỏi tin học.

Khi đó A = B + C, với B và C là hai biến cố không xung khắc

Ta có: P(A) = P(B + C) = P(B) + P(C) – P(BC)

$$= \frac{30}{100} + \frac{40}{100} - \frac{20}{100} = \frac{50}{100}$$

Ví dụ 3: *Chọn ngẫu nhiên 6 cây bài từ bộ bài có 52 cây bài. Tính xác suất để ít nhất có 2 cây At.*

Gọi A là biến cố chọn ít nhất 2 cây At từ 6 cây bài chọn ra.

A_i là biến cố chọn được i cây At từ 6 cây bài chọn ra $(i = \overline{0,4})$.

Suy ra: $A = A_2 + A_3 + A_4$

Ta có: Hệ các biến cố $\{A_2, A_3, A_4\}$ xung khắc từng đôi, nên:

$$P(A) = P(A_2 + A_3 + A_4) = P(A_2) + P(A_3) + P(A_4)$$

$$= \frac{C_4^2 C_{48}^4}{C_{52}^6} + \frac{C_4^3 C_{48}^3}{C_{52}^6} + \frac{C_4^4 C_{48}^2}{C_{52}^6} \approx 0,06$$

Nhận xét: Trong dãy n biến cố $A_1, A_2, \ldots, A_n$:

+ Nếu từng đôi một các biến cố mà độc lập với nhau thì dãy này gọi là độc lập từng đôi;

+ Nếu dãy độc lập toàn phần thì độc lập từng đôi nhưng điều ngược lại không đúng.

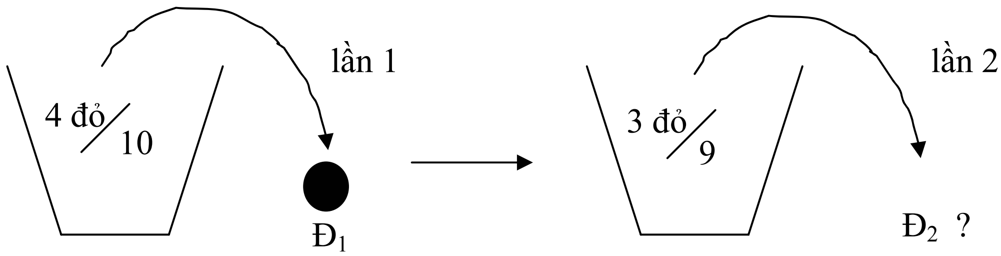
4.3 Công thức nhân xác suất:

4.3.1 Xác suất có điều kiện:

Định nghĩa: *Xác suất của biến cố A với điều kiện biến cố B xảy ra được gọi là xác suất có điều kiện của biến cố A. Ký hiệu P(A/B).*

Ví dụ 4: *Hộp có 10 viên bi trong đó có 4 viên màu đỏ, 6 viên màu trắng. Lần lượt rút không hoàn lại 2 viên bi. Giả sử lần thứ nhất rút được bi màu đỏ, tính xác suất để lần thứ hai rút được bi màu đỏ.*

Gọi A_i là biến cố rút được bi màu đỏ lần thứ i.



Hình 5

Ta có: P(A_2/A_1) = $\frac{3}{9}$

Chú ý: Cho A, B là hai biến cố với P(B) > 0. Ta còn có công thức:

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

Ví dụ 5: *Một bộ bài có 52 lá. Rút ngẫu nhiên 1 lá bài. Tính xác suất để rút được con "át", biết rằng lá bài rút ra là lá bài màu đen.*

Gọi A là biến cố rút được con "át".

B là biến cố rút được lá bài màu đen.

Ta thấy trong bộ bài có 26 lá bài màu đen nên $P(B) = \frac{26}{52} = \frac{1}{2}$

một con át đen nên $P(AB) = \frac{2}{52}$

Do đó ta có: $P(A/B) = \frac{P(AB)}{P(B)} = \frac{2/52}{26/52} = \frac{1}{13}$

Ví dụ 6: *Thi 2 môn, xác suất đậu một thứ nhất là 0,6. Nếu môn thứ nhất đậu thì khả năng sinh viên đó đậu môn thứ hai là 0,8. Nếu môn thứ nhất không đậu thì khả năng sinh viên đó đậu môn thứ 2 chỉ là 0,6. Tính xác suất trong các trường hợp sau:*

a) Sinh viên đó đậu chỉ một môn.

b) Sinh viên đó đậu 2 môn.

Giải

a) Sinh viên đó đậu chỉ một môn:

Gọi A là biến cố sinh viên đó đậu chỉ một môn.

A_i là biến cố sinh viên đó đậu môn thứ i (i =1, 2).

Ta có: $A = A_1\overline{A_2} + \overline{A_1}A_2$

Suy ra: $P(A) = P(A_1\overline{A_2} + \overline{A_1}A_2) = P(A_1\overline{A_2}) + P(\overline{A_1}A_2)$

$= P(A_1)P(\overline{A_2}/A_1) + P(\overline{A_1})P(A_2/\overline{A_1})$ = (0,6.(0,2) + (0,4).(0,6) = 0,36

b) Sinh viên đó đậu 2 môn:

Gọi B là biến cố sinh viên đậu hai môn.

Ta có: $B = A_1A_2$

Suy ra: $P(B) = P(A_1A_2) = P(A_1)P(A_2/A_1) = (0,6).(0,8) = 0,48$

4.3.2 Công thức nhân xác suất:

Cho A và B là hai biến cố bất kỳ của một phép thử. Ta luôn có:

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

- Nếu A và B độc lập, có: $P(AB) = P(A) \cdot P(B)$
- Mở rộng: $P(A_1.A_2\ldots\ldots A_n) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1A_2) \ldots P(A_n/A_1A_{n-1})$
- Nhóm các biến cố độc lập toàn phần: $A_1, A_2, \ldots, A_n$ được gọi là độc lập toàn phần khi và chỉ khi: $P(A_1A_2\ldots A_n) = P(A_1). P(A_2)\ldots P(A_n)$

Ví dụ 7: *Tung đồng thời hai con xúc xắc. Tính xác suất để cả 2 con xúc xắc đều xuất hiện mặt 6 chấm.*

Gọi A là biến cố cả hai xúc xắc đều xuất hiện mặt 6 chấm.

A_i là biến cố xúc xắc thứ i xuất hiện mặt 6 chấm (i = 1, 2)

Ta có: A= A_1A_2

Do A_1 và A_2 độc lập nhau, nên: $P(A) = P(A_1A_2) = P(A_1)P(A_2) = \frac{1}{6}\frac{1}{6} = \frac{1}{36}$

Ví dụ 8: *Hai xạ thủ mỗi người bắn một phát đạn vào bia. Xác suất bắn trúng của người thứ nhất là p = 0,9; của người thứ hai là p = 0,7. Tính xác suất:*

a) Cả hai đều bắn trúng.

b) Có đúng một viên đạn trúng bia.

c) Bia bị trúng đạn.

Biết rằng hai người bắn độc lập với nhau.

Gọi A là biến cố xạ thủ I bắn trúng bia.

B là biến cố xạ thủ II bắn trúng bia.

C là biến cố cả hai xạ thủ trúng bia.

D là biến cố có một viên đạn trúng bia.

E là biến cố bia bị trúng đạn.

a) Xác suất để cả hai đều bắn trúng: Ta có C = AB

$P(C) = P(AB) = P(A) \cdot P(B) = 0,9 \cdot 0,7 = 0,63$

b) Xác suất để có một viên đạn trúng bia:

Ta có: $D = \overline{A}B + \overline{B}A$

Vì $\overline{A}B$ và $\overline{B}A$ là xung khắc với nhau

$\Rightarrow P(D) = P(\overline{A}B) + P(\overline{B}A) = P(\overline{A}).P(B) + P(A).P(\overline{B})$

$\Rightarrow P(D) = 0,9 \cdot 0,3 + 0,1 \cdot 0,7 = 0,34$

c.) Xác suất để bia bị trúng đạn:

Ta có: $\overline{E} = \overline{A}\overline{B} \Rightarrow$

$$P(\overline{E}) = P(\overline{A}\overline{B}) = P(\overline{A}).P(\overline{B}) = 0{,}3.0{,}1 = 0{,}03$$

P(E) = 1 – 0,03 = 0,97

Bước học 5: CÔNG THỨC XÁC SUẤT ĐẦY ĐỦ VÀ CÔNG THỨC BAYES
5.1 Công thức xác suất đầy đủ:

Định nghĩa: *Giả sử $A_1, A_2, \ldots, A_n$ là nhóm biến cố đầy đủ xung khắc từng đôi và B là biến cố bất kỳ có thể xảy ra đồng thời với một trong các biến cố A_i (i= 1, .. , n). Khi đó xác suất B được tính bởi công thức:*

$$P(B) = \sum_{i=1}^{n} P(A_i).P(B/A_i)$$

Khi B xảy ra thì có một và chỉ một biến cố A_i cùng xảy ra với B.

Chú ý: Vận dụng công thức xác suất đầy đủ để giải một bài toán, vấn đề quan trọng là phải chỉ ra được nhóm biến cố đầy đủ và xung khắc từng đôi. Trong thực tế việc này thường gặp ở 2 hình thức sau:

- ✓ Công việc tiến hành trải qua 2 phép thử. Thực hiện phép thử thứ nhất ta có một trong n khả năng xảy ra là các biến cố: $A_1, A_2, ..., A_n$. Sau khi thực hiện phép thử thứ nhất ta thực hiện phép thử thứ hai. Trong phép thử thứ hai ta quan tâm đến biến cố B. Khi đó biến cố B sẽ được tính theo công thức xác suất toàn phần với nhóm biến cố đầy đủ và xung khắc từng đôi là các biến cố $A_i \ (i = \overline{1,n})$.

- ✓ Một tập hợp chứa n nhóm phần tử. Mỗi nhóm phần tử có một tỉ lệ phần tử có tính chất P nào đó. Lấy ngẫu nhiên từ tập hợp ra 1 phần tử. Gọi A là biến cố chọn được phần tử thuộc nhóm thứ i. Khi đó xác suất của biến cố chọn được phần tử có tính chất P trong phép thử sẽ được tính theo công thức xác suất toàn phần với nhóm biến cố đầy đủ và xung khắc từng đôi là $A_i \ (i = \overline{1,n})$.

Ví dụ 1: *Xét một lô sản phẩm, trong đó có sản phẩm của nhà máy 1 sản phẩm chiếm 20%, nhà máy 2 sản phẩm chiếm 30%, nhà máy 3 sản phẩm chiếm 50%. Xác suất phế phẩm của nhà máy 1, 2, 3 lần lượt là 0,001; 0,005; 0,006. Lấy ngẫu nhiên 1 sản phẩm từ lô hàng. Tính xác suất để sản phẩm lấy ra là phế phẩm.*

Gọi B là biến cố lấy được sản phẩm là phế phẩm.

A_1, A_2, A_3 lần lượt là biến cố lấy được sản phẩm của nhà máy 1, 2, 3.

Ta có: A_1, A_2, A_3 là nhóm biến cố đầy đủ, xung khắc từng đôi. Theo công thức xác suất đầy đủ, ta có:

$$P(B) = \sum_{i=1}^{3} P(Ai).P(B/Ai) = P(A_1) . P(B/A_1) + P(A_2) . P(B/A_2) + P(A_3) . P(B/A_3)$$

$$= 20/100 . 0{,}001 + 30/100 . 0{,}005 + 50/100 . 0{,}006 = 0{,}0065.$$

5.2 Công thức Bayes:

Từ giả thuyết, để tính xác suất đầy đủ, nếu B xảy ra thì xác suất biến cố A_i bằng bao nhiêu?

Định nghĩa: *Giả sử $A_1, A_2, .., A_n$ là nhóm biến cố đầy đủ xung khắc từng đôi và B là biến cố bất kỳ có thể xảy ra đồng thời với một trong các biến cố A_i. Khi đó ta có công thức:*

$$P(A_i / B) = \frac{P(A_i).P(B / A_i)}{P(B)} \qquad \text{(Công thức Bayes)}$$

Với $P(B) = \sum_{i=1}^{n} P(A_i).P(B / A_i)$

Ví dụ 2: *Một phân xưởng sản xuất chi tiết máy có hai máy: Máy I sản xuất 60% sản phẩm của phân xưởng; Máy II sản xuất 40% sản phẩm của phân xưởng. Tỉ lệ phế phẩm của máy I là 0,1 và tỉ lệ phế phẩm của máy II là 0,05. Sản phẩm của phân xưởng sau khi sản xuất được đem trộn lẫn với nhau. Lấy ngẫu nhiên một sản phẩm của phân xưởng thì thấy sản phẩm đó là phế phẩm. Tính xác suất để phế phẩm lấy ra do máy I sản xuất.*

Gọi B_1 là biến cố sản phẩm lấy ra do máy I sản xuất.

B_2 là biến cố sản phẩm lấy ra do máy II sản xuất.

A là biến cố sản phẩm lấy ra là phế phẩm.

$\Rightarrow B_1, B_2$ lập thành nhóm đầy đủ các biến cố.

Theo công thức xác suất toàn phần: $P(A) = P(B_1).P(A/B_1)+P(B_1).P(A/B_2) = 0,08$.

Theo công thức Bayer: $P(B_1 / A) = \frac{P(B_1).P(A.B_1)}{P(A)} = \frac{0,6.0,1}{0,08} = 0,75$.

Vậy xác suất để phế phẩm do máy I sản xuất là $P(B_1/A) = 0,75$.

Ví dụ 3: *Có 3 hộp đựng sản phẩm, mỗi hộp có 10 sản phẩm, trong đó số phế phẩm lần lượt là 2, 3, 4. Chọn ngẫu nhiên một hộp, rồi từ hộp đã chọn, rút ra ngẫu nhiên một sản phẩm.*

a) Tính xác suất để sản phẩm chọn ra là phế phẩm.

b) Nếu sản phẩm rút ra là phế phẩm, thì theo bạn phế phẩm đó có khả năng thuộc hộp nào nhiều nhất, tại sao?

Giải

a) Tính xác suất để sản phẩm chọn ra là phế phẩm:

Gọi B là biến cố rút được sản phẩm là phế phẩm.

A_i là biến cố chọn được hộp thứ i ($i = \overline{1,3}$).

Theo công thức xác suất toàn phần, ta có:

$$P(B) = P(A_1)P(B / A_1) + P(A_2)P(B / A_2) + P(A_3)P(B / A_3)$$

$$= \frac{1}{3}\frac{2}{10} + \frac{1}{3}\frac{3}{10} + \frac{1}{3}\frac{4}{10} = \frac{1}{3}\frac{9}{10} = \frac{3}{10} = 0,3$$

b) Theo công thức Bayes, ta có:

$$P(A_1/B) = \frac{P(A_1)P(B/A_1)}{P(B)} = \frac{\frac{1}{3}\frac{2}{10}}{\frac{3}{10}} = \frac{2}{9}$$

$$P(A_2/B) = \frac{P(A_2)P(B/A_2)}{P(B)} = \frac{\frac{1}{3}\frac{3}{10}}{\frac{3}{10}} = \frac{1}{3} = \frac{3}{9}$$

$$P(A_3/B) = \frac{P(A_3)P(B/A_3)}{P(B)} = \frac{\frac{1}{3}\frac{4}{10}}{\frac{3}{10}} = \frac{4}{9}$$

So sánh các kết quả, ta thấy phế phẩm rút ra có khả năng thuộc hộp thứ III nhiều nhất.

5.3 Công thức Bernoulli:

Định nghĩa: *Ta tiến hành n phép thử độc lập. Giả sử trong mỗi phép thử chỉ xảy ra hai trường hợp: Hoặc biến cố A xảy ra với xác suất p hoặc biến cố A không xảy ra với xác suất q = 1 – p.*

Các bài toán thỏa mãn các điều kiện trên thì được gọi là tuân theo lược đồ Bernoulli. Khi đó xác suất để trong n phép thử độc lập biến cố A xuất hiện k lần được ký hiệu: $P_n(k)$ và được tính $P_n(k) = C_n^k.p^k.q^{n-k}$, công thức này mang tên là **công thức Bernoulli**.

Ví dụ 4: *Hộp có 10 viên bi, trong đó có 6 viên bi màu đỏ. Lần lượt rút có hoàn lại 5 viên bi. Gọi A là biến cố rút được viên bi màu đỏ trong mỗi lần rút, ta được một lược đồ Bernoulli với:*

* Số phép thử độc lập: n = 5.
* P(A) = 6/15.

Ví dụ 5: *Trong một phân xưởng có 5 máy hoạt động độc lập, xác suất để một máy bị hư trong một ca sản xuất là bằng nhau và bằng p = 0,1. Tính xác suất để trong 1 ca có hai máy bị hư.*

Ta thấy 5 máy hoạt động độc lập cho nên ta có thể coi như tiến hành 5 phép thử độc lập và mỗi phép thử chỉ có hai kết cục máy hoạt động tốt hoặc máy bị hư với xác suất p = 0,1.

$\Rightarrow$ bài toán tuân theo lược đồ Bernoulli.

Do đó xác suất để trong một ca có hai máy bị hư. $P_5(2) = C_5^2.(0,1)^2.(0,9)^3$

Ví dụ 6: *Một sinh viên thi trắc nghiệm môn Ngoại Ngữ gồm có 10 câu hỏi. Mỗi câu có 4 phần để lựa chọn trả lời, trong đó chỉ có 1 phần đúng. Giả sử sinh viên làm bài bằng cách chọn ngẫu nhiên các phần của câu hỏi. Tính xác suất trong các trường hợp sau:*

a) *Sinh viên vừa đủ điểm đậu (5 điểm).*

b) *Sinh viên chọn đúng ít nhất 1 câu hỏi.*

Giải

a) Sinh viên vừa đủ điểm đậu:

Gọi A là biến cố sinh viên vừa đủ điểm đậu.

Xem việc chọn câu trả lời ở mỗi câu hỏi của sinh viên là 1 phép thử thì trong mỗi phép thử có 1 trong 2 khả năng xảy ra :

- ✓ Sinh viên trả lời đúng với xác suất là p = $\frac{1}{4}$.
- ✓ Sinh viên trả lời sai với xác suất là q = $\frac{3}{4}$.

Vậy: $P(A) = P(10,5) = C_{10}^{5}(\frac{1}{4})^5(\frac{3}{4})^5 \approx 0{,}058$

b) Sinh viên chọn đúng ít nhất 1 câu hỏi:

Gọi B là biến cố sinh viên chọn đúng ít nhất 1 câu hỏi.

$\Rightarrow \overline{B}$ là biến cố sinh viên không chọn đúnh câu hỏi nào.

Ta có: $P(\overline{B}) = P(10,0) = C_{10}^{0}(\frac{1}{4})^0(\frac{3}{4})^{10} = (\frac{3}{4})^{10}$

$$\Rightarrow P(B) = 1 - P(\overline{B}) = 1 - (\frac{3}{4})^{10} = 0{,}056$$

Ví dụ 7: *Một bác sĩ có xác suất chữa khỏi bệnh là 0,8. Có người nói rằng cứ 10 người đến chữa bệnh thì chắc chắn có 8 người khỏi bệnh. Điều khẳng định đó có đúng không?*

Điều khẳng định trên là sai. Ta có thể xem việc chữa bệnh cho 10 người là một dãy của một phép thử độc lập. Nếu gọi A là biến cố chữa khỏi bệnh cho một người thì P(A) = 0,8

Do đó: Xác suất để trong 10 người đến chữa bệnh thì có 8 người khỏi bệnh là:

$$P_{10}(8) = C_{10}^{8}.(0{,}8)^8.(0{,}2)^2 \approx 0{,}3108.$$

5.4 Công Thức Bernoulli Mở Rộng:

5.4.1 Lược đồ Bernoulli mở rộng:

Định nghĩa: Một lược đồ Bernoulli mở rộng gồm :

- ✓ Dãy n phép thử độc lập.
- ✓ Hệ biến cố $\{A_1, A_2, ..., A_k\}$ đầy đủ, xung khắc.

Trong đó: $P(A_1) = p_1, P(A_2) = p_2, ..., P(A_k) = p_k$ và $p_1 + p_2 + ... + p_k = 1$.

5.4.2 Công thức Bernoulli mở rộng:

Công thức: *Xác suất để trong n phép thử độc lập, biến cố A_1 xảy ra m_1 lần, biến cố A_2 xảy ra m_2 lần , ..., biến cố A_k xảy ra m_k lần (trong đó $m_1 + m_2 + ... + m_k = n$) là:*

$$P(n; m_1, m_2, ..., m_k) = \frac{n!}{m_1!m_2!...m_k!} p_1^{m_1}.p_2^{m_2}...p_k^{m_k}$$

Ví dụ 8: *Lô hàng có 100 sản phẩm trong đó có 30 sản phẩm loại A, 50 sản phẩm loại B và 20 sản phẩm loại C. Lần lượt rút có hoàn lại 9 sản phẩm để kiểm tra. Tính xác suất để trong 9 lần rút đó có 3 lần rút được sản phẩm loại A, 4 lần rút được sản phẩm loại B và 2 lần rút được sản phẩm loại C.*

Gọi A, B, C lần lượt là các biến cố rút được sản phẩm loại A, B, C trong mỗi lần rút.

Rõ ràng hệ $\{A, B, C\}$ đầy đủ và xung khắc từng đôi.

Và $P(A) = \frac{30}{100}$, $P(B) = \frac{50}{100}$, $P(A) = \frac{20}{100}$

Do đó: $\text{P}(9;3\text{A},4\text{B},2\text{C}) = \frac{9!}{3!4!2!}(\frac{30}{100})^3(\frac{50}{100})^4(\frac{20}{100})^2 = 0{,}086$

BÀI TẬP

1. Một tổ gồm có 8 nam và 6 nữ. Chọn ngẫu nhiên một nhóm 5 người. Tính xác suất để trong nhóm:

a. Có ít nhất một nữ.

b. Số nữ nhiều hơn số nam.

2. Ở một hội đồng nhân dân tỉnh có 20 đại biểu trong đó có một người nữ. Để điều hành một công việc nào đó cần thành lập một tiểu ban gồm 5 người. Tính xác suất sao cho tiểu ban đó có số lượng nam nhiều hơn số lượng nữ khi chọn ngẫu nhiên các đại biểu.

3. Một lớp có 30 học sinh, gồm: 10 học sinh giỏi toán, 10 học sinh giỏi văn, 10 học sinh giỏi ngoại ngữ. Trong đó có 5 học sinh vừa giỏi ngoại ngữ và toán, 3 học sinh vừa giỏi ngoại ngữ và văn, không có học sinh nào giỏi văn và toán hoặc giỏi cả 3 môn. Chọn ngẫu nhiên một học sinh, tính xác suất để được học sinh giỏi ít nhất 1 trong 3 môn nói trên.

4. Bắn liên tiếp vào một mục tiêu cho đến khi viên đạn đầu tiên trúng mục tiêu hoặc hết đạn thì ngưng. Xác suất bắn trúng mục tiêu của mỗi lần bắn là 0,6.

a. Nếu người đó có 4 viên đạn. Tính xác suất để bắn đến viên đạn thứ 4.

b. Nếu người đó có số viên đạn không hạn chế. Tính xác suất để việc bắn ngưng lại ở lần thứ tư.

5. Một lô hàng gồm 10 sản phẩm trong đó có lẫn lộn 1 phế phẩm. Người ta lấy lần lượt từng sản phẩm từ lô hàng để tìm phế phẩm đó.

a. Tìm xác suất sao cho phế phẩm đó lấy ra ở lần sau cùng.

b. Giả sử lô hàng có 2 phế phẩm. Người ta lấy lần lượt từng sản phẩm cho đến khi phát hiện hết 2 phế phẩm thì dừng. Tính xác suất sao cho việc kiểm tra dừng lại ở lần kiểm tra thứ 4.

6. Một sinh viên thi vào trường ngoại ngữ phải thi 5 môn với xác suất đậu của mỗi môn tương ứng là: 0,7; 0,6; 0,4; 0,8; 0,5. Tìm xác suất để sinh viên đó:

a. Đậu cả 5 môn.

b. Đậu ít nhất 1 môn.

c. Đậu nhiều nhất 1 môn.

7. Một trận không chiến giữa máy bay ta và máy bay địch. Máy bay ta đã bắn trước với xác suất trúng là 0,5. Nếu bị trượt máy bay địch bắn trả lại với xác suất trúng là 0,4. Nếu không bị trúng đạn máy bay ta lại bắn trả lại với xác suất trúng là 0,3. Trận không chiến đến đây kết thúc, và máy bay sẽ bị rơi nếu như bị trúng. Tìm xác suất:

a. Máy bay địch bị rơi trong cuộc không chiến trên.

b. Máy bay ta bị rơi trong cuộc không chiến.

8. Trong một kỳ thi mỗi sinh viên phải thi 2 môn. Giả sử bạn ước lượng rằng: Bạn có hy vọng đậu 80% môn thứ nhất. Nếu đạt môn thứ nhất, điều này làm bạn phấn khởi và do bạn phấn khởi sẽ có hy vọng 60% đạt yêu cầu môn thứ hai. Nếu không đạt môn thứ nhất, điều này làm bạn nản lòng làm cho hy vọng đạt môn thứ hai chỉ còn 30%. Hãy tìm xác suất để bạn:

a. Đạt cả hai môn.

b. Đạt môn thứ hai.

c. Đạt ít nhất một môn.

d. Không đạt cả hai môn.

9. Nếu dùng 3 loại thuốc A, B, C riêng lẻ để điều trị bệnh phổi thì tỉ lệ kháng thuốc theo thứ tự là: 15%, 20%, 25%. Dùng phối hợp cả 3 loại thuốc trên thì khả năng kháng thuốc của vi trùng là bao nhiêu.

10. Chọn ngẫu nhiên một vé số có 5 chữ số. Tính xác suất để được vé số không có số 1 hoặc không có số 5.

11. Chọn ngẫu nhiên một vé số có 5 chữ số. Tính xác suất để được vé số có số 5 và số chẵn.

12. Một người bỏ ngẫu nhiên 3 lá thư vào 3 phong bì đã ghi địa chỉ. Tính xác suất để có ít nhất một lá thư bỏ đúng phong bì của nó.

13. Trong một hộp đựng 30 ấm trà, trong đó có 7 ấm bị sứt vòi, 5 ấm bị mẻ miệng, 6 ấm bị bể nắp, 3 ấm vừa sứt vòi vừa bể nắp, 2 ấm vừa sứt vòi vừa mẻ miệng, 1 ấm vừa sứt vừa bể nắp vừa mẻ miệng.

a. Lấy ngẫu nhiên một ấm từ hộp. Tính xác suất để ấm ấy có nhược điểm.

b. Tìm xác suất để lấy ra một ấm sẽ là ấm bị sứt vòi khi nó đã bị bể nắp.

c. Lấy ngẫu nhiên ra 4 ấm. Tính xác suất để trong 4 ấm này có 2 ấm có nhược điểm.

14. Biết rằng một người có nhóm máu AB có thể nhận máu bất kỳ nhóm máu nào. Nếu người nào đó có nhóm máu còn lại (A hoặc B hoặc O) thì chỉ nhận máu của người cùng nhóm với mình hoặc người có nhóm máu O. Cho biết tỉ lệ người có nhóm máu A, B, O và AB tương ứng là: 33,7%; 37,5%; 20,9%; 7,9%.

a. Chọn ngẫu nhiên một người cần tiếp máu và một người cho máu. Tính xác suất để sự truyền máu thực hiện được.

b. Chọn ngẫu nhiên một người cần tiếp máu và hai người cho máu. Tính xác suất để sự truyền máu thực hiện được.

15. Có 2 lô sản phẩm. Mỗi lô có 10 sản phẩm, trong đó số lượng phế phẩm của mỗi lô lần lượt là: 2 và 3.

a. Lấy ngẫu nhiên mỗi lô một sản phẩm.

b. Lấy ngẫu nhiên một lô, rồi từ lô đó lấy ra 2 sản phẩm.

Hãy đánh giá xem phương thức nào chọn được một phế phẩm lớn hơn.

16. Một người có 3 con gà mái và 2 con gà trống nhốt trong chuồng. Một người đến mua, người bán gà bắt ngẫu nhiên ra một con. Người mua chấp nhận mua con đó.

a. Tìm xác suất để bắt được gà trống.

b. Người thứ 2 đến mua, người bán bắt ra ngẫu nhiên một con. Tính xác suất để được gà mái.

c. Xác suất này sẽ bằng bao nhiêu nếu người thứ hai đến mua, biết rằng người bán gà quên mất đã bán cho người thứ nhất là gà trống hay gà mái.

17. Một tổ sinh viên gồm có 4 người nam và 6 người nữ. Giả sử tổ được Đoàn trường cho 3 vé xem phim.

a. Có bao nhiêu cách phân phối sao cho nữ có 2 vé và nam có 1 vé.

b. Nếu việc phân phối thực hiện bằng cách bốc thăm ngẫu nhiên mỗi người lần lượt lấy một vé từ 10 vé, trong đó có 3 vé có dấu hiệu đặc biệt mà người bốc trúng sẽ được xem phim. Theo bạn nên chọn việc bốc thăm lần thứ mấy để có lợi nhất, tại sao?

18. Một hộp có 3 bi trắng và 5 bi đỏ.

a. Lấy 2 bi không chú ý màu của nó, rồi bỏ vào hộp 2 bi trái màu với nó. Sau đó lấy tiếp một bi. Tính xác suất để bi lấy ra lần sau là đỏ.

b. Lấy ra lần đầu một bi, sau đó lấy tiếp một bi nữa. Tính xác suất để 2 bi này cùng màu.

19. Có 3 lô hàng 1, 2, 3 theo thứ tự có tỉ lệ phế phẩm là: 3/10, 6/15, 4/20. Chọn ngẫu nhiên một lô hàng, rồi từ đó lấy tiếp ra một sản phẩm.

a. Tính xác suất để sản phẩm lấy ra là phế phẩm.

b. Giả sử sản phẩm lấy ra là chính phẩm, nó có thể là của hộp nào nhiều nhất, tại sao?

20. Một nhóm gồm có 10 người, trong đó có 6 người có nhóm máu O. Chọn ngẫu nhiên 3 người, rồi từ nhóm 3 người chọn ngẫu nhiên một người.

a. Tính xác suất để chọn được người có nhóm máu O.

b. Giả sử chọn được người có nhóm máu O. Tính xác suất để 3 người chọn ra trước đó có 2 người có nhóm máu O.

21. Có 4 chiến sĩ độc lập bắn vào một chiếc xe, mỗi người bắn một viên với xác suất trúng là: 0,8; 0,4; 0,6; 0,5. Biết rằng có k viên đạn bắn trúng xe thì xe bị tiêu diệt với xác suất là: $p_k = 1 - \frac{1}{2^k}$. Tìm xác suất để xe bị tiêu diệt.

22. Có 2 hộp: Hộp 1 có 3 bi đỏ và 7 bi trắng.

Hộp 2 có 6 bi đỏ và 4 bi trắng.

a. Lấy 2 viên bi từ hộp 1 bỏ vào hộp 2, sau đó rút lần lượt hộp 2 ra 2 viên. Tính xác suất để 2 viên này đều trắng.

b. Lấy mỗi hộp 2 viên. Tính xác suất để được 3 viên trắng.

c. Nếu lấy được 3 viên trắng, 1 viên đen ở câu (b). Tính xác suất để viên bi đen là của hộp 2.

23. Một công ty bảo hiểm cho người bị tai nạn. Công ty chia khách hàng của mình ra thành 3 nhóm: Người ít bị rủi ro, người bị rủi ro trung bình và người thường xuyên bị rủi ro với tỉ lệ là: 60%, 30%, 10%. Xác suất bị rủi ro của các nhóm lần lượt là: 0,01; 0,05; 0,1.

a. Tính tỉ lệ người bị tai nạn trong năm.

b. Nếu người không bị tai nạn trong năm, họ có khả năng thuộc nhóm nào nhiều nhất, tại sao?

24. Một hộp đựng 3 đồng xu trong đó có 1 đồng xu thiên vị ngữa (luôn lật mặt ngữa khi tung) và 2 đồng xu công bằng. Chọn ngẫu nhiên một đồng xu trong hộp rồi tung. Nếu ngữa thì tung tiếp đồng xu đó một lần nửa. Nếu sấp thì rút một đồng xu khác trong hộp và tung.

a. Tìm xác suất để 2 lần tung đều xuất hiện mặt ngữa.

b. Nếu một đồng xu được tung 2 lần. Tìm xác suất để đó là đồng xu thiên vị ngữa.

25. Hai nhà máy cùng sản xuất ra một loại chi tiết. Năng suất của máy I gấp đôi máy II. Tỉ lệ chi tiết đạt tiêu chuẩn của máy I là 64%, của máy II là 80%. Lấy ngẫu nhiên một chi tiết từ lô hàng do 2 máy sản xuất thì được chi tiết đạt tiêu chuẩn. Tính xác suất để chi tiết đó do máy I sản xuất.

26. Hộp A: có 15 lọ thuốc tốt, 5 lọ thuốc hỏng.

Hộp B: có 17 lọ thuốc tốt, 3 lọ thuốc hỏng.

Hộp C: có 10 lọ thuốc tốt, 10 lọ thuốc hỏng.

a. Lấy ở mỗi hộp 1 lọ. Tính xác suất để có một lọ thuốc hỏng.

b. Chọn ngẫu nhiên 1 hộp, rồi từ hộp đã chọn lấy ra 3 lọ. Tính xác suất được 2 lọ tốt và 1 lọ hỏng.

c. Trộn chung 3 hộp lại, rồi từ đó lấy ra 3 lọ. Tính xác suất để được 3 lọ thuốc tốt.

d. Kiểm tra từng lọ ở hộp B cho đến khi phát hiện đủ 3 lọ thuốc hỏng thì dừng. Tính xác suất để việc kiểm tra dừng lại ở lần lấy thứ 5.

27. Tỉ lệ lọ thuốc hỏng trong các lô thuốc A, B lần lượt là: 0,1; 0,07. Giả sử các lô thuốc này có rất nhiều lọ.

a. Lấy ngẫu nhiên 2 lọ ở mỗi lô thuốc. Tính xác suất để có một lọ thuốc hỏng.

b. Chọn ngẫu nhiên 1 trong 2 lô, rồi từ đó lấy ra 4 lọ. Tính xác suất để được 1 lọ thuốc hỏng.

c. Cửa hàng nhận 600 lọ thuốc ở lô thứ nhất và 400 lọ thuốc ở lô thứ hai. Ta mua ngẫu nhiên 1 lọ. Tính xác suất để lọ này là lọ hỏng.

28. Ở hội chợ có 3 cửa hàng. Cửa hàng loại I phục vụ những người "may mắn" bán hàng với tỉ lệ phế phẩm là 1%. Cửa hàng loại II phục vụ bán hàng với tỉ lệ phế phẩm là 5%. Cửa hàng loại III phục vụ những người "rủi ro" bán hàng với tỉ lệ phế phẩm là 10%. Một người vào hội chợ phải gieo 2 đồng xu. Người đó là may mắn nếu cả 2 đều sấp, là rủi ro nếu cả 2 đều ngữa.

a. Tính xác suất để một người vào hội chợ mua phải hàng xấu.

b. Nếu một người mua phải hàng xấu, theo ý bạn người đó là may mắn hay rủi ro.

a. Một bệnh nhân nghi là có thể mắc một trong 3 bệnh A, B, C với xác suất tương ứng là: 0,3; 0,4 và 0,3. Người đó đến khám bệnh ở 4 bác sĩ một cách độc lập. Bác sĩ thứ nhất chuẩn đoán bệnh A, bác sĩ thứ hai chuẩn đoán bệnh B, bác sĩ thứ ba chuẩn đoán bệnh C và bác sĩ thứ tư chuẩn đoán bệnh A. Hỏi khi khám bệnh xong, người bệnh đánh giá lại xác suất mắc bệnh A, B, C của mình là bao nhiêu. Biết rằng xác suất chuẩn đoán đúng của mỗi ông bác sĩ là

29. Một loại sản phẩm được gia công qua 3 giai đoạn độc lập với nhau, với tỉ lệ khuyết tật của mỗi công đoạn theo thứ tự là: 5%, 4%, 2%. Nếu sản phẩm bị khuyết tật ở 3 công đoạn thì nó trở thành phế phẩm. Nếu sản phẩm bị khuyết tật ở 2 công đoạn thì nó trở thành phế phẩm với tỉ lệ 50%. Nếu sản phẩm bị khuyết tật ở 1 công đoạn thì nó trở thành phế phẩm với tỉ lệ 30%. Tính tỉ lệ phế phẩm của nhà máy đó.

30. Một lô hàng gồm 5 sản phẩm không rõ chất lượng cụ thể. Lấy ngẫu nhiên 2 sản phẩm từ lô hàng thì được cả 2 chính phẩm.

a. Nếu lấy tiếp 1 sản phẩm nữa từ lô hàng theo ý bạn sẽ được chính phẩm hay phế phẩm, tại sao?

b. Theo ý bạn khả năng số sản phẩm tốt trong hộp có khả năng nhất là bao nhiêu trong 3 sản phẩm còn lại, tại sao?

31. Một cuộc thi có 3 vòng. Vòng 1 lấy 90% thí sinh. Vòng 2 lấy 80% thí sinh của vòng 1 và vòng 3 lấy 90% thí sinh của vòng 2.

a. Tính xác suất để thí sinh lọt qua 3 vòng thi.

b. Tính xác suất để thí sinh đó bị loại ở vòng 2 nếu biết rằng thí sinh đó bị loại.

32. Tung một con xúc xắc liên tục cho đến khi mặt 6 chấm xuất hiện 4 lần thì ngưng. Tính xác suất sao cho việc tung xúc xắc ngưng ở lần thứ 6.

33. Một sinh viên thi trắc nghiệm môn Vật Lý gồm 10 câu hỏi. Mỗi câu gồm có 4 phần để chọn. Giả sử sinh viên đó chỉ biết rõ 3 câu hỏi, còn lại thì chọn một cách ngẫu nhiên.

a. Tính xác suất để sinh viên đó chọn đúng tất cả những câu hỏi trên.

b. Nếu chọn đúng từ phân nửa trở đi sinh viên đó sẽ đậu. Tính xác suất để sinh viên đó đậu.

34. Phải tung xúc xắc ít nhất bao nhiêu lần để có ít nhất một lần nhận mặt 4 chấm không bé hơn 0,95.

35. Theo kết quả điều tra, tỉ lệ bệnh lao ở một vùng là: 0,001. Tính xác suất để khi khám cho 10 người:

a. Không ai bệnh lao.

b. 5 người bệnh lao.

c. Có ít nhất 1 người bệnh lao.

36. Một cầu thủ có tiếng về đá phạt đền. Xác suất cho banh vào lưới của cầu thủ đó trong mỗi lần đá là 0,8. Một người nói cầu thủ đó cứ đá 10 lần đá chắc chắn có 8 lần bóng vào lưới, điều đó đúng hay sai? Tại sao?

37. Một sọt cam rất lớn được phân loại theo cách sau: Chọn ngẫu nhiên 20 quả cam làm mẫu đại diện. Nếu mẫu không có quả cam nào bọ hỏng thì sọt cam được xếp loại I. Nếu mẫu có 1 hoặc 2 quả cam bị hỏng thì sọt cam được xếp loại II. Trong trường hợp còn lại thì sọt cam được xếp loại III. Giả sử tỉ lệ cam hỏng của sọt là 3%. Hãy tính xác suất để:

a. Sọt cam được xếp loại I.

b. Sọt cam được xếp loại II.

c. Sọt cam được xếp loại III.

38. Tính xác suất khi rút có hoàn lại 10 lần từ bộ bài 52 cây ta được 4 cây chuồng, 2 cây pít, 3 cây rô, 1 cây cơ.

KQHT 2: GIẢI CÁC BÀI TOÁN LIÊN QUAN ĐẾN ĐẠI LƯỢNG NGẪU NHIÊN VÀ PHÂN PHỐI XÁC SUẤT

Bước học 1: ĐẠI LƯỢNG NGẪU NHIÊN

1.1 Các định nghĩa:

Định nghĩa: *Đại lượng ngẫu nhiên là đại lượng biến đổi biểu thị các giá trị kết quả của một phép thử ngẫu nhiên.*

Ta thường dùng các kí hiệu: X, Y, Z,…... để biểu thị cho đại lượng ngẫu nhiên.

Ví dụ 1: *Tung một con xúc sắc, gọi X là số chấm xuất hiện trên mặt con xúc xắc. Khi đó, X là đại lượng ngẫu nhiên.*

Gọi Y là số học sinh vắng trong một buổi học $\Rightarrow$ Y = 0, 1, 2, . . .

Y là đại lượng ngẫu nhiên.

Gọi Z là điểm rơi của hạt cát trên đoạn [0;1] thì Z cũng là đại lượng ngẫu nhiên.

Đo chiều cao của các sinh viên ở một trường đại học. Gọi Y là chiều cao đo được của các sinh viên. Giả sử Y$\in$ [0.5m ; 1.2m]. Vậy Y là đại lượng ngẫu nhiên.

♥ Có hai loại đại lượng ngẫu nhiên:

+ Đại lượng ngẫu nhiên rời rạc: *Đại lượng ngẫu nhiên được gọi là rời rạc nếu nó có một số hữu hạn hoặc vô hạn đếm được các giá trị.*

$\Rightarrow$ X, Y là các đại lượng ngẫu nhiên rời rạc.

Các giá trị có thể của đại lượng ngẫu nhiên X được ký hiệu x_1, x_2, …, hay y_1, y_2, …

+ Đại lượng ngẫu nhiên liên tục: *Đại lượng ngẫu nhiên được gọi là liên tục nếu các giá trị có thể có của nó lấp đầy một khoảng trên trục số.*

$\Rightarrow$ Z là đại lượng ngẫu nhiên liên tục.

Ta không thể liệt kê các giá trị của đại lượng ngẫu nhiên liên tục.

Các đại lượng ngẫu nhiên chỉ nhiệt độ, diện tích, thể tích, thời gian, … là liên tục.

1.2 Phân phối xác suất của đại lượng ngẫu nhiên:

Định nghĩa: *Luật phân phối xác suất của đại lượng ngẫu nhiên là biểu đồ (bảng, đồ thị,...) trong đó chỉ ra:*

- ✓ *Các giá trị có thể nhận được của đại lượng ngẫu nhiên.*
- ✓ *Xác suất tương ứng của đại lượng ngẫu nhiên nhận giá trị đó.*

1.2.1 Bảng phân phối xác suất:

Bảng phân phối xác suất dùng để thiết lập luật phân phối xác suất của đại lượng ngẫu nhiên rời rạc.

Bảng gồm 2 dòng: Dòng trên ghi các giá trị có thể có của đại lượng ngẫu nhiên là: x_1, x_2, .. , x_n; dòng dưới ghi các xác suất tương ứng là: P_1, P_2, .. , P_n.

X	x_1	x_2	x_3	. . .	x_n
P	P_1	P_2	P_3	. . .	P_n

Trong đó: $\sum_{i=1}^{n} P_i = 1$

<u>*Ghi chú*</u>: $X = x_i$: Đại lượng ngẫu nhiên X nhận giá trị x_i.

$P(X = x_i)$: Xác suất để đại lượng ngẫu nhiên X nhận giá trị x_i.

Ví dụ 2: *Tung 1 con xúc sắc, gọi X là số chấm xuất hiện trên mặt của một con xúc xắc. Khi đó bảng phân phối xác suất của X là:*

X	1	2	3	4	5	6
P	1/6	1/6	1/6	1/6	1/6	1/6

Ví dụ 3: *Tiến hành thử độ bền của 3 loại vật liệu, với điều kiện vật liệu thử trước phải vượt qua được phép thử mới thử tiếp vật liệu sau. Biết rằng khả năng vượt qua phép thử của các vật liệu đều bằng 0,8. Hãy tìm luật phân phối xác suất của số vật liệu vượt qua phép thử.*

Gọi X là số vật liệu vượt qua phép thử.

A_i là biến cố vật liệu thứ i vượt qua phép thử $\left(i = \overline{1,3}\right)$.

Ta có:

$P(X = 0) = P(\overline{A_1}) = 0,2$

$P(X = 1) = P(A_1\overline{A_2}) = P(A_1)P(\overline{A_2}) = (0,8)(0,2) = 0,16$

$P(X = 2) = P(A_1A_2\overline{A_3}) = P(A_1)P(A_2)P(\overline{A_3}) = (0,8)(0,8)(0,2) = 0,128$

$P(X = 3) = P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3) = (0,8)(0,8)(0,8) = 0,512$

Bảng phân phối xác suất của X là:

X	0	1	2	3
P	0,2	0,16	0,128	0,512

Ví dụ 4: *Hộp có 10 viên bi, trong đó có 6 viên màu đỏ, còn lại màu trắng. Rút đồng thời 4 viên bi và gọi X là số viên bi màu đỏ được rút ra. Lập luật phân phối xác suất của X.*

Gọi A_i là biến cố rút được i viên bi màu đỏ $(i = \overline{1,4})$.

Các xác suất được tính theo nguyên tắc hộp kín như sau:

$$P(X = 0) = P(A_0) = \frac{C_6^0C_4^4}{C_{10}^4} = \frac{1}{210} = 0,005$$

$$P(X = 1) = P(A_1) = \frac{C_6^1C_4^3}{C_{10}^4} = \frac{24}{210} = 0,114$$

$$P(X = 2) = P(A_2) = \frac{C_6^2C_4^2}{C_{10}^4} = 0,429$$

$$P(X = 3) = P(A_3) = \frac{C_6^3C_4^1}{C_{10}^4} = 0,318$$

$$P(X = 4) = P(A_4) = \frac{C_6^4C_4^0}{C_{10}^4} = 0,071$$

Vậy ta có bảng phân phối xác suất của X là:

X	0	1	2	3	4
P	0,005	0,114	0,429	0,381	0,071

1.2.2 Hàm mật độ xác suất:

Định nghĩa: *Luật phân phối xác suất của đại lượng ngẫu nhiên liên tục X được biểu thị bởi hàm số y = f(x) xác định trên (-∞, +∞) thỏa mãn:*

i) $f(x) \geq 0, \forall x$

ii) $\int_{-\infty}^{+\infty} f(x)dx = 1$

- **Tính chất:**

i) $P(X = x_0) = 0$.

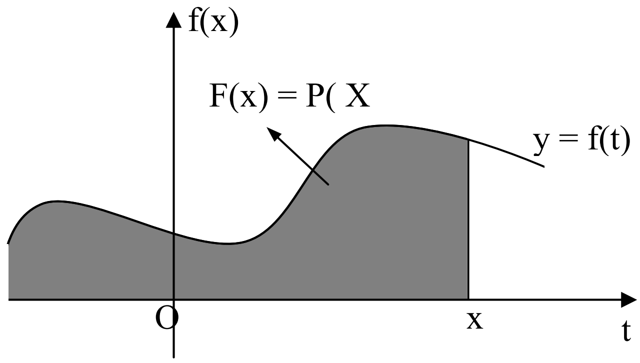
ii) $P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = \int_a^b f(x)dx$

(Diện tích hình thang cong cạnh trái x = a, cạnh phải x = b (xem hình 7)).

iii) $P(X < \alpha) = P(-\infty < X < \alpha) = \int_{-\infty}^{\alpha} f(x)dx$

iv) $P(X > \alpha) = P(\alpha < X < +\infty) = \int_{\alpha}^{+\infty} f(x)dx$

v) <u>*Đặc biệt:*</u> f(x) chỉ nhận giá trị trên [a; b] thì: $\int_a^b f(x)dx = 1$



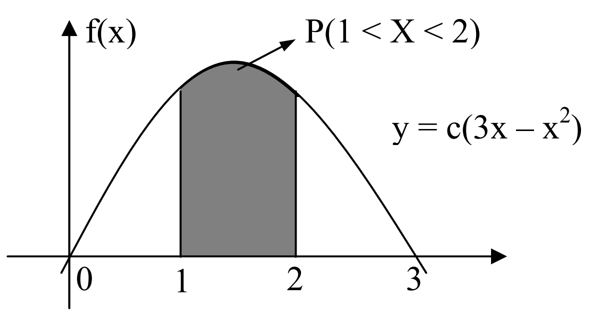
Hình

Ví dụ 5: *Cho đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất*

$$f(x) = \begin{cases} c(3x - x^2) & \text{nếu } x \in [0,3] \\ 0 & \text{nếu } x \notin [0,3] \end{cases}$$

a) Xát định hằng số c.

b) Tính $P(1 < X < 2)$.



Giải

a) Ta có:

$$1 = \int_{-\infty}^{+\infty} f(x).dx$$

$$= \int_{-\infty}^{0} f(x).dx + \int_{0}^{3} f(x).dx + \int_{3}^{+\infty} f(x).dx$$

$$= \int_{-\infty}^{0} 0.dx + \int_{0}^{3} c(3x - x^2) + \int_{3}^{+\infty} 0.dx \quad = 0 + \frac{9}{2}c + 0 = \frac{9}{2}c$$

Vậy: $c = \dfrac{2}{9}$

b) Ta có:

$$P\,(1 < X < 2) = \int_{1}^{2} f(x)\,dx \; = \int_{1}^{2} \frac{2}{9}(3x - x^2)\,dx \; = \frac{13}{27}.$$

1.2.3 Hàm phân phối xác suất:

Định nghĩa: *Hàm phân phối xác suất của đại lượng ngẫu nhiên X (liên tục hoặc rời rạc), ký hiệu F(x), là hàm được xác định như sau:*

F(x) = P(X < x)

Cụ thể :

- ✓ X là đại lượng ngẫu nhiên rời rạc: $F(x) = \sum_{x_i < x} p_i$
- ✓ X là đại lượng ngẫu nhiên liên tục: $F(x) = \int_{-\infty}^{x} f(x)dx$

(Bằng diện tích hình thang cong, cạnh trái t = -∞, cạnh phải t = x (xem hình 9)).

Tính chất:

i) $0 \le F(x) \le 1 \;, \forall x$

ii) F(x) là hàm không giảm

iii) F(-∞) = 0 F(+∞) = 1

iv) P(a ≤ X < b) = F(b) - F(a)

v) Nếu X là đại lượng ngẫu nhiên rời rạc thì F(x) có dạng bậc thang

vi) Nếu X là đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất f(x) thì $F^{/}(x) = f(x)$

Ý nghĩa: *Hàm phân phối xác suất F(x) phản ánh mức độ tập trung xác suất về phía bên trái của điểm x.*

Ví dụ 6: *Cho X có:*

X	*1*	*2*	*3*
P	*0,5*	*0,2*	*0,3*

Tìm F(x) và vẽ đồ thị.

Giải

Ta có: $F(x) = \sum_{i<x} p_i$

$+\, x \le 1 : F(x) = 0$

$+\, 1 < x \le 2 : F(x) = 0{,}5$

$+\, 2 < x \le 3 :$

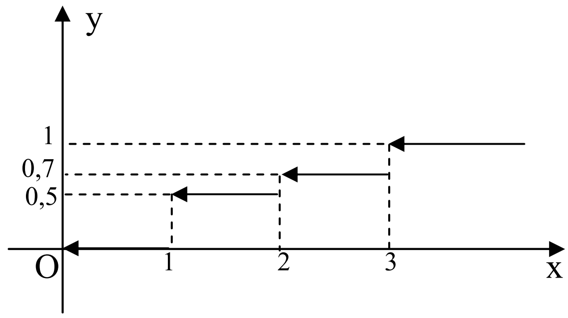
$F(x) = 0{,}5 + 0{,}2 = 0{,}7$

$+\, x > 3 :$

$F(x) = 0{,}5 + 0{,}2 + 0{,}3 = 1$

Vậy: $F(x) = \begin{cases} 0 & \text{nếu } x \le 1 \\ 0{,}5 & \text{nếu } 1 < x \le 2 \\ 0{,}7 & \text{nếu } 2 < x \le 3 \\ 1 & \text{nếu } x > 3 \end{cases}$

Đồ thị F(x)



Đồ thị hàm số có dạng bậc thang

Hình 10

Ví dụ 7: *Cho đại lượng ngẫu nhiên X có:* $f(x) = \begin{cases} 0 & \text{nếu } x \le 0 \\ x & \text{nếu } 0 < x \le 1 \\ 2 - x & \text{nếu } 1 < x \le 2 \\ 0 & \text{nếu } x > 2 \end{cases}$

Tìm hàm phân phối xác suất F(x) và vẽ đồ thị của nó .

Ta có: $F(x) = \int_{-\infty}^{x} f(x)dx$

$$+\, x \le 0 : F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} 0dx = 0$$

$$+\, 0 < x \le 1 : F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{x} f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{x} xdx = \left.\frac{x^2}{2}\right|_0^x = \frac{x^2}{2}$$

$$+\, 1 < x \le 2 : F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{x} f(x)dx =$$

$$= \int_{-\infty}^{0} 0dx + \int_{0}^{1} x.dx + \int_{1}^{x} (2-x)dx =$$

$$= \frac{1}{2} + 2x - \frac{x^2}{2} - 2 + \frac{1}{2} = -\frac{x^2}{2} + 2x + -1$$

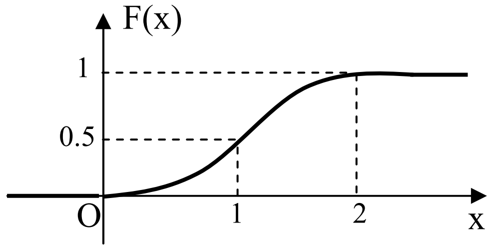
$$+\, x > 2 : F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{2} f(x)dx + \int_{2}^{x} f(x)dx =$$

$$= \int_0^1 x dx + \int_1^2 (2-x) dx = \frac{1}{2} + 4 - 2 - 2 + \frac{1}{2} = 1$$

Vậy:

$$F(x) = \begin{cases} 0 & \text{nếu } x \le 0 \\ \dfrac{x^2}{2} & \text{nếu } 0 < x \le 1 \\ -\dfrac{x^2}{2} + 2x - 1 & \text{nếu } 1 < x \le 2 \\ 1 & \text{nếu } x > 2 \end{cases}$$

Đồ thị



Hình 11

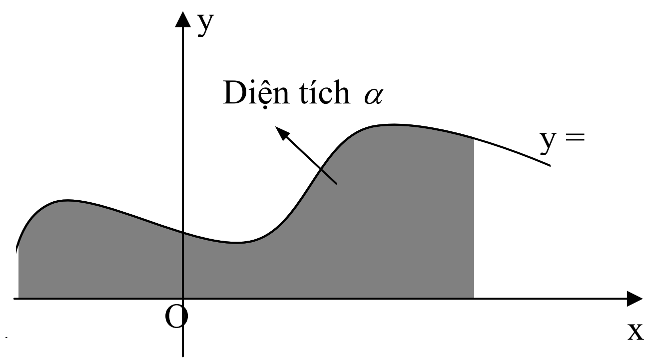
1.2.4. Phân vị mức xác suất α:

Định nghĩa: *Phân vị mức xác suất α của đại lượng ngẫu nhiên liên tục X là số X_α sao cho:*

$$P(X < X_\alpha) = \alpha \qquad (*)$$

Hệ thức (*) tương đương với: $\int_{-\infty}^{X_\alpha} f(x)\,dx = \alpha$

Như vậy, X_α là cận trên của tích phân sao cho tích phân bằng α hay X_α là vị trí cạnh phải của hình thang cong sao cho diện tích hình thang cong bằng α (xem hình 12)).

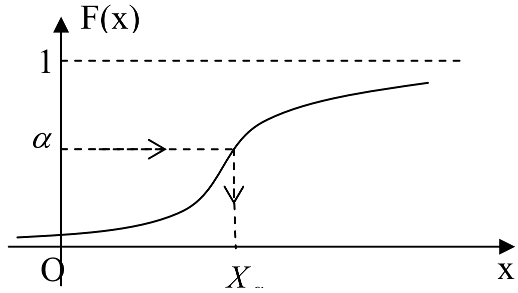


Hình

Mặt khác, từ hệ thức (*) suy ra:

$$F(X_\alpha) = \alpha \text{ hay } X_\alpha = F^{-1}(\alpha)$$

Như vậy, X_α là giá trị ngược của hàm phân phối xác suất F(x) tại α cho trước (xem hình 13).



Hình 13

Ví dụ 8: *Cho đại lượng ngẫu nhiên X liên tục có hàm mật độ xác suất:*

$$f(x) = \begin{cases} (3x - x^2) & \text{nếu } x \in [0,3] \\ 0 & \text{nếu } x \notin [0,3] \end{cases}$$

a) Xác định hàm phân phối xác suất F(x) của X.

b) Tìm $X_{7/27}$.

Giải

a) Ta có: $F(x) = \int_{-\infty}^{x} f(x)\,dx$

Do đó:

- $x < 0$, $F(x) = \int_{-\infty}^{x} 0\,dx = 0.$
- $0 \le x \le 3$, $F(x) = \int_{-\infty}^{0} 0\,dx + \int_{0}^{x} \frac{2}{9}(3x - x^2)\,dt = \frac{x^2(9 - 2x)}{27}$.
- $x > 3$, $F(x) = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{3} f(x)\,dx + \int_{3}^{x} f(x)\,dx = 1.$

Vậy: $F(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ \dfrac{x^2(9-2x)}{27} & \text{nếu } 0 \le x \le 3 \\ 1 & \text{nếu } x > 3 \end{cases}$

b) Tính $X_{7/27}$: Ta có: $\alpha = \frac{7}{27}$.

Đặt u = $X_{7/27}$, sử dụng công thức trên, ta có $F(u) = \frac{7}{27}$. Từ câu a), suy ra

$$\frac{u^2(9-2u)}{27} = \frac{7}{27}, \quad 0 < u < 3$$

$$\Rightarrow 2u^3 - 9u^2 + 7 = 0 \quad \Rightarrow \quad u = 1 \;; \quad u = \frac{7 \pm \sqrt{105}}{4} \quad \text{(loại)}$$

Vậy: $X_{7/27} = 1$.

Bước học 2: CÁC THAM SỐ ĐẶC TRƯNG CỦA ĐẠI LƯỢNG NGẪU NHIÊN:

2.1 Kỳ vọng: (expectation)

Định nghĩa: *Giả sử X là đại lượng ngẫu nhiên rời rạc có thể nhận các giá trị x_1, x_2, .. , x_n với các xác suất tương ứng P_1, P_2, .. , P_n*

Khi đó kỳ vọng của X , kí hiệu là E(X) hay M(X) được xác định bởi công thức:

$$E(X) = \sum_{i=1}^{n} x_i P_i$$

Nếu X là đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất là f(x) thì kỳ vọng của X là:

$$E(X) = \int_{-\infty}^{+\infty} x.f(x)dx$$

Ví dụ 1: *Cho X là đại lượng ngẫu nhiên rời rạc có bảng phân phối xác suất sau:*

X	5	6	7	8	9	10	11
P	1/12	2/12	3/12	2/12	2/12	1/12	1/12

Tìm E(X)?

Ta có: $E(X) = \sum_{i=1}^{7} x_i P_i = 5.\frac{1}{12} + 6.\frac{2}{12} + 7.\frac{3}{12} + 8.\frac{2}{12} + 9.\frac{2}{12} + 10.\frac{1}{12} + 11.\frac{1}{12} = \frac{93}{12} = 7{,}75$

Ví dụ 2: *Cho X là đại lượng ngẫu nhiên rời rạc có luật phân phối:*

X	0	1	3	4	7	8
P	$\frac{1}{30}$	$\frac{3}{30}$	$\frac{12}{30}$	$\frac{8}{30}$	$\frac{4}{30}$	$\frac{2}{30}$

Tính E(X).

Ta có: $E(X) = \sum_{i=1}^{n} x_i p_i = 0\frac{1}{30} + 1\frac{3}{30} + 3\frac{12}{30} + 4\frac{8}{30} + 7\frac{4}{30} + 8\frac{2}{30} = \frac{125}{30} = \frac{25}{6} \approx 4{,}17$

Ví dụ 3: *Cho đại lượng ngẫu nhiên liên tục X có hàm mật độ xác suất:*

$$f(x) = \begin{cases} c(4x - x^2) & \text{nếu } x \in [0,4] \\ 0 & \text{nếu } x \notin [0,4] \end{cases}$$

Tính E(X).

Trước tiên, ta xác định hằng số c:

Ta có:

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{0} 0.dx + \int_{0}^{4} f(x)dx + \int_{4}^{+\infty} 0.dx$$

$$= \int_{0}^{4} c(4x - x^2)dx = c\left(\frac{4x^2}{2} - \frac{x^3}{3}\right)\Bigg|_0^4$$

$$= c(2.4^2 - \frac{4^3}{3}) = c\frac{3.2.4^2 - 4.4^2}{3} = c.\frac{32}{2}$$

$$\Rightarrow c = \frac{3}{32}$$

Do đó:

$$E(X)=\int_{+\infty}^{-\infty}x.f(x)dx=\int_0^4 x.\frac{3}{32}(4x-x^2)dx=\frac{3}{32}\int_0^4(4x^2-x^3)dx=\frac{3}{32}\left[4\frac{x^3}{3}-\frac{x^4}{4}\right]_0^4$$

$$=\frac{3}{32}\left[\frac{4^4}{3}-\frac{4^4}{4}\right]=\frac{3}{2.4^2}\frac{4.4^4-3.4^4}{3.4}=\frac{4^4}{2.4^3}=2$$

♦ ***Tính chất:***

i) E(C) = C

ii) E(C.X) = C.E(X) , với C là hằng số.

iii) E(X + Y) = E(X) + E(Y)

iv) Nếu X, Y là hai đại lượng ngẫu nhiên độc lập thì:

E(X.Y) = E(X).E(Y).

❖ **Chú ý:** Tính chất iii) và iv) có thể mở rộng cho nhiều đại lượng ngẫu nhiên.

Ý nghĩa: Kỳ vọng của 1 đại lượng ngẫu nhiên chính là giá trị trung bình (theo xác suất) của đại lượng ngẫu nhiên đó. Nó là trung tâm điểm của phân phối mà các giá trị cụ thể của X sẽ tập trung quanh đó.

Ví dụ 4: *Giả sử ta có cái bình lớn đựng 10 quả cầu giống nhau nhưng khác nhau về trọng lượng: 5 quả nặng 1 kg, 2 quả nặng 2 kg, 3 quả nặng 3 kg. Ta lấy ngẫu nhiên từ bình ra 1 quả cầu và gọi X là trọng lượng của quả cầu đó. Tính E(X) và so sánh E(X) với trọng lượng trung bình của 1 quả cầu trong hộp.*

✓ Bảng phân phối xác suất của X:

X	1	2	3
P	$\frac{5}{10}$	$\frac{2}{10}$	$\frac{3}{10}$

$$\Rightarrow E(X)=\sum_{x=1}^{3}x_ip_i==1.\frac{5}{10}+2.\frac{2}{10}+3.\frac{3}{10}=\frac{18}{10}$$

$\Rightarrow$ E(X) = 1,8

✓ Gọi M là trọng lượng trung bình của các quả cầu trong bình.

Ta có: $M=\frac{5.1+2.2+3.3}{10}=\frac{18}{10}=1,8$

Vậy: E(X) = M

2.2 Phương sai: (Variance)

Định nghĩa: *Phương sai (độ lệch bình phương trung bình) của đại lượng ngẫu nhiên X, kí hiệu Var(X) (hay V(x) hoặc D(X)) được xác định bởi công thức:*

$$Var(X) = E\{[X - E(X)]^2\}$$

Nếu X là đại lượng ngẫu nhiên rời rạc có thể nhận các giá trị là x_1, x_2, .., x_n với các xác suất tương ứng là P_1, P_2, .. , P_n thì:

$$Var(X) = \sum_{i=1}^{n} \left[x_i - E(X)\right]^2 .P_i$$

Nếu X là đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất là f(x) thì:

$$Var(X) = \int_{-\infty}^{+\infty} \left[x - E(X)\right]^2 f(x)dx$$

Chú ý: Trong thực tế ta thường tính phương sai bằng công thức:

$Var(X) = E(X^2) – [E(X)]^2$

Ví dụ 5: *Cho X là đại lượng ngẫu nhiên rời rạc có bảng phân phối xác suất sau:*

X	1	3	5
P	0,1	0,4	0,5

Tìm phương sai của X.

Dễ dàng ta có: E(X) = 3,8

$Var(X) = E(X^2) – [E(X)]^2 = 1,76$

Ví dụ 6: *Cho X là đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất sau:*

$$f(x)=\begin{cases} c.x^3 & \text{nếu } 0 \le x \le 3 \\ 0 & x \notin [0;3] \end{cases}$$

Tìm hằng số c, E(X), Var(X)

Ta có: $1 = \int_0^3 c.x^3dx = c.\left[\frac{x^4}{4}\right]_0^3 = \frac{81c}{4}$

Dễ dàng tính được c = 4/81, E(X) = 2,4; Var(X) = 0,24

♦ ***Tính chất:***

i) Var(C) = 0

ii) $Var(C.X) = C^2.Var(X)$

iii) Nếu X, Y là 2 đại lượng ngẫu nhiên độc lập thì:

Var(X+Y) = Var(X) + Var(Y); Var(X – Y) = Var(X) + Var(Y)

iv) Var(C+X) = Var(X)

Ý nghĩa: *Ta thấy X - E(X) là độ lệch khỏi giá trị trung bình. Do đó phương sai $Var(X) = E\{[X – E(X)]^2\}$ gọi là độ lệch bình phương trung bình. Nên phương sai phản ánh mức độ phân tán các giá trị của đại lượng ngẫu nhiên xung quanh giá trị trung bình.*

Như vậy, phương sai phản ánh mức độ phân tán các giá trị của đại lượng ngẫu nhiên chung quanh kỳ vọng. Đại lượng ngẫu nhiên có phương sai càng lớn thì các giá trị càng phân tán và ngược lại.

Ứng dụng: Trong công nghiệp, phương sai biểu thị độ chính xác của sản xuất. Trong chăn nuôi, nó biểu thị độ đồng đều của các con gia súc. Trong trồng trọt, nó biểu thị mức độ ổn định của năng suất, .v.v..

Ví dụ 7: *Giả sử X là khối lượng các gói bột giặt của phân xưởng I, Y là khối lượng các gói bột giặt của phân xưởng II. Trong đó: E(X) = E(Y) = 500g và Var(X) >Var(Y). Khi đó, các gói bột giặt của phân xưởng II có khối lượng tập trung hơn xung quanh khối lương 500g. Nói cách khác, hệ thống đóng gói của phân xưởng II hoạt động tốt hơn phân xưởng I.*

2.3 Độ lệch tiêu chuẩn:

Định nghĩa: *Độ lệch tiêu chuẩn của đại lượng ngẫu nhiên X, kí hiệu σ(X) được xác định bởi công thức:* $\sigma(X) = \sqrt{Var(X)}$

2.4 Môment:

Môment cấp k của đại lượng ngẫu nhiên X là số $m_k = E(X^k)$

Môment quy tâm cấp k của đại lượng ngẫu nhiên X là số: $\alpha_k = \mathrm{E}\{[\mathrm{X} - \mathrm{E(X)}]^k\}$

- *Nhận xét:* Môment cấp 1 của X là kỳ vọng của X

  Môment quy tâm cấp 2 của X là phương sai của X

2.5 Mode:

Mod(X) là giá trị của đại lượng ngẫu nhiên X có xác suất lớn nhất.

Đối với đại lượng ngẫu nhiên rời rạc, mod(X) là giá trị của X ứng với xác suất lớn nhất. Còn đối với đại lượng ngẫu nhiên liên tục thì mod(X) là giá trị của X tại đó hàm mật độ đạt giá trị cực đại.

Chú ý: Một đại lượng ngẫu nhiên có thể có 1 mode hoặc nhiều mode.

Ví dụ 8: *Gọi X là điểm thi của sinh viên thì mod(X) là điểm mà có nhiều sinh viên đạt được nhất.*

Ví dụ 9: X là đại lượng ngẫu nhiên rời rạc có luật phân phối:

X	0	1	3	4	7	8
P	$\frac{1}{30}$	$\frac{3}{30}$	$\frac{12}{30}$	$\frac{8}{30}$	$\frac{4}{30}$	$\frac{2}{30}$

Ta thấy $P(x=3) = \frac{12}{30} \to \max$

=> mod(X) = 3.

Ví dụ 10: *Cho đại lượng ngẫu nhiên X liên tục có hàm mật độ:*

$$f(x) = \begin{cases} 0 & \text{nếu } x \le 0 \\ \frac{x}{2} e^{-\frac{x^2}{4}} & \text{nếu } x > 0 \end{cases}$$

Hãy tìm mod(X).

Xét: $f(x) = \frac{x}{2} e^{-\frac{x^2}{4}}$

Có: $f'(x) = \frac{1}{2} e^{-\frac{x^2}{4}} - \frac{x^2}{4} e^{-\frac{x^2}{4}}$

$$\Rightarrow f'(x) = 0 \Leftrightarrow \frac{1}{2}\mathrm{e}^{-\frac{x^2}{4}} - \frac{x^2}{4}\mathrm{e}^{-\frac{x^2}{4}} = 0$$

$$\Leftrightarrow \frac{1}{2}(1 - \frac{x^2}{2})\mathrm{e}^{-\frac{x^2}{4}} = 0$$

$$\Leftrightarrow (1 - \frac{x^2}{2}) = 0$$

$$\Leftrightarrow x = \pm\sqrt{2}$$

Và: $$f''(x) = -\frac{x}{4}e^{-\frac{x^2}{4}} - \left(\frac{x}{2}e^{-\frac{x^2}{4}} - \frac{x^3}{8}e^{-\frac{x^2}{4}}\right) = -\frac{3x}{4}e^{-\frac{x^2}{4}} + \frac{x^3}{8}e^{-\frac{x^2}{4}}$$

$$= (\frac{x^2}{2} - 3)\frac{x}{4}e^{-\frac{x^2}{4}}$$

Suy ra:

$$+ x = \sqrt{2} : f''(\sqrt{2}) = (2-3)\frac{\sqrt{2}}{4}e^{-1} = -\frac{\sqrt{2}}{4e} < 0$$

$$\Rightarrow f(\sqrt{2}) \to \max$$

$$+ x = -\sqrt{2} : f''(-\sqrt{2}) = (2-3)\frac{-\sqrt{2}}{4}e^{-1} = \frac{\sqrt{2}}{4e} > 0$$

$$\Rightarrow f(\sqrt{2}) \to \min$$

Vậy: $\mathrm{mod}(X) = \sqrt{2} = 1{,}414$

2.6 Trung vị:

Định nghĩa: *Trung vị của đại lượng ngẫu nhiên X là giá trị của X chia phân phối xác suất thành 2 phần có xác suất giống nhau.*

Kí hiệu: med(X).

Công thức: $P(X < med(X)) = P(X \geq med(X)) = \frac{1}{2}$

Nhận xét: *Từ định nghĩa ta thấy để tìm trung vị chỉ cần giải phương trình* $F(med(X)) = \frac{1}{2}$. *Trong ứng dụng, trung vị là đặc trưng vị trí tốt nhất, nhiều khi tốt hơn cả kỳ vọng, nhất là khi trong số liệu có nhiều sai sót. Trung vị còn gọi là phân vị 50% của phân phối.*

Ví dụ 11: *Cho X như trong* ***ví dụ 10****. Hãy xác định med(X).*

Med(X) là nghiệm của phương trình:

$$F(med(X)) = \int_{-\infty}^{med(X)} f(x)dx = \frac{1}{2}$$

$$\Rightarrow \int_{0}^{med(X)} f(x)dx = \int_{0}^{med(X)} \frac{x}{2}\mathrm{e}^{-\frac{x^2}{4}}dx = \frac{1}{2}$$

$$\Rightarrow -\int_{0}^{med(X)} e^{-\frac{x^2}{4}} d(-\frac{x^2}{4}) = \frac{1}{2}$$

$$\Rightarrow -e^{-\frac{x^2}{4}}\Big|_{0}^{med(X)} = \frac{1}{2}$$

$$\Rightarrow 1 - e^{-\frac{[med(X)]^2}{4}} = \frac{1}{2}$$

$$\Rightarrow e^{-\frac{[med(X)]^2}{4}} = \frac{1}{2}$$

$$\Rightarrow -\frac{[med(X)]^2}{4} = \ln\frac{1}{2} = -0{,}693$$

$$\Rightarrow [med(X)]^2 = 2{,}772$$

$$\Rightarrow med(X) = 1{,}665 \qquad (do\, med(X) > 0)$$

Vậy: med(X) = 1.665

Chú ý: Nói chung, ba số đặc trưng: E(X), mod(X), med(X) không trùng nhau. Chẳng hạn, từ các ***ví dụ 10*** và ***11*** và ta tính thêm kỳ vọng ta có: E(X) = 1,772, mod(X) = 1,414 và med(X) = 1,665. Tuy nhiên nếu phân phối đối xứng chỉ có một mod thì 3 đặc trưng đó trùng nhau.

BÀI TẬP
ĐẠI LƯỢNG NGẪU NHIÊN RỜI RẠC

1.

a. Một lô hàng gồm N sản phẩm, trong đó có M sản phẩm tốt, còn lại là sản phẩm xấu. Ba người khách hàng lần lượt đến mua mỗi người một sản phẩm bằng cách lấy ngẫu nhiên. Xác suất chọn sản phẩm tốt của người thứ nhất, thứ hai, thứ ba có khác nhau không, tại sao?

b. Nếu lô hàng có 10 sản phẩm trong đó có 4 sản phẩm xấu. Mỗi sản phẩm tốt nặng 3 kg, mỗi sản phẩm xấu chỉ nặng 2 kg. Chọn ngẫu nhiên 3 sản phẩm từ lô hàng thì tổng 3 sản phẩm đó nặng bao nhiêu là có khả năng tin chắc nhất, tại sao?

2. Trong một cái bát có để 5 hạt đậu, trong đó có hai hạt đỏ. Lấy ngẫu nhiên ra 2 hạt. Gọi X là số hạt đậu đỏ được lấy ra.

a. Lập bảng phân phối xác suất của X.

b. Viết biểu thức hàm phân phối của X.

c. Tính E(X) và Var(X).

3. Có 3 hộp mỗi hộp đựng 10 sản phẩm, trong đó số phế phẩm có trong mỗi hộp lần lượt là: 2, 3, 5.

a. Lấy ngẫu nhiên từ mỗi hộp ra 1 sản phẩm. Tìm quy luật phân phối xác suất của số sản phẩm tốt có trong 3 sản phẩm lấy ra.

b. Chọn ngẫu nhiên một hộp rồi từ hộp lấy ra 3 sản phẩm. Lập bảng phân phối xác suất của số sản phẩm tốt có trong 3 sản phẩm lấy ra.

4. Một xạ thủ có 5 viên đạn bắn vào một cái bia. Anh ta bắn từng viên một vào bia với xác suất trúng tâm mỗi lần bắn là 0,9.

a. Nếu có 3 viên đạn liên tiếp trúng tâm hoặc hết đạn thì không bắn nữa.

b. Nếu có 3 viên đạn trúng tâm hoặc hết đạn thì không bắn nữa.

Gọi X, Y là số đạn mà anh ta dùng tương ứng theo hai quy tắc trên. Lập bảng phân phối xác suất của X và Y.

5. Có hai hộp bi: Hộp I có 3 bi trắng và 1 bi đỏ. Hộp II có 2 bi trắng và 2 bi đỏ. Lấy ngẫu nhiên từ hộp I ra 2 bi bỏ vào hộp II. Sau đó lấy ngẫu nhiên từ hộp II ra 2 bi bỏ vào hộp I. Gọi X và Y là số bi trắng ở hộp I và hộp II sau hai lần chuyển bi như trên. Lập bảng phân phối xác suất của X và Y.

6. Hộp I có 4 bi đỏ và 8 bi trắng. Hộp II có 3 bi đỏ và 5 bi trắng. Lấy ngẫu nhiên 2 bi ở hộp I bỏ vào hộp II, rồi lấy không hoàn lại 3 bi ở hộp II. Gọi X là số đỏ lấy được từ hộp II.

a. Tìm luật phân phối xác suất của X.

b. Tính E(X), Var(X) và $P(1 \le X \le 10)$.

c. Giải bài tập trên với giả thiết lấy có hoàn lại 3 bi ở hộp II.

7. Một kiện hàng có 15 sản phẩm, nhưng chưa biết chất lượng cụ thể. Gọi X là số sản phẩm loại A có trong hộp và cho biết X có luật phân phối xác suất như sau:

X	10	12	14	15
P	0,1	0,3	0,4	0,2

Lấy ngẫu nhiên từ kiện ra 3 sản phẩm để kiển tra (lấy không hoàn lại).

a. Tìm luật phân phối xác suất của số sản phẩm loại A có trong 3 sản phẩm lấy ra kiểm tra.

b. Tìm luật phân phối trên nếu quá trình lấy 3 sản phẩm theo phương thức có hoàn lại.

8. Một hộp có 6 sản phẩm. Mọi giả thiết về số sản phẩm tốt có trong hộp lúc đầu đều đồng khả năng. Lấy ngẫu nhiên ra 3 sản phẩm thì thấy cả 3 sản phẩm đều tốt. Tìm luật phân phối xác suất của số sản phẩm tốt có trong 3 sản phẩm còn lại trong hộp.

9. Trong ngày hội thi, một công nhân nào đó dự thi sẽ sản xuất 2 sản phẩm. Mỗi sản phẩm loại 1 sẽ được thưởng 10.000 đồng, nhưng mỗi sản phẩm không phải loại 1 sẽ bị phạt 5.000 đồng. Giả sử xác suất để công nhân đó sản xuất được sản phẩm loại 1 là 0,4. Tìm luật phân phối xác suất của số tiền công nhân thu được qua cuộc thi trên. Tìm số tiền trung bình mà công nhân dự thi đó có thể có.

10. Một hộp có 5 sản phẩm trong đó có 2 phế phẩm. Ta chọn ngẫu nhiên từng sản phẩm để kiểm tra (kiểm tra không hoàn lại) cho đến khi hết 2 phế phẩm thì dừng. Lập bảng phân phối xác suất của số lần kiểm tra. Tính xem trong việc kiểm tra trên trung bình ta phải kiểm tra bao nhiêu lần.

ĐẠI LƯỢNG NGẪU NHIÊN LIÊN TỤC

11. Cho hàm $f(x) = \begin{cases} \dfrac{c}{x^4} & \text{nếu } x \in [1,2] \\ 0 & \text{nếu } x \notin [1,2] \end{cases}$

a. Tính c, E(X), Var(X).

b. Tìm F(x).

c. Tính $P(\frac{3}{2} < X \le 4)$.

12. Cho đại lượng ngẫu nhiên X có hàm mật độ:

$$f(x) = \begin{cases} k.(1-x^2) & khi|x| \le 1 \\ 0 & khi|x| > 1 \end{cases}$$

a. Tính k.

b. Tính kỳ vọng và phương sai của đại lượng ngẫu nhiên $Y = 2X^2$.

c. Tính $P(-\frac{1}{2} \le X < \frac{1}{2})$.

13. Biến ngẫu nhiên X nhận giá trị tập trung trong $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ với hàm mật độ có dạng f(x) = c. cosx.

a. Xác định hằng số c.

b. Viết biểu thức hàm phân phối của X.

c. Tìm $P(0 < X < \frac{\pi}{4})$.

d. Nếu quan sát X 5 lần thì có bao nhiêu lần X nhận giá trị trong khoảng $(0, \frac{\pi}{4})$ là có khả năng nhất. Tính xác suất đó.

14. Cho X và Y là hai đại lượng ngẫu nhiên có hàm mật độ xác suất như sau:

$$f(x) = \begin{cases} \dfrac{c+d}{x^4} & \text{nếu } x > 2 \\ 0 & \text{nếu } x \le 2 \end{cases} \quad ; \quad f(y) = \begin{cases} 0 & \text{nếu } x \notin [0,2] \\ \dfrac{-c+d}{96}\left(4y - y^3\right) & \text{nếu } x \in [0,2] \end{cases}$$

a. Tính các hằng số c, d.

b. Tính E(X), Var(X).

c. Tính $P(1 \le Y \le 2)$.

15. Tuổi thọ trung bình của một loại côn trùng nào đó là một đại lượng ngẫu nhiên X (đơn vị là tháng) với hàm mật dộ như sau:

$$f(x) = \begin{cases} kx^2\left(4-x\right) & \text{nếu } 0 \le x \le 4 \\ 0 & \text{nếu trái lại} \end{cases}$$

a/ Tính k. b/ Tìm mod(X).

c/ Tính xác suất để côn trùng chết trước khi nó được một tháng tuổi.

16. Cho hàm mật độ của đại lượng ngẫu nhiên X là:

$$f(x) = \begin{cases} A.x.e^{-2x} & \text{với } x \ge 0 \\ 0 & \text{với } x < 0 \end{cases}$$

a. Tìm A.

b. Tìm hàm phân phối xác suất F(x) của X.

c. Tìm kỳ vọng của X.

17. Cho hàm $F(x) = \begin{cases} 0 & \text{nếu } x < -2 \\ \frac{1}{2} + \frac{1}{\pi}\arcsin\left(\frac{x}{2}\right) & \text{nếu } -2 \le x \le 2 \\ 1 & \text{nếu } x > 2 \end{cases}$

a. Tính P(-1 < X < 1).

b. Tính xác suất sao cho trong 4 lần quan sát độc lập về X có hai lần X nhận giá trị thuộc khoảng (-1,1).

c. Tìm hàm mật độ xác suất f(x) của X.

18. Cho đại lượng ngẫu nhiên X có hàm phân phối xác suất:

$$F(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ \frac{1}{4}x(4-3x) & \text{nếu } 0 \le x \le 1 \\ m-1 & \text{nếu } x > 1 \end{cases}$$

a. Tính hằng số m.

b. Tìm c sao cho $P(0 < X \le c) = \frac{1}{4}$.

c. Tính E(X).

19. Năng suất của 3 loại máy cùng sản xuất 1 loại sản phẩm là đại lượng ngẫu nhiên X_1, X_2, X_3 có luật phân phối xác suất như sau:

X_1	1	2	3	4
P	0,1	0,2	0,5	0,2

X_2	2	4	5
P	0,4	0,3	0,3

X_3	2	3	4	5
P	0,1	0,4	0,4	0,1

Giả bạn cần mua 1 trong 3 loại máy này với giả thiết giá của 3 loại máy này như nhau thì bạn sẽ mua loại máy nào, tại sao?

Bước học 3: MỘT SỐ QUI LUẬT PHÂN PHỐI XÁC SUẤT THÔNG DỤNG

3.1 Phân phối nhị thức:

Định nghĩa: *Đại lượng ngẫu nhiên rời rạc có thể nhận 1 trong các giá trị 0, 1, 2,…,n với các xác suất tương ứng được tính theo công thức Bernoulli là:*

$P_x = P(X = x) = C_n^x p^x q^{n-x}$ *được gọi là có phân phối nhị thức với tham số n và p*

Phân phối nhị thức, kí hiệu: **B(n;p)**

Đại lượng ngẫu nhiên X có phân phối nhị thức kí hiệu là $X \in B(n,p)$ hay $X \sim B(n,p)$

Công thức: Với h là số nguyên dương thỏa $h \le n - x$ thì:

$$P(x \le X \le x+h) = P_x + P_{x+1} + .. + P_{x+h} \quad \text{với } P_x = C_n^x p^x q^{n-x}$$

Ví dụ 1: *Tỷ lệ phế phẩm trong 1 lô hàng là 3%. Lấy ngẫu nhiên lần lượt 100 sản phẩm ra để kiểm tra. Tính xác suất để:*

a) *Có 3 phế phẩm.*

b) *Có không quá 3 phế phẩm.*

Mỗi lần kiểm tra một sản phẩm là thực hiện một phép thử. Do đó lấy lần lượt 100 sản phẩm ra để kiểm tra, ta xem như thực hiện 100 phép thử độc lập, khi đó n = 100.

Gọi A là biến cố sản phẩm lấy ra là phế phẩm

$$P(A) = P = 3\% = 0{,}03$$

Gọi X là số phế phẩm có trong 100 sản phẩm lấy ra, có: $X \in [0;100]$, X là đại lượng ngẫu nhiên rời rạc $\Rightarrow X \in B(100; 0{,}03)$

a) $P(X = 3) = C_{100}^{3}(0{,}03)^3(0{,}97)^{97}$

b) $P(0 \le X \le 3) = P_0 + P_1 + P_2 + P_3$

$= C_{100}^{0}.(0{,}03)^0.(0{,}97)^{100} + C_{100}^{1}.(0{,}03)^1.(0{,}97)^{99} + C_{100}^{2}.(0{,}03)^2.(0{,}97)^{98} + C_{100}^{3}.(0{,}03)^3.(0{,}97)^{97}$

= 0,647

+ ***Nhận xét***: *Trong phân phối nhị thức, nếu n khá lớn và xác suất p không quá gần 0 và 1 thì ta có công thức xấp xỉ sau:*

i) $P(X=x) = C_n^x p^x q^{n-x} \approx \dfrac{1}{\sqrt{npq}} f(u)$ với $u = \dfrac{x - np}{\sqrt{npq}}$, $f(u) = \dfrac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}}$

Công thức trên được gọi là **công thức địa phương Laplace.**

**** Chú ý:*** Các giá trị của hàm f(u) đã tính thành bảng (được tính trong bảng giá trị hàm Gauss).

ii) $P(x \le X \le x+h) = \varphi(u_2) - \varphi(u_1)$

Với $u_1 = \dfrac{x - np}{\sqrt{npq}}$, $u_2 = \dfrac{x + h - np}{\sqrt{npq}}$, $\varphi(u) = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_0^u e^{\frac{-t^2}{2}} dt$

Công thức trên được gọi là **công thức tích phân Laplace.**

- ***Chú ý:*** Hàm f(u) là hàm chẵn, hàm φ(u) là hàm lẻ.

Các giá trị của hàm φ(u) đã tính thành bảng (được tính trong bảng giá trị hàm Laplace).

⊕ Các tham số đặc trưng:

Nếu $X \in B(n,p)$ thì E(X) = np

Var(X) = npq

np - q ≤ mod(X) ≤ np + p

Ví dụ 2: *Một máy sản xuất được 200 sản phẩm trong một ngày. Xác suất để máy sản xuất ra phế phẩm là 0,05. Tìm số phế phẩm trung bình và số phế phẩm có khả năng tin chắc của máy đó trong một ngày.*

Gọi X là số phế phẩm của máy trong một ngày thì $X \in B(200; 0{,}05)$

Số phế phẩm trung bình của máy trong một ngày là:

E(X) = np = 200.0,05 = 10

Số phế phẩm tin chắc trong một ngày là mod(X). Ta có:

np – q = 200.0,05 – 0,95 = 9,05

np + p = 200.0,05 + 0,05 = 10,05

$\Rightarrow$ 9,05 $\leq$ mod(X) $\leq$ 10,05

Vì X $\in$ B(200; 0,05) nên mod(X) $\in$ **Z**. Do đó mod(X) = 10

Ví dụ 3: *Một nhà máy sản xuất sản phẩm với tỉ lệ phế phẩm là 20%.*

a) *Nếu lấy từ nhà máy ra 5 sản phẩm. Tính xác suất để được 2 phế phẩm.*

b) *Nếu lấy từ nhà máy ra 400 sản phẩm:*

i) *Tính xác suất được 80 phế phẩm.*

ii) *Tính xác suất được từ 60 đến 80 phế phẩm.*

iii) *Tính xem trung bình có bao nhiêu phế phẩm.*

Giải

a) Gọi X là số phế phẩm trong 5 sản phẩm chọn ra.

Ta có: X $\in$ B(5;0,2)

Suy ra: $P(X=2)=C_5^2(0,2)^2(0,8)^3=10.(0,04).(0,512)=0,2048$

b) Gọi Y là số phế phẩm có trong 400 sản phẩm chọn ra.

Ta có: Y $\in$ B(400 ;0,2)

Do n = 400, 0 << p = 0,2 << 1 nên ta có thể áp dụng công thức xấp xỉ:

i) $P(Y=80)=C_{400}^{80}(0,2)^{80}(0,8)^{320}$

$$=\frac{1}{\sqrt{400.(0,2)(0,8)}}l\left(\frac{80-400(0,2)}{\sqrt{400.(0,2)(0,8)}}\right)$$

$$=\frac{1}{8}l(0)=\frac{1}{8}0,3989=0,0499$$

ii) $P(60\leq Y\leq 80)=\varphi\left(\frac{80-400.(0,2)}{\sqrt{400.(0,2).(0,8)}}\right)-\varphi\left(\frac{60-400.(0,2)}{\sqrt{400.(0,2).(0,8)}}\right)$

$$=\varphi(0)-\varphi(-2,5)=\varphi(0)+\varphi(2,5)$$

$$=0+0,4938=0,4938$$

iii) E(Y) = n.p = 400.(0,2) = 80

Vậy trung bình có 80 phế phẩm trong 400 sản phẩm chọn ra.

3.2 Phân phối Poison:

i) **Công thức:** *Giả sử X là đại lượng ngẫu nhiên có phân phối nhị thức với tham số n và p. Khi n khá lớn và* ***np = a*** *(hằng số) thì từ công thức Bernoulli, ta có công thức xấp xỉ:*

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

Khi đó ta sẽ dùng công thức: $P(x=k)=\frac{a^k}{k!}e^{-a}$ thay cho công thức Bernoulli và được gọi là **công thức Poison.**

Bảng phân phối xác suất:

X	0	1	…	k	…
P				$e^{-a}\frac{a^k}{k!}$	

ii) **Định nghĩa:** *Đại lượng ngẫu nhiên rời rạc X có thể nhận 1 trong các giá trị 0,1,2,.., n với các xác suất tương ứng được tính theo công thức Poison được gọi là có phân phối Poison với tham số là a, kí hiệu là* $X \in \boldsymbol{\rho(a)}$ *hay* $X \sim \boldsymbol{\rho(a)}$

⊕ Chú ý:

$$P(k \le X \le k+h) = P_k + P_{k+1} + .. + P_{k+h} \quad \text{với } P_k = \frac{a^k}{k!}e^{-a}$$

Ví dụ 4: *Một nhà máy dệt có 1000 ống sợi. Xác suất để trong 1 giờ máy hoạt động có 1 ống sợi bị đứt là 0,002. Tính xác suất để trong 1 giờ máy hoạt động:*

a) Có 2 ống sợi bị đứt.

b) *Có không quá 2 ống sợi bị đứt.*

Vì n khá lớn n =1000; p = 0,002 $\Rightarrow$ np = 2

Việc quan sát ống sợi xem như là một phép thử, mà ta có n = 1000 ống sợi nên có 1000 phép thử độc lập.

Gọi A là biến cố ống sợi bị đứt và X là số ống sợi bị đứt trong 1 giờ máy hoạt động, ta có:

$P = P(A) = 0,002$

$\Rightarrow X \in B(1000; 0,002)$

Nhưng vì n khá lớn và np = 2 = a (hằng số) $\Rightarrow X \in \rho(2)$

a) Ta cần tính P(X=2)

b) Xác suất để có không quá 2 ống sợi bị đứt là $P(0 \le X \le 2)$

Ta có: $P(0 \le X \le 2) = P_0 + P_1 + P_2$, trong đó:

$P_0 = P(X=0) = \frac{2^0}{0!}.e^{-2}$

$P_1 = P(X=1) = \frac{2^1}{1!}.e^{-2}$

$P_2 = P(X=2) = \frac{2^2}{2!}.e^{-2}$

Do đó: $P(0 \le X \le 2) = (1+2+2)e^{-2} = 0,6808$

♥ Các tham số đặc trưng:

Nếu $X \in \rho(a)$ thì $E(X) = Var(X) = a$

và a – 1 ≤ mod(X) ≤ a

♥ ***Các ứng dụng của đại lượng ngẫu nhiên có phân phối Poison:***

Số lỗi in sai trong một trang (hoặc một số trang) của một cuốn sách, số người trong một cộng đồng sống cho tới 100 tuổi, số cuộc điện thoại gọi sai trong một ngày, số transitor hư trong ngày đầu tiên sử dụng, số khách hàng vào bưu điện trong một ngày, số hạt α phát ra từ các hạt phóng xạ trong một chu kỳ,..

3.3 Phân phối siêu bội:

i). **Bài toán:** *Cho 1 tập hợp gồm N phần tử trong đó có M phần tử có tính chất A. Lấy ngẫu nhiên ra n phần tử. Gọi X là số phần tử có tính chất A có trong n phần tử lấy ra. Khi đó, X là đại lượng ngẫu nhiên rời rạc có thể nhận các giá trị 0,1,2,.. ,n với các xác suất tương ứng là:*

$$P(X=k)=\frac{C_M^K.C_{N-M}^{n-k}}{C_N^n}$$, gọi là **công thức Siêu bội.**

ii) **Định nghĩa:** *Đại lượng ngẫu nhiên rời rạc X có thể nhận các giá trị 0, 1, 2,.. ,n với các xác suất tương ứng được tính theo* ***công thức Siêu bội*** *được gọi là có phân phối Siêu bội với tham số N, M, n.*

Kí hiệu: X ∈ **H(N, M, n)**

Bảng phân phối xác suất:

X	0	1	2	…	k	…	n
P				…	$\frac{C_M^k.C_{N-M}^{n-k}}{C_N^n}$	…	

Ví dụ 5: *Một lô hàng gồm có 10 sản phẩm, trong đó có 4 phế phẩm. Lấy ngẫu nhiên 4 sản phẩm từ lô hàng. Tính xác suất để trong 4 sản phẩm lấy ra chỉ có 1 phế phẩm.*

Gọi X là số sản phẩm tốt trong 4 sản phẩm lấy ra thì X là đại lượng ngẫu nhiên có phân phối siêu bội với tham số N = 10, M = 6 và n = 4

Xác suất để trong 4 sản phẩm lấy ra chỉ có 1 phế phẩm là:

$$P(X=3)=\frac{C_6^3.C_4^1}{C_{10}^4}=\frac{8}{21}=0{,}3809$$

Chú ý: Nếu n << N thì $\frac{C_M^K.C_{N-M}^{n-k}}{C_N^n}\approx C_n{}^k p^k q^{n-k}$ với p = $\frac{M}{N}$

Như vậy: Khi n << N, ta có thể xem như X∈ B(n;p) và p = $\frac{M}{N}$

♥ ***Các tham số đặc trưng:***

Nếu X ∈ H(N;M;n) thì:

E(X) = np , với p = $\frac{M}{N}$

$$Var(X) = npq\,\frac{N-n}{N-1}$$

Ví dụ 6: *Gọi X là số cây At trong 3 cây bài lấy ra từ bộ bài 52 cây. Hãy tính: E(X), Var(X) và $\sigma(X)$.*

Ta có: X ~ H(N, M, n). Với N = 52, M = 4, n = 3.

$\Rightarrow$ p = $\frac{M}{N} = \frac{4}{25} = \frac{1}{13}$ $\Rightarrow$ q = 1 – p = 1 - $\frac{1}{13} = \frac{12}{13}$

Ta được: E(X) = n.p = $3.\frac{1}{13} = 0{,}231$.

Var(X) = npq.$\frac{N-n}{N-1} = 3.\frac{1}{13}.\frac{12}{13}.\frac{52-3}{52-1} = 0{,}051$.

$\sigma(X) = \sqrt{0{,}051} = 0{,}226$

Ví dụ 7: *Một trường gồm có 10000 sinh viên, trong đó có 1000 học kém. Một Đoàn thanh tra đến trường, chọn ngẫu nhiên 100 sinh viên để kiểm tra. Tính xác suất để có 20 sinh viên học kém.*

Gọi X là số sinh viên học kém trong 100 sinh viên được chọn ra.

Ta có: X $\in$ H(10000; 1000; 100)

Suy ra: $P(X=20) = \frac{C_{1000}^{20}C_{9000}^{80}}{C_{10000}^{100}}$

Vì N = 10000 rất lớn, n = 100 << 10000 = N nên X xấp xỉ phân phối nhị thức: X $\in$ B(100; 0,1) với p = $\frac{M}{N} = \frac{1000}{10000} = 0{,}1$.

Mặt khác, do n = 100 rất lớn và 0 << p = 0,1 << 1 nên ta có thể áp dụng công thức xấp xỉ sau:

$$P(X=20) = C_{100}^{20}(0{,}1)^{20}(0{,}9)^{80} = \frac{1}{\sqrt{100.(0{,}1).(0{,}9)}}\,f\left(\frac{20-100.(0{,}1)}{\sqrt{100.(0{,}1).(0{,}9)}}\right)$$

$$= \frac{1}{3}.f\left(\frac{10}{3}\right) = \frac{1}{3}.f(3{,}33) = \frac{1}{3}.0{,}0017 = 0{,}00057.$$

Bảng tổng kết các phân phối rời rạc

Phân phối	Kí hiệu	Xác suất P(X = k)	E(X)	Var(X)
Nhị thức	B(n , p)	$C_n^k p^k (1-p)^{n-k}$	np	npq
Siêu bội	H(N, M, n)	$\frac{C_M^k . C_{N-M}^{n-k}}{C_N^n}$	np (p = $\frac{M}{N}$)	npq$\frac{N-n}{N-1}$
Poisson	ρ(a)	$e^{-a}\frac{a^k}{k!}$	a	a

3.4 Phân phối chuẩn:

3.4.1 Phân phối chuẩn:

Định nghĩa: *Đại lượng ngẫu nhiên liên tục X nhận giá trị từ $-\infty$ đến $+\infty$ với hàm mật độ xác suất:*

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

Trong đó: μ là hằng số, $0 < \delta$: hằng số, $-\infty < x < +\infty$.

được gọi là có luật phân phối chuẩn với tham số μ, σ.

Kí hiệu: $X \in N(\mu;\sigma)$ hay $X \sim N(\mu;\sigma)$.

Khảo sát hàm số f(x):

+ $f'(x) = -f(x)\left(\frac{x-\mu}{\sigma^2}\right) \quad \Rightarrow f'(x) = 0 \Leftrightarrow x = \mu$.

+ $f''(x) = f(x)\left[\left(\frac{x-\mu}{\sigma}\right)^2 - 1\right] \quad \Rightarrow f''(x) = 0 \Leftrightarrow x = \mu \pm \sigma$

+ $\lim\limits_{x\to\infty} f(x) = 0 \quad \Rightarrow$ Đồ thị hàm số f(x) nhận trục Ox làm tiệm cận ngang.

+ Bảng biến thiên:

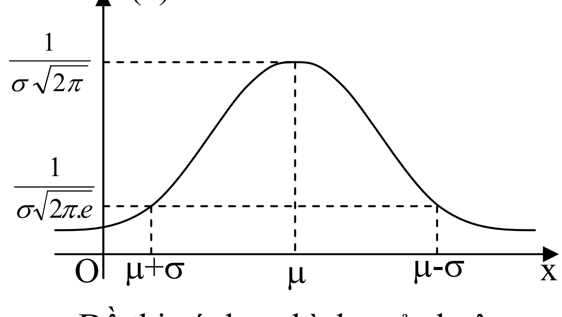
X	$-\infty$		$\mu-\sigma$		μ		$\mu+\sigma$		$+\infty$
f'(x)		+		+	0	-		-	
f''(x)		+	0	-		-	0	+	
f(x)	0		$\frac{1}{\sigma\sqrt{2\pi.e}}$		$\frac{1}{\sigma\sqrt{2\pi}}$		$\frac{1}{\sigma\sqrt{2\pi.e}}$		0

Đồ thị đạt giá trị cực đại $M\left(\mu, \frac{1}{\sigma\sqrt{2\pi}}\right)$.

Và có 2 điểm uốn : $U_1\left(\mu-\sigma, \frac{1}{\sigma\sqrt{2\pi.e}}\right)$, $U_2\left(\mu+\sigma, \frac{1}{\sigma\sqrt{2\pi.e}}\right)$.

Có trục đối xứng : $x = \mu$.

. Đồ thị (hình 19):



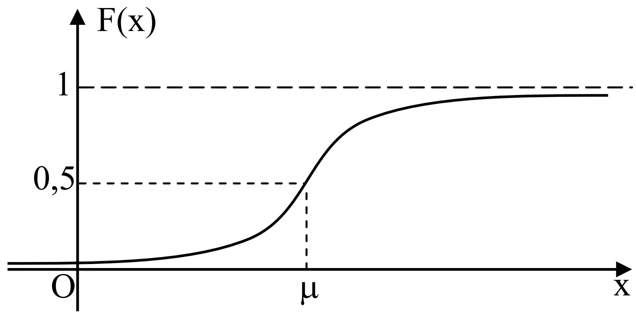
Đồ thị có dạng hình quả chuông

Hình 19

Hàm phân phối xác suất F(x)**:**

Hàm phân phối xác suất F(x) không biểu diễn được thành hàm sơ cấp:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$



Hình 20

Đồ thị hàm sin F(x) có tâm đối xứng: (μ ; 0,5) (hình 20)

Các tham số đặc trưng:

a) Mod(X) = μ.

b) E(X) = μ.

c) Var(X) = σ^2.

Ví dụ 8: *Cho X ∈ N(5,9). Hãy viết hàm mật độ xác suất f(x), chỉ ra tọa độ đỉnh của đồ thị hàm số f(x) và xác định các đặt trưng số: E(X), Var(X), σ(X), mod(X).*

Ta có: $X \in N(5;9) \Rightarrow \begin{cases} \mu = 5 \\ \sigma = \sqrt{9} = 3 \end{cases}$

Hàm mật độ f(x): f(x) = $\frac{1}{3\sqrt{2\pi}}.e^{-\frac{(x-5)^2}{18}}$.

Tọa độ đỉnh: $M\begin{cases} x = 5 \\ y = \frac{1}{3\sqrt{2\pi}} \approx 0{,}133 \end{cases}$

Đặc trưng số:

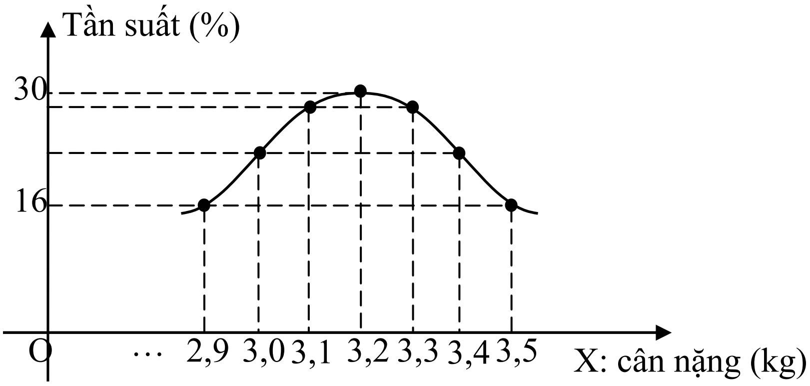
. Mod(X) = 5; E(X) = 5; Var(X) = 9; σ(X) = 3.

Ứng dụng: *Phân phối chuẩn có ý nghĩa rất lớn trong thực tế. Rất nhiều đại lượng ngẫu nhiên có luật phân phối chuẩn. Những đại lượng ngẫu nhiên có liên quan đến số lượng lớn, chịu ảnh hưởng của các yếu tố cân bằng nhau thường có luật phân phối chuẩn. Chẳng hạn:*

- ✓ *Các chỉ số sinh học (cân bằng, chiều cao,...) của người cùng giới tính và cùng độ tuổi.*
- ✓ *Các chỉ số sinh học của các loài cây, loài vật cùng độ tuổi.*

✓ *Khối lượng, kích thước của các sản phẩm do cùng 1 hệ thống máy sản xuất ra.*

Ví dụ 9: *Gọi X là cân nặng của trẻ sơ sinh ở khu vực dân cư lớn. Khảo sát đại lượng ngẫu nhiên X sau 1 khoảng thời gian như hình vẽ. Với khoảng chia 0,1 kg, nối các điểm biểu diễn tần suất, ta được đường gấp khúc có điểm cao nhất tại X = 3,2 kg. Nếu các điểm chia được mịn hơn, đường tần số có dạng "trơn" dần, xấp xỉ 1 phần đường hình chuông của phân phối chuẩn với đỉnh tại X = 3,2 kg, do đó có thể nói rằng trung bình căn nặng trẻ sơ sinh khu vực đó là E(X) = 3,2 kg.*



Hình 21

3.4.2 Phân phối chuẩn tắc:

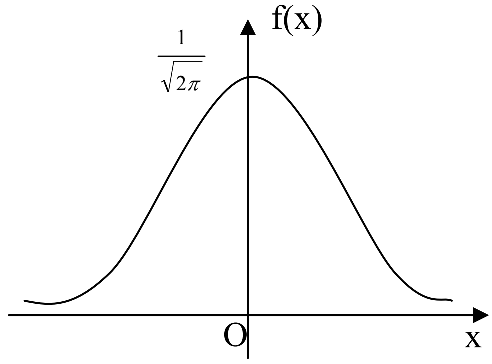
Định nghĩa: *Đại lượng ngẫu nhiên liên tục T có luật phân phối chuẩn với $\mu = 0$ và $\sigma = 1$, được gọi là có luật phân phối chuẩn tắc.*

Kí hiệu: $T \in N(0;1)$ hay $T \sim N(0;1)$.

Hàm mật độ xác suất :

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$ được gọi là hàm Laplace.

Đồ thị hàm mật độ xác suất: có dạng đường cong hình chuông đối xứng qua trục tung (hình 22).



Hình 22

+ Cực đại : M $\left(0, \frac{1}{\sqrt{2\pi}}\right)$

+ 2 điểm uốn :

$U_1\left(-1, \frac{1}{\sqrt{2\pi.e}}\right)$

$U_2\left(1, \frac{1}{\sqrt{2\pi.e}}\right)$.

+ Tiệm cận ngang: trục Ox.

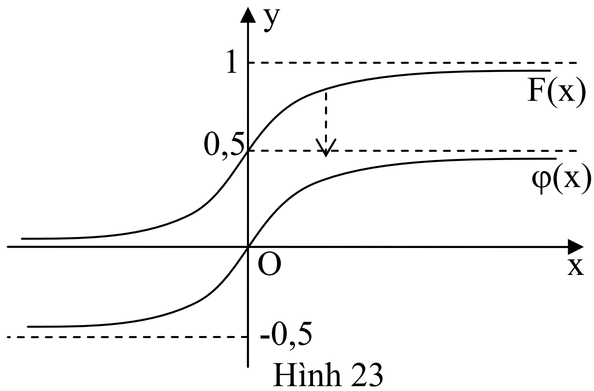
+ Hàm f(x) là hàm chẵn: f(x) = f(-x).

+ Giá trị của hàm f(x) được cho trong bảng phụ lục.

Hàm phân phối xác suất: Hàm phân phối xác suất, kí hiệu: F(x), được gọi là hàm Gauss:

$$F(x) = \int_{-\infty}^{x} l(t)dt = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$$

Đồ thị hàm phân phối xác suất F(x) đối xứng qua điểm có tọa độ: x = 0, y = 0,5 (hình 23).



Hình 23

Hàm tích phân Laplace:

+ Đặt :

$$\varphi(x) = \int_{0}^{x} l(t)dt = \frac{1}{\sqrt{2\pi}}\int_{0}^{x} e^{-\frac{t^2}{2}}dt$$

+ φ(x) được gọi là tích phân Laplace. Giá trị của hàm φ(x) được cho trong bảng phụ lục.

+ F(x) = φ(x) + 0,5.

+ φ(x) là hàm lẻ : φ(x) = - φ(-x).

+ $P(\alpha \le T \le \beta) = \varphi(\beta) - \varphi(\alpha)$.

Thật vậy : $P(\alpha \le T \le \beta) = \int_{\alpha}^{\beta} f(t)dt = \int_{\alpha}^{0} f(t)dt + \int_{0}^{\beta} f(t)dt$

$= \int_{0}^{\beta} f(t)dt - \int_{0}^{\alpha} f(t)dt = \varphi(\beta) - \varphi(\alpha)$ (đpcm)

Các tham số đặc trưng:

a) Mod(T) = 0.
b) E(T) = 0.
c) Var(T) = 1.

Ví dụ 10: *Cho đại lượng ngẫu nhiên T có luật phân phối chuẩn tắc, Tính* $P(-1 \le T \le 2)$*, mod(T), E(T), Var(T).*

Ta có T ∈ N(0;1).

Suy ra: $P(-1 \le T \le 2)$ = φ(2) - φ(-1) = φ(2) + φ(1) = 0,4772 + 0,3413 = 0,8185.

+ Mod(T) = 0; E(T) = 0; Var(T) = 1.

Định lý: *$X \in N(\mu, \sigma^2) \Rightarrow T = \frac{X-\mu}{\sigma} \in N(0,1)$. Tức là nếu X phân phối chuẩn với tham số μ và σ^2 thì $\frac{X-\mu}{\sigma}$ có phân phối chuẩn tắc.*

Hệ quả: *Cho $X \in N(\mu, \sigma^2)$, ta có:*

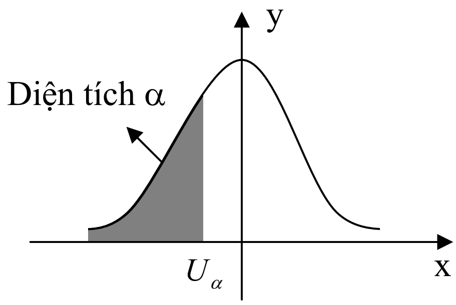
a) $P(x_1 \le X \le x_2) = \varphi\left(\frac{x_2-\mu}{\sigma}\right) - \varphi\left(\frac{x_1-\mu}{\sigma}\right)$.

b) $P(|X-\mu| \le \varepsilon) = 2\varphi\left(\frac{\varepsilon}{\sigma}\right)$.

c) $P(|X-\mu| \le 3\sigma) \approx 1$

Phân vị chuẩn: *Phân vị chuẩn mức α, kí hiệu U_α, là giá trị của đại lượng ngẫu nhiên U có phân phối chuẩn hóa thỏa mãn điều kiện: $P(U < U_\alpha) = \alpha$.*

Với α cho trước có thể tính được các giá trị của U_α. Các giá trị của U_α được cho trong bảng phụ lục.



Hình 24

3.5 Phân phối mũ:

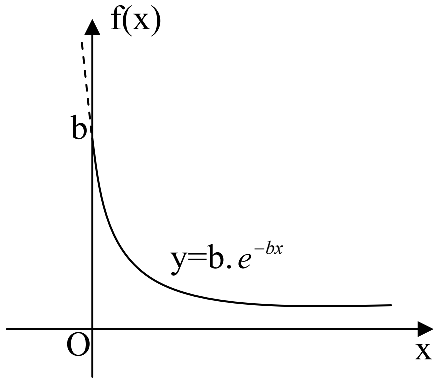
Định nghĩa: *Đại lượng ngẫu nhiên liên tục X có hàm mật độ xác suất:*

$$f(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ b.e^{-bx} & \text{nếu } x \ge 0 \end{cases} \quad (b > 0)$$

được gọi là có luật phân phối mũ với tham số b.

Kí hiệu: $X \in E(b)$ hay $X \sim E(b)$.

Đồ thị hàm f(x): (hình 17)



Hình 17

Hàm phân phối xác suất:

Ta có: $F(x) = \int_{-\infty}^{x} f(x)dx$

✓ $x < 0$: $F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} 0dx = 0$

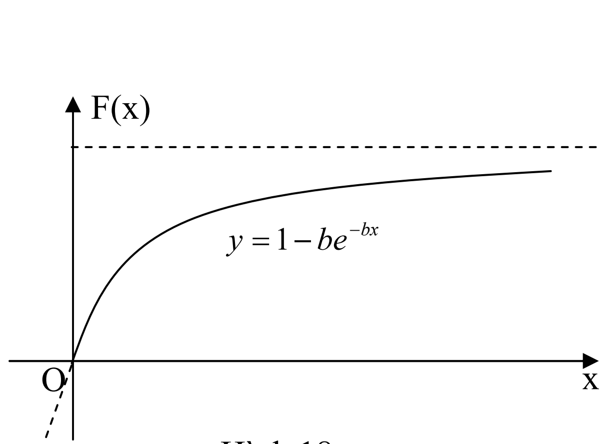
✓ $x \ge 0$: $F(x) = \int_{-\infty}^{x} f(x)dx =$

$\int_{0}^{x} b.e^{-bx}dx = -e^{-bx}\Big|_0^x = -e^{-bx} + 1 = 1 - e^{-bx}$

Vậy ta có:

$$F(x) = \begin{cases} 0 & , x < 0 \\ 1\text{- be}^{-bx} & , x \ge 0 \end{cases}$$

Đồ thị hàm F(x): (hình 18)



Hình 18

Các tham số đặc trưng:

a) $E(X) = \frac{1}{b}$.

b) $Var(X) = \frac{1}{b^2}$.

Ví dụ 11: *Giả sử tuổi thọ (tính bằng năm) của 1 mạch điện tử trong máy tính là đại lượng ngẫu nhiên có luật phân phối mũ với kỳ vọng là 6,25. Thời gian bảo hành của mạch điện tử này là 5 năm. Hỏi có bao nhiêu % mạch điện tử bán ra phải thay thế trong thời gian bảo hành?*

Gọi X là tuổi thọ của mạch điện tử.

Ta có: E(X) = 6,25.

Thời gian bảo hành 5 năm.

Mặt khác: $E(X) = \frac{1}{b} = 6{,}25 \Rightarrow b = \frac{1}{6{,}25}$

Và hàm phân phối xác suất của X:

$$F(x) = \begin{cases} 0 & , x < 0 \\ 1 - be^{-bx} & , x \geq 0 \end{cases}$$

Suy ra: $P(X \leq 5) = F(5) = 1 - e^{-b5}$

$$= 1 - e^{-\frac{1}{6{,}25}5} = 1 - e^{-0{,}8} = 1 - 0{,}449 = 0{,}551$$

Vậy có khoảng 55,1% mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

Ứng dụng: *Khoảng thời gian giữa hai lần xuất hiện của một biến có luật phân phối mũ. Chẳng hạn khoảng thời gian giữa hai ca cấp cứu ở một bệnh viện, giữa hai lần hỏng hóc của một cái máy, giữa hai trận lụt hay động đất là những đại lượng ngẫu nhiên có luật phân phối mũ.*

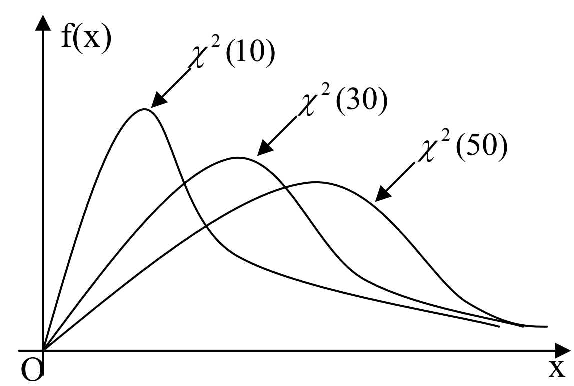
3.6 Phân phối χ^2:

Định nghĩa: *Cho các đại lượng ngẫu nhiên X_i, $i = \overline{1,n}$ độc lập với nhau cùng có luật phân phối chuẩn tắc. Khi đó đại lượng ngẫu nhiên $\chi^2 = \sum_{i=1}^{n} X_i^2$ được gọi là có luật phân phối khi bình phương, bậc tự do n.*

Kí hiệu: $\chi^2 \in \chi^2(n)$ hay $\chi^2 \sim \chi^2(n)$.

Hàm mật độ xác suất:

$$f(x) = \begin{cases} 0 & \text{nếu } x \leq 0 \\ \dfrac{e^{-\frac{x}{2}}.x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}.\Gamma(\frac{n}{2})} & \text{nếu } x > 0 \end{cases}$$



Hình 25

Trong đó hàm $\Gamma(u) = \int_0^{+\infty} t^{u-1}.e^{-t}dt$, có tên gọi là hàm Gamma, Γ(1) = 1, Γ (u + 1) = u. Γ (u).

Chú ý: Đại lượng ngẫu nhiên χ^2 nhận giá trị không âm.

Đồ thị hàm mật độ f(x): *là đường cong không đối xứng (hình 25). Khi bậc tự do n ≥ 30, đồ thị hàm f(x) gần đối xứng (dạng hình chuông), phân phối χ^2 tiệm cận phân phối chuẩn.*

Các đặc trưng số:

a) $E(\chi^2) = n$.

b) $Var(\chi^2) = 2n$.

Phân vị χ^2 : Phân vị χ^2 mức α, kí hiệu χ^2_α **,** là giá trị của đại lượng ngẫu nhiên χ^2 có phân phối "khi bình phương" với bậc tự do n thoả mãn: $P(\chi^2 < \chi^2_\alpha) = \alpha$

Các giá trị của χ^2_α được cho trong bảng phụ lục.

3.7 Phân phối Student:

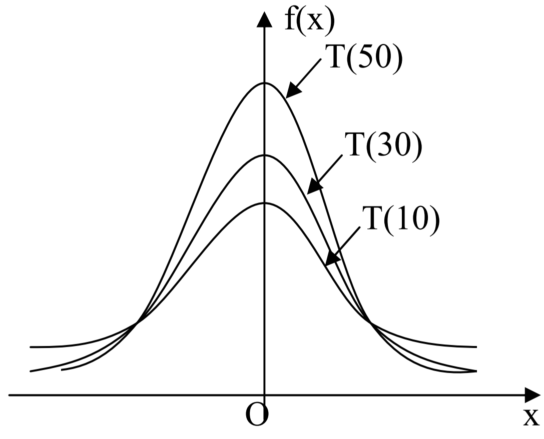
Định nghĩa: *Cho đại lượng ngẫu nhiên U ∈ N(0,1), $\chi^2 \in \chi^2(n)$, trong đó U và χ^2 độc lập nhau. Khi đó đại lượng ngẫu nhiên:*

$$T = \frac{U}{\sqrt{\frac{X^2}{n}}} = \frac{\sqrt{n}.U}{\sqrt{X^2}}$$ được gọi là có luật phân phối Student bậc tự do n.

Kí hiệu: T ∈ T(n).

Hàm phân phối xác suất:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}.\Gamma\left(\frac{n}{2}\right)}.\left(1+\frac{x^2}{2}\right)^{-\frac{n+1}{2}}$$



Hình 26

Đồ thị hàm f(x) là đường cong đối xứng qua trục tung (hình 26). Khi bậc tự do n ≥ 30, đồ thị hàm f(x) tiệm cận đồ thị hàm Laplace. Khi n ≥ 30 , phân phối Student xấp xỉ phân phối chuẩn tắc.

Các đặc trưng số:

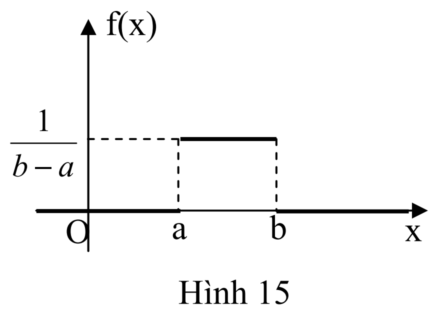
a) E(T) = 0.

b) $Var(T) = \frac{n}{n-2}$.

8. Phân phối đều:

Định nghĩa: *Đại lượng ngẫu nhiên liên tục X có hàm mật độ xác suất:*

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



Hình 15

được gọi là có luật phân phối đều trên đoạn [a; b].

Kí hiệu: $X \in R[a;b]$ hay $X \sim R[a;b]$

Đồ thị hàm f(x): (hình 15)

Hàm phân phối xác suất:

Ta có: $F(x) = \int_{-\infty}^{x} f(x)dx$

- ✓ $x < a$: $F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} 0dx = 0$.
- ✓ $a \le x \le b$: $F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{a} f(x)dx + \int_{a}^{x} f(x)dx$

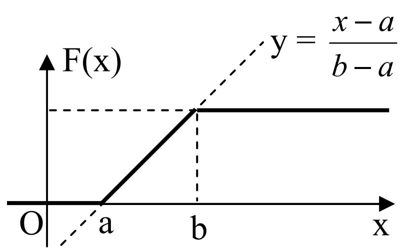
$$= \int_{-\infty}^{a} 0dx + \int_{a}^{x} \frac{1}{b-a}dx = \frac{1}{b-a}x\Big|_a^x = \frac{x-a}{b-a}$$

- ✓ $x > b$: $F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{a} f(x)dx + \int_{a}^{b} f(x)dx + \int_{b}^{x} f(x)dx$

$$= \int_{-\infty}^{a} 0dx + \int_{a}^{b} f(x)dx + \int_{b}^{x} 0dx = \int_{a}^{b} \frac{1}{b-a}dx = 1$$

Vậy ta có:

$$F(x) = \begin{cases} 0 & \text{nếu} \quad x < a \\ \dfrac{x - a}{b - a} & \text{nếu} \quad a \le x \le b \\ 1 & \text{nếu} \quad x > b \end{cases}$$



Hình 16

Đồ thị của hàm F(x) Hình 16):

Các tham số đặc trưng :

a) $E(X) = \dfrac{a+b}{2}$

b) $Var(X) = \dfrac{(a-b)^2}{12}$

Ví dụ 12: *Lịch chạy của xe buýt tại một trạm xe buýt như sau: chiếc xe buýt đầu tiên trong ngày sẽ khởi hành từ trạm này vào lúc 7 giờ, cứ sau mỗi 15 phút sẽ có một xe khác đến trạm. Giả sử một hành khách đến trạm trong khoảng thời gian từ 7 giờ đến 7 giờ 30. Tìm xác suất để hành khách này chờ:*

a) *Ít hơn 5 phút.*

b) *Ít nhất 12 phút.*

Gọi X là số phút sau 7 giờ mà hành khách đến trạm.

Ta có: $X \in R[0;30]$.

a) Hành khách sẽ chờ ít hơn 5 phút nếu đến trạm giữa 7 giờ 10 và 7 giờ 15 hoặc giữa 7 giờ 25 và 7 giờ 30. Do đó xác suất cần tìm là:

$$P(10 < X < 15) + P(25 < X < 30) = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}$$

b) Hành khách chờ ít nhất 12 phút nếu đến trạm giữa 7 giờ và 7 giờ 3 phút hoặc giữa 7 giờ 15 phút và 7 giờ 18 phút. Xác suất cần tìm là:

$$P(0 < X < 3) + P(15 < X < 18) = \frac{3}{30} + \frac{3}{30} = \frac{1}{5}$$

Bảng tóm tắt các phân phối liên tục

Phân phối	**Kí hiệu**	**Hàm mật độ f(x)**	**E(X)**	**Var(X)**
Đều	R[a;b]	$\frac{1}{b-a}(a \le x \le b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Mũ	E(b)	$b.e^{-bx}\,(x > 0)$	$\frac{1}{b}$	$\frac{1}{b^2}$
Chuẩn	N(μ;σ)	$\frac{1}{\sigma\sqrt{2\pi}}.e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Chuẩn tắc	N(0;1)	$\frac{1}{\sqrt{2\pi}}.e^{-\frac{x^2}{2}}$	0	1
Khi bình phương	$\chi^2(n)$	$\frac{e^{-\frac{x}{2}}.x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}.\Gamma(\frac{n}{2})}\quad (x > 0, n > 0)$	n	2n
Student	T(n)	$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}.\Gamma\left(\frac{n}{2}\right)}.\left(1+\frac{x^2}{2}\right)^{-\frac{n+1}{2}}\quad$ (n > 0)	0 (n > 1)	$\frac{n}{n-2}$

BÀI TẬP

1. Xác suất để 1 con gà đẻ mỗi ngày là 0,6.

a. Trong chuồng có 10 con, tính xác suất để một ngày có 8 con đẻ.

b. Phải nuôi ít nhất bao nhiêu con để mỗi ngày trung bình thu được không ít hơn 30 trứng.

2. Sản phẩm xuất xưởng của nhà máy có tới 70% sản phẩm loại A. Lấy ngẫu nhiên 10 sản phẩm.

a. Tính xác suất để có 8 sản phẩm loại A.

b. Nếu muốn có trung bình 15 sản phẩm loại A thì phả kiểm tra bao nhiêu sản phẩm?

3. Một loại sản phẩm do 3 nhà máy sản xuất với tỉ lệ là 20%, 30%, 50%. Tỉ lệ phế phẩm của các nhà máy lần lượt là: 0,1; 0,2; 0,3.

a. Lấy ngẫu nhiên 1 sản phẩm. Tính xác suất để được sản phẩm tốt.

b. Nếu lấy lần lượt (có hoàn lại) 4 sản phẩm. Gọi X là số sản phẩm xấu. Tìm qui luật phân phối xác suất của X.

c. Tím xác suất sao cho trong 20 sản phẩm lấy ra có 4 sản phẩm xấu.

4. Ba phân xưởng cùng sản xuất 1 loại sản phẩm. Tỉ lệ sản phẩm loại II của các phân xưởng tương ứng là: 10%, 20%, 30%. Từ lô hàng gồm 10.000 sản phẩm (trong đó có 3.000 sản phẩm của phân xưởng I, 4.000 sản phẩm của phân xưởng II và 3.000 sản phẩm của phân xưởng III). Người ta lấy ngẫu nhiên ra 100 sản phẩm để kiểm tra (lấy có hoàn lại). Nếu thấy có không quá 24 sản phẩm loại II thì nhận lô hàng. Tìm xác suất để nhận lô hàng đó?

5. Sản phẩm được đóng thành hộp. Mỗi hộp có 10 sản phẩm trong đó có 7 sản phẩm loại A. Người mua hàng qui định cách kiểm tra như sau: Từ hộp lấy ngẫu nhiên 3 sản phẩm nếu thấy cả 3 sản phẩm đều loại A thì nhận hộp đó. Nếu ngược lại thì loại hộp.

Giả sử kiểm tra 100 hộp (trong rất nhiều hộp). Tính xác suất để:

a. Có 25 hộp được nhận.

b. Có không quá 30 hộp được nhận.

c. Phải kiểm tra ít nhất bao nhiêu hộp để xác suất có ít nhất một hộp được nhận không nhỏ hơn 95%?

6. Hai nhà máy cùng sản xuất 1 loại sản phẩm. Tỉ lệ sản phẩm loại I của nhà máy A là 85%, của nhà máy B là 90%. Một người mua 50 sản phẩm của nhà máy A và 40 sản phẩm của nhà máy B. Tìm số sản phẩm loại I tin chắc nhất mà người đó có thể mua được.

7. Một nhà máy theo công thức thiết kế sẽ sản xuất được 80% sản phẩm loại I. Nhưng trong thực tế sản phẩm loại I chỉ bằng 90% thiết kế. Tính xác suất để khi lấy 125 sản phẩm do nhà máy đó sản xuất có ít nhất 100 sản phẩm loại I.

8. Một sinh viên thi trắc nghiệm môn Vật Lý gồm 100 câu hỏi. Mỗi câu có 4 phần để chọn, trong đó chỉ có 1 câu đúng. Giả sử sinh viên chỉ chọn ngẫu nhiên các phần trả lời của câu hỏi.

a. Tìm xác suất sao cho sinh viên đó trả lời đúng 40 câu hỏi.

b. Tìm xác suất sao cho sinh viên đó trả lời đúng từ 40 đến 60 câu hỏi.

c. Tính xem số câu hỏi trung bình mà sinh viên đó trả lời đúng là bao nhiêu.

9. Giả sử mỗi cặp vợ chồng trong một xã nào đó sinh 3 con và khả năng có con trai và con gái trong mỗi lần sinh là như nhau.

a. Gọi X là số con gái trong mỗi gia đình. Lập bảng phân phối xác suất của X.

b. Tính xác suất để trong 100 gia đình có 50 gia đình có số con gái nhiều hơn con trai.

c. Theo bạn thì trong 100 gia đình có trung bình bao nhiêu gia đình có duy nhất 1 đứa con gái.

10. Khi tiêm truyền một loại huyết thanh trung bình có 1 trường hợp bị phản ứng trên 1000 ca. Ta dùng loại huyết thanh trên tiêm cho 2000 người. Tìm xác suất để:

a. Có 3 ca bị phản ứng.

b. Nhiều nhất 3 ca bị phản ứng.

c. Hơn 3 ca bị phản ứng.

11. Xác suất bắn trúng máy bay của một khẩu súng là 0,001. Có 5000 khẩu súng bắn lên một lượt. Người ta biết rằng máy bay chắc chắn bị hạ nếu có ít nhất hai viên trúng. Nếu có 1 viên trúng thì xác suất bị hạ 80%. Tính xác suất để máy bay bị hạ?

12. Hằng ngày ở phòng cấp cứu trung bình có 5 ca tới cấp cứu. Tính xác suất để:

a. Có hơn 10 ca tới cấp cứu.

b. Có 11 ca tới cấp cứu.

c. Có không ít hơn 12 ca tới cấp cứu.

13. Tại 1 trạm giao thông trung bình 1 phút có 2 xe ô tô đi qua. Tính xác suất để có đúng 6 xe ô tô đi qua trong vòng 3 phút. (Số xe ô tô đi qua trạm giao thông là đại lượng ngẫu nhiên có luật phân phối Poisson).

14. Một mạch điện gồm 1000 bóng đèn mắc song song. Xác suất để mỗi bóng đèn bị hư tại mỗi thời điểm là 0,002. Tính xác suất để tại một thời điểm:

a. Không có bóng đèn nào bị hư.

b. Có nhiều hơn 5 bóng đèn bị hư.

c. Hãy cho biết số bóng đèn bị hư trung bình tại một thời điểm.

15. Một cái máy gồm 5000 bộ phận. Xác suất để mỗi bộ phận không hoạt động tại một thời điểm là 0,001. Biết rằng nếu có từ hai bộ phận trở lên không hoạt động thì máy không hoạt động. Nếu có một bộ phận không hoạt động thì máy sẽ không hoạt động với xác suất là 50%. Tính xác suất để máy không hoạt động.

16.

a. Một lô hàng có tỉ lệ phế phẩm thật sự là 0,02. Theo qui định nếu lô hàng có tỉ lệ phế phẩm nhỏ hơn 0,06 thì nhận lô hàng đó. Một người kiểm tra lô hàng bằng cách làm như sau: Lấy từ lô hàng 100 sản phẩm (có hoàn lại) nếu có không quá 3 phế phẩm thì nhận lô hàng đó. Nếu lớn hơn hoặc bằng 9 thì không nhận. Nếu có từ 4 – 8 thì lấy thêm một mẫu khác gồm 50 sản phẩm. Trong 50 sản phẩm nếu có không quá 2 phế phẩm thì nhận, nếu có từ 3 phế phẩm trở lên thì không nhận. Tính xác suất mắc phải sai lầm của cách kiểm tra của người đó.

b. Nếu lô hàng thật sự có tỉ lệ phế phẩm là 0,08. Khi kiểm tra 1000 lô hàng bằng cách làm như trên thì trung bình có bao nhiêu lô hàng mắc phải sai lầm.

17. Độ dài của một chi tiết máy là một đại lượng ngẫu nhiên có phân phối chuẩn với trung bình E(X) = 20 cm, Var(X) = 0,04 cm^2.

a. Tính xác suất để lấy được một chi tiết máy thì độ dài chi tiết máy nằm trong khoảng (19,8cm ; 20,1cm).

b. Những chi tiết sai lệch so với trung bình nhỏ hơn 0,3 cm được coi là loại tốt. Tính tỉ lệ chi tiết loại tốt của máy đó.

c. Nếu muốn tỉ lệ chi tiết loại tốt là 90% thì độ dài chi tiết sai lệch so với trung bình là bao nhiêu?

18. Trọng lượng trẻ sơ sinh là đại lượng ngẫu nhiên X có phân phối chuẩn với trọng lượng trung bình là 3 kg. Độ lệch chuẩn $\sigma = 0,2$ kg. Biết trọng lượng đứa trẻ sinh ra có trọng lượng tối thiểu là 1,5 kg.

a. Tính trọng lượng trẻ sơ sinh cân nặng từ 3 kg đến 3,4 kg.

b. Trẻ sơ sinh thiếu cân nếu có trọng lượng nhỏ hơn 2,5 kg. Tính tỉ lệ trẻ thiếu cân.

c. Người ta muốn có chế độ chăm sóc đặc biệt cho 10% tổng số trẻ nhẹ cân nhất. Tính trọng lượng tối đa cho những đứa trẻ được chăm sóc đặc biệt.

19. Giả sử trọng lượng của những hộp sữa do một nhà máy sản xuất phân phối theo qui luật chuẩn N(400;100). Một hộp sữa được xem như đạt yêu cầu nếu trọng lượng của chúng sai lệch so với trung bình nhỏ hơn 5 gam.

a. Tính tỉ lệ những hộp sữa dưới mức yêu cầu.

b. Giả sử muốn có 80% hộp sữa sai lệch so với trung bình m gam. Hãy xác định m.

20. Thời gian đi từ nhà tới trường của sinh viên A là một đại lượng ngẫu nhiên T (Đơn vị là phút) có phân phối chuẩn. Biết rằng 65% số ngày A đến trường mất 20 phút và 8% số ngày mất hơn 30 phút.

a. Tính thời gian trung bình đến trường của A và độ lệch tiêu chuẩn.

b. Giả sử A xuất phát trước giờ học 25 phút. Tính xác suất để A bị trễ học.

c. A cần phải xuất phát trước giờ học bao nhiêu phút để xác suất bị trễ học của A bé hơn 2%.

Bước học 4: ĐẠI LƯỢNG NGẪU NHIÊN HAI CHIỀU

4.1 Định nghĩa:

Một cặp (X, Y) gồm hai biến X và Y được gọi là đại lượng ngẫu nhiên hai chiều X, Y. (X, Y được gọi là các thành phần của đại lượng ngẫu nhiên hai chiều).

Đại lượng ngẫu nhiên hai chiều X, Y được gọi là rời rạc nếu X, Y là các đại lượng ngẫu nhiên rời rạc.

Đại lượng ngẫu nhiên hai chiều X, Y được gọi là liên tục nếu X, Y là các đại lượng ngẫu nhiên liên tục.

4.2 Phân phối xác suất của đại lượng ngẫu nhiên hai chiều:

4.2.1 Bảng phân phối xác suất:

Giả sử X, Y là đại lượng ngẫu nhiên 2 chiều nhận các giá trị lần lượt là x_i (i=1,..,n) và y_j (j=1,..,m) với các xác suất tương ứng $P_{ij} = P(x=x_i, y=y_j)$ là xác suất để X nhận giá trị x_i và Y nhận giá trị y_j. Khi đó, bảng phân phối xác suất của X, Y là:

Y / X	y_1	y_2	. . .	y_j	. . .	y_m
x_1	P_{11}	P_{12}	. . .	P_{1j}	. . .	P_{1n}
x_2	P_{21}	P_{22}	. . .	P_{2j}	. . .	P_{2n}

. . .	. . .	. . .	. . .	. . .	. . .	. . .
x_i	P_{i1}	P_{i2}	. . .	P_{ij}	. . .	P_{in}
. . .	. . .	. . .	. . .	. . .	. . .	. . .
x_n	P_{n1}	P_{n2}	. . .	P_{nj}	. . .	P_{nm}

Trong đó:

- $x_i(i=\overline{1,n})$ là các giá trị có thể của thành phần X.
- $y_j(j=\overline{1,m})$ là các giá trị có thể của thành phần Y.
- $P(x_i,y_j)=P((X,Y)=(x_i,y_j))=P(X=x_i,Y=y_j)$, $i=\overline{1,n}$, $j=\overline{1,m}$.
- $\sum_{i=1}^{n}\sum_{j=1}^{m}P(x_i,y_j)=1$.

Từ bảng phân phối xác suất của (X, Y), ta lập được bảng phân phối xác suất của X và Y với công thức:

$$P(X=x_i)=\sum_{j=1}^{m}P_{ij};P(Y=y_{\hat{i}})=\sum_{i=1}^{n}P_{ij}$$

4.2.2 Hàm phân phối xác suất:

Định nghĩa: *Hàm phân phối xác suất của đại lượng ngẫu nhiên 2 chiều (X,Y) là hàm của hai đối số X và Y được xác định bởi công thức:*

F(X,Y) = P(X < x, Y < y)

Trong đó: X < x , Y < y là các biến cố đồng thời.

Tính chất:

i) $\lim\limits_{y\to+\infty} F(x,y) = F_1(x)$ $\qquad$ $\lim\limits_{x\to+\infty} F(x,y) = F_2(y)$

$F_1(x)$, $F_2(y)$ lần lượt là hàm phân phối xác suất của X và Y.

ii) $\lim\limits_{\substack{x\to+\infty\\ y\to+\infty}} F(x,y) = 1$ $\qquad$ $\lim\limits_{\substack{x\to-\infty\\ y\to-\infty}} F(x,y) = 0$

4.2.3 Hàm mật độ xác suất:

Định nghĩa: *Hàm mật độ xác suất của đại lượng ngẫu nhiên 2 chiều liên tục (X,Y), kí hiệu f(x,y) là đạo hàm riêng hỗn hợp của hàm F(x,y)*

$$f(x,y)=\frac{\partial^2F(x,y)}{\partial x.\partial y}$$

♣ ***Tính chất:***

i) $\quad f(x,y) \geq 0$

ii) $$P[(X,Y)\in D]=\iint_D f(x,y)dxdy$$

iii) $$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$$

iv) $$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdy$$

$$f_1(x)=\int_{-\infty}^{+\infty} f(x,y)dy\,; \qquad f_2(y)=\int_{-\infty}^{+\infty} f(x,y)dx$$

4.3 Các tham số đặc trưng của hàm một biến ngẫu nhiên:

4.3.1 Trường hợp (X,Y) rời rạc:

Công thức:

$$E(X)=\sum_{i=1}^{n}\sum_{j=1}^{m} x_i P(x_i,y_j)\ ; \qquad E(Y)=\sum_{j=1}^{m}\sum_{i=1}^{n} y_j P(x_i,y_j)$$

$$Var(X)=\sum_{i=1}^{n}\sum_{j=1}^{m} x_i^2 P(x_i,y_j)-\left[E(X)\right]^2;\quad Var(Y)=\sum_{j=1}^{m}\sum_{i=1}^{n} y_j^2 P(x_i,y_j)-\left[E(Y)\right]^2$$

Ví dụ 1: *Ta có 4 lô sản phẩm, mỗi lô đều có 10 sản phẩm. Lô thứ i có i sản phẩm hỏng* $(i=\overline{1,4})$. *Ta tung một khối tứ diện đều (các mặt của khối có ghi số chấm từ 1 đến 4). Nếu mặt có i chấm chạm bàn thì ta chọn lô thứ i và từ đó lấy ra 1 sản phẩm. Gọi X là số chấm của tứ diện chạm bàn và Y là số sản phẩm xấu lấy được. Hãy lập bảng phân phối xác suất của đại lượng ngẫu nhiên N hai chiều (X,Y). Tính kỳ vọng và phương sai các thành phần.*

Ta có: X có thể nhận các giá trị: 1, 2, 3, 4.

Y có thể nhận các giá trị: 0, 1.

Suy ra:

$$P(1,0) = P(X = 1,Y = 0) = \frac{1}{4}.\frac{9}{10}=0,225.$$

$$P(1,1) = P(X = 1,Y = 1) = \frac{1}{4}.\frac{1}{10}=0,025.$$

$$P(2,0) = P(X = 2,Y = 2) = \frac{1}{4}.\frac{8}{10}=0,200.$$

$$P(2,1) = P(X = 2,Y = 1) = \frac{1}{4}.\frac{2}{10}=0,050.$$

$$P(3,0) = P(X = 3,Y = 0) = \frac{1}{4}.\frac{7}{10}=0,175.$$

$$P(3,1) = P(X = 3,Y = 1) = \frac{1}{4}.\frac{3}{10}=0,075.$$

$$P(4,0) = P(X = 4, Y = 0) = \frac{1}{4}.\frac{6}{10} = 0,150.$$

$$P(4,1) = P(X = 4, Y = 1) = \frac{1}{4}.\frac{4}{10} = 0,100.$$

Vậy bảng phân phối xác suất:

X\Y	0	1
1	0,225	0,025
2	0,200	0,050
3	0,175	0,075
4	0,150	0,100

Kỳ vọng thành phần:

$$E(X) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i P(x_i, y_j)$$
$$= 1.(0,225 + 0,025) + 2.(0,2 + 0,05) + 3.(0,175 + 0,075) + 4.(0,15 + 0,1)$$
$$= 0,25 + 0,5 + 0,75 + 1 = 2,5$$

$$E(Y) = \sum_{j=1}^{m}\sum_{i=1}^{n} y_j P(x_i, y_j)$$
$$= 0.(0,225 + 0,2 + 0,175 + 0,15) + 1.(0,025 + 0,05 + 0,075 + 0,1) = 0,25$$

Phương sai thành phần:

$$Var(X) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i^2 P(x_i, y_j) - [E(X)]^2$$
$$= 1^2.(0,225 + 0,025) + 2^2.(0,2 + 0,05) + 3^2.(0,175 + 0,075) + 4^2.(0,15 + 0,1) - (2,5)^2$$
$$= 0,25 + 1 + 2,25 + 4 - 6,25 = 7,5 - 6,25 = 1,25$$

$$Var(Y) = \sum_{j=1}^{m}\sum_{i=1}^{n} y_j^2 P(x_i, y_j) - [E(Y)]^2$$
$$= 0^2.(0,225 + 0,2 + 0,175 + 0,15) + 1^2.(0,025 + 0,05 + 0,075 + 0,1) - (0,25)^2$$
$$= 0,25 - 0,0625 = 0,1875$$

4.3.2 Trường hợp (X,Y) liên tục:

Công thức:

$$E(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)dxdy; \qquad E(Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)dxdy$$

$$Var(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2 f(x,y)dxdy - [E(X)]^2; \quad Var(Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^2 f(x,y)dxdy - [E(Y)]^2$$

Ví dụ 2: *Cho đại lượng ngẫu nhiên hai chiều (X,Y) có hàm mật độ xác suất như sau:*

$$f(x,y)=\begin{cases}\dfrac{A}{\sqrt{xy}} & 0<x\le y<1\\ 0 & \text{trường hợp khác}\end{cases} \qquad (A>0)$$

a) *Xác định A.*

b) *Tính: E(X), E(Y), Var(X), Var(Y).*

Giải

a) Theo giả thiết f(x,y) là hàm mật độ xác suất của đại lượng ngẫu nhiên hai chiều (X,Y) nên suy ra:

$$1=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)dxdy=\int_0^1\left(\int_0^y\frac{A}{\sqrt{xy}}dx\right)dy=\int_0^1 A\left(\int_0^y (xy)^{-\frac{1}{2}}dx\right)dy$$

$$=A\int_0^1 y^{-\frac{1}{2}}\left(\int_0^y x^{-\frac{1}{2}}dx\right)dy=A\int_0^1 y^{-\frac{1}{2}}\left(2x^{\frac{1}{2}}\Big|_0^y\right)dy=2A\int_0^1 y^{-\frac{1}{2}}.y^{\frac{1}{2}}dy\ =2A\int_0^1 dy=2A$$

$$\Rightarrow A=\frac{1}{2}$$

Vậy: $f(x,y)=\begin{cases}\dfrac{1}{2\sqrt{xy}} & 0<x\le y<1\\ 0 & \text{trường hợp khác}\end{cases}$

b) Tính:

$$E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy=\int_0^1\left(\int_0^y x\frac{1}{2\sqrt{xy}}dx\right)dy=\frac{1}{2}\int_0^1\left(\int_0^y xx^{-\frac{1}{2}}y^{-\frac{1}{2}}dx\right)dy$$

$$=\frac{1}{2}\int_0^1 y^{-\frac{1}{2}}\left(\int_0^y x^{\frac{1}{2}}dx\right)dy=\frac{1}{2}\int_0^1 y^{-\frac{1}{2}}\left(\frac{2}{3}x^{\frac{3}{2}}\Big|_0^y\right)dy=\frac{1}{3}\int_0^1 y^{-\frac{1}{2}}y^{\frac{3}{2}}dy$$

$$=\frac{1}{3}\int_0^1 ydy=\frac{1}{3}\frac{y^2}{2}\Big|_0^1=\frac{1}{6}\approx 0{,}167$$

$$E(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)dxdy=\int_0^1\left(\int_0^y y\frac{1}{2\sqrt{xy}}dx\right)dy=\frac{1}{2}\int_0^1\left(\int_0^y yx^{-\frac{1}{2}}y^{-\frac{1}{2}}dx\right)dy$$

$$=\frac{1}{2}\int_0^1 y^{\frac{1}{2}}\left(\int_0^y x^{-\frac{1}{2}}dx\right)dy=\frac{1}{2}\int_0^1 y^{\frac{1}{2}}\left(2x^{\frac{1}{2}}\Big|_0^y\right)dy=\int_0^1 y^{\frac{1}{2}}y^{\frac{1}{2}}dy$$

$$=\int_0^1 ydy=\frac{y^2}{2}\Big|_0^1=\frac{1}{2}\approx 0{,}5$$

$$Var(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy-\left[E(X)\right]^2=\int_0^1\left(\int_0^y x^2\frac{1}{2\sqrt{xy}}dx\right)dy-\left(\frac{1}{6}\right)^2$$

$$=\frac{1}{2}\int_0^1\left(\int_0^y x^2x^{-\frac{1}{2}}y^{-\frac{1}{2}}dx\right)dy-\frac{1}{36}=\frac{1}{2}\int_0^1 y^{-\frac{1}{2}}\left(\int_0^y x^{\frac{3}{2}}dx\right)dy-\frac{1}{36}$$

$$= \frac{1}{2}\int_0^1 y^{-\frac{1}{2}}\left(\frac{2}{5}x^{\frac{5}{2}}\Big|_0^y\right)dy - \frac{1}{36} = \frac{1}{5}\int_0^1 y^{-\frac{1}{2}}y^{\frac{5}{2}}dy - \frac{1}{36}$$

$$= \frac{1}{5}\int_0^1 y^2 dy - \frac{1}{36} = \frac{1}{5}\frac{y^3}{3}\Big|_0^1 - \frac{1}{6} = \frac{1}{15} - \frac{1}{6} = \frac{7}{210} = 0{,}033$$

$$E(Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^2 f(x,y)dxdy - [E(X)]^2 = \int_0^1\left(\int_0^y y^2 \frac{1}{2\sqrt{xy}}dx\right)dy - (0{,}5)^2$$

$$= \frac{1}{2}\int_0^1\left(\int_0^y y^2 x^{-\frac{1}{2}} y^{-\frac{1}{2}}dx\right)dy - 0{,}25 = \frac{1}{2}\int_0^1 y^{\frac{3}{2}}\left(\int_0^y x^{-\frac{1}{2}}dx\right)dy - 0{,}25$$

$$= \frac{1}{2}\int_0^1 y^{\frac{3}{2}}\left(2x^{\frac{1}{2}}\Big|_0^y\right)dy - 0{,}25 = \int_0^1 y^{\frac{3}{2}}y^{\frac{1}{2}}dy - 0{,}25$$

$$= \int_0^1 y^2 dy - 0{,}25 = \frac{y^3}{3}\Big|_0^1 - 0{,}25 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \approx 0{,}083$$

4.4. Hàm của các đại lượng ngẫu nhiên:

4.4.1 Hàm một biến ngẫu nhiên:

Định nghĩa: *Cho các đại lượng ngẫu nhiên X, Y, Z, ... , T. Hàm số Q = φ(X, Y, Z, ... , T) được gọi là hàm của các đại lượng ngẫu nhiên.*

Trong đó:

- ✓ Mỗi bộ X = x, Y = y, Z = z, ... , T = t thì qui tắc φ xác định 1 giá trị duy nhất của đại lượng ngẫu nhiên Q.
- ✓ Qui tắc φ đươc lập thành từ các phép toán và các hàm toán học thông thường.

Định nghĩa: *Nếu X, Y, Z, ... , T là các đại lượng ngẫu nhiên rời rạc (liên tục) thì Q được gọi là hàm của các đại lượng ngẫu nhiên rời rạc (liên tục).*

Ví dụ 3: Xét đại lượng ngẫu nhiên Z = $X_1^2 + 3X_1X_2 - \ln X_3$ là hàm của các đại lượng ngẫu nhiên độc lập X_1, X_2, X_3. Giả sử $X_1 = 2$, $X_2 = 4, X_3 = 1$ với các xác suất tương ứng là $p_1 = 0{,}2$; $p_2 = 0{,}3$; $p_3 = 0{,}4$ thì đại lượng ngẫu nhiên Z nhận giá trị 28 với xác suất p = (0,2)(0,3)(0,4) = 0,024.

Việc nghiên cứu hàm Z của các đại lượng ngẫu nhiên trên rất khó khăn, đặc biệt là các giá xác suất của Z. Các phần tiếp theo chỉ ra luật phân phối xác suất của Z trong một số trường hợp riêng.

4.4.2 Hàm của các đại lượng ngẫu nhiên rời rạc:

Định nghĩa: *Cho đại lượng ngẫu nhiên rời rạc X, Y = φ(X) được gọi là hàm của 1 đại lượng ngẫu nhiên rời rạc.*

Bảng phân phối xác suất:

Bảng phân phối xác suất của Y = φ(X) có dạng:

Y	y_1	y_2	y_3	.....	y_n

P	p_1	p_2	p_3	.....	p_n

Trong đó:

. y_i: Các giá trị có thể có của Y được tìm từ các giá trị có thể có của X thông qua hàm φ.

. p_i: Xác suất tương ứng của y_i: $p_i = \sum\{p(X = x_i) / \varphi(x_i) = y_i\}$.

Ví dụ 4: *Cho đại lượng ngẫu nhiên rời rạc X có bảng phân phối xác suất như sau:*

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

Hãy lập bảng phân phối xác suất của Y = φ(X) = $X^2 + 1$.

Với các giá trị của X được cho trong bảng và hàm số Y = φ(X) = $X^2 + 1$ ta suy ra các giá trị Y có thể nhận là:1, 2, 5.

Ta có: P(Y = 1) = P(X = 0) = 0,3.

P(Y = 2) = P(X = -1) + P(X = 1) = 0,3.

P(Y = 5) = P(X = 2) = 0,4.

Bảng phân phối xác suất của Y là:

X	1	2	5
P	0,3	0,3	0,4

Các tham số đặc trưng:

i) Ta có thể xem Y như 1 đại lượng ngẫu nhiên, rồi áp dụng các công thức đã biết:

. E(Y) = $\sum y_i p_i$.

. Var(Y) = $\sum y_i^2 p_i - [E(Y)]^2$.

ii) Công thức:

. E(Y) = $\sum \varphi(x_i).p_i$.

. Var(Y) = $\sum \varphi(x_i)^2.p_i - [E(Y)]^2$.

Ví dụ 5: *Với giả thiết cho trong* ***ví dụ 4****, hãy tìm E(Y), Var(Y).*

Ta có:

E(Y) = $\sum y_i p_i$ = 1.(0,3) + 2.(0,3) + 5.(0,4) = 2,9

(= $\sum \varphi(x_i).p_i$ = 2.(0,2) + 1.(0,3) + 2.(0,1) +5.(0,4) = 2,9).

Var(Y) = $\sum y_i^2 p_i - [E(Y)]^2 = (1^2.(0,3) + 2^2.(0,3) + 5^2.(0,4) - (2,9)^2 = 3,09$

(= $\sum \varphi(x_i)^2.p_i - [E(Y)]^2 = 2^2.(0,2) + 1^2.(0,3) + 2^2.(0,1) + 5^2.(0,4) = 3,09)$.

4.4.3 Hàm của hai đại lượng ngẫu nhiên rời rạc độc lập:

Định nghĩa: *Cho 2 đại lượng ngẫu nhiên rời rạc độc lập X, Y có luật phân phối xác suất như sau:*

X	x_1	x_2	x_3	.....	x_n
P	p_1	p_2	p_3	.....	p_n

Y	y_1	y_2	y_3	.....	y_m
P	q_1	q_2	q_3	.....	q_m

Đại lượng ngẫu nhiên Z = φ(X,Y) là hàm của 2 đại lượng ngẫu nhiên rời rạc độc lập.

Bảng phân phối xác suất:

Bảng phân phối xác suất của Z có dạng:

Z	z_1	z_2	z_3	.....	z_k
P	r_1	r_2	r_3	.....	r_k

Trong đó:

. z_k: Là các giá trị có thể có của Z được tìm từ các giá trị có thể có của X và Y thông qua hàm φ.

. r_k: Xác suất tương ứng $Z = z_k$:

$$p_k = P(Z = z_k) = \sum\left\{P(X = x_n).P(Y = y_m) / \varphi(x_n, y_m) = z_k\right\}.$$

Ví dụ 6: *Cho X, Y là các đại lượng ngẫu nhiên rời rạc độc lập lần lượt có bảng phân phối xác suất sau:*

X	*0*	*1*	*2*
P	*0,5*	*0,3*	*0,2*

Y	*1*	*2*
P	*0,6*	*0,4*

Lập bảng phân phối xác suất của đại lượng ngẫu nhiên Z = X.Y.

Ta lập bảng sau:

X\Y	0	1	2
1	0	1	2
2	0	2	4

Dựa vào bảng trên ta thấy Z có thể nhận các giá trị: 0, 1, 2, 4.

Ta có:

P(Z = 0) = P(X = 0).P(Y = 1) + P(X = 0).P(Y = 2) = (0,5).(0,6) + (0,5).(0,4) = 0,5

P(Z = 1) = P(X = 1).P(Y = 1) = (0,3).(0,6) = 0,18

P(Z = 2) = P(X = 2).P(Y= 1) + P(X = 1).P(Y = 2) = (0,2).(0,6) + (0,3).(0,4) = 0,24

P(Z = 4) = P(X = 2).P(Y = 2) = (0,2).(0,4) = 0,08

Vậy bảng phân phối xác suất của Z là:

Z	0	1	2	4
P	0,5	0,18	0,24	0,08

Các tham số đặc trưng: *Để tính kỳ vọng và phương sai của Z ta xem Z như là 1 đại lượng ngẫu nhiên rời rạc rồi áp dụng các công thức tính đã biết. Hoặc áp dụng các tính chất của kỳ vọng và phương sai để tính.*

Ví dụ 7: *Với giả thiết trong **ví dụ 6**, hãy tính E(Z) và Var(Z).*

Ta có:

$$E(Z) = \sum z_i p_i = 0.(0,5) + 1.(0,18) + 2.(0,24) + 4.(0,08) = 0,98$$

$$Var(Z) = \sum z_i^2 p_i - [E(Z)]^2 = 0^2.(0,5) + 1^2.(0,18) + 2^2.(0,24) + 4^2.(0,08) - [0,98]^2 = 1,46.$$

4.4.4 Hàm của các đại lượng ngẫu nhiên liên tục:

Định nghĩa: *Cho đại lượng ngẫu nhiên liên tục X, Y = φ(X) là hàm của 1 đại lượng ngẫu nhiên liên tục.*

Hàm mật độ xác suất: *Khi đại lượng ngẫu nhiên liên tục X có hàm mật độ xác suất f(x), Y = φ(X) là hàm khả vi, đơn điệu tăng hoặc giảm, có hàm ngược là X = W(Y). Khi đó ta có hàm mật độ xác suất của Y là:*

$$g(y) = f(W(y)).W'(y)$$

Ví dụ 8: *Cho đại lượng ngẫu nhiên liên tục X có hàm mật độ xác suất:* $f(x) = \frac{1}{\sqrt{2\pi}}.e^{-\frac{x^2}{2}}$. *Đại lượng ngẫu nhiên Y được xác định như sau: Y = φ(X) = 2X. Tìm hàm mật độ xác suất của Y.*

Ta có Y = 2X là hàm khả vi, luôn tăng và có hàm ngược là: $X = \frac{Y}{2}$.

Suy ra hàm mật độ của Y:

$$g(y) = f\left(\frac{y}{2}\right)\left(\frac{y}{2}\right)' = \frac{1}{\sqrt{2\pi}}.e^{-\frac{y^2}{8}}.\frac{1}{2} = \frac{1}{2\sqrt{2\pi}}e^{-\frac{y^2}{8}}.$$

Các tham số đặc trưng:

$$E(Y) = \int_{-\infty}^{+\infty} \varphi(x) f(x)dx.$$

Var(Y) = $\int_{-\infty}^{+\infty}[\varphi(x)]^2.f(x)dx-[E(Y)]^2$.

Ví dụ 9: *Cho đại lượng ngẫu nhiên liên tục X có hàm mật độ xác suất:*

$$f(x)=\begin{cases}\dfrac{2}{9}(3x-x^2) & \text{nếu } x\in[0,3]\\ 0 & \text{nếu } x\notin[0,3]\end{cases}$$

Và Y = $6X^2(3-X)$*. Tính E(Y), Var(Y).*

Ta có:

$$\text{E(Y)} = \int_{-\infty}^{+\infty}\varphi(x)f(x)dx=\int_0^3 6x^2(3-x).\frac{2}{9}(3x-x^2)dx=\int_0^3\frac{4}{3}(3x^2-x^3)(3x-x^2)dx$$

$$=\int_0^3\frac{4}{3}(9x^3-3x^4-3x^4-x^5)dx=\int_0^3(12x^3-8x^4+\frac{4}{3}x^5)dx$$

$$=\left[3x^4-\frac{8}{5}x^5+\frac{2}{9}x^6\right]_0^3=16{,}2$$

$$\text{Var(Y)} = \int_{-\infty}^{+\infty}[\varphi(x)]^2f(x)dx-[E(x)]^2=\int_0^3[6x^2(3-x)]^2.\frac{2}{9}(3x-x^2)dx-(16{,}2)^2$$

$$=\int_0^3\left(216x^6-\frac{216}{7}x^7+9x^8-\frac{8}{9}x^9\right)dx-262{,}44$$

$$=312{,}43-262{,}44=49{,}99$$

4.4.5 Hàm tổng của hai đại lượng ngẫu nhiên liên tục độc lập nhau:

Định nghĩa: *Cho X, Y là 2 đại lượng ngẫu nhiên liên tục độc lập có hàm mật độ xác suất f(x), g(y) và Z = X + Y. Khi đó, hàm mật độ xác suất của Z được xác định như sau:*

$$H(z)=\int_{-\infty}^{+\infty}f(x).g(z-x)dx=\int_{-\infty}^{+\infty}g(y).f(z-y)dy$$

Ví dụ 10: *Cho X* $\in$ *R[0,2]; Y* $\in$ *R[0,5] và Z = X + Y. Tìm hàm mật độ xác suất của Z.*

Ta có:

$$\text{X} \in \text{R[0,2]} \Rightarrow \begin{cases}\dfrac{1}{2} & \text{nếu } x\in[0,2]\\ 0 & \text{nếu } x\notin[0,2]\end{cases}$$

$$\text{Y} \in \text{R[0,5]} \Rightarrow \begin{cases}\dfrac{1}{5} & \text{nếu } x\in[0,5]\\ 0 & \text{nếu } x\notin[0,5]\end{cases}$$

Hàm mật độ của Z: h(x) = $\int_{-\infty}^{+\infty}f(x).g(z-x)dx$

$$=\int_{-\infty}^{0}0.g(z-x)dx+\int_0^2\frac{1}{2}.g(z-x)dx+\int_2^{+\infty}0.g(z-x)dx=\int_0^2\frac{1}{2}.g(z-x)dx$$

Đặt: $t = z - x \Rightarrow dt = -dx \Rightarrow \left| \begin{array}{l} x=2 \Rightarrow t=z-2 \\ x=0 \Rightarrow t=z \end{array} \right. \Rightarrow h(x) = \frac{1}{2}\int_{z}^{z-2} g(t).(-dt) = \frac{1}{2}\int_{z-2}^{z} g(t).dt$

Xét:

. $z < 0$: $h(z) = \frac{1}{2}\int_{z-2}^{z} 0.dt = 0$.

. $\begin{cases} z \geq 0 \\ z-2 < 0 \end{cases} \Leftrightarrow 0 \leq z < 2$: $h(z) = \frac{1}{2}\int_{z-2}^{0} 0.dt + \frac{1}{2}\int_{0}^{z}\frac{1}{5}dt = \frac{1}{10}t\Big|_{0}^{z} = \frac{z}{10}$

. $\begin{cases} z < 5 \\ z-2 \geq 0 \end{cases} \Leftrightarrow 2 \leq z < 5$: $h(z) = \frac{1}{2}\int_{z-2}^{z}\frac{1}{5}dt = \frac{1}{10}t\Big|_{z-2}^{z} = \frac{z-(z-2)}{10} = \frac{1}{5}$

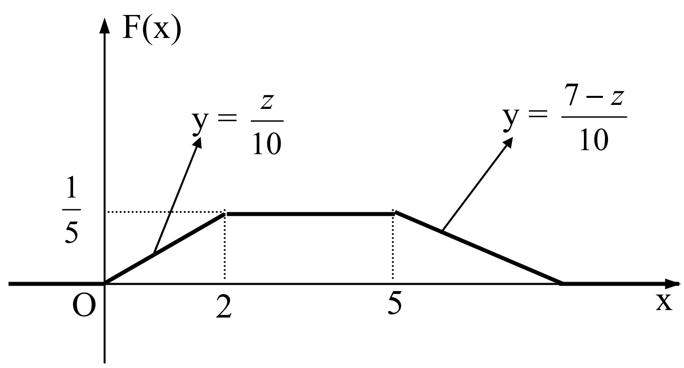
. $\begin{cases} z \geq 5 \\ z-2 < 5 \end{cases} \Leftrightarrow 5 \leq z < 7$: $h(x) = \frac{1}{2}\int_{z-2}^{5}\frac{1}{5}dt + \frac{1}{2}\int_{5}^{z} 0.dt$

$$= \frac{1}{10}t\Big|_{z-2}^{5} = \frac{5-(z-2)}{10} = \frac{7-z}{10}$$

. $z - 2 \geq 5 \Leftrightarrow z \geq 7$: $h(z) = \frac{1}{2}\int_{z-2}^{z} 0.dt = 0$

Vậy: $f(x) = \begin{cases} 0 & \text{nếu } z<0 \\ \frac{z}{10} & \text{nếu } 0\leq z<2 \\ \frac{1}{5} & \text{nếu } 2\leq z<5 \\ \frac{7-z}{10} & \text{nếu } 5\leq z<7 \\ 0 & \text{nếu } z\geq 7 \end{cases}$

Minh họa đồ thị :



Hình 27

BÀI TẬP

1. Cho X, Y là 2 đại lượng ngẫu nhiên có bảng phân phối xác suất đồng thời như sau:

Y \ X	-1	0	1
-1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
1	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

Hãy tính E(X), E(Y).

2. Cho X, Y là 2 đại lượng ngẫu nhiên có bảng phân phối xác suất đồng thời như sau:

Y \ X	-1	0	1
-1	$\frac{4}{15}$	$\frac{1}{15}$	$\frac{4}{15}$
0	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{15}$
1	0	$\frac{2}{15}$	0

a. Hãy tính E(X), E(Y).

b. Đại lượng ngẫu nhiên X và Y có độc lập nhau không?

HÀM CÁC ĐẠI LƯỢNG NGẪU NHIÊN

3. Cho X, Y là 2 đại lượng ngẫu nhiên độc lập nhau có bảng phân phối xác suất như sau:

X	-3	-1	0	2	3
P	0,2	0,1	0,3	0,2	0,2

Y	-2	3	5
P	0,3	0,5	0,2

a. Lập bảng phân phối xác suất của đại lượng ngẫu nhiên X và vẽ đồ thị của nó.

b. Lập bảng phân phối xác suất của $Z = 2X^2 + Y - 2$.

c. Tính E(X), E(X + 3Y), Var(4Z), Var(X^2).

d. Lập bảng phân phối xác suất của Z = X.Y. Tính E(Z) bằng hai cách.

4. Tung đồng thời hai khối vuông cân đối và đồng chất mà trên các mặt của khối vuông được đánh các số: 1, 1, 1, 2, 3, 4.

a. Gọi X_1, X_2 lần lượt là các con số xuất hiện trên mặt của hai khối vuông I và II. Hãy lập bảng phân phối xác suất của Y = $X_1 + X_2$.

b. Tính E(Y), Var(Y).

c. Tìm số lần tung tối thiểu hai khối vuông để có ít nhất một lần con số xuất hiện trên hai mặt khối vuông ấy bằng nhau không bé hơn 0,9.

5. Xạ thủ A có hai viên đạn, xác suất bắn trúng bia là 0,6. Xạ thủ B có 3 viên đạn, ,xác suất bắn trúng bia là 0,7. Họ lần lượt thay phiên nhau A bắn 1 viên rồi B bắn 1 viên ... (Tất nhiên nếu một người đã ngưng bắn do hết đạn thì ngươi kia bắn liên tiếp không cần chờ đến phiên mình) vào một cái bia. Cả hai cùng bắn cho đến khi hoặc hết đạn hoặc bia bị trúng đạn mới thôi. Gọi X_1 là số viên đạn A bắn và X_2 là số viên đạn B bắn, $X = 2X_1 - 3X_2 + 12$.

a. Tìm luật phân phối xác suất của X_1, X_2, X.

b. Tìm luật phân phối xác suất của X_1, X_2, X trong trường hợp hai người bắn vào hai bia khác nhau.

6. Cho X là đại lượng ngẫu nhiên có phân phối mũ với tham số b = 2. Tìm kỳ vọng và độ lệch tiêu chuẩn của đại lượng ngẫu nhiên $Y = e^{-X}$.

7. Cho đại lượng ngẫu nhiên có hàm mật độ xác suất như sau:

$$f(x) = \begin{cases} a - \frac{3}{2} & \text{nếu } x \in [0,1] \\ 0 & \text{nếu } x \notin [0,1] \end{cases}$$

a. Tính a.

b. Lập hàm mật độ xác suất của $Y = 3X^2 + 1$.

8. Giả sử X $\in$ B(2;0,4) và Y $\in$ B(2;0,7). X và Y là 2 đại lượng ngẫu nhiên độc lập.

a. Tìm bảng phân phối xác suất của Z = X + Y.

b. Tính P(X + Y = 4).

c. Chứng minh rằng Z không có phân phối nhị thức.

Bước học 5: LUẬT SỐ LỚN

5.1 Bất đẳng thức Markov:

Định lý: *Nếu X là đại lượng ngẫu nhiên nhận giá trị không âm thì $\forall a > 0$ ta có:*

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Chứng minh: Ta chứng minh trong trường hợp X là đại lượng ngẫu nhiên liên tục có hàm mật độ f(x).

$$E(X) = \int_0^{+\infty} xf(x)dx = \int_0^a xf(x)dx + \int_a^{+\infty} xf(x)dx \geq \int_a^{+\infty} xf(x)dx \geq \int_a^{+\infty} af(x)dx = a\int_a^{+\infty} f(x)dx = aP(X \geq a).$$

$$\Rightarrow P(X \geq a) \leq \frac{E(X)}{a}$$

5.2 Bất đẳng thức Tchebyshev:

Định lý: *Nếu X là đại lượng ngẫu nhiên có kỳ vọng μ và phương sai σ^2 hữu hạn thì $\forall$ $\varepsilon > 0$ bé tùy ý ta có:*

$$P(|X - \mu| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2} \quad \textit{hay} \quad P(|X - \mu| \geq \varepsilon) > 1 - \frac{Var(X)}{\varepsilon^2}$$

Chứng minh:

Ta thấy $(X - \mu)^2$ là đại lượng ngẫu nhiên nhận giá trị không âm.

Áp dụng bất đẳng thức Tchebyshev với a = ε^2 ta được:

$$P\left[(X - \mu)^2 \geq \varepsilon^2\right] \leq \frac{E\left[(X - \mu)^2\right]}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2}$$

Vì $(X - \mu)^2 \geq \varepsilon^2$ khi và chỉ khi $|X - \mu| \geq \varepsilon$ nên $P(|X - \mu| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$

Chú ý: *Bất đẳng thức Markov và Tchebyshev giúp ta phương tiện thấy được giới hạn của xác suất khi kỳ vọng và phương sai của phân phối xác suất chưa biết.*

Ví dụ: *Giả sử số sản phẩm được sản xuất của 1 nhà máy trong một tuần là một đại lượng ngẫu nhiên với kỳ vọng μ = 50.*

a) *Có thể nói gì về xác suất sản phẩm của tuần này vượt quá 75.*

b) *Nếu phương sai của sản phẩm trong tuần này là* $\sigma^2 = 25$ *thì có thể nói gì về xác suất sản phẩm tuần này sẽ ở giữa 40 và 60.*

Giải

a) Theo bất đẳng Markov: $P(X > 75) \geq \dfrac{E(X)}{75} = \dfrac{50}{75} = \dfrac{2}{3}$

b) Theo bất đẳng thức Tchebyshev: $P\left(|x - 50| \geq 10\right) \leq \dfrac{\sigma^2}{10^2} = \dfrac{25}{100} = \dfrac{1}{4}$

Do đó: $P(40 < X < 60) = P\left(|X - 50| < 10\right) > 1 - \dfrac{1}{4} = \dfrac{3}{4}$

5.3 Định lý Tchebyshev:

Định lý*: Nếu các đại lượng ngẫu nhiên* $X_1, X_2, ..., X_n$ *độc lập từng đôi, có kỳ vọng hữu hạn và các phương sai đều bị chặn trên bởi số C thì* $\forall \varepsilon > 0$ *bé tùy ý ta có:*

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right) = 1$$

Đặc biệt, khi $E(X_i) = a; \quad \left(i = \overline{1,n}\right)$ thì $\lim\limits_{n\to\infty}\left(\left|\dfrac{1}{n}\sum\limits_{i=1}^{n} X_i - a\right| < \varepsilon\right) = 1$

Chứng minh: Ta chứng minh trong trường hợp đặc biệt $E(X_i) = \mu,\ Var(X_i) = \sigma^2$ (i = 1, 2, . . ., n). Ta có:

$$E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \mu,$$

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$$

Theo bất đẳng thức Tchebyshev: $P\left(\left|\dfrac{1}{n}\sum\limits_{i=1}^{n} X_i - \mu\right| < \varepsilon\right) \geq 1 - \dfrac{\sigma^2}{n\varepsilon^2}$

Khi n $\to \infty$ ta có: $\lim\limits_{n\to\infty} P\left(\left|\dfrac{1}{n}\sum\limits_{i=1}^{n} X_i - \dfrac{1}{n}\sum\limits_{i=1}^{n} E(X_i)\right| < \varepsilon\right) = 1$

Ý nghĩa: *Mặt dù từng đại lượng ngẫu nhiên độc lập có thể nhận giá trị sai khác nhiều so với kỳ vọng của chúng, nhưng trung bình số học của một số lớn đại lượng ngẫu nhiên lại nhận giá trị gần bằng trung bình số học của các kỳ vọng của chúng. Điều này cho phép ta dự đoán giá trị trung bình số học của các đại lượng ngẫu nhiên.*

5.4 Định lý Bernoulli:

Định lý: *Nếu* f_n *là tần số xuất hiện biến cố A trong n phép thử độc lập và p là xác suất xuất hiện biến cố A trong mỗi phép thử thì* $\forall \varepsilon > 0$ *bé tùy ý ta có*

$$\lim_{n\to\infty} P\left(|f_n - p| < \varepsilon\right) = 1$$

Ý nghĩa: *Tần suất xuất hiện biến cố trong n phép thử độc lập dần về xác suất xuất hiện biến cố trong mỗi phép thử khi số phép thử tăng lên vô hạn.*

KQHT 3: KHÁI NIỆM TỔNG THỂ VÀ MẪU

Bước học 1: TỔNG THỂ VÀ MẪU

1.1 Tổng thể:

Khi nghiên cứu một vấn đề, người ta thường quan tâm đến một dấu hiệu nào đó, các dấu hiệu này thể hiện trên nhiều phần tử và tập hợp tất cả các phần tử mang dấu hiệu được gọi là tổng thể hay đám đông (population).

Ví dụ 1: Khi nghiên cứu về tập hợp gà trong một trại chăn nuôi thì quan tâm đến vấn đề trọng lượng của nó.

Khi nghiên cứu về chất lượng học tập của sinh viên của một trường thì ta quan tâm đến dấu hiệu điểm.

Nghiên cứu các trục máy do một máy tự động tiện ra ta chỉ quan tâm đến dấu hiệu độ dài đường kính của trục máy, khi đó tổng thể là tập hợp các độ dài đường kính của các trục máy….

Chú ý: Trong phần này ta sử dụng một số kí hiệu như sau:

N: Số phần tử của tổng thể hay kích thước của tổng thể.

X*: Dấu hiệu ta quan tâm.

x_i: Giá trị của dấu hiệu X* đo được trên phần tử thứ i

N_i: Tần số của X_i (số phần tử có chung giá trị X_i)

P_i: Tần suất của X_i ($P_i = \dfrac{N_i}{N}$)

♦ ***Các tham số đặc trưng của tổng thể:***

Sự tương ứng giữa các giá trị X_i và tần suất P_i được biểu diễn thành bảng cơ cấu của tổng thể như sau:

Giá trị của X* (x_i)	x_1	x_2	. . . .	x_n
P_i	P_1	P_2	. . . .	P_n

Khi đó trung bình của tổng thể:

$$m = \sum_{i=1}^{n} x_i P_i$$

Phương sai của tổng thể:

$$\sigma^2 = \sum_{i=1}^{n} (x_i - m)^2 P_i$$

Độ lệch tiêu chuẩn của tổng thể:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{n} (x_i - m)^2 P_i}$$

Ví dụ 2: *Khảo sát điểm số của toàn bộ 5000 bài thi cuối năm ở một trường phổ thông trung học ta thu được kết quả như sau:*

Điểm	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Số học sinh có điểm tương ứng N_i	*100*	*150*	*300*	*400*	*450*	*850*	*700*	*750*	*750*	*300*	*250*
Tần suất p_i	*0,02*	*0,03*	*0,06*	*0,08*	*0,09*	*0,17*	*0,14*	*0,15*	*0,15*	*0,06*	*0,05*

Khi đó:

$$+\ \mu = \sum_{i=1}^{k} x_i p_i = 0.(0{,}02) + 1.(0{,}03) + 2.(0{,}06) + 3.(0{,}08) + 4.(0{,}09) + 5.(0{,}17)$$

$$+ 6.(0{,}14) + 7.(0{,}15) + 8.(0{,}15) + 9.(0{,}06) + 10.(0{,}05)$$

$$= 5{,}73$$

$$+\ \sigma^2 = \sum_{i=1}^{k} (x_i - \mu)^2 p_i = \sum_{i=1}^{k} x_i^{\ 2} p_i - \mu^2$$

$$= 0^2.(0{,}02) + 1^2.(0{,}03) + 2^2.(0{,}06) + 3^2.(0{,}08) + 4^2.(0{,}09) + 5^2.(0{,}17)$$

$$+ 6^2.(0{,}14) + 7^2.(0{,}15) + 8.^2(0{,}15) + 9.^2(0{,}06) + 10^2.(0{,}05) - (5{,}73)^2$$

$$= 38{,}23 - 32{,}83 = 5{,}4$$

$$+\ \sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{k} (x_i - \mu)^2 p_i} = \sqrt{5{,}4} = 2{,}3$$

1.2 Mẫu:

Từ tổng thể, ta lấy ra n phần tử và đo lường dấu hiệu X của chúng. Khi đó, n phần tử này lập nên một mẫu và số phần tử của mẫu được gọi là kích thước của mẫu* (*sample*).

Vì từ mẫu, ta kết luận cho tổng thể nên mẫu phải được chọn một cách khách quan để đại diện cho tổng thể.

Có hai phương pháp lấy mẫu:

➢ *Lấy có hoàn lại*: Lấy từ tổng thể ra một phần tử, xem xét xong trả lại tổng thể rồi mới lấy tiếp một phần tử khác để xem xét. Các kết quả ở từng lần lấy sẽ độc lập với nhau.

➢ *Lấy không hoàn lại*: Lấy đồng thời tất cả các phần tử cần xem xét. Khi đó các kết quả xem xét được ở mỗi phần tử sẽ phụ thuộc nhau. Tuy nhiên trong nhiều trường hợp phương pháp lấy mẫu có hoàn lại khá phức tạp, do đó nếu kích thước tổng thể N rất lớn và kích thước mẫu n rất nhỏ (n << N) thì ta có thể thay thế phương pháp lấy mẫu có hoàn lại bằng phương pháp lấy mẫu không hoàn lại, tức lấy đồng thời n phần tử. Khi đó các kết quả cũng có thể xem như độc lập.

1.3 Mô hình xác suất của tổng thể và mẫu:

Từ tổng thể, ta lấy ra một phần tử. Gọi X là giá trị của dấu hiệu X* đo được trên phần tử lấy ra. Khi đó, X là đại lượng ngẫu nhiên có bảng phân phối xác suất như sau:

X	x_1	x_2	....	x_n
P	P_1	P_2	....	P_n

Ta thấy, dấu hiệu X* được mô hình hóa bởi đại lượng ngẫu nhiên X nên X được gọi là đại lượng ngẫu nhiên gốc và phân phối xác suất của X được gọi là phân phối gốc.

♥ **Các tham số đặc trưng của đại lượng ngẫu nhiên X:**

$$E(X) = \sum_{i=1}^{k} x_i P_i \qquad ; \qquad Var(X) = \sum_{i=1}^{n} [x_i - E(X)]^2 . P_i$$

♥ **Mẫu ngẫu nhiên:**

Lấy n phần tử của tổng thể theo phương pháp có hoàn lại để quan sát. Gọi X_i là giá trị (dấu hiệu) của X đo được trên phần tử thứ i (i= 1,..,n), thì $X_1, X_2, .., X_n$ cũng là các đại lượng ngẫu nhiên có cùng phân phối xác suất như đại lượng ngẫu nhiên gốc X. Khi đó, bộ $(X_1, X_2, .. , X_n)$ được gọi là mẫu ngẫu nhiên kích thước n được tạo nên từ đại lượng ngẫu nhiên gốc X và kí hiệu: $W_X = (X_1, X_2, .. , X_n)$.*

Nếu giả sử X_i nhận giá trị x_i thì $(x_1, x_2, .. , x_n)$ được gọi là một mẫu cụ thể của mẫu ngẫu nhiên W_X

Kí hiệu: $w_x = (x_1, x_2, .. , x_n)$

Ví dụ 3: Kết quả điểm môn Toán của một lớp gồm 100 sinh viên cho bởi bảng sau:

Điểm	3	4	5	6	7
Số sinh viên có điểm tương ứng	25	20	40	10	5

Gọi X là điểm môn toán của một sinh viên được chọn ngẫu nhiên trong danh sách lớp thì X là đại lượng ngẫu nhiên có phân phối:

X	3	4	5	6	7
P	0,25	0,2	0,4	0,1	0,05

Chọn ngẫu nhiên 5 sinh viên trong danh sách lớp để xem điểm. Gọi X_i là điểm của sinh viên thứ i. Ta có mẫu ngẫu nhiên kích thước n = 5 được xây dựng từ đại lượng ngẫu nhiên X.

$$W_X = (X_1, X_2, .. , X_5)$$

Giả sử sinh viên thứ nhất được 4 điểm, thứ hai được 3 điểm, thứ ba được 6 điểm, thứ tư được 7 điểm và thứ năm được 5 điểm thì ta được mẫu cụ thể:

$$w_x = (4, 3, 6, 7, 5)$$

Bước học 2: THỐNG KÊ

Trong thống kê, việc tổng hợp mẫu ngẫu nhiên $W_X = (X_1, X_2, .. , X_n)$ được thực hiện dưới dạng hàm $G = f(X_1, X_2, .. , X_n)$, với $X_1, X_2, .. , X_n$ là các giá trị của mẫu ngẫu nhiên W_X. Khi đó G được gọi là một thống kê. Ta có các thống kê như sau:

2.1 Trung bình của mẫu ngẫu nhiên:

Định nghĩa: *Trung bình của mẫu ngẫu nhiên $W_X = (X_1, X_2, ..., X_n)$ là một thống kê, ký hiệu $\overline{X}$, được xác định bởi:* $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Chú ý:

i) Vì $X_1, X_2, ..., X_n$ là các đại lượng ngẫu nhiên nên $\overline{X}$ cũng là đại lượng ngẫu nhiên .

ii) Nếu mẫu ngẫu nhiên $W_X = (X_1, X_2, ..., X_n)$ có mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$ thì $\overline{X}$ sẽ nhận giá trị $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ và $\overline{x}$ được gọi là trung bình của mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$.

Tính chất: Nếu X có $\begin{cases} E(X) = \mu \\ Var(X) = \sigma^2 \end{cases}$ thì $\overline{X}$ có $\begin{cases} E(\overline{X}) = \mu \\ Var(\overline{X}) = \dfrac{\sigma^2}{n} \end{cases}$

Phân phối xác suất của $\overline{X}$:

i) Nếu X $\in$ B(n,p) thì $\overline{X} \in$ B(n,p).

ii) Nếu X $\in$ P(a) thì $\overline{X} \in$ P(a).

iii) Nếu X $\in N(\mu, \sigma^2)$ thì $\overline{X} \in N(\mu, \frac{\sigma^2}{n})$.

iv) Nếu X $\in \chi^2(n)$ thì $\overline{X} \in \chi^2(n)$.

2.2 Phương sai của mẫu ngẫu nhiên:

Định nghĩa*: Phương sai của mẫu ngẫu nhiên $W_X = (X_1, X_2, ..., X_n)$ là một thống kê, ký hiệu S^2, được xác định bởi:* $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$, trong đó $\overline{X}$ là trung bình của mẫu ngẫu nhiên.

Chú ý:

i) Vì $X_1, X_2, ..., X_n$ là các đại lượng ngẫu nhiên nên S^2 cũng là đại lượng ngẫu nhiên.

ii) Nếu mẫu ngẫu nhiên $W_X = (X_1, X_2, ..., X_n)$ có mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$ thì S^2 sẽ nhận giá trị $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$. Khi đó s^2 được gọi là phương sai của mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$.

Tính chất: Nếu X có $Var(X) = \sigma^2$ thì S^2 có $E(S^2) = \frac{n-1}{n}\sigma^2$.

Phân phối xác suất: Giả sử $W_X = (X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên được xây dựng bởi đại lượng ngẫu nhiên gốc X có luật phân phối chuẩn X $\in N(\mu, \sigma^2)$. Khi đó:

i) $\dfrac{nS^2}{\sigma^2} = \sum_{i=1}^{n}\dfrac{(X_i - \overline{X})^2}{\sigma^2} \in \chi^2(n-1)$.

ii) $\sum_{i=1}^{n}\dfrac{(X_i - \mu)^2}{\sigma^2} \in \chi^2(n)$.

2.3 Phương sai điều chỉnh của mẫu ngẫu nhiên:

Định nghĩa*: Phương sai điều chỉnh của mẫu ngẫu nhiên $W_X = (X_1, X_2, ..., X_n)$ là một thống kê, ký hiệu S'^2, được xác định bởi: $S'^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, trong đó S^2 là phương sai của mẫu ngẫu nhiên.*

Chú ý:

i) S'^2 cũng là đại lượng ngẫu nhiên.

ii) Nếu mẫu ngẫu nhiên $W_X = (X_1, X_2, ..., X_n)$ có mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$ thì S'^2 sẽ nhận giá trị $s'^2 = \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$. Khi đó s'^2 được gọi là phương sai điều chỉnh của mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$.

Tính chất: Nếu X có $Var(X) = \sigma^2$ thì S'^2 có $E(S'^2) = \sigma^2$.

Phân phối xác suất: Giả sử $W_X = (X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên được xây dựng bởi đại lượng ngẫu nhiên gốc X có luật phân phối chuẩn X $\in N(\mu,\sigma^2)$. Khi đó:

$$\frac{(n-1)S'^2}{\sigma^2} \in \chi^2(n-1)$$

2.4 Độ lệch tiêu chuẩn và độ lệch tiêu chuẩn điều chỉnh:

Định nghĩa: *Độ lệch tiêu chuẩn của mẫu ngẫu nhiên* $W_X = (X_1, X_2, ..., X_n)$ *là một thống kê, ký hiệu* S*, được xác định bởi:* $S = \sqrt{S^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$*, trong đó* S^2 *là phương sai của mẫu ngẫu nhiên.*

Chú ý:

i) S cũng là đại lượng ngẫu nhiên.

ii) Nếu mẫu ngẫu nhiên $W_X = (X_1, X_2, ..., X_n)$ có mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$ thì S sẽ nhận giá trị $s = \sqrt{s^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$. Khi đó s được gọi là độ lệch tiêu chuẩn của mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$.

Định nghĩa: *Độ lệch tiêu chuẩn điều chỉnh của mẫu ngẫu nhiên* $W_X = (X_1, X_2, ..., X_n)$ *là một thống kê, ký hiệu* S'*, được xác định bởi:* $S' = \sqrt{S'^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$*, trong đó* S'^2 *là phương sai điều chỉnh của mẫu ngẫu nhiên.*

Chú ý:

i) S' cũng là đại lượng ngẫu nhiên.

ii) Nếu mẫu ngẫu nhiên $W_X = (X_1, X_2, ..., X_n)$ có mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$ thì S' sẽ nhận giá trị $s' = \sqrt{s'^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$. Khi đó s' được gọi là độ lệch tiêu chuẩn điều chỉnh của mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$.

Bước học 3: THU THẬP SỐ LIỆU VÀ SẮP XẾP SỐ LIỆU

Quá trình nghiên cứu thống kê thường trải qua hai khâu: Thu thập số liệu liên quan đến nghiên cứu và xử lý số liệu.

3.1 Thu thập số liệu:

Dữ liệu thống kê: *Là những điều cụ thể của quy luật ngẫu nhiên, nó có thể là quy luật của một loạt các phép cân, đong, đo, đếm,...và nó được tập hợp lại để phản ánh một số tính chất nào đó của các hiện tượng ngẫu nhiên.*

3.2 Sắp xếp số liệu:

Để việc xử lý được thuận lợi ta cần phải sắp xếp lại số liệu.

Trường hợp mẫu có kích thước nhỏ:

Giả sử mẫu có kích thước n và đại lương ngẫu nhiên gốc X nhận các giá trị có thể x_i (i = 1, ..., k) với số lần lặp lại (tần số) n_i (i = 1,..,k). Ta thường lập bảng sau:

x_i	n_i
x_1	n_1
x_2	n_2
.	.
.	.
x_k	n_k

Chú ý rằng: $\sum_{i=1}^{k} n_i = n$

Trường hợp mẫu có kích thước lớn:

Ta chia mẫu thành các khoảng (lớp), trong mỗi khoảng ta chọn một giá trị đại diện. Người ta thường chia thành các khoảng đều nhau và chọn giá trị đại diện là giá trị trung tâm của khoảng. Ở đây quy ước rằng đầu mút bên phải của mỗi khoảng thuộc khoảng đó mà không thuộc khoảng tiếp theo khi tính tần số của mỗi khoảng.

Trong thống kê, phương sai điều chỉnh mẫu được tính từ phương sai mẫu theo công thức $s'^2 = \frac{n}{n-1}s^2$, độ lệch tiêu chuẩn và độ lêch tiêu chuẩn điều chỉnh mẫu được tính từ phương sai và phương sai điều chỉnh mẫu theo công thức $s = \sqrt{s^2}$ và $s' = \sqrt{s'^2}$. Vì vậy ta chỉ cần quan tâm đến cách tính trung bình mẫu $\overline{x}$ và phương sai mẫu s^2 mà thôi.

*** Tính trực tiếp:**

- Ta lập bảng tính dạng sau:

x_i	n_i	$n_i x_i$	$n_i x_i^2$
x_1	n_1	$n_1 x_1$	$n_1 x_1^2$
x_2	n_2	$n_2 x_2$	$n_2 x_2^2$
. . .	. . .	. . .	. . .
x_k	n_k	$n_k x_k$	$n_k x_k^2$
$\sum$	$\sum_{i=1}^{k} n_i = n$	$\sum_{i=1}^{k} n_i x_i$	$\sum_{i=1}^{k} n_i x_i^2$

- Ta dùng các công thức:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i , \qquad s^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \left(\overline{x}\right)^2$$

Ví dụ: *Số xe hơi bán được trung bình trong một tuần ở mỗi đại lý trong 45 đại lý cho bởi:*

Số xe hơi được bán trong một tuần / đại lý	n_i	Số xe hơi được bán trong một tuần / đại lý	n_i
1	15	4	5
2	12	5	3
3	9	6	1

Ta lập bảng sau:

x_i	n_i	$n_i x_i$	$n_i x_i^2$
1	15	15	15
2	12	24	48
3	9	27	81
4	5	20	80
5	3	15	75
6	1	6	36
$\sum$	n = 45	107	335

Ta có:

$$\overline{x} = \frac{107}{45} = 2.38$$

$$s^2 = \frac{335}{45} - (2.38)^2 = 7.444 - 5.664 = 1.78$$

**** Tính bằng máy tính***

3.3 Thực hành tính các giá trị $\overline{x}$, s^2:

KQHT 4: ƯỚC LƯỢNG THAM SỐ CỦA ĐẠI LƯỢNG NGẪU NHIÊN

Bước học 1: GIỚI THIỆU CÁC PHƯƠNG PHÁP

1.1 Mô tả phương pháp:

Giả sử đại lượng ngẫu nhiên X có tham số đặc trưng θ chưa biết. Ước lượng tham số θ là dựa vào mẫu ngẫu nhiên $W_X = (X_1, X_2,..., X_n)$. Ta chọn thống kê $G = \theta(X_1, X_2,..., X_n)$ để ước lượng θ (dự đoán θ).

Có hai phương pháp để ước lượng:

1. *Ước lượng điểm: Chỉ ra $\theta = \theta_0$ nào đó để ước lượng.*
2. *Ước lượng khoảng: Chỉ ra khoảng (θ_1, θ_2) chứa θ sao cho $P(\theta_1 < \theta < \theta_2) = 1 - \alpha$ (với α là độ tin cậy cho trước)*

1.2 Các phương pháp ước lượng điểm:

a) Mô tả phương pháp:

Ước lượng điểm cho đặc trưng số θ là chỉ ra một giá trị θ_0 (chính xác hoặc gần đúng) của θ. Giá trị θ_0 được tìm như sau :

i) Từ đại lượng ngẫu nhiên gốc X ta lập mẫu ngẫu nhiên $W_X = (X_1, X_2,..., X_n)$ kích thước n.

ii) Chọn thống kê $G(X_1, X_2,..., X_n)$ thích hợp làm hàm ước lượng cho θ.

iii) Thực hiện phép thử ta được mẫu cụ thể $w_x = (x_1, x_2,..., x_n)$, thay thế các giá trị $x_1, x_2,..., x_n$ vào hàm ước lượng G, tính được giá trị G_0, là giá trị của thống kê G tương ứng với mẫu cụ thể $w_x = (x_1, x_2,..., x_n)$.

iv) Lấy $\theta_0 = G_0$ là giá trị ước lượng cần tìm của θ.

b) Các tiêu chuẩn ước lượng:

Có thể chọn nhiều hàm ước lượng khác nhau để cùng ước lượng cho đặc trưng số θ, do đó có thể tìm thấy nhiều giá trị ước lượng θ_0 khác nhau. Vì vậy, phải có các tiêu chuẩn để so sánh các hàm ước lượng. Cùng một tiêu chuẩn so sánh, hàm ước lượng nào cho giá trị gần nhất so với θ được coi là tốt hơn. Tuy nhiên, một hàm ước lượng có thể là tốt hơn đối với tiêu chuẩn này nhưng không tốt hơn đối với tiêu chuẩn khác nhau.

Sau đây là một số tiêu chuẩn để đánh giá các hàm ước lượng:

i) Ước lượng không chệch:

Định nghĩa: Thống kê G = $G(X_1, X_2,..., X_n)$ được gọi là ước lượng không chệch của tham số θ nếu E(G) = θ hay $\lim\limits_{n\to\infty} E(G) = \theta$.

Lưu ý: Nếu E(G) ≠ θ thì G được gọi là ước lượng chệch.

Ý nghĩa: Giả sử G là ước lượng không chệch của tham số θ. Ta có:

$$E(G - \theta) = E(G) - E(\theta) = \theta - \theta = 0$$

Vậy ước lượng không chệch là ước lượng có sai số trung bình bằng 0.

Nhận xét:

i) Trung bình của mẫu ngẫu nhiên $\overline{X}$ là ước lượng không chệch của trung bình của tổng thể $\theta = E(X) = \mu$ vì $E(\overline{X}) = \mu$.

ii) Phương sai điều chỉnh của mẫu ngẫu nhiên S'^2 là ước lượng không chệch của phương sai của tổng thể $\theta = Var(X) = \sigma^2$ vì $E(S'^2) = \sigma^2$.

Ví dụ 1: *Chiều cao của 500 cây lim được cho bởi bảng sau:*

Khoảng chiều cao (mét)	Số cây lim
[6,25 – 6,75)	1
[6,75 – 7,25)	2
[7,25 – 7,75)	5
[7,75 – 8,25)	11
[8,25 – 8,75)	18
[8,75 – 9,25)	9
[9,25 – 9,75)	3
[9,75 – 10,25)	1

Gọi X là chiều cao của cây lim.

a) Hãy chỉ ra ước lượng điểm cho chiều cao trung bình của các cây lim.

b) Hãy chỉ ra ước lượng điểm cho độ tản mát của các chiều cao cây lim so với chiều cao trung bình.

c) Gọi $p = P(7{,}75 \le X \le 8{,}75)$. *Hãy chỉ ra ước lượng điểm cho p.*

Ta lập bảng tính:

Chọn $x_0 = 8{,}5;\ h = 0{,}5$

Thực hiện phép biến đổi $u_i = \dfrac{x_i^0 - x_0}{h} = \dfrac{x_i^0 - 8{,}5}{0{,}5}$

$x_i - x_{i+1}$	n_i	x_i^0	u_i	$n_i u_i$	$n_i u_i^2$
[6,25 – 6,75)	1	6,5	-4	-4	16
[6,75 – 7,25)	2	7,0	-3	-6	18
[7,25 – 7,75)	5	7,5	-2	-10	20
[7,75 – 8,25)	11	8	-1	-11	11
[8,25 – 8,75)	18	8,5	0	0	0
[8,75 – 9,25)	9	9	1	9	9
[9,25 – 9,75)	3	9,5	2	6	12
[9,75 – 10,25)	1	10	3	3	9
$\sum$	50			-13	95

Ta có: $\overline{u} = -\dfrac{13}{50} = -0{,}26$. Suy ra:

$$\overline{x} = h\overline{u} + x_0 = 0{,}5.(-0{,}26) + 8{,}5 = 8{,}37$$

$$s^2 = h^2 s_u^2 = h^2\left[\frac{1}{n}\sum_{i=1}^{k} n_i u_i^2 - \left(\overline{u}\right)^2\right] = (0{,}5)^2\left[\frac{95}{50} - (-0{,}26)^2\right] = 0{,}4581$$

$$\Rightarrow s'^2 = \frac{n}{n-1}s^2 = \frac{50}{49}0{,}4581 = 0{,}4674 \Rightarrow s' = \sqrt{0{,}4674} \approx 0{,}684$$

a) Chiều cao trung bình được ước lượng là 8,37 mét.

b) Độ tản mát được ước lượng là s’ = 0,684 mét.

c) Trong 50 quan sát đã có 11 + 18 = 29 quan sát cho chiều cao cây lim thuộc khoảng [7.5–8.5).

Vậy ước lượng điểm cho p là $p^* = \frac{29}{50} = 0,58$.

ii) Ước lượng hiệu quả:

Nhận xét: Giả sử G là ước lượng không chệch của θ, tức là E(G) = θ. Theo bất đẳng thức Tchebyshev, ta có:

$$P(|G - E(G)| < \varepsilon) > 1 - \frac{Var(G)}{\varepsilon^2}$$

Vì $E(G) = \theta$ nên $P(|G - \theta| < \varepsilon) > 1 - \frac{Var(G)}{\varepsilon^2}$

Ta thấy nếu Var(G) càng nhỏ thì $P(|G - \theta| < \varepsilon)$ càng gần 1. Do đó ta sẽ chọn G sao cho Var(G) nhỏ nhất.

Định nghĩa: *Ước lượng không chệch G được gọi là ước lượng hiệu quả của tham số θ nếu G có phương sai Var(G) nhỏ nhất trong các ước lượng không chệch của θ.*

Như vậy ước lượng hiệu quả của θ là ước lượng không chệch mà các giá trị tính được thông qua ước lượng đó với nhiều mẫu cụ thể khác nhau có mật độ tập trung nhất xung quanh θ.

***Định lý Ramer – Rao**: Giả sử đại lượng ngẫu nhiên X có hàm mật độ xác suất f(x,θ), trong đó θ là 1 đặc trưng số của X và G là 1 ước lượng không chệch của θ, khi đó:*

$$Var(G) \geq \frac{1}{n.E\left[\left(\frac{\partial \ln[f(x)]}{\partial \theta}\right)^2\right]}$$

Bất đẳng thức được gọi là bất đẳng thức Gramer – Rao, cho biết cận dưới của phương sai các ước lượng không chệch.

Chú ý: Nếu G là ước lượng không chệch của θ, G có phương sai thỏa mãn dấu bằng trong bất đẳng thức Gramer – Rao thì G là ước lượng hiệu quả của θ.

Nhận xét: Nếu đại lượng ngẫu nhiên gốc X có luật phân phối chuẩn $X \in N(\mu, \sigma^2)$ thì trung bình mẫu là ước lượng hiệu quả của kỳ vọng E(X) = μ.

Thật vậy, ta biết $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \in N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Var(\overline{X}) = \frac{\sigma^2}{n}$

Mặt khác, do X có phân phối chuẩn nên nếu f(x) là hàm mật độ xác suất của X thì:

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

Ta có:

$$\frac{\partial \ln[f(x)]}{\partial \mu} = \left[\ln\left(\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\left(\frac{x-\mu}{\sigma}\right)^2}{2}}\right)\right]'_{\mu} = \left[\frac{-\left(\frac{x-\mu}{\sigma}\right)^2}{2}\right]'_{\mu}$$

$$= -\frac{1}{2\sigma^2}\left[(x-\mu)^2\right]'_{\mu} = -\frac{1}{2\sigma^2}.2.(x-\mu)(-1) = \frac{x-\mu}{\sigma^2}$$

$$\Rightarrow n.E\left[\left(\frac{\partial \ln[f(x)]}{\partial \mu}\right)^2\right] = n.E\left[\left(\frac{x-\mu}{\sigma^2}\right)^2\right] = n.E\left[\frac{(x-\mu)^2}{\sigma^4}\right]$$

$$= \frac{n}{\sigma^4}.E\left[(x-\mu)^2\right] = \frac{n}{\sigma^4}.\sigma^2 = \frac{n}{\sigma^2}$$

$$\Rightarrow \frac{1}{n.E\left[\left(\frac{\partial \ln[f(x)]}{\partial \mu}\right)^2\right]} = \frac{1}{\frac{n}{\sigma^2}} = \frac{\sigma^2}{n}$$

Suy ra: $Var(\overline{X}) = \dfrac{1}{nE\left[\left(\frac{\partial \ln[f(x)]}{\partial \mu}\right)^2\right]}$

Vậy: $\overline{X}$ là ước lượng hiệu quả của E(X) = μ.

iii) Ước lượng vững:

Định nghĩa: *Thống kê* $G = G(X_1, X_2, ..., X_n)$ *được gọi là ước lượng vững của tham số* θ *nếu* $\lim\limits_{n\to\infty} P(|G-\theta| < \varepsilon) = 1$.

Như vậy, G là 1 ước lượng vững của θ nếu sự kiện sai số tuyệt đối nhỏ tùy ý trở thành như chắc chắn.

Định lý: *Nếu* $\lim\limits_{n\to\infty} E(G) = \theta$ *và* $\lim\limits_{n\to\infty} Var(G) = 0$ *thì G là ước lượng vững của* θ.

Hay nói cách khác G là 1 ước lượng vững của θ nếu G là ước lượng không chệch và có phương sai giảm dần về 0

Bước học 2: ƯỚC LƯỢNG CÁC THAM SỐ

2.1 Mô tả phương pháp:

Giả sử tổng thể có tham số θ *chưa biết. Ta tìm khoảng* (θ_1, θ_2) *chứa* θ *sao cho* $P(\theta_1 < \theta < \theta_2) = 1-\alpha$ *cho trước (α khá nhỏ).*

Phương pháp thường dùng như sau:

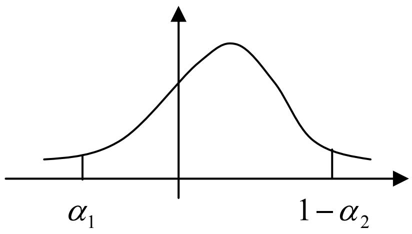
i) Từ đại lượng ngẫu nhiên gốc lập mẫu ngẫu nhiên kích thước n $W_X = (X_1, X_2, ..., X_n)$.

ii) Chọn 1 thống kê G thích hợp chứa θ: $G = G(X_1, X_2, ..., X_n, \theta)$ *có luật phân phối xác định và không phụ thuộc vào θ.*

iii) Với α khá nhỏ cho trước ta tìm được 1 cặp α_1, α_2 *sao cho:* $\alpha_1 + \alpha_2 = \alpha$. *Xác định các phân vị* $G_{\alpha_1}, G_{\alpha_2}$ *:* $P(G < G_{\alpha_1}) = \alpha_1$, $P(G > G_{\alpha_2}) = \alpha_2$.

Khi đó ta có:

$$P(G_{\alpha_1} < G < G_{1-\alpha_2}) = 1 - \alpha_1 - \alpha_2 = 1 - \alpha$$



Trong trường hợp đặc biệt $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$ ta có: $P(G_{\frac{\alpha}{2}} < G < G_{1-\frac{\alpha}{2}}) = 1 - \alpha$ (1)

Bằng phép biến đổi tương ứng đưa bất đẳng thức trên về dạng: $P(\theta_1 < \theta < \theta_2) = 1 - \alpha$

iv) Thực hiện phép thử ta được mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$, ta tính được giá trị cụ thể của θ_1, θ_2. Khi đó $(\theta_1;\theta_2)$ là khoảng ước lượng cần tìm của θ.

**** Một số khái niệm được sử dụng trong phần này:***

- Khoảng (θ_1, θ_2) được gọi là *khoảng ước lượng*.
- Số α được gọi là *mức ý nghĩa*.
- 1 - α được gọi là *độ tin cậy*.
- $|\theta_1 - \theta_2|$ được gọi là *độ dài của khoảng ước lượng*.
- Thông thường, khoảng tin cậy là khoảng đối xứng $(\theta_1, \theta_2) = (\theta_0 - \varepsilon, \theta_0 + \varepsilon)$, trong đó θ_0 là ước lượng điểm của tham số θ, ε được gọi là *độ chính xác* (*bán kính* hay *sai số*) của ước lượng và khi đó độ dài khoảng ước lượng là 2ε.

2.2 Ước lượng trung bình:

Giả sử tổng thể có trung bình E(X) = m chưa biết, ước lượng trung bình là ta chỉ ra khoảng (m_1 , m_2) chứa m sao cho xác suất P(m_1< m < m_2) = 1 - α

Từ X lập mẫu ngẫu nhiên $W_X = (X_1, X_2, ..., X_n)$ và xét các trường hợp sau:

Ta xét 3 trường hợp:

i) Var(X) = σ^2 đã biết và n ≥ 30 (n < 30 và X có phân phối chuẩn)

Chọn thống kê: $U = \dfrac{(\overline{X} - m).\sqrt{n}}{\sigma} \sim N(0,1)$

Trong đó: . m = E(X) chưa biết,

. $\sigma = \sqrt{Var(X)}$ đã biết,

. n là kích thước mẫu,

. $\overline{X}$ là thống kê nhận giá trị bằng trung bình mẫu.

. N(0,1): phân phối chuẩn tắc.

Từ (1) ta có:

$$P\left(U_{\frac{\alpha}{2}} < U = \frac{\left(\overline{X} - m\right)\sqrt{n}}{\sigma} < U_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Trong đó: $U_{\frac{\alpha}{2}}, U_{1-\frac{\alpha}{2}}$ lần lượt là phân vị chuẩn tắc mức xác suất $\frac{\alpha}{2}, 1 - \frac{\alpha}{2}$.

Vì phân vị chuẩn có tính chất $U_{\frac{\alpha}{2}} = -U_{1-\frac{\alpha}{2}}$ nên:

$$P\left(\overline{X} - U_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < m < \overline{X} + U_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Thực hiện phép thử để có được mẫu cụ thể $w_x = \left(x_1, x_2, ..., x_n\right)$, tính được $\overline{x}$ và dựa vào công thức trên, ta tìm được khoảng ước lượng (m_1, m_2) của m với độ tin cậy $1 - \alpha$, trong đó:

$m_1 = \overline{x} - \varepsilon$; $m_2 = \overline{x} + \varepsilon$

với độ chính xác ε là: $\varepsilon = U_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

Ví dụ 1: *Khối lượng sản phẩm là đại lượng ngẫu nhiên X có luật phân phối chuẩn, biết rằng phương sai σ^2 = 4g. Kiểm tra 25 sản phẩm, tính được trung bình mẫu $\overline{x}$ = 20g.*

a) Ước lượng trung bình của khối lượng sản phẩm với độ tin cậy 95%.

b) Nếu cho bán kính của ước lượng $\varepsilon = 0{,}4g$ thì độ tin cậy của ước lượng là bao nhiêu?

c) Với bán kính ước lượng $\varepsilon < 0{,}4g$, muốn có độ tin cậy $1 - \alpha = 95\%$ thì phải kiểm tra ít nhất bao nhiêu sản phẩm?

Giải

a) Đặt m = E(X) chưa biết.

Chọn thống kê $U = \frac{\left(\overline{X} - m\right)\sqrt{n}}{\sigma} \in N(0{,}1)$ để ước lượng trung bình m, trong đó: $\sigma = 2g$, n = 25 , $\overline{x} = 20g$.

Độ tin cậy 1 - α = 95% = 0,95 ⇒ α = 0,05 $\Rightarrow \frac{\alpha}{2} = 0{,}025 \Rightarrow 1 - \frac{\alpha}{2} = 0{,}975$

Do đó: $\varepsilon = U_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = U_{0.975}\frac{2}{\sqrt{25}} = \left(1{,}96\right)\frac{2}{5} = 0{,}78$

Suy ra: $m_1 = 20 - 0{,}78 = 1{,}22$; $m_2 = 20 + 0{,}78 = 20{,}78$

Vậy: khoảng ước lượng trung bình khối lượng sản phẩm với độ tin cậy 95% là (19.22g ; 20.78g).

b) Với ε = 0,4g, sử dụng công thức:

$$\varepsilon = U_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow U_{1-\frac{\alpha}{2}} = \frac{\varepsilon\sqrt{n}}{\sigma} = \frac{(0{,}4).\sqrt{25}}{2} = 1 \approx 0{,}994 = U_{0{,}84}$$

$$\Rightarrow 1 - \frac{\alpha}{2} = 0{,}84 \Rightarrow \frac{\alpha}{2} = 0{,}16 \Rightarrow \alpha = 0{,}32$$

$\Rightarrow 1 - \alpha = 0,68$

Vậy: độ tin cậy là 68%.

c) Với $\varepsilon < 0,4$g, $1 - \alpha = 95\% = 0,95 \Rightarrow 1 - \frac{\alpha}{2} = 0,975$.

Từ công thức trên, suy ra:

$$n = U^2_{0,975}.\frac{\sigma^2}{\varepsilon^2} > U^2_{0,975}.\frac{2^2}{(0,4)^2} = (1,96)^2.\frac{4}{(0,4)^2} = 96,04$$

Vì n là số nguyên $\Rightarrow$ n > 97 $\Rightarrow n_{\min} = 97$.

Vậy phải kiểm tra ít nhất 97 sản phẩm.

<u>Nhận xét</u>: *Công thức độ chính xác cho thấy độ tin cậy 1 - α càng lớn thì bán kính ε càng lớn, do đó khoảng ước lượng ($\bar{x}$ - ε; $\bar{x}$ + ε) cho giá trị thông tin thấp. Kết quả câu b cho thấy nếu giảm bán kính ε thì khoảng ước lượng ($\bar{x}$ - ε; $\bar{x}$ + ε) có giá trị thông tin cao nhưng độ tin cậy của ước lượng giảm xuống. Như vậy, muốn có bán kính ε nhỏ và độ tin cậy 1 - α lớn thì tăng kích thước mẫu (kết quả câu c).*

ii) Var(X) = σ^2 chưa biết và n ≥ 30

Chọn thống kê: $U = \frac{(\overline{X} - m).\sqrt{n}}{S'} \sim N(0,1)$

Trong trường hợp n ≥ 30, ta dùng ước lượng S'^2 thay cho σ^2

Lập luận hoàn toàn tương tự như trên. Khi đó, ta cũng tìm được các khoảng ước lượng là $(x_1, x_2) = (\bar{x} - \varepsilon, \bar{x} + \varepsilon)$

với $\varepsilon = U_\gamma \frac{s'}{\sqrt{n}}$

U_γ là phân vị chuẩn mức $\gamma = 1 - \frac{\alpha}{2}$

Ví dụ 2: *Khảo sát chiều cao của cây cùng độ tuổi thu được kết quả như sau :*

Chiều cao (cm)	*Số cây*
< 180	*3*
180 – 190	*12*
190 – 200	*35*
200 – 210	*70*
210 – 220	*62*
220 – 230	*32*
> 230	*6*
	220

Ước lượng trung bình μ của chiều cao cây với độ tin cậy 99%.

Chọn thống kê: $U = \dfrac{(\overline{X} - m)\sqrt{n}}{S'} \in$ N(0,1) để ước lượng trung bình m.

Trong đó: . n = 220,

. $\overline{X}$, S' lần lượt là thống kê nhận giá trị trung bình mẫu và độ lệch tiêu chuẩn điều chỉnh mẫu.

Khoảng ước lượng trung bình m là (m_1, m_2) trong đó :

$$m_1 = \overline{x} - \varepsilon \quad ; \quad m_2 = \overline{x} + \varepsilon \quad \text{với} \quad \varepsilon = U_{1-\frac{\alpha}{2}} \cdot \frac{s'}{\sqrt{n}}$$

Với mẫu cho trong bảng trên, các khoảng chiều cao được thay thế bởi điểm giữa, riêng khoảng <180 được thay thế bởi 175cm, còn khoảng >230 được thay thế bởi 235cm. Ta tính được $\overline{x}$= 208,455cm, s'=12,233.

Với độ tin cậy: 1 - α = 99% $\Rightarrow$ α = 0,01 $\Rightarrow \dfrac{\alpha}{2} = 0{,}005 \Rightarrow 1 - \dfrac{\alpha}{2} = 0{,}995$

Do đó: $\varepsilon = U_{0.995} \cdot \dfrac{s'}{\sqrt{n}} = (2{,}576) \cdot \dfrac{12{,}233}{\sqrt{220}} = 2{,}125$ (cm)

Suy ra: $m_1 = 208{,}455 - 2{,}125 = 206{,}33$ (cm)

$m_2 = 208{,}455 + 2{,}125 = 210{,}58$ (cm)

Vậy: khoảng ước lượng trung bình chiều cao của cây với độ tin cậy 99% là (206,33 cm ; 210,58 cm).

iii) Var(X) = σ^2 chưa biết và n < 30, X có phân phối chuẩn

Chọn thống kê: $T = \dfrac{(\overline{X} - m)\sqrt{n}}{S'} \in T(n-1)$

Trong đó: . m = E(X) chưa biết.

. n: kích thước mẫu.

. $\overline{X}$, S' lần lượt là các thống kê nhận giá trị bằng trung bình mẫu và độ lệch tiêu chuẩn điều chỉnh mẫu.

. T(n – 1): phân phối Student bậc tự do (n–1).

Từ (1) ta có:

$$P\left(T_{\frac{\alpha}{2}}(n-1) < U = \frac{(\overline{X} - m)\sqrt{n}}{S'} < T_{1-\frac{\alpha}{2}}(n-1) \right) = 1 - \alpha$$

Trong đó: $T_{\frac{\alpha}{2}}(n-1), T_{1-\frac{\alpha}{2}}(n-1)$ lần lượt là phân vị Student, bậc tự do (n - 1), mức xác suất $\dfrac{\alpha}{2}, 1 - \dfrac{\alpha}{2}$.

Vì phân vị Student có tính chất $T_{\frac{\alpha}{2}}(n-1) = -T_{1-\frac{\alpha}{2}}(n-1)$ nên ta có:

$$P\left(\overline{X} - T_{1-\frac{\alpha}{2}}(n-1)\frac{S'}{\sqrt{n}} < m < \overline{X} + T_{1-\frac{\alpha}{2}}(n-1)\frac{S'}{\sqrt{n}} \right) = 1 - \alpha$$

Thực hiện phép thử để có được mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$, tính được $\overline{x}$, s' và ta tìm được khoảng ước lượng (m_1; m_2) của m với độ tin cậy $1-\alpha$, trong đó:

$$m_1 = \overline{x} - \varepsilon \;; \qquad m_2 = \overline{x} + \varepsilon \quad \text{với độ chính xác } \varepsilon \text{ là: } \quad \varepsilon = T_{1-\frac{\alpha}{2}}(n-1)\frac{s'}{\sqrt{n}}$$

Ví dụ 3: *Lượng chi phí 1 loại nguyên liệu cho đơn vị sản phẩm là đại lượng ngẫu nhiên X có luật phân phối chuẩn. khảo sát 25 sản phẩm tính được trung bình mẫu $\overline{x}$ = 50g, độ lệch tiêu chuẩn điều chỉnh s' = 8,25g. Hãy ước lượng trung bình m của chi phí nguyên liệu với độ tin cậy 95%.*

Chọn thống kê: $T = \dfrac{(\overline{X} - m)\sqrt{n}}{S'}$ ∈ T(n – 1) để ước lượng trung bình m.

Trong đó:

. n: kích thước mẫu.

. $\overline{X}$, S' lần lượt là thống kê nhận giá trị bằng trung bình mẫu, độ lệch tiêu chuẩn điều chỉnh mẫu.

Khoảng ước lượng cho trung bình m là (m_1; m_2) trong đó :

$$m_1 = \overline{x} - \varepsilon \;; \qquad m_2 = \overline{x} + \varepsilon \quad \text{với} \quad \varepsilon = T_{1-\frac{\alpha}{2}}(n-1).\frac{s'}{\sqrt{n}}$$

Với mẫu có n = 25, $\overline{x}$ = 50g, s' = 8,25g và 1 - α = 95% $\Rightarrow 1 - \dfrac{\alpha}{2} = 0{,}975$

$$\Rightarrow \varepsilon = T_{0,975}(24).\frac{8{,}25}{\sqrt{25}} = (2{,}064).\frac{8{,}25}{5} = 3{,}406 \text{ (g)}$$

Suy ra: $m_1 = 50 - 3{,}406 = 46{,}594$ (g)

$m_2 = 50 + 3{,}406 = 53{,}406$ (g)

Vậy: khoảng ước lượng trung bình của m với độ tin cậy 95% là (46,594g;53,406g).

2.3 Ước lượng tỉ lệ:

Bài toán: *Giả sử tổng thể có hai loại phần tử, tỉ lệ phần tử có tính chất A là p chưa biết. Ước lượng tỉ lệ phần tử có tính chất A là tìm ra khoảng (f_1, f_2) chứa p sao cho $P(f_1 < p < f_2) = 1 - \alpha$*

Phương pháp:

- Từ đại lượng ngẫu nhiên X, ta lập mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ có kích thước n và tính tỉ lệ f các phần tử có tính chất A.
- Chọn thống kê $U = \dfrac{(f-p).\sqrt{n}}{\sqrt{pq}} \sim N(0,1)$

Trong đó: . p là tỉ lệ chưa biết,

. q = 1 – p,

. n : kích thước mẫu khá lớn,

. f: thống kê nhận giá trị bằng tỉ lệ phần tử có tính chất A trong mẫu.

và với mẫu cụ thể, ta tìm được khoảng ước lượng là $(f_1, f_2) = (f - \varepsilon, f + \varepsilon)$

Trong đó $\varepsilon = U_\gamma \sqrt{\dfrac{f(1-f)}{n}}$

U_γ là phân vị chuẩn mức $\gamma = 1 - \dfrac{\alpha}{2}$

Từ công thức trên ta có: $U_\gamma = \varepsilon\sqrt{\dfrac{n}{f(1-f)}}$, $n = \dfrac{{U_\gamma}^2.f(1-f)}{\varepsilon^2}$

Khi cần n chẵn, có: $n = \left[\dfrac{{U_\gamma}^2.f(1-f)}{\varepsilon^2}\right] + 1$

Ví dụ 4: *Kiểm tra 100 sản phẩm từ lô hàng thì thấy có 20 phế phẩm.*

a) *Ước lượng tỉ lệ của lô hàng với độ tin cậy 99%*

b) *Nếu muốn độ tin cậy chính xác khi ước lượng ε = 0,04 thì độ tin cậy của ước lượng là bao nhiêu?*

c) *Nếu muốn độ tin cậy là 99% và độ chính xác là 0,04 thì cần điều tra bao nhiêu sản phẩm?*

Giải

a) Ta có: n = 100, f = $\dfrac{20}{100}$ = 0,2.

Xét $U = \dfrac{(f-p).\sqrt{100}}{\sqrt{pq}} \sim N(0,1)$.

Ta có: 1 - α = 0,99 $\Rightarrow \alpha$ = 0,01 $\Rightarrow \gamma$ = 1 - $\dfrac{\alpha}{2}$ = 0,995. Tra bảng phân vị chuẩn, ta có: U_γ = 2,576.

$$\varepsilon = U_\gamma\sqrt{\frac{f(1-f)}{n}} = 2{,}576.\frac{\sqrt{0{,}2.0{,}8}}{\sqrt{100}} = 0{,}1$$

f_1 = f - ε = 0,2 – 0,1 = 0,1.

f_2 = f + ε = 0,2 + 0,1 = 0,3.

Vậy khoảng tin cậy là: (0,1 ; 0,3)

b) $U_{1-\frac{\alpha}{2}} = \varepsilon\sqrt{\dfrac{n}{f(1-f)}} = 0{,}04.\sqrt{\dfrac{100}{0{,}2.0.8}} = 1$

Ta tìm được: 1 - $\dfrac{\alpha}{2}$ = 0,84 $\Rightarrow$ 1 - α = 0,68.

Vậy độ tin cậy là 68%.

c) Ta có: 1 - α = 0,99 $\Rightarrow \alpha$ = 0,01 $\Rightarrow$ 1 - $\dfrac{\alpha}{2}$ = 0,995. Tra bảng phân vị chuẩn, ta có: U_γ = 2,576.

Do đó: $n = \left[\dfrac{U_\gamma^{\;2}.f(1-f)}{\varepsilon^2}\right] + 1 = \left[\dfrac{(2{,}576)^2.0{,}2.0{,}8}{(0{,}04)^2}\right] + 1 = 664$

Vậy: n = 664.

2.4 Ước lượng về phương sai:

Đại lượng ngẫu nhiên X có luật phân phối chuẩn $X \in N(\mu,\sigma^2)$ *trong đó phương sai* $Var(X) = \sigma^2$ *chưa biết. Cho số* α *khá nhỏ, ước lượng phương sai* σ^2 *với mức ý nghĩa* α *là chỉ ra khoảng* (σ_1,σ_2) *sao cho:* $P(\sigma_1 < \sigma < \sigma_2) = 1-\alpha$.

Từ X lập mẫu ngẫu nhiên $W_X = (X_1, X_2, ..., X_n)$ và xét các trường hợp sau:

a. Trường hợp 1: Biết E(X) = μ.

Chọn thống kê: $\chi^2 = \sum_{i=1}^{n} \dfrac{(X_i - \mu)^2}{\sigma^2} \in \chi^2(n)$

Trong đó: . $\chi^2(n)$: phân phối khi bình phương bậc tự do n.

Từ (1) ta có:

$$P\left(\chi^2_{\frac{\alpha}{2}}(n) < \chi^2 = \sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma^2} < \chi^2_{1-\frac{\alpha}{2}}(n)\right) = 1-\alpha$$

Trong đó: $\chi^2_{\frac{\alpha}{2}}(n), \chi^2_{1-\frac{\alpha}{2}}(n)$ lần lượt là phân vị khi bình phương, bậc tự do n, mức xác suất $\dfrac{\alpha}{2}, 1-\dfrac{\alpha}{2}$.

Hệ thức trên được biến đổi thành: $P\left(\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)} < \sigma^2 < \dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)}\right) = 1-\alpha$

Thực hiện phép thử để có được mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$, tính được $\sum_{i=1}^{k}(x_i-\mu)^2 n_i$ và dựa vào công thức ta tìm được khoảng ước lượng (σ_1^2, σ_2^2) của σ^2 với độ tin cậy $1-\alpha$, trong đó:

$$\sigma_1^2 = \frac{\sum_{i=1}^{k}(x_i-\mu)^2 n_i}{\chi^2_{1-\frac{\alpha}{2}}(n)} \;;\; \sigma_2^2 = \frac{\sum_{i=1}^{k}(x_i-\mu)^2 n_i}{\chi^2_{\frac{\alpha}{2}}(n)}$$

Ví dụ 5: *Kiểm tra 25 sản phẩm của một công ty sản xuất thức ăn đóng gói ta được kết quả sau:*

Trọng lượng (g)	*195*	*100*	*205*
Số sản phẩm	*5*	*18*	*2*

Biết trọng lượng trung bình μ *= 200g. Hãy ước lượng sai của trọng lượng các sản phẩm với độ tin cậy 90%.*

Chọn thống kê: $\chi^2 = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \in \chi^2(n)$

Từ số liệu trên ta có bảng sau:

x_i	$x_i - 200$	n_i
195	-5	5
200	0	18
205	5	2

Ta tính được: $\sum (x_i - 200)^2 n_i = 175$

Độ tin cậy 1 - α = 90% = 0,9 $\Rightarrow \alpha = 0,1 \Rightarrow \frac{\alpha}{2} = 0,05 \Rightarrow 1 - \frac{\alpha}{2} = 0,95$

Tra bảng ta được:

$$\chi^2_{\frac{\alpha}{2}}(n) = \chi^2_{0.05}(25) = 14,611$$

$$\chi^2_{1-\frac{\alpha}{2}}(n) = \chi^2_{0.95}(25) = 37,652$$

Suy ra:

$$\sigma_1^2 = \frac{\sum (x_i - 200)^2 n_i}{\chi^2_{0.95}(25)} = \frac{175}{37,652} = 4,65$$

$$\sigma_2^2 = \frac{\sum (x_i - 200)^2 n_i}{\chi^2_{0,05}(25)} = \frac{175}{14,611} = 11,98$$

Vậy: khoảng ước lượng của σ với độ tin cậy 90% là: (4,65 g^2 ; 11,98 g^2)

b. Trường hợp 2: E(X) chưa biết.

Chọn thống kê: $\chi^2 = \frac{(n-1)S'^2}{\sigma^2} \in \chi^2(n-1)$

Trong đó:

. n: kích thước mẫu.

. S'^2 : thống kê nhận giá trị bằng phương sai điều chỉnh của mẫu.

. $\chi^2(n-1)$:phân phối khi bình phương bậc tự do n – 1.

Từ (1) ta có:

$$P\left(\chi^2_{\frac{\alpha}{2}}(n-1) < \chi^2 = \frac{(n-1)S'^2}{\sigma^2} < \chi^2_{1-\frac{\alpha}{2}}(n-1)\right) = 1 - \alpha$$

Trong đó $\chi^2_{\frac{\alpha}{2}}(n-1)$ và $\chi^2_{1-\frac{\alpha}{2}}(n-1)$ lần lượt là phân vị khi bình phương, bậc tự do (n -1), mức xác suất $\frac{\alpha}{2}$ và $1-\frac{\alpha}{2}$.

Hệ thức trên được biến đổi thành:

$$P\left(\frac{(n-1)S'^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} < \sigma^2 < \frac{(n-1)S'^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right) = 1-\alpha$$

Thực hiện phép thử để có được mẫu cụ thể $w_x = (x_1, x_2, ..., x_n)$, tính được s'^2 và dựa vào công thức ta tìm được khoảng ước lượng (σ_1^2, σ_2^2) của phương sai σ^2 với độ tin cậy 1 - α là , trong đó:

$$\sigma_1^2 = \frac{(n-1)s'^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \quad ; \; \sigma_2^2 = \frac{(n-1)s'^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}$$

Ví dụ 6: *Với giả thuyết đã cho trong ví dụ 6 ở trên, hãy ước lượng phương sai $Var(X) = \sigma^2$ với độ tin cậy 95%.*

Chọn thống kê: $\chi^2 = \frac{(n-1)S'^2}{\sigma^2} \in \chi^2(n-1)$, có luật phân phối khi bình phương bậc tự do n – 1, để ước lượng cho σ^2.

Trong đó: . n: là kích thước mẫu,

. S'^2 là thống kê nhận giá trị bằng phương sai điều chỉnh mẫu.

Khoảng ước lượng σ^2 với độ tin cậy 1 - α là (σ_1^2, σ_2^2), trong đó:

$$\sigma_1^2 = \frac{(n-1)s'^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \quad ; \; \sigma_2^2 = \frac{(n-1)s'^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}$$

Với mẫu đã cho có n = 25, s'= 8,25g, độ tin cậy 1 - α = 95% $\Rightarrow 1-\frac{\alpha}{2} = 0{,}975$, ta có:

$$\sigma_1^2 = \frac{24.(8{,}25)^2}{\chi^2_{0,795}(24)} = \frac{24.(8{,}25)^2}{39{,}364} = 41{,}497.$$

$$\sigma_2^2 = \frac{24.(8{,}25)^2}{\chi^2_{0,025}(24)} = \frac{24.(8{,}25)^2}{12{,}401} = 131{,}723.$$

Tóm lại, khoảng ước lượng phương sai với độ tin cậy 95% là (41,497g^2;131,723g^2).

BÀI TẬP
ƯỚC LƯỢNG ĐIỂM

1. Theo dõi số áo thun thể thao bán được ở một cửa hàng bán dụng cụ thể thao trong ngày trong một tháng (30 ngày), ta có kết quả sau:

Số áo bán được	12	14	18	21	25	28
Số ngày	1	4	7	9	4	5

a) Hãy ước lượng không chệch cho số áo thun trung bình bán được trong ngày.

b) Nếu ta qui ước những ngày bán được không quá 15 cái áo là ngày "ế hàng", hãy ước lượng không chệch cho tỉ lệ những ngày ế hàng.

2. Để tìm hiểu số lượng mũ cao su mỗi cây cao su cho ta trong một ngày trong năm đầu khai thác (tính bằng g), ghi nhận ở 100 cây ta có kết quả sau:

Trọng lượng (g)	Số cây
200 – 210	2
210 – 220	8
220 – 230	14
230 – 240	30
240 – 250	25
250 – 260	12
260 – 270	9

a) Hãy ước lượng không chệch cho lượng mũ trung bình mỗi cây cho trong ngày.

b) Hãy ước lượng không chệch cho độ biến động của lượng mũ đó.

c) Giả sử ta qui định những cây cho một ngày trên 220g mũ là cây loại I, còn lại là cây loại II. Hãy ước lượng không chệch cho tỉ lệ cây loại II.

3. Quan sát thu nhập của một số người ở một công ty ta có kết quả cho ở bảng sau:

Thu nhập (ngàn đồng/tháng)	Số người	Thu nhập (ngàn đồng/tháng)	Số người
500 – 550	5	750 – 800	47
550 – 600	9	800 – 850	24
600 – 650	12	850 – 900	18
650 – 700	35	900 – 950	6
700 – 750	6	950 - 1000	3

a) Tìm ước lượng không chệch cho thu nhập trung bình trong một tháng của một người ở công ty này.

b) Những người có thu nhập trên 800 ngàn đồng/tháng là những người có thu nhập cao. Tìm ước lượng không chệch những người có thu nhập cao.

ƯỚC LƯỢNG KHOẢNG

4. Trọng lượng của các trái cây ở một nông trường là đại lượng ngẫu nhiên có phương sai $400g^2$. Cân thử 100 trái cây của nông trường đó ta có kết quả sau:

Trọng lượng (g)	Số trái	Trọng lượng (g)	Số trái
35 – 55	3	115 – 135	20
55 – 75	10	135 – 155	6
75 – 95	25	155 – 175	1
95 – 115	35		

a) Hãy ước lượng trọng lượng trung bình của các trái cây ở nông trường với độ tin cậy 99%.

b) Nếu độ chính xác của ước lượng là 2,35g thì độ tin cậy của ước lượng là bao nhiêu?

c) Với độ chính xác của ước lượng bé hơn 2,35g và độ tin cậy 99% thì phải cân thêm ít nhất bao nhiêu trái?

5. Một kỹ sư nông nghiệp muốn ước lượng tỉ lệ nảy mầm của một loại hạt giống. Người kỹ sư đó đem gieo thử 1000 hạt giống thì thấy có 640 hạt nảy mầm.

a) Hãy ước lượng khoảng tin cậy 90% cho tỉ lệ những hạt nảy mầm.

b) Nếu muốn có độ tin cậy 90% và bán kính ước lượng không vượt quá 0,02 thì cần lấy mẫu có kích thước bé nhất là bao nhiêu?

6. Mức hao phí nhiên liệu cho một đơn vị sản phẩm là đại lượng ngẫu nhiên X có luật phân phối chuẩn. Quan sát 28 sản phẩm thu được kết quả sau:

Lượng nhiên liệu hao phí (g)	19	19,5	20	20,5
Số sản phẩm	5	6	14	3

Với độ tin cậy 90% hãy ước lượng phương sai của X trong hai trường hợp sau:

a) Biết E(X) = 20g.

b) E(X) chưa biết.

7. Ở một nhà máy dệt, kiểm tra một số tấm vải (dài 30m), thấy kết quả như sau:

Số khuyết tật/tấm	Số tấm	Số khuyết tật/tấm	Số tấm
0	8	4	30
1	20	5	25
2	12	6	15
3	40		

a) Hãy ước lượng số khuyết tật trung bình của mỗi tấm vải với độ tin cậy 95%.

b) Nếu độ chính xác của ước lượng là 0,14 (khuyết tật) thì độ tin cậy của ước lượng là bao nhiêu?

c) Với độ chính xác của ước lượng là bé hơn 0,14, độ tin cậy 95% thì phải kiểm tra thêm ít nhất bao nhiêu tấm vải nữa.

d) ếu gọi vải loại I là loại ở mỗi tấm vải có không quá 2 khuyết tật, hãy ước lượng tỉ lệ vải loại I với độ tin cậy 99%.

8. Kiểm tra ngẫu nhiên khối lượng của 28 sản phẩm cùng loại do một máy sản xuất, ta thu được kết quả sau:

Khối lượng (kg)	Số sp
3,9 – 3,94	2
3,94 – 3,98	7
3,98 – 4,02	10
4,02 – 4,06	6
4,06 – 4,1	3

Biết rằng khối lượng sản phẩm có luật phân phối chuẩn.

a) Với độ tin cậy 0,95 hãy tìm khoảng ước lượng của khối lượng trung bình của sản phẩm.

b) Những sản phẩm có khối lượng lớn hơn 4,02 kg được xem là sản phẩm loại I. Hãy ước lượng cho tỉ lệ sản phẩm loại I với mức ý nghĩa 1%.

c) Hãy ước lượng phương sai của khối lượng sản phẩm do máy sản xuất độ tin cậy 95%.

9. Để nghiên cứu nhu cầu của một loại hàng ở một khu vực người ta tiến hành điều tra về nhu cầu của mặt hàng đó ở 400 gia đình. Kết quả điều tra cho ở bảng sau:

Nhu cầu (kg/tháng)	Số gđ	Nhu cầu (kg/tháng)	Số gđ
<1	10	4 – 5	78
1 – 2	35	5 – 6	31
2 – 3	86	6 – 7	18
3 – 4	132	7 - 8	10

Giả sử khu vực đó có 4000 hộ gia đình.

a) Ước lượng nhu cầu trung bình về mặt hàng này của khu vực trong một năm với độ tin cậy 95%.

b) Khi ước lượng nhu cầu trung bình về mặt hàng này của khu vực trong một năm, nếu ta muốn có độ tin cậy đạt được 95% và độ chính xác 4,8 tấn thì cần điều tra về nhu cầu mặt hàng này ở bao nhiêu hộ gia đình.

10. Giả sử trái cây của nông trường đã được đóng thành sọt, mỗi sọt có 10 trái. Kiểm tra 50 sọt được kết quả như sau:

Số trái hỏng/sọt	Số sọt	Số trái hỏng/sọt	Số sọt
0	0	6	4
1	2	7	7
2	3	8	0
3	7	9	0
4	20	10	1
5	6		

a) Tìm ước lượng khoảng cho tỉ lệ trái cây hỏng trong nông trường với độ tin cậy 95%.

b) Tìm ước lượng khoảng cho tỉ lệ trái cây hỏng trung bình ở mỗi sọt với mức ý nghĩa 5%.

c) Tìm ước lượng khoảng tin cậy cho phương sai của tỉ lệ trái cây hỏng ở mỗi sọt với độ tin cậy 95%. Biết tỉ lệ trái cây hỏng trung bình E(X) = 40%.

KQHT 5: KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

Bước học 1: GIỚI THIỆU CÁC KHÁI NIỆM

1.1 Các khái niệm:

1.1.1 Bài toán kiểm định trên giả thiết thống kê:

Giả thiết thống kê là dự đoán về :

- *Tham số đặc trưng của đại lượng ngẫu nhiên, như: giả thiết về trung bình, phương sai, tỉ lệ.*
- *Luật phân phối xác suất của đại lượng ngẫu nhiên, chẳng hạn, giả thiết đại lượng ngẫu nhiên có luật phân phối chuẩn.*
- *Tính độc lập của hai đại lượng ngẫu nhiên, chẳng hạn, giả thiết đại lượng ngẫu nhiên X độc lập với đại lượng ngẫu nhiên Y.*

Giả sử đại lượng ngẫu nhiên X có tham số đặc trưng θ chưa biết. Giả thiết về θ được phát biểu (H) : $\theta = \theta_0$, kèm theo đối thiết ($\overline{H}$) là một và chỉ một trong các trường hợp sau: $\theta > \theta_0$, $\theta < \theta_0$, $\theta \neq \theta_0$.

Kiểm định giả thiết thống kê là kết luận giả thiết (đối thiết) đúng hay sai dựa trên số liệu mẫu ngẫu nhiên. Kết luận nói trên thường đúng với xác suất khá lớn và có thể sai với xác suất khá nhỏ.

1.1.2 Sai lầm loại I và sai lầm loại II:

Giả thiết liên quan đến toàn tổng thể. Nhưng việc ta chỉ căn cứ vào một mẫu cụ thể để kết luận chấp nhận hay bác bỏ giả thiết (H) theo cách như trên có thể dẫn đến sai lầm. Có hai loại sai lầm:

a) Sai lầm loại I: bác bỏ giả thiết trong khi (H) đúng.

b) Sai lầm loại II: chấp nhận giả thiết trong khi (H) sai.

Hai loại sai lầm này có tính chất đối kháng, tức là muốn hạn chế khả năng phạm sai lầm loại I, ta có xu hướng làm tăng khả năng phạm sai lầm loại II và ngược lại. Vì muốn hạn chế sai lầm loại I ta có xu hướng dè dặt trong việc bác bỏ và sẽ có khuynh hướng dễ dãi trong việc chấp nhận. Khi đó lại dễ phạm sai lầm loại II. Còn muốn giảm sai lầm loại II, ta dè dặt trong việc chấp nhận và dẫn đến dễ dãi trong việc bác bỏ. Điều này làm cho nguy cơ phạm sai lầm loại I tăng lên! Tức là:

$$\text{P(sai lầm loại I)} \downarrow \Rightarrow \text{P(sai lầm loại II)}\uparrow$$

$$\text{P(sai lầm loại II)} \downarrow \Rightarrow \text{P(sai lầm loại I)}\uparrow.$$

(Tất nhiên có một cách làm giảm cả hai xác suất sai lầm nếu tăng kích thước mẫu n lên. Nhưng khi đó chi phí cũng tăng lên và đôi khi ta không phải trực tiếp làm ra được số liệu).

Giải quyết mâu thuẫn này bằng cách nào?

Thực ra sai lầm loại I và loại II rất tương đối, nó không có sẵn từ đầu, mà chỉ xác định khi ta đã đặt giả thiết. Chẳng hạn đối với một bác sĩ khám bệnh, ông ta có thể sai phải một trong hai tình huống sai lầm sau:

i/. Người có bệnh, sau khi thử nghiệm, ông kết luận không có bệnh.

ii/. Người không bệnh, sau khi thử nghiệm, ông kết luận: nhập viện!

Sai lầm nào là loại I? Sai lầm nào là loại II? Tất nhiên là chưa thể nói được.

Nếu bác sĩ đặt giả thiết (H): "người này có bệnh" thì trường hợp i) là sai lầm loại I còn ii) là sai lầm loại II. Còn nếu bác sĩ đặt giả thiết (H): "người này không bệnh" thì trường hợp i) là sai lầm loại II còn ii) là sai lầm loại I.

Nên đặt giả thiết thế nào?

Muốn vậy người ta phải xem xét sai lầm nào quan trọng hơn, tức là khi phạm phải sẽ chịu tổn thất lớn hơn, thì ta sẽ đặt bài toán để sai lầm đó là loại I.

Chẳng hạn bác sĩ điều trị bệnh lao phổi. Đó là bệnh mà nếu phát hiện để điều trị gần như chắc chắn sẽ khỏi, còn nếu không được phát hiện kịp thời để điều trị thì bệnh sẽ nặng dần và dẫn đến tử vong. Khi đó sai lầm i) "có bệnh bảo không" là quan trọng hơn, nó có thể dẫn đến tử vong, còn sai lầm ii) "không bệnh bảo có" cũng gây tổn hại, nhưng ít tổn hại hơn sai lầm i). Vì vậy với trường hợp này ta nên đặt giả thiết (H): “người này có bệnh”.

1.1.3 Mức ý nghĩa α:

Sau khi đã đặt bài toán và xác định được sai lầm loại I, ta phải đưa qui tắc kiểm định sao cho sai lầm loại I không *vượt quá một số α nhỏ không đáng kể nào đó.*

P(sai lầm loại I) ≤ α.

α bằng bao nhiêu được xem là nhỏ không đáng kể? Điều đó còn tuỳ thuộc vào mức độ quan trọng của sai lầm này, α được gọi là mức ý nghĩa của tiêu chuẩn.

Trên cơ sở đảm bảo được mức ý nghĩa α, ta sẽ cố gắng hạn chế thấp nhất có thể được xác suất phạm sai lầm loại II.

1.2 Phương pháp kiểm định giả thiết thống kê:

Các bước kiểm định một giả thiết thống kê với mức ý nghĩa α khá nhỏ được tiến hành như sau:

i/. Thành lập giả thiết (H) và đối thiết ($\overline{H}$) căn cứ vào yêu cầu thực tế.

ii/. Chọn thống kê $G(G_1, G_2, ..., G_n)$ thích hợp sao cho: nếu giả thiết (H) đúng thì thống kê G có luật phân phối xác định. Thống kê G được gọi là tiêu chuẩn kiểm định cho giả thiết (H).

iii/. Dựa vào luật phân phối xác suất của G, tìm miền W_α sao cho:

$$P(G \in W_\alpha / (H) \text{ đúng}) = \alpha$$

Như vậy: $P(G \notin W_\alpha) = 1 - \alpha$

Miền W_α được gọi là miền bác bỏ của giả thiết (H) và được thành lập dựa vào các phân vị của G, sẽ được chỉ ra trong các bài toán kiểm định cụ thể.

Số 1 - α được gọi là độ tin cậy của ước lượng.

iv/. Dựa vào mẫu cụ thể kích thước n, tính các thông số của mẫu cần thiết, thay thế vào thống kê G tính được giá trị G_0 và gọi là giá trị quan sát thực tế hay giá trị thực nghiệm của thống kê G tương với mẫu.

v/. Kết luận về giả thiết (H) và đối thiết ($\overline{H}$):

- *Nếu $G \in W_\alpha$ thì giả thiết (H) bị bác bỏ, đối thiết ($\overline{H}$) được chấp nhận.*
- *Nếu $G \notin W_\alpha$ thì chấp nhận giả thiết (H), khi đó đối thiết ($\overline{H}$) bị bác bỏ.*

Như vậy, việc chấp nhận hay bác bỏ giả thiết (H) và đối thiết ($\overline{H}$) phụ thuộc vào mức ý nghĩa α (hay độ tin cậy 1 - α) cho trước. Cùng một tiêu chuẩn kiểm định G và cùng một số liệu mẫu, giả thiết được chấp nhận hay bác bỏ tuỳ thuộc vào độ tin cậy 1 - α.

Bước học 2: KIỂM ĐỊNH CÁC THAM SỐ

2.1 Kiểm định về trung bình:

Giả sử đại lượng ngẫu nhiên X có trung bình E(X) = m chưa biết. Ta đưa ra giả thuyết để kiểm định là H: $m = m_0$ và các đối thuyết tương ứng là $\overline{H}: m \neq m_0$

hoặc $\overline{H}: m < m_0$ hoặc $\overline{H}: m > m_0$

Ta xét 3 trường hợp:

i). *Trường hợp 1:* Var(X) = σ^2 đã biết và n ≥ 30 (hoặc n < 30, X có phân phối chuẩn)

Chọn thống kê: $U = \dfrac{(\overline{X} - m_0)\sqrt{n}}{\sigma}$

Nếu H đúng thì U có phân phối chuẩn hóa, tức là U ~ N(0,1).

Với mức ý nghĩa α cho trước, ta tìm được miền bác bỏ W_α theo các đối thuyết như sau:

Nếu $\overline{H}$: $m \neq m_0$ thì $W_\alpha = (-\infty; -U_{1-\frac{\alpha}{2}}) \cup (U_{1-\frac{\alpha}{2}}; +\infty)$

Nếu $\overline{H}$: $m < m_0$ thì $W_\alpha = (-\infty; -U_{1-\alpha})$

Nếu $\overline{H}$: $m > m_0$ thì $W_\alpha = (U_{1-\alpha}; +\infty)$

Với mẫu cụ thể, ta tính được giá trị quan sát là: $U_0 = \dfrac{(\overline{x} - m_0)\sqrt{n}}{\sigma}$

Kết luận: *Nếu $U_0 \in W_\alpha$ thì bác bỏ giả thuyết H, chấp nhận đối thuyết $\overline{H}$*

Nếu $U_0 \notin W_\alpha$ thì chấp nhận giả thuyết H, bác bỏ đối thuyết $\overline{H}$

Ví dụ 1: *Khối lượng sản phẩm của đại lượng ngẫu nhiên X có trung bình qui định m = 100g, độ lệch chuẩn σ = 0,8g. Sau một thời gian sản xuất, người ta nghi ngờ khối lượng sản phẩm có xu hướng tăng lên. Kiểm tra 60 sản phẩm tính được trung bình mẫu $\overline{x}$ = 100,2g.*

a) Với độ tin cậy 95%, hãy kết luận về nghi ngờ trên.

b) Câu hỏi tương tự với độ tin cậy 99%.

c) *Với độ tin cậy lớn nhất có thể được là bao nhiêu để kết luận điều nghi ngờ nói trên là đúng?*

Giải

a) Xét giả thiết (H): m = 100g. Đối thiết ($\overline{H}$): m > 100g.

Chọn thống kê $U = \dfrac{(\overline{X} - m_0)\sqrt{n}}{\sigma}$ làm tiêu chuẩn kiểm định cho giả thiết (H).

Trong đó: σ = 0,8g, m_0 = 100g, n = 60; $\overline{X}$: thống kê nhận giá trị bằng trung bình mẫu kích thước n.

Nếu giả thiết (H) đúng thì U ∈ N(0,1).

Độ tin cậy 95% nên 1 - α = 0,95.

Miền bác bỏ: $W_\alpha = (U_{1-\alpha}, +\infty) = (U_{0.95}, +\infty) = (1{,}645. + \infty)$.

Với mẫu đã cho có: n = 60, $\overline{x}$ = 100,2g, ta có giá trị quan sát thực tế của U là:

$$U_0 = \frac{(\overline{x} - m_0)\sqrt{n}}{\sigma} = \frac{(100{,}2 - 100)\sqrt{60}}{0{,}8} = 1{,}93$$

Kết luận: $U_0 \in W_\alpha \Rightarrow$ Giả thiết (H) bị bác bỏ, chấp nhận đối thiết $(\overline{H})$ đúng. Vậy, điều nghi ngờ khối lượng sản phẩm tăng lên là đúng.

b) Lời giải tương tự câu a) nhưng độ tin cậy 99% nên 1 - α = 0,95

Ta có miền bác bỏ là: $W_\alpha = (U_{1-\alpha}, +\infty) = (U_{0.99}, +\infty) = (2{,}326,. + \infty)$.

Kết luận: $U_0 \notin W_\alpha$ nên chấp nhận giả thiết (H), bác bỏ đối thiết ($\overline{H}$).

Vậy, điều nghi ngờ khối lượng tăng lên là sai.

c) Để kết luận điều nghi ngờ khối lượng tăng lên là đúng thì phải bác bỏ được giả thiết (H), nghĩa là: U_0 = 1,93 $\in$ $W_\alpha = (U_{1-\alpha}, +\infty)$.

Tìm giá trị 1 - α lớn nhất có thể được để $U_{1-\alpha}$ < 1,93. Dựa bảng phân vị chuẩn tắc, ta có 1 - α = 0,973.

ii) *Trường hợp 2:* Var(X) = σ^2 chưa biết và n ≥ 30

Chọn thống kê: $U = \dfrac{(\overline{X} - m_0)\sqrt{n}}{S'}$

Nếu H đúng thì U có phân phối chuẩn hóa tức là U ~ N(0,1).

Với mức ý nghĩa α cho trước, ta tìm được miền bác bỏ W_α tương ứng với các đối thuyết giống như trường hợp 1.

Với mẫu cụ thể, ta tính được giá trị quan sát là: $U_0 = \dfrac{(\overline{x} - m_0)\sqrt{n}}{s'}$

Kết luận: *Giống như trường hợp 1*

iii) *Trường hợp 3:* Var(X) = σ^2 chưa biết và n < 30, X có phân phối chuẩn .

Chọn thống kê: $T = \dfrac{(\overline{X} - m_0)\sqrt{n}}{S'}$

Nếu H đúng thì U có phân phối Student với n -1 bậc tự do, T ~ T(n-1)

Với mức ý nghĩa α cho trước, ta tìm được miền bác bỏ W_α tương ứng với các đối thuyết như sau:

Nếu $\overline{H}$: m ≠ m_0 thì $W_\alpha = (-\infty; -t_{n-1;1-\frac{\alpha}{2}}) \cup (t_{n-1;1-\frac{\alpha}{2}}; +\infty)$

Nếu $\overline{H}$: m < m_0 thì $W_\alpha = (-\infty; -t_{n-1;1-\alpha})$

Nếu $\overline{H}$: m > m_0 thì $W_\alpha = (t_{n-1;1-\alpha}; +\infty)$

Với mẫu cụ thể, ta tính được giá trị quan sát là: $t_0 = \dfrac{(\overline{x} - m_0)\sqrt{n}}{s'}$

Kết kuận: *Nếu $t_0 \in W_\alpha$ thì bác bỏ giả thuyết H, chấp nhận đối thuyết $\overline{H}$*

Nếu $t_0 \notin W_\alpha$ thì chấp nhận giả thuyết H, bác bỏ đối thuyết $\overline{H}$

Ví dụ 2: *Độ dài chi tiết máy là đại lượng ngẫu nhiên X có luật phân phối chuẩn. Kiểm tra 28 sản phẩm thu được số liệu như sau: (đơn vị tính cm)*

20,10	20,05	20,03	19,98	20,00	20,02	20,01
20,00	20,02	19,99	19,97	20,02	19,99	19,96
19,97	20,00	20,00	20,02	20,03	19,97	20,00
20,01	20,04	19,99	20,03	20,02	20,00	20,04

Với độ tin cậy 95%, có thể cho rằng trung bình độ dài chi tiết máy bằng 20cm hay không?

Giải

Xét giả thiết (H): m = 20cm. Đối thiết ($\overline{H}$): m ≠ 20cm.

Chọn thống kê: $T = \dfrac{(\overline{X} - m_0)\sqrt{n}}{S'}$ làm tiêu chuẩn kiểm định cho giả thiết (H).

Trong đó: m_0 = 20, n = 28, $\overline{X}$, s': lần lượt là các thống kê nhận giá trị bằng trung bình mẫu và độ lệch tiêu chuẩn điều chỉnh mẫu.

Nếu giả thiết (H) đúng thì T có luật phân phối Student bậc tự do n – 1 = 27, T ∈ T(27).

Độ tin cậy 95% nên 1 - α = 0,95 suy ra $1 - \dfrac{\alpha}{2} = 0{,}975$.

Miền bác bỏ:

$$\begin{aligned} W_\alpha &= \left(-\infty, -T_{1-\frac{\alpha}{2}}(n-1)\right) \cup \left(T_{1-\frac{\alpha}{2}}(n-1), +\infty\right) \\ &= (-\infty, -T_{0.975}(27)) \cup (T_{0.975}(27), +\infty) \\ &= (-\infty, -2{,}052) \cup (0{,}052, +\infty) \end{aligned}$$

Với mẫu đã cho: n = 28, $\overline{x}$ = 20,01cm, s'= 0,024cm. Ta có giá trị thực nghiệm của T tương ứng với mẫu là:

$$T_0 = \frac{(20{,}01 - 20)\sqrt{28}}{0{,}024} = 2{,}205$$

Kết luận: $T_0 \in W_\alpha \Rightarrow$ Giả thiết (H) bị bác bỏ, đối thiết $(\overline{H})$ được chấp nhận. Vậy, không thể cho rằng trung bình độ dài chi tiết máy bằng 20cm.

Ví dụ 3: *Một nhóm người nghiên cứu tuyên bố rằng trung bình một người vào siêu thì X tiêu hết 140 nghìn đồng. Chọn ngẫu nhiên 50 người mua hàng, tính được số tiền trung bình họ tiêu là 154 nghìn đồng với độ lệch tiêu chuẩn điều chỉnh của mẫu là s'=62. Với mức ý nghĩa 0,02 hãy kiểm định xem tuyên bố của nhóm người nghiên cứu có đúng hay không?*

Ví dụ 4: *Trọng lượng của các bao gạo là đại lượng ngẫu nhiên X có phân phối chuẩn với trọng lượng trung bình là E(X) = 50 kg. Sau một khoảng thời gian hoạt động, người ta nghi ngờ trọng lượng các bao gạo có thay đổi. Cân thử 25 bao và thu được kết quả như sau:*

X (khối lượng)	n_i (số bao)
48 – 48,5	2
48,5 – 49	5
49 – 49,5	10
49,5 – 50	6
50 – 50,5	2

Hãy kết luận điều nghi ngờ nói trên.

2.2 Kiểm định về tỉ lệ:

Giả sử tổng thể có hai loại phần tử (phần tử có tính chất A và không có tính chất A). Gọi p là tỉ lệ phần tử có tính chất A của tổng thể. Ta đưa ra giả thuyết về kiểm định H: p = p_0. Khi đó, H sẽ nhận một trong các đối thuyết tương ứng là:

$\overline{H}: p \neq p_0$

hoặc $\overline{H}: p < p_0$ hoặc $\overline{H}: p > p_0$

Chọn thống kê $U = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0 q_0}}$

Nếu H đúng thì U có phân phối chuẩn hóa tức là U ~ N(0,1).

Với mức ý nghĩa α cho trước, ta tìm được miền bác bỏ W_α tương ứng với các đối thuyết như sau:

Nếu $\overline{H}: p \neq p_0$ thì $W_\alpha = (-\infty; -U_{1-\frac{\alpha}{2}}) \cup (U_{1-\frac{\alpha}{2}}; +\infty)$

Nếu $\overline{H}: p < p_0$ thì $W_\alpha = (-\infty; -U_{1-\alpha})$

Nếu $\overline{H}: p > p_0$ thì $W_\alpha = (U_{1-\alpha}; +\infty)$

Với mẫu cụ thể, ta tính được giá trị quan sát là: $U_0 = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0 q_0}}$

Trong đó f là tỉ lệ phần tử có tính chất A.

Kết kuận: *Nếu $U_0 \in W_\alpha$ thì bác bỏ giả thuyết H, chấp nhận đối thuyết $\overline{H}$*

Nếu $U_0 \notin W_\alpha$ thì chấp nhận giả thuyết H, bác bỏ đối thuyết $\overline{H}$

Ví dụ 5: *Tỉ lệ phế phẩm của máy là p = 5%. Sau khi cải tiến kỹ thuật, kiểm tra 400 sản phẩm có 12 phế phẩm. Với độ tin cậy 99%, có thể kết luận việc cải tiến kỹ thuật có hiệu quả hay không?*

Xét giả thiết (H): p = 0,05. Đối thiết $(\overline{H})$: p < 0,05.

Chọn thống kê: $U = \dfrac{(f - p)\sqrt{n}}{\sqrt{pq}}$ làm tiêu chuẩn kiểm định giả thiết (H).

Trong đó: $p_0 = 0,05$, $q_0 = 1 - 0,05 = 0,95$, n = 400, f là thống kê nhận giá trị bằng tỉ lệ mẫu.

Độ tin cậy 99% nên 1 - α = 0,99.

Miền bác bỏ: $W_\alpha = \left(-\infty, -U_{,99}\right) = \left(-\infty, -2,326\right)$.

Với mẫu có kích thước n = 400 và tỉ lệ mẫu $f = \frac{12}{400} = 0,03$. Ta có giá trị quan sát thực tế của U là:

$$U_0 = \frac{(0,03 - 0,05)\sqrt{400}}{\sqrt{(0,05)(0,95)}} = -1,835$$

Kết luận: $U_0 \notin W_\alpha \Rightarrow$ Chấp nhận giả thiết (H), bác bỏ đối thiết $\left(\overline{H}\right)$. Vậy, chưa thể cho rằng việc cải tiến kỹ thuật có hiệu quả.

2.3 Kiểm định về phương sai:

Giả sử đại lượng ngẫu nhiên X có phân phối chuẩn với phương sai Var(X) = σ^2 chưa biết. Ta đưa ra giả thuyết để kiểm định H: $\sigma^2 = \sigma^2_0$. Khi đó, H sẽ nhận một trong các đối thuyết tương ứng là:

$\overline{H} : \sigma^2 \neq \sigma^2_0$

hoặc $\overline{H} : \sigma^2 < \sigma^2_0$ hoặc $\overline{H} : \sigma^2 > \sigma^2_0$

Chọn thống kê $\chi^2 = \frac{(n-1)s'^2}{\sigma_0^2}$

Nếu H đúng thì χ^2 có phân phối $\chi^2 \sim \chi^2$ (n -1).

Với mức ý nghĩa α cho trước, ta tìm được miền bác bỏ W_α tương ứng với các đối thuyết như sau:

Nếu $\overline{H} : \sigma^2 \neq \sigma^2_0$ thì $W_\alpha = (-\infty; \chi^2_{n-1;\frac{\alpha}{2}}) \cup (\chi^2_{n-1;1-\frac{\alpha}{2}}; +\infty)$

Nếu $\overline{H} : \sigma^2 < \sigma^2_0$ thì $W_\alpha = (-\infty; \chi^2_{n-1;\alpha})$

Nếu $\overline{H} : \sigma^2 > \sigma^2_0$ thì $W_\alpha = (\chi^2_{n-1;1-\alpha}; +\infty)$

Với mẫu cụ thể, ta tính được giá trị quan sát là: $\chi_0^2 = \frac{(n-1)s'^2}{\sigma_0^2}$

Kết kuận: *Nếu $\chi^2_0 \in W_\alpha$ thì bác bỏ giả thuyết H, chấp nhận đối thuyết $\overline{H}$*

Nếu $\chi^2_0 \notin W_\alpha$ thì chấp nhận giả thuyết H, bác bỏ đối thuyết $\overline{H}$

Ví dụ 6: *Khối lượng sản phẩm do hệ thống máy sản xuất là đại lượng ngẫu nhiên X có luật phân phối chuẩn, phương sai Var(X) = 15 g^2. Sau một thời gian sản xuất, người ta nghi ngờ rằng khối lượng các sản phẩm được sản xuất ra không ổn định. Kiểm tra 25 sản phẩm, tính được phương sai điều chỉnh $s'^2 = 26g^2$. Với độ tin cậy 99%, hãy kết luận về nghi ngờ trên.*

Xét giả thiết (H): $\sigma^2 = 15g^2$. Đối thiết $\left(\overline{H}\right)$: $\sigma^2 > 15g^2$.

Chọn thống kê: $\chi^2 = \frac{(n-1)S'^2}{\sigma_0^2}$ làm tiêu chuẩn kiểm định cho giả thiết (H).

Trong đó: $\sigma_0^2 = 15g^2$, n = 25, S'^2 là thống kê nhận giá trị bằng phương sai điều chỉnh mẫu.

Nếu giả thiết (H) đúng thì χ^2 có luật phân phối khi bình phương bậc tự do n - 1 = 24, $\chi^2 \in \chi^2(24)$

Với độ tin cậy 99% nên 1 - α = 0,99

Miền bác bỏ: $W_\alpha = \left(\chi^2_{1-\alpha}(n-1), +\infty\right) = \left(\chi^2_{0,99}(24), +\infty\right) = (42{,}98, +\infty)$

Với mẫu cụ thể có n = 25, $s'^2 = 26$. Ta có giá trị quan sát thực tế của χ^2 là:

$$\chi_0^2 = \frac{(n-1)s'^2}{\sigma_0^2} = \frac{(25-1).(26)}{15} = 41{,}6$$

Kết luận: $\chi_0^2 \notin W_\alpha \Rightarrow$ Chấp nhận giả thiết (H), bác bỏ đối thiết $\left(\overline{H}\right)$.Vậy, điều nghi ngờ là sai.

2.4 Kiểm đinh về sự bằng nhau của hai trung bình:

Giả sử hai đại lượng ngẫu nhiên X và Y độc lập có luật phân phối chuẩn với hai tham số trung bình E(X) và E(Y) chưa biết.

Xét giả thiết (H): E(X) = E(Y). Đối thiết $\left(\overline{H}\right)$ là một và chỉ một trong các trường hợp sau: E(X) > E(Y), E(X) < E(Y), E(X) ≠ E(Y).

Với số α khá nhỏ, hãy kiểm định giả thiết (H) với mức ý nghĩa α.

Ta có các trường hợp sau:

1. **Trường hợp 1:**

$$\text{Phương sai } \mathrm{Var(X)} = \sigma_X^2, \mathrm{Var(Y)} = \sigma_Y^2 \text{ đã biết}.$$

Lập mẫu ngẫu nhiên $W_X = \left(X_1, X_2, ..., X_{n_X}\right)$, $W_Y = \left(Y_1, Y_2, ..., Y_{n_Y}\right)$ đối với X và Y.

Thống kê được chọn làm tiêu chuẩn kiểm định cho giả thiết (H) là:

$$U = \frac{\overline{X} - \overline{Y} - \left(E(X) - E(Y)\right)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}$$

Trong đó: + $\overline{X}, \overline{Y}$: lần lượt là các thống kê nhận giá trị bằng trung bình mẫu của đại lượng ngẫu nhiên X và Y.

+ n_X, n_Y : lần lượt là kích thước mẫu của đại lượng ngẫu nhiên X và Y.

+ σ_X^2, σ_Y^2 : lần lượt là phương sai của đại lượng ngẫu nhiên X và Y.

Nếu giả thiết (H) đúng thì $U = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}$ có luật phân phối chuẩn tắc: U ∈ N(0,1).

Với mẫu cụ thể $w_x = \left(x_1, x_2, ..., x_{n_X}\right)$, $w_y = \left(y_1, y_2, ..., y_{n_Y}\right)$ của đại lượng ngẫu nhiên X và Y, tính được $\overline{x}, \overline{y}$ lần lượt là trung bình mẫu cụ thể của đại lượng ngẫu nhiên X và Y. Ta có giá trị quan sát thực tế của thống kê U tương ứng với mẫu cụ thể là:

$$U_0 = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$$

Miền bác bỏ W_α được thành lập theo dạng U như sau:

- Nếu ($\overline{H}$) có dạng E(X) > E(Y) thì $W_\alpha = \left(U_{1-\alpha}, +\infty\right)$.
- Nếu ($\overline{H}$) có dạng E(X) < E(Y) thì $W_\alpha = \left(-\infty, -U_{1-\alpha}\right)$.
- Nếu ($\overline{H}$) có dạng E(X) ≠ E(Y) thì: $W_\alpha = \left(-\infty, -U_{1-\frac{\alpha}{2}}\right) \cup \left(U_{1-\frac{\alpha}{2}}, +\infty\right)$

Ví dụ 7: *Trọng lượng sản phẩm do hai nhà máy sản xuất là đại lượng ngẫu nhiên có luật phân phối chuẩn và có cùng độ lệch tiêu chuẩn là $\sigma = 1kg$. Với mức ý nghĩa α = 0,05, có thể xem trọng lượng trung bình của sản phẩm do hai nhà máy sản xuất ra là như nhau hay không? Nếu cần thử 25 sản phẩm của nhà máy A ta tính được $\overline{x} = 50kg$, cân 20 sản phẩm của nhà máy B thì tính được $\overline{y} = 50{,}6kg$.*

Gọi trọng lượng của nhà máy A là X, trọng lượng của nhà máy B là Y.

Ta có X, Y là các đại lượng ngẫu nhiên có luật phân phối chuẩn.

Và $Var(X) = Var(Y) = 1$

Xét giả thiết (H): E(X) = E(Y). Đối thiết $\left(\overline{H}\right)$: E(X) ≠ E(Y).

Chọn thống kê: $U = \dfrac{\overline{X} - \overline{Y} - (E(X) - E(Y))}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$ làm tiêu chuẩn kiểm định cho giả thiết (H).

Trong đó: + $\overline{X}, \overline{Y}$: lần lượt là các thống kê nhận giá trị bằng trung bình mẫu của đại lượng ngẫu nhiên X và Y.

+ n_X, n_Y: lần lượt là kích thước mẫu của đại lượng ngẫu nhiên X và Y.

+ σ_X^2, σ_Y^2: lần lượt là phương sai của đại lượng ngẫu nhiên X và Y.

Với giả thiết (H) đúng thì $U = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$ có luật phân phối chuẩn tắc.

Với mức ý nghĩa α = 0,05 $\Rightarrow 1 - \dfrac{\alpha}{2} = 0{,}975 \Rightarrow U_\alpha = 1{,}96$

Miền bác bỏ: $W_\alpha = \left(-\infty, -U_{1-\frac{\alpha}{2}}\right) \cup \left(U_{1-\frac{\alpha}{2}}, +\infty\right)$

$$= \left(-\infty, -U_{0.975}\right) \cup \left(U_{0.975} + \infty\right) = \left(-\infty, -1{,}96\right) \cup \left(1{,}96, +\infty\right)$$

Với mẫu cụ thể có $n_X = 25$, $n_Y = 20$, $\overline{x} = 50kg$, $\overline{y} = 50,6kg$, ta tính được giá trị quan sát thực tế của U là:

$$U_0 = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} = \frac{50 - 50,6}{\sqrt{\frac{1}{25} + \frac{1}{20}}} = -2$$

Kết luận: $U_0 \in W_\alpha \Rightarrow$ bác bỏ giả thiết (H), chấp nhận đối thiết $(\overline{H})$. Vậy trọng lượng trung bình của sản phẩm sản xuất ở hai nhà máy là khác nhau.

2. Trường hợp 2:

$$\begin{cases} \text{Phương sai Var(X), Var(Y) chưa biết,} \\ \text{Kích thước mẫu } n_X \geq 30, n_Y \geq 30. \end{cases}$$

Lập mẫu ngẫu nhiên $W_X = (X_1, X_2, ..., X_{n_X})$, $W_Y = (Y_1, Y_2, ..., Y_{n_Y})$ đối với X và Y.

Thống kê được chọn làm tiêu chuẩn kiểm định cho giả thiết (H) là:

$$U = \frac{\overline{X} - \overline{Y} - (E(X) - E(Y))}{\sqrt{\frac{S'^2_X}{n_X} + \frac{S'^2_Y}{n_Y}}}$$

Trong đó: + $\overline{X}$, $\overline{Y}$: lần lượt là các thống kê nhận giá trị bằng trung bình mẫu của đại lượng ngẫu nhiên X và Y.

+ n_X, n_Y : lần lượt là kích thước mẫu của đại lượng ngẫu nhiên X và Y.

+ S'^2_X, S'^2_Y : lần lượt là thống kê nhận giá trị bằng phương sai điều chỉnh mẫu của đại lượng ngẫu nhiên X và Y.

Nếu giả thiết (H) đúng thì $U = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\frac{S'^2_X}{n_X} + \frac{S'^2_Y}{n_Y}}}$ có luật phân phối chuẩn tắc: U ∈ N(0,1).

Với mẫu cụ thể $w_x = (x_1, x_2, ..., x_{n_X})$, $w_y = (y_1, y_2, ..., y_{n_Y})$ của đại lượng ngẫu nhiên X và Y, tính được $\overline{x}$, $\overline{y}$, s'^2_X, s'^2_Y lần lượt là trung bình, phương sai điều chỉnh của mẫu cụ thể của đại lượng ngẫu nhiên X và Y. Ta có giá trị quan sát thực tế của thống kê U tương ứng với mẫu cụ thể là:

$$U_0 = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s'^2_X}{n_X} + \frac{s'^2_Y}{n_Y}}}$$

Miền bác bỏ W_α được thành lập theo dạng U giống (5 - 3) của trường hợp 1.

Ví dụ 8: *Theo một tài liệu của viện nghiên cứu phát triển gia cầm thì hai giống gà X và Y có trọng lượng trung bình ở 3 tháng tuổi là như nhau. Ta nuôi thử mỗi giống 100 con và ở 3 tháng tuổi cân lại ta tính được kết quả tương ứng là:*

$$\overline{x} = 1825g, \quad s'^2_X = 1628g^2, \quad \overline{y} = 1937g, \quad s'^2_Y = 1876g^2$$

Hãy căn cứ vào mẫu đó cho nhận xét về tài liệu trên với mức ý nghĩa 1%

Xét giả thiết (H): E(X) = E(Y). Đối thiết $\left(\overline{H}\right)$: E(X) ≠ E(Y).

Chọn thống kê: $U = \dfrac{\overline{X} - \overline{Y} - \left(E(X) - E(Y)\right)}{\sqrt{\dfrac{S'^2_X}{n_X} + \dfrac{S'^2_Y}{n_Y}}}$ làm tiêu chuẩn kiểm định cho giả thiết (H).

Trong đó: + $n_X = 100$, $n_Y = 100$.

+ $\overline{X}, \overline{Y}, S'^2_X, S'^2_Y$: lần lượt là các thống kê nhận giá trị bằng trung bình mẫu, phương sai điều chỉnh mẫu của đại lượng ngẫu nhiên X và Y.

Nếu giả thiết (H) đúng thì $U = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S'^2_X}{n_X} + \dfrac{S'^2_Y}{n_Y}}}$ có luật phân phối chuẩn tắc.

Mức ý nghĩa 1% nên α = 0.01 ⇒ 1 - $\dfrac{\alpha}{2}$ = 0,995.

Miền bác bỏ: $W_\alpha = \left(-\infty, -U_{1-\frac{\alpha}{2}}\right) \cup \left(U_{1-\frac{\alpha}{2}}, +\infty\right) = \left(-\infty, -U_{0.995}\right) \cup \left(U_{0.995} + \infty\right)$

$$= \left(-\infty, -2{,}576\right) \cup \left(2{,}576, +\infty\right)$$

Với mẫu đã cho có $n_X = 100, n_Y = 100, \overline{x} = 1825g$, $\overline{y} = 1937g$, $s'^2_X = 1628g^2$, $s'^2_Y = 1876g^2$. Ta có giá trị thực tế của U là:

$$U_0 = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{s'^2_X}{n_X} + \dfrac{s'^2_Y}{n_Y}}} = \frac{1825 - 1973}{\sqrt{\dfrac{1628}{100} + \dfrac{1876}{100}}} = -25$$

Kết luận: $U_0 \notin W_\alpha \Rightarrow$ chấp nhận giả thiết (H), bác bỏ đối thiết $\left(\overline{H}\right)$. Vậy tài liệu của viện nghiên cứu là chính xác.

3. Trường hợp 3:

$$\begin{cases} \text{Phương sai Var(X)} = \sigma^2_{\text{X}}, \text{Var(Y)} = \sigma^2_{\text{Y}} \text{ chưa biết,} \\ \text{Kích thước mẫu } \text{n}_{\text{X}} < 30, \text{n}_{\text{Y}} < 30, \\ \text{Giả sử } s'_X \approx s'_Y \ (\dfrac{s'_X}{s'_Y} < 1{,}5 \text{ nếu } s'_X > s'_Y \text{ hoặc } \dfrac{s'_Y}{s'_X} < 1{,}5 \text{ nếu } s'_Y > s'_X\ ) \end{cases}$$

Lập mẫu ngẫu nhiên $W_X = \left(X_1, X_2, ..., X_{n_X}\right)$, $W_Y = \left(Y_1, Y_2, ..., Y_{n_Y}\right)$ đối với X và Y.

Thống kê được chọn làm tiêu chuẩn kiểm định cho giả thiết (H) là:

$$T = \frac{\overline{X} - \overline{Y} - \left(E(X) - E(Y)\right)}{S.\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}}$$

Trong đó: + $\overline{X}, \overline{Y}$: lần lượt là các thống kê nhận giá trị bằng trung bình mẫu của đại lượng ngẫu nhiên X và Y.

+ n_X, n_Y : lần lượt là kích thước mẫu của đại lượng ngẫu nhiên X và Y.

+ S'^2_X, S'^2_Y : lần lượt là thống kê nhận giá trị bằng phương sai điều chỉnh mẫu của đại lượng ngẫu nhiên X và Y.

+ $S = \sqrt{\dfrac{(n_X - 1)S'^2_X + (n_Y - 1)S'^2_Y}{m_X + n_Y - 2}}$ được gọi là phương sai gọp.

Nếu giả thiết (H) đúng thì $T = \dfrac{\overline{X} - \overline{Y}}{S.\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}}$ có luật phân phối Student bậc tự do $(n_X + n_Y - 2)$: T $\in$ T($n_X + n_Y$ -2).

Với mẫu cụ thể $w_x = (x_1, x_2, ..., x_{n_X})$, $w_y = (y_1, y_2, ..., y_{n_Y})$ của đại lượng ngẫu nhiên X và Y, tính được $\overline{x}$, $\overline{y}$, s'^2_X, s'^2_Y lần lượt là trung bình, phương sai điều chỉnh của mẫu cụ thể của đại lượng ngẫu nhiên X và Y. Ta có giá trị quan sát thực tế của thống kê U tương ứng với mẫu cụ thể là:

$$T_0 = \frac{\overline{x} - \overline{y}}{s.\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}} \qquad \text{với} \qquad s = \sqrt{\frac{(n_X - 1)s'^2_X + (n_Y - 1)s'^2_Y}{n_X + n_Y - 2}}$$

Miền bác bỏ W_α được thành lập theo dạng T (với bậc tự do $(n_X + n_Y - 2)$) như sau:

- Nếu ($\overline{H}$) có dạng E(X) > E(Y) thì $W_\alpha = \left(T_{1-\alpha}(n_X + n_Y - 2), +\infty\right)$.
- Nếu ($\overline{H}$) có dạng E(X) < E(Y) thì $W_\alpha = \left(-\infty, -T_{1-\alpha}(n_X + n_Y - 2)\right)$.
- Nếu ($\overline{H}$) có dạng E(X) $\neq$ E(Y) thì

$$W_\alpha = \left(-\infty, -T_{1-\frac{\alpha}{2}}(n_X + n_Y - 2)\right) \cup \left(T_{1-\frac{\alpha}{2}}(n_X + n_Y - 2), +\infty\right)$$

Ví dụ 9: *Dùng hai phương pháp để cùng làm một loại sản phẩm. Phương pháp A được một nhóm 12 người thực hiện có năng suất trung bình là 45 sản phẩm trong một ca làm việc, với độ lệch tiêu chuẩn điều chỉnh mẫu là 5 sản phẩm. Phương pháp B được một nhóm 15 người khác thực hiện, có năng suất trung bình là 53 sản phẩm trong một ca làm việc, với độ lệch tiêu chuẩn điều chỉnh mẫu là 6 sản phẩm. Với mức ý nghĩa α = 0,05, hãy kiểm tra hiệu quả của hai phương pháp này có bằng nhau không?*

Gọi X, Y lần lượt là số sản phẩm được sản xuất ra từ phương pháp A và B.

Xét giả thiết (H): E(X) = E(Y). Đối thiết $(\overline{H})$: E(X) $\neq$ E(Y).

Theo giả thiết bài toán ta có: $s'_X = 5$, $s'_Y = 6$

$$\Rightarrow \frac{s'_Y}{s'_X} = \frac{6}{5} = 1{,}2 < 1{,}5$$

$$\Rightarrow s'_X \approx s'_Y$$

Chọn thống kê làm tiêu chuẩn kiểm định cho giả thiết (H) là:

$$T=\frac{\overline{X}-\overline{Y}-\left(E(X)-E(Y)\right)}{S.\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}}, \text{ trong đó } S=\sqrt{\frac{(n_X-1)S'^2_X+(n_Y-1)S'^2_Y}{m_X+n_Y-2}}$$

Nếu giả thiết (H) đúng thì $T=\dfrac{\overline{X}-\overline{Y}}{S.\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}}$ có luật phân phối Student bậc tự do (n_X+n_Y-2): T $\in$ T(n_X+n_Y -2).

Với mức ý nghĩa $\alpha=0{,}05\Rightarrow\dfrac{\alpha}{2}=0{,}025\Rightarrow 1-\dfrac{\alpha}{2}=0{,}975$

$$\Rightarrow T_{1-\frac{\alpha}{2}}(n_X+n_Y-2)=T_{1-\frac{\alpha}{2}}(12+15-2)=T_{1-\frac{\alpha}{2}}(25)=2{,}060$$

Miền bác bỏ: $W_\alpha=\left(-\infty,-T_{1-\frac{\alpha}{2}}(n_X+n_Y-2)\right)\cup\left(T_{1-\frac{\alpha}{2}}(n_X+n_Y-2),+\infty\right)$

$$=(-\infty,-2{,}060)\cup(2{,}060,+\infty)$$

Với mẫu đã cho có $n_X=12, n_Y=15, \overline{x}=45, \overline{y}=53, s'_X=5, s'_Y=6$, ta tính được giá trị của phương sai gọp tương ứng với mẫu là:

$$s=\sqrt{\frac{(n_X-1)s'^2_X+(n_Y-1)s'^2_Y}{n_X+n_Y-2}}=\sqrt{\frac{11{,}25+14{,}36}{12+15-2}}\approx 5{,}58$$

Suy ra giá trị quan sát thực tế của T là:

$$T_0=\frac{\overline{x}-\overline{y}}{s.\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}}=\frac{45-53}{(5{,}58).\sqrt{\frac{1}{12}+\frac{1}{15}}}\approx 3{,}7$$

Kết luận: $T_0\in W_\alpha\Rightarrow$ Bác bỏ giả thiết (H), chấp nhận đối thiết $(\overline{H})$. Vậy hiệu quả của hai phương pháp này không bằng nhau.

4. Trường hợp 4: So sánh cặp.

Giả sử hai đại lượng ngẫu nhiên X và Y có hai mẫu ngẫu nhiên cùng kích thước n: $W_X=\left(X_1,X_2,...,X_{n_X}\right)$, $W_Y=\left(Y_1,Y_2,...,Y_{n_Y}\right)$ với các X_i và $\mathrm{Y_i}$ $\left(i=\overline{1,n}\right)$ từng đôi một tương ứng. Cần kiểm định giả thiết (H): E(X) = E(Y).

Để giải bài toán này ta xét hiệu số: D = X – Y.

Suy ra D cũng là đại lượng ngẫu nhiên.

Gọi μ_X, μ_Y, μ lần lượt là trung bình của đại lượng ngẫu nhiên X, Y, D.

Với mẫu cụ thể $w_x=\left(x_1,x_2,...,x_{n_X}\right)$, $w_y=\left(y_1,y_2,...,y_{n_Y}\right)$ của hai đại lượng ngẫu nhiên X và Y, ta tính được mẫu cụ thể của đại lượng ngẫu nhiên D: $w_d=(d_1,d_2,...,d_n)$ với $d_i=x_i-y_i$, $\left(i=\overline{1,n}\right)$.

Giả thiết ta muốn kiểm định: (H): $\mu_X=\mu_Y$ được quy về bài toán kiểm định giả thiết: (H): $\mu_X-\mu_Y=0$ hay (H): $\mu=0$.

Như vậy, ta đã đưa bài toán so sánh về bài toán kiểm định giả thiết về trung bình ở §2. Tuy nhiên, trong trường hợp này kích thước mẫu n của các đại lượng ngẫu nhiên thường nhỏ hơn 30 (n <30) và phương sai σ_X^2, σ_Y^2 là chưa biết, nên bài toán kiểm định rơi vào trường hợp 3. Ta thực hiện như sau:

Thống kê được chọn làm tiêu chuẩn kiểm định cho giả thiết (H) là:

$$T = \frac{\overline{D}\sqrt{n}}{S'_D}$$, trong đó: . n: kích thước mẫu.

. $\overline{D}, S'_D$ lần lượt là thống kê nhận giá trị bằng trung bình và độ lệch tiêu chuẩn điều chỉnh mẫu cụ thể của đại lượng ngẫu nhiên D.

Nếu giả thiết (H) đúng thì T có luật phân phối Student bậc tự do (n – 1): T ∈ T(n – 1).

Với mẫu cụ thể $w_x = (x_1, x_2, ..., x_{n_X})$, $w_y = (y_1, y_2, ..., y_{n_Y})$ X và Y, ta tìm được mẫu cụ thể $w_d = (d_1, d_2, ..., d_n)$ của D. Từ đó ta tính được trung bình mẫu $\overline{d}$ và độ lệch tiêu chuẩn điều chỉnh s'_D. Ta có giá trị quan sát thực tế của T là:

$$T_0 = \frac{\overline{d}.\sqrt{n}}{s'_D}$$

Miền bác bỏ W_α được thành lập theo dạng T (với bậc tự do $(n_X + n_Y - 2)$) như sau:

- Nếu ($\overline{H}$) có dạng E(X) > E(Y) thì $W_\alpha = (T_{1-\alpha}(n-1), +\infty)$.
- Nếu ($\overline{H}$) có dạng E(X) < E(Y) thì $W_\alpha = (-\infty, -T_{1-\alpha}(n-1))$.
- Nếu ($\overline{H}$) có dạng E(X) ≠ E(Y) thì: $W_\alpha = \left(-\infty, -T_{1-\frac{\alpha}{2}}(n-1)\right) \cup \left(T_{1-\frac{\alpha}{2}}(n-1), +\infty\right)$

Ví dụ 10: *Người ta tiến hành một cuộc khảo sát về giá cả của hai cửa hiệu thực phẩm lớn trong thành phố, 12 mặt hàng thông dụng nhất được chọn ngẫu nhiên và giá của chúng bán ở hai cửa hiệu được ghi lại như sau:*

Mặt hàng	1	2	3	4	5	6
Cửa hiệu A	0,89	0,59	1,29	1,50	2,49	0,65
Cửa hiệu B	0,95	0,55	1,49	0,69	2,39	0,79

Mặt hàng	7	8	9	10	11	12
Cửa hiệu A	0,99	1,99	2,25	0,50	1,99	1,79
Cửa hiệu B	0,99	1,79	2,39	0,59	2,19	1,99

Với mức ý nghĩa α = 2%, hãy kiểm định xem có sự khác nhau về giá cả trung bình của các mặt hàng ở hai cửa hiệu hay không?

Gọi X, Y lần lượt là giá của các mặt hàng ở cưả hiệu A và B.

Xét giả thiết (H): E(X) = E(Y), đối thiết ($\overline{H}$): E(X) ≠ E(Y).

Chọn thống kê làm tiêu chuẩn kiểm định cho giả thiết (H) là:

$$T = \frac{\overline{D}\sqrt{n}}{S'_D}$$

trong đó: . n: kích thước mẫu.

. $\overline{D}, S'_D$ lần lượt là thống kê nhận giá trị bằng trung bình và độ lệch tiêu chuẩn điều chỉnh mẫu cụ thể của đại lượng ngẫu nhiên D.

Nếu giả thiết (H) đúng thì T có luật phân phối Student bậc tự do (n – 1): T ∈ T(n–1).

Với mức ý nghĩa α = 2% = 0,02 $\Rightarrow \frac{\alpha}{2} = 0{,}01 \Rightarrow 1 - \frac{\alpha}{2} = 0{,}99$

$$\Rightarrow T_{1-\frac{\alpha}{2}}(n-1) = T_{0,99}(11) = 2{,}718$$

Miền bác bỏ: $W_\alpha = \left(-\infty, -T_{1-\frac{\alpha}{2}}(n-1)\right) \cup \left(T_{1-\frac{\alpha}{2}}(n-1), +\infty\right) = (-\infty, 2{,}718) \cup (2{,}718, +\infty)$

Với mẫu cụ thể đã cho trong giả thiết, ta lập bảng các giá trị của hiệu số D = X – Y:

Mặt hàng	X	Y	D = X – Y
1	0,89	0,95	- 0,06
2	0,59	0,55	0,04
3	1,29	1,49	- 0,2
4	1,50	1,69	- 0,19
5	2,49	2,39	0,1
6	0,65	0,79	- 0,14
7	0,99	0,99	0
8	1,99	1,79	0,2
9	2,25	2,39	- 0,14
10	0,50	0,59	- 0,09
11	1,99	2,19	- 0,2
12	1,79	1,99	- 0,2

Từ bảng này ta tính được: $\overline{d} = -0{,}073$; $s'_D = 0{,}133$

Suy ra: $T_0 = \frac{\overline{d}.\sqrt{n}}{s'_D} = \frac{(-0{,}073)\sqrt{12}}{0{,}133} = -1{,}901$

Kết luận: $T_0 \notin W_\alpha \Rightarrow$ Chấp nhận giả thiết (H). Vậy giá cả trung bình của các mặt hàng bán ở hai cửa hiệu là không khác nhau.

2.5 Kiểm định về sự bằng nhau của hai tỉ lệ:

Giả sử hai đại lượng ngẫu nhiên X và Y có tỉ lệ phần tử có tính chất A là p_X, p_Y chưa biết.

Xét giả thiết (H): $p_X = p_Y = p_0$. Đối thiết $\left(\overline{H}\right)$ là một và chỉ một trong các trường hợp sau: $p_X > p_Y$, $p_X < p_Y$, $p_X \neq p_Y$

Cho số α khá nhỏ, hãy kiểm định giả thiết (H) vơi mức ý nghĩa α.

Lấy mẫu ngẫu nhiên $W_X = \left(X_1, X_2, ..., X_{n_X}\right)$, $W_Y = \left(Y_1, Y_2, ..., Y_{n_Y}\right)$ đối với X, Y.

Chọn thống kê làm tiêu chuẩn kiểm định cho giả thiết (H) là:

$$U = \frac{f_X - f_Y - (p_X - p_Y)}{\sqrt{p_0 q_0 \left(\dfrac{1}{n_X} + \dfrac{1}{n_Y}\right)}}$$

Trong đó:

+ f_X, f_Y: lần lượt là thống kê nhận giá trị bằng tỉ lệ phần tử có tính chất A của đại lượng ngẫu nhiên X và Y.

+ p_0: giá trị trong giả thiết (H).

+ $q_0 = 1 - p_0$

+ n_X, n_Y: lần lượt là kích thước mẫu của đại lượng ngẫu nhiên X và Y (n_X, n_Y khá lớn).

Nếu giả thiết (H) đúng thì $U = \dfrac{f_X - f_Y}{\sqrt{p_0 q_0 \left(\dfrac{1}{n_X} + \dfrac{1}{n_Y}\right)}}$ có luật phân phối chuẩn:U ∈ N(0,1).

Với mẫu cụ thể $w_x = \left(x_1, x_2, ..., x_{n_X}\right)$, $w_y = \left(y_1, y_2, ..., y_{n_Y}\right)$ có f_X, f_Y lần lượt là tỉ lệ phần tử có tính chất A của đại lượng ngẫu nhiên X và Y. Ta có giá trị quan sát thực tế của thống kê U tương ứng với mẫu như sau:

$$U = \frac{f_X - f_Y}{\sqrt{p_0 q_0 \left(\dfrac{1}{n_X} + \dfrac{1}{n_Y}\right)}}$$

Chú ý: Nếu giả thiết chưa cho p_0 thì ta thế p_0 bằng p^*, với p^* được tính như sau:

$$p^* = \frac{n_X.\overline{f_X} + n_Y.\overline{f_Y}}{n_X + n_Y}$$

$$\Rightarrow q^* = 1 - p^* \text{ thay thế cho } q_0.$$

Miền bác bỏ W_α được thành lập theo dạng U như sau:

- Nếu ($\overline{H}$) có dạng $p_X > p_Y$ thì $W_\alpha = \left(U_{1-\alpha}, +\infty\right)$.
- Nếu ($\overline{H}$) có dạng $p_X < p_Y$ thì $W_\alpha = \left(-\infty, -U_{1-\alpha}\right)$.
- Nếu ($\overline{H}$) có dạng $p_X \neq p_Y$ thì: $W_\alpha = \left(-\infty, -U_{1-\frac{\alpha}{2}}\right) \cup \left(U_{1-\frac{\alpha}{2}}, +\infty\right)$

Ví dụ 11: *Từ hai tổng thể tiến hành hai mẫu với $n_X = 100$, $n_Y = 100$ quan sát. Tính được $f_X = 0{,}2$; $f_Y = 0{,}3$. Hãy kiểm định giả thiết (H): $p_X = p_Y$ với mức ý nghĩa 1%.*

Xét giả thiết (H): $p_X = p_Y$. Đối thiết $(\overline{H})$: $p_X \neq p_Y$

Chọn thống kê làm tiêu chuẩn kiểm định cho giả thiết (H) là:

$$U = \frac{F_X - F_y - (p_X - p_Y)}{\sqrt{p^* q^* \left(\frac{1}{n_X} + \frac{1}{n_Y} \right)}}$$

Trong đó:

$+F_X$, F_Y: lần lượt là thống kê nhận giá trị bằng tỉ lệ phần tử có tính chất A của đại lượng ngẫu nhiên X và Y.

$+\ p^* = \dfrac{n_X.\overline{f_X} + n_Y.\overline{f_Y}}{n_X + n_Y}$, $q^* = 1 - p^*$

$+\ n_X$, n_Y: lần lượt là kích thước mẫu của đại lượng ngẫu nhiên X và Y.

Nếu giả thiết (H) đúng thì $U = \dfrac{F_X - F_Y}{\sqrt{p_0 q_0 \left(\frac{1}{n_X} + \frac{1}{n_Y} \right)}}$ có luật phân phối chuẩn: U $\in$ N(0,1).

Với mức ý nghĩa 1% $\Rightarrow$ α = 0,01 $\Rightarrow$ 1 - $\dfrac{\alpha}{2}$ = 0,995.

Miền bác bỏ W_α được thành lập theo dạng U như sau:

$$W_\alpha = \left(-\infty, -U_{1-\frac{\alpha}{2}} \right) \cup \left(U_{1-\frac{\alpha}{2}}, +\infty \right) = (-\infty, -U_{0.995}) \cup (U_{0.995}, +\infty)$$
$$= (-\infty, -0{,}576) \cup (2{,}576, +\infty)$$

Với mẫu cụ thể có $n_X = 100$, $n_Y = 120$, $f_X = 0{,}2$, $f_Y = 0{,}3$

Suy ra: $p^* = \dfrac{100.(0{,}2) + 120.(0{,}3)}{100 + 120} = 0{,}255$

Ta có giá trị thực tế của U là: $U_0 = \dfrac{f_X - f_Y}{\sqrt{p^* q^* \left(\frac{1}{n_X} + \frac{1}{n_Y} \right)}}$

$$= \frac{0{,}2 - 0{,}3}{\sqrt{(0{,}255)(0{,}745)\left(\frac{1}{100} + \frac{1}{120} \right)}} \approx 1{,}695$$

Kết luận: $U_0 \notin W_\alpha \Rightarrow$ Chấp nhận giả thiết (H), bác bỏ đối thiết $(\overline{H})$.

2.6 Kiểm định về sự bằng nhau của hai phương sai:

Giả sử hai đại lượng ngẫu nhiên X và Y độc lập, cùng có luật phân phối chuẩn với các tham số tương ứng σ_X^2, σ_Y^2 chưa biết.

Xét giả thiết (H): $\sigma_X^2 = \sigma_Y^2$. Đối thiết $(\overline{H})$: $\sigma_X^2 > \sigma_Y^2$.

Cho số α khá nhỏ, hãy kiểm định giả thiết (H) vơi mức ý nghĩa α.

Lập mẫu ngẫu nhiên $W_X = (X_1, X_2, ..., X_{n_X})$, $W_Y = (Y_1, Y_2, ..., Y_{n_Y})$ đối với X, Y.

Chọn thống kê làm tiêu chuẩn kiểm định cho giả thiết (H) là:

$$F = \frac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2}$$

Trong đó: S_X^2, S_Y^2 : lần lượt là thống kê nhận giá trị bằng phương sai mẫu của đại lượng ngẫu nhiên X và Y.

Nếu giả thiết (H) đúng thì $F = \frac{S_X^2}{S_Y^2}$ có luật phân phối Fisher – Snedecor bậc tự do $(n_X - 1,\ n_Y - 1)$: $F \in F(n_X - 1,\ n_Y - 1)$.

Với mẫu cụ thể $w_x = (x_1, x_2, ..., x_{n_X})$, $w_y = (y_1, y_2, ..., y_{n_Y})$ lần lượt có s_X^2, s_Y^2 là phương sai mẫu của đại lượng ngẫu nhiên X và Y. Ta có giá trị quan sát thực tế của thống kê F tương ứng với mẫu như sau: $F_0 = \frac{s_X^2}{s_Y^2}$

Miền bác bỏ W_α được thành lập theo dạng F (bậc tự do $(n_X - 1,\ n_Y - 1)$) như sau:

$$W_\alpha = \left(F_{1-\alpha}(n_X - 1,\ n_Y - 1), +\infty\right)$$

Ví dụ 12: *Một phản ứng hoá học có thể được kích thích bởi hai chất xúc tác A và B khác nhau. Người ta nghi ngờ rằng tốc độ xảy ra phản ứng do chất xúc tác A kích thích không ổn định bằng chất xúc tác B kích thích. Lấy mẫu gồm 12 nhóm phản ứng dùng cho chất xúc tác A, tính được phương sai điều chỉnh là 0,35* s^2. *Lấy mẫu gồm 10 nhóm phản ứng dùng cho chất xúc tác B, tính được phương sai điều chỉnh là 0,14* s^2. *Với mức ý nghĩa* $\alpha = 5\%$, *hãy kiểm định điều nghi ngờ trên. Biết rằng tốc độ xảy ra các phản ứng có luật phân phối chuẩn.*

Gọi X, Y lần lượt là tốc độ xảy ra phản ứng do chất xúc tác A, B kích thích cùng có luật phân phối chuẩn và Var(X), Var(Y) chưa biết.

Xét giả thiết (H): $\sigma_X^2 = \sigma_Y^2$. Đối giả thiết $(\overline{H})$: $\sigma_X^2 > \sigma_Y^2$.

Chọn thống kê làm tiêu chuẩn kiểm định cho giả thiết (H) là:

$$F = \frac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2}$$

Trong đó: S_X^2, S_Y^2 : lần lượt là thống kê nhận giá trị bằng phương sai mẫu của đại lượng ngẫu nhiên X và Y.

Nếu giả thiết (H) đúng thì $F = \frac{S_X^2}{S_Y^2}$ có luật phân phối Fisher – Snedecor bậc tự do $(n_X - 1,\ n_Y - 1)$: $F \in F(n_X - 1,\ n_Y - 1)$.

Với mẫu cụ thể có $s_X^2 = 0{,}35s^2$, $s_Y^2 = 0{,}14s^2$, ta có giá trị quan sát thực tế của thống kê F tương ứng với mẫu như sau:

$$F_0 = \frac{s_X^2}{s_Y^2} = \frac{0{,}35}{0{,}14} = 2{,}5$$

Với mức ý nghĩa $\alpha = 5\% = 0{,}05 \Rightarrow 1 - \alpha = 0{,}95$ và $n_X = 12,\ n_Y = 10$ ta được:

$$F_{1-\alpha}(n_X - 1,\ n_Y - 1) = F_{0.95}(11,\ 9) = 3{,}1$$

Miền bác bỏ: $W_\alpha = \left(F_{1-\alpha}(n_X - 1,\ n_Y - 1), +\infty\right) = (3{,}1, +\infty)$

Kết luận: $F_0 \notin W_\alpha \Rightarrow$ Chấp nhận giả thiết (H). Vậy, chưa thể cho rằng tốc độ xảy ra phản ứng do chất xúc tác A kích thích không ổn định bằng chất xúc tác B kích thích.

BÀI TẬP

1. Tại một khu vườn trồng xoài cát Hoà Lộc, để điều tra trọng lượng của các trái xoài, một người đã cân thử một 100 trái xoài và kết quả được cho ở bảng sau:

Trọng lượng (g)	Số trái	Trọng lượng (g)	Số trái
450 – 500	2	650 – 700	22
500 – 550	10	700 – 750	7
550 – 600	24	750 – 800	1
600 – 650	34		

a) Nếu người đó cho biết trọng lượng trung bình của các trái xoài là 610g với độ tin cậy 95% thì có thể chấp nhận được không?

b) Những trái xoài có trọng lượng từ 650g trở lên được xem là loại I. Người đó cho biết tỉ lệ loại I là 25% với độ tin cậy 99% thì có đúng hay không?

c) Những trái xoài không phải là loại I thì là loại II. Với độ tin cậy 95% có thể khẳng định trọng lượng trung bình của các trái xoài loại II là 580g được không?

2. Hàm lượng dầu trung bình trong một trái cây lúc đầu là 5%. Người ta chăm sóc bằng một loại phân N và sau một thời gian, kiểm tra một số trái ta được kết quả như sau:

Hàm lượng dầu (%)	Số trái	Hàm lượng dầu (%)	Số trái
1 – 5	51	21 – 25	8
5 – 9	47	25 – 29	7
9 – 13	39	29 – 33	3
13 – 17	36	33 – 37	2
17 – 21	32		

a) Cho biết kết luận về loại phân N trên với mức ý nghĩa 1%.

b) Tìm một ước lượng cho hàm lượng dầu trung bình của loại trái cây đó sau chăm bón với độ tin cậy 99,6%.

c) Giả sử với số liệu điều tra ở trên, muốn ước lượng hàm lượng dầu trung bình với độ chính xác 0,8 (%) thì độ tin cậy đạt được là bao nhiêu?

d) Những trái có hàm lượng dầu từ 21% trở lên là loại A. Có thể xem tỉ lệ loại A là 5% được không với mức ý nghĩa 5%.

e) Hãy ước lượng cho tỉ lệ loại A với độ tin cậy 95%.

f) Có thể xem phương sai của hàm lượng dầu là 5% được không với mức ý nghĩa 5% (giả thiết hàm lượng này có luật phân phối chuẩn).

3. Trong một nhà máy sản xuất bánh kẹo, một máy tự động sản xuất ra các thanh socola với trọng lượng trung bình quy định là 250g, biết rằng trọng lượng các thanh socola được sản xuất ra có luật phân phối chuẩn $N(\mu\ ,\ 5^2)$. Sau một thời gian người ta nghi ngờ rằng trọng lượng trung bình của các thanh socola được sản xuất ra từ máy tự động nhỏ hơn quy định. Kiểm tra 16 thanh socola ta có kết quả sau:

Trọng lượng (g)	Số thanh	Trọng lượng (g)	Số thanh
236 – 240	2	248 – 252	4
240 – 244	4	252 – 256	2
244 – 248	3	256 – 260	1

Với mức ý nghĩa $\alpha = 0.012$, hãy kiểm định điều nghi ngờ trên.

4. Trọng lượng trung bình khi xuất chuồng ở một trại chăn nuôi gà trước đây là 3,3 kg/con. Năm nay người ta sử dụng một loại thức ăn mới, cân thử 15 con gà khi xuất chuồng ta được các số liệu sau:

3,25	2,50	4,00	3,75	3,80	3,90	4,02	3,60
3,80	3,20	3,82	3,40	3,75	4,00	3,50	

a) Với mức ý nghĩa $\alpha = 0,05$, hãy cho kết luận về tác dụng của loại thức ăn này.

b) Nếu trại chăn nuôi báo cáo trọng lượng trung bình khi xuất chuồng là 3,5 kg/con thì có chấp nhận được hay không?

5. Đo chỉ số mỡ sữa của 1340 con lai Hà – Ấn F_1 ta được bảng số liệu:

Chỉ số mỡ sữa	Số bò	Chỉ số mỡ sữa	Số bò
3,0 – 3,6	2	5,4 – 6,0	22
3,6 – 4,2	8	6,0 – 6,6	15
4,2 – 4,8	35	6,6 – 7,2	5
4,8 – 5,4	43		

a) Hãy ước lượng chỉ số mỡ sữa trung bình của giống bò lai trên với độ tin cậy 99%.

b) Biết rằng chỉ số mỡ sữa của giống bò Hà Lan thuần chủng là 4,95. Với mức ý nghĩa 1%, hãy cho kết luận về hiệu quả của việc lai giống.

6. Để nghiên cứu tác dụng của một chất kích thích sinh trưởng đối với năng suất ngô, người ta ghi lại kết quả ở 5 mảnh ruộng thí nghiệm và 5 mảnh ruộng đối chứng được bảng số liệu sau (tính theo tạ/ha):

Năng suất ngô trên các mảnh ruộng thí nghiệm: X	60	58	29	39	47
Năng suất ngô trên các mảnh ruộng đối chứng : Y	55	53	30	37	49

Với mức ý nghĩa 5%, hãy cho kết luận về hiệu quả của chất kích thích trên, coi năng suất ngô là đại lượng có luật phân phối chuẩn.

7. Cân thử 100 trái cây ở nông trường I, ta có kết quả như sau:

Trọng lượng X (g)	Số trái
15 – 35	12
35 – 55	26

55 – 75	35
75 – 95	22
95 – 115	5

Cân thử 150 trái cây ở nông trường II, ta có kết quả như sau:

Trọng lượng Y (g)	Số trái	Trọng lượng Y (g)	Số trái
45 – 50	2	70 – 75	18
50 – 55	7	75 – 80	12
55 – 60	15	80 – 85	8
60 – 65	32	85 – 90	5
65 – 70	47	90 – 95	4

a) Hãy ước lượng trọng lượng trung bình của các trái cây ở hai nông trường trên với độ tin cậy 95%.

b) Có thể xem trọng lượng trung bình của trái cây ở hai nông trường này bằng nhau được không với mức ý nghĩa 1%.

c) Những trái cây có trọng lượng lớn hơn 75g được xem là loại I. Hãy ước lượng tỉ lệ trái cây loại I của nông trường với mức ý nghĩa 3%.

d) Có thể cho rằng tỉ lệ trái cây loại I của nông trường I lớn hơn tỉ lệ trái cây loại I của nông trường II được không với mức ý nghĩa 5%.

e) Nếu cho rằng trọng lượng trung bình của trái cây loại I ở nông trường I lớn hơn trọng lượng trung bình của trái cây loại I ở nông trường II với độ tin cậy 95% thì có được không? Biết rằng trọng lượng của các trái cây loại I có luật phân phối chuẩn.

f) Tương tự câu e) với trọng lượng của trái cây loại I có phân phối chuẩn $N(\mu\ , 6^2)$.

8. Trước và sau dịp Tết giá của mặt hàng A tại 8 cửa hiệu trong thành phố như sau:

Cửa hiệu	Trước Tết	Sau Tết	Cửa hiệu	Trước Tết	Sau Tết
1	95	98	5	105	109
2	109	105	6	99	105
3	99	99	7	109	115
4	98	99	8	102	110

Với mức ý nghĩa 5%, hãy kiểm định xem có phải có khuynh hướng tăng giá sau Tết đối với mặt hàng A hay không?

9. Có hai lô chuột thí nghiệm tăng trọng với hai khẩu phần ăn khác nhau. Lô thứ nhất cho ăn khẩu phần ăn nhiều đạm. Lô thứ hai cho ăn khẩu phần ăn ít đạm hơn. Sự tăng trọng của hai lô chuột sau một thời gian được ghi lại như sau (đv: mg):

Lô thứ nhất: 123, 134, 146, 104, 119, 124, 161, 107, 83, 113, 129, 97.

Lô thứ hai : 70, 118, 85, 107, 132, 94, 101, 100.

a) Với mức ý nghĩa 5%, hãy nhận định việc cho ăn đạm có tác dụng tăng trọng hay không?

b) Với mức ý nghĩa 5%, có thể xem việc cho ăn đạm làm cho chuột tăng trọng không đồng đều hay không?

10. Một thầy giáo dạy Toán cho rằng việc cho học sinh ôn tập một vài buổi trước khi thi có tác dụng tốt tới kết quả học tập của các em. Một mẫu gồm 21 học sinh được chọn để theo dõi điểm thi của các em trước và sau khi ôn tập. Kết cho ở bảng sau đây:

Học sinh	Điểm thi trước ôn tập	Điểm thi sau ôn tập	Học sinh	Điểm thi trước ôn tập	Điểm thi sau ôn tập
1	22	21	11	28	27
2	26	29	12	24	25
3	17	15	13	27	27
4	20	20	14	18	20
5	28	26	15	20	23
6	31	32	16	14	16
7	23	25	17	24	26
8	13	14	18	15	20
9	19	19	19	19	20
10	25	27	20	18	17
			21	27	19

Với mức ý nghĩa 5%, có thể kết luận rằng sau khi được ôn tập kết quả thi của học sinh tốt hơn hay không?

KQHT6: XÁC ĐỊNH TƯƠNG QUAN VÀ HỒI QUI

Bước học 1: TƯƠNG QUAN

1.1 Mối quan hệ giữa hai đại lượng ngẫu nhiên:

Khi khảo sát hai đại lượng ngẫu nhiên X và Y ta thấy giữa chúng có thể có một số quan hệ sau:

i) X và Y độc lập nhau, tức là việc nhận giá trị của đại lượng ngẫu nhiên này không ảnh hưởng đến việc nhận giá trị của đại lượng ngẫu nhiên kia.

ii) X và Y có mối quan hệ phụ thuộc hàm số $Y = \varphi(X)$.

iii) X và Y có sự phụ thuộc tương quan và không tương quan.

1.2 Hệ số tương quan:

1.2.1 Moment tương quan (Covarian):

Định nghĩa: *Moment tương quan (hiệp phương sai) của hai đại lượng ngẫu nhiên X và Y, ký hiệu cov(X,Y) hay* μ_{XY}*, là số dược xác định như sau:*

$$\operatorname{cov}(X,Y) = E\{[X - E(X)][Y - E(Y)]\}$$

Chú ý:

$$\operatorname{cov}(X,Y) = E(XY) - E(X).E(Y)$$

Thật vậy:

$$\begin{aligned}\operatorname{cov}(X,Y) &= E[X.Y - X.E(Y) - Y.E(X) + E(X).E(Y)] \\ &= E(X.Y) - E(X).E(Y) - E(X).E(Y) + E(X).E(Y) \\ &= E(X.Y) - E(X).E(Y)\end{aligned}$$

Cụ thể:

i) Nếu (X,Y) rời rạc thì: $\operatorname{cov}(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j P(x_i, y_j) - E(X)E(Y)$

ii) Nếu (X,Y) liên tục thì: $\operatorname{cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy - E(X)E(Y)$

Nhận xét:

i) X và Y độc lập $\Leftrightarrow$ cov(X,Y) = 0: khi đó ta nói rằng X, Y không tương quan.

ii) Cov(X,X) = Var(X).

1.2.2 Hệ số tương quan:

Định nghĩa*: Hệ số tương quan của hai đại lượng ngẫu nhiên X và Y, ký hiệu* r_{XY}*, là số được xác định như sau:*

$r_{XY} = \dfrac{\operatorname{cov}(X,Y)}{S_X.S_Y}$*, trong đó* S_X*,* S_Y *là độ lệch tiêu chuẩn của X và Y.*

Ước lượng hệ số tương quan:

Lập mẫu ngẫu nhiên $W_{XY} = [(X_1, Y_1),\ (X_2, Y_2),\ \ldots,\ (X_n, Y_n)]$

Để ước lượng hệ số tương quan ta dùng thống kê: $R = \dfrac{\overline{XY} - \overline{X}.\overline{Y}}{S_X.S_Y}$, trong đó:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i\ ,\quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i\ ,\quad \overline{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i\ ,$$

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\ ,\quad S_Y^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$$

Với mẫu cụ thể ta tính được giá trị cụ thể của R là: $r_{XY} = \dfrac{\overline{xy} - \overline{x}.\overline{y}}{s_x.s_y}$

trong đó: $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$, $\overline{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{xy} = \dfrac{1}{n}\sum_{i=1}^{n} x_i y_i$

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2\ ,\quad s_Y^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$$

Ta có:

$$r_{XY} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\left(\sum_{i=1}^{n} y_i^2\right) - \left(\sum_{i=1}^{n} y_i\right)^2}}$$

Ví dụ 1: *Từ số liệu được cho bởi bảng sau, hãy xác định hệ số tương quan của Y và X:*

X	1	3	4	6	8	9	11	14
Y	1	2	4	4	5	7	8	9

Ta lập bảng sau:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	1	1	1	1
3	2	9	4	6
4	4	16	16	16
6	4	36	16	64
8	5	64	25	40
9	7	81	49	63
11	8	121	64	88
14	9	196	81	126
$\sum_{i=1}^{n} x_i = 56$	$\sum_{i=1}^{n} y_i = 40$	$\sum_{i=1}^{n} x_i^2 = 524$	$\sum_{i=1}^{n} y_i^2 = 256$	$\sum_{i=1}^{n} x_i y_i = 364$

Hệ số tương quan của X và Y là:

$$r_{XY} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\left(\sum_{i=1}^{n} y_i^2\right) - \left(\sum_{i=1}^{n} y_i\right)^2}}$$

$$= \frac{(8).(364) - (56)(40)}{\sqrt{(8)(524) - (56)^2}\sqrt{(8)(256) - (40)^2}} = \frac{672}{687.81} = 0.977$$

Tính chất và ý nghĩa của hệ số tương quan:

Hệ số tương quan r được dùng để đánh giá mức độ chặt chẽ của sự phụ thuộc tương quan tuyến tính giữa hai đại lượng ngẫu nhiên X và Y, nó có các tính chất sau đây:

i) $|r| \le 1$.

ii) Nếu $|r|$ *= 1 thì X và Y có quan hệ tuyến tính.*

iii) Nếu $|r|$ *càng lớn thì sự phụ thuộc tương quan tuyến tính giữa X và Y càng chặt chẽ.*

iv) Nếu $|r|$ *= 0 thì giữa X và Y không có phụ thuộc tuyến tính tương quan.*

v) Nếu r > 0 thì X và Y có tương quan thuận (X, Y cùng tăng hoặc cùng tăng). Nếu r < 0 thì X và Y có tương quan nghịch (X giảm thì Y tăng hoặc ngược lại).

1.3 Tỷ số tương quan:

Định nghĩa: *Tỷ số tương quan của hai đại lượng ngẫu nhiên Y và X, ký hiệu $\eta_{Y/X}$, là số được xác định như sau:*

$$\eta_{Y/X} = \frac{s_{\bar{y}}}{s_y}$$

trong đó: $s_{\bar{y}} = \sqrt{\frac{1}{n}\sum n_i.\left(\overline{y_{x_i}} - \bar{y}\right)^2}$, $s_y = \sqrt{\frac{1}{n}\sum m_j\left(y_j - \bar{y}\right)^2}$

Ý nghĩa của tỷ số tương quan: *Tỷ số tương quan đo mức độ chặt chẽ của sự phụ thuộc tương quan phi tuyến tính giữa X và Y.*

Tính chất của tỷ số tương quan:

i) $0 \le \eta_{Y/X} \le 1$.

ii) $\eta_{Y/X} = 0$ khi và chỉ khi Y và X không phụ thuộc tương quan.

iii) $\eta_{Y/X} = 1$ khi và chỉ khi Y và X phụ thuộc hàm số.

iv) $\eta_{Y/X} \ge |r|$.

Nếu $\eta_{Y/X} = |r|$ thì sự phụ thuộc tương quan của Y và X có dạng tuyến tính.

Bước học 2: TÌM HÀM HỒI QUI

2.1 Kỳ vọng có điều kiện:

a) Đại lượng ngẫu nhiên rời rạc:

i) Kỳ vọng có điều kiện của đại lượng ngẫu nhiên rời rạc Y với điều kiện X = x là:

$$E(Y/x) = \sum_{j=1}^{m} y_j.P\left(X = x, Y = y_j\right)$$

ii) Kỳ vọng có điều kiện của đại lượng ngẫu nhiên rời rạc X với điều kiện Y = y là:

$$E(X/y) = \sum_{i=1}^{n} x_i.P\left(X = x_i, Y = y\right)$$

b) Đại lượng ngẫu nhiên liên tục:

i) Kỳ vọng có điều kiện của đại lượng ngẫu nhiên liên tục Y với điều kiện X = x là:

$E(Y/x) = \int_{-\infty}^{+\infty} yf(y/x)dy$, với $f(y/x) = f(x,y)$ với x không đổi.

ii) Kỳ vọng có điều kiện của đại lượng ngẫu nhiên liên tục X với điều kiện Y = y là:

$E(X/y) = \int_{-\infty}^{+\infty} xf(x/y)dx$, với $f(x/y) = f(x,y)$ với y không đổi.

2.2 Hàm hồi qui:

Trong thực tế ta thường gặp hai đại lượng ngẫu nhiên X, Y có mối quan hệ với nhau, trong đó việc khảo sát X thì dễ còn khảo sát Y thì khó hơn thậm chí không thể khảo sát được. Người ta muốn tìm mối quan hệ $\varphi(X)$ nào đó giữa X và Y để biết X có thể dự đoán được Y.

Giả sử biết X, nếu dự đoán Y bằng $\varphi(X)$ thì sai số phạm phải là $E\{[Y-\varphi(X)]^2\}$. Vấn đề được đặt ra là tìm $\varphi(X)$ như thế nào để $E\{[Y-\varphi(X)]^2\}$ là nhỏ nhất.

Ta sẽ chứng minh khi chọn $\varphi(X)=E(Y/X)$ thì $E\{[Y-\varphi(X)]^2\}$ sẽ nhỏ nhất.

Thật vậy ta có:

$$E\{[Y-\varphi(X)]^2\}=E\{([Y-E(Y/X)]+[E(Y/X)-\varphi(X)])^2\}$$
$$=\{E[Y-E(Y/X)]^2\}+E\{[E(Y/X)-\varphi(X)]^2\}+2E\{[Y-E(Y/X)][E(Y/X)-\varphi(X)]\}$$

Ta thấy E(Y/X) chỉ phụ thuộc vào X nên có thể đặt T(X) = E(Y/X) - $\varphi(X)$.

Vì E[E(Y/X)T(X)] = E[YT(X)] nên

$$2E\{[Y-E(Y/X)][E(Y/X)-\varphi(X)]\}=2E\{[Y-E(Y/X)]T(X)\}$$
$$=2E[YT(X)]-2E[E(Y/X)T(X)]=0$$

Do đó: $E\{[Y-\varphi(X)]^2\}=E\{[Y-E(Y/X)]^2\}+E\{[E(Y/X)-\varphi(X)]^2\}$

nhỏ nhất khi $E\{[E(Y/X)-\varphi(X)]^2\}=0$

Ta chỉ cần chọn $\varphi(X)=E(Y/X)$

Ta gọi $\varphi(x)=E(Y/x)$ là hàm hồi qui của Y đối với X.

Tương tự, ta gọi $\varphi(y)=E(X/y)$ là hàm hồi qui của X đối với Y.

Nếu $\varphi(x)$ [hoặc $\varphi(y)$] là hàm bậc nhất thì ta nói rằng Y (hoặc X) là hồi qui tuyến tính đơn đối với X (hoặc Y).

2.3 Xác định hàm hồi qui tuyến tính mẫu (thực nghiệm):

Giả sử giữa hai đại lượng X và Y có liên quan tuyến tính, tức là: E(Y/X) = AX + B

Dựa vào n cặp giá trị $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ của (X,Y) ta tìm hàm $\overline{y_x}=y=ax+b$ (*) để ước lượng hàm Y = AX + B.

Và (*) được gọi là hàm hồi qui tuyến tính mẫu.

Vì các cặp giá trị trên là trị xấp xỉ của x và y nên thỏa (*) một cách xấp xỉ.

Do đó: $y_i=ax_i+b+\varepsilon_i$ hay $\varepsilon_i=y_i-ax_i-b$.

Tìm a, b sao cho các sai số ε_i $(i=\overline{1,n})$ có trị tuyệt đối nhỏ nhất hay hàm

$$S(a,b)=\sum_{i=1}^{n}(y_i-ax_i-b)^2$$

đạt cực tiểu. Phương pháp tìm này được gọi là phương pháp bình phương bé nhất.

Ta thấy S sẽ đạt giá trị nhỏ nhất tại điểm dừng thoả mãn

$$\begin{cases}0=\dfrac{\partial S}{\partial a}=-2\sum_{i=1}^{n}x_i(y_i-ax_i-b)\\[2ex] 0=\dfrac{\partial S}{\partial b}=-2\sum_{i=1}^{n}(y_i-ax_i-b)\end{cases}\Leftrightarrow\begin{cases}\left(\sum_{i=1}^{n}x_i^2\right).a+\left(\sum_{i=1}^{n}x_i\right).b=\sum_{i=1}^{n}x_iy_i\\[2ex] \left(\sum_{i=1}^{n}x_i\right).a+nb=\sum_{i=1}^{n}y_i\end{cases}$$

Hệ trên có định thức:

$$D = \begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{vmatrix} = n.\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2$$

$$D_x = \begin{vmatrix} \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i & n \end{vmatrix} = n.\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$

$$D_y = \begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i \end{vmatrix} = \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i y_i\right)$$

Vì các x_i khác nhau nên theo bất đẳng thức Bunhiakovski ta có:

$$\left(\sum_{i=1}^{n} x_i\right)^2 < n.\sum_{i=1}^{n} x_i^2 \qquad \Rightarrow D > 0$$

Suy ra hệ có nghiệm duy nhất:

$$a = \frac{n.\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n.\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}; \qquad b = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i y_i\right)}{n.\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Nếu đặt: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$, $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$

thì nghiệm của hệ có thể việt lại:

$$a = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{s_x^2}; \qquad b = \frac{\overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy}}{\overline{x^2} - \left(\overline{x}\right)^2} = \frac{\overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy}}{s_x^2}$$

Tóm lại, ta có thể tìm hàm $\overline{y_x} = ax + b$ từ các công thức:

$$a = \frac{\overline{xy} - \overline{x}\,\overline{y}}{s_x^2} = \frac{n.\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n.\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}; \qquad b = \overline{y} - a.\overline{x}$$

Chú ý:

i) Đường gấp khúc nối các điểm (x_1, y_1), (x_2, y_2), . . . , (x_n, y_n) được gọi là hàm hồi qui thực nghiệm.

ii) Đường thẳng y = ax + b nhận được bởi công thức bình phương bé nhất không đi qua tất cả các điểm nhưng là đường thẳng "gần" các điểm đó nhất được gọi là đường thẳng hồi qui và thủ tục làm thích hợp đường thẳng thông qua các dự liệu cho trước được gọi là hồi qui tuyến tính.

iii) Theo trên ta có $b = \overline{y} - a.\overline{x}$, do đó điểm $\left(\overline{x}, \overline{y}\right)$ luôn nằm trên đường thẳng hồi qui.

iv) Ta có: $r_{xy} = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{s_x . s_y} \Rightarrow a = r_{xy} \dfrac{s_y}{s_x}$

Ví dụ 2: *Ước lượng hàm hồi qui tuyến tính mẫu của Y theo X trên cơ sở bảng tương quan cặp sau:*

X	15	38	23	16	16	13	20	24
Y	145	228	150	130	160	114	142	265

Ta lập bảng sau:

x_i	y_i	x_i^2	$x_i y_i$
15	145	225	3175
28	228	1444	8664
23	150	529	3450
16	130	256	2080
16	160	2556	2560
13	114	169	1482
20	142	400	2840
24	265	576	6360
$\sum x_i = 165$	$\sum y_i = 1334$	$\sum x_i^2 = 3855$	$\sum x_i y_i = 29611$

Ta có:

$$a = \frac{n.\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{n.\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} = \frac{8.(29611) - (165)(1334)}{8.(3855) - (165)^2} = \frac{16778}{3615} = 4{,}64$$

$$b = \overline{y} - a\overline{x} = \frac{1334}{8} - \left(\frac{16778}{3615}\right)\left(\frac{165}{8}\right) = 71$$

Vậy hàm hồi qui tuyến tính mẫu là $\overline{y_x} = 4{,}64x + 71$

Ví dụ 3: *Độ ẩm của không khí ảnh hưởng đến sự bay hơi của nước trong sơn khi phun ra. Người tiến hành nghiên cứu mối liên hệ giữa độ ẩm của không khí X và độ bay hơi Y. Sự hiểu biết về mối liên hệ này sẽ giúp ta tiết kiệm được lượng sơn bằng cách chỉnh súng phun sơn một cách thích hợp. Tiến hành 25 quan sát ta được các số liệu sau:*

Quan sát	**Độ ẩm** (%)	**Độ bay hơi** (%)	**Quan sát**	**Độ ẩm** (%)	**Độ bay hơi** (%)
1	35.3	11.0	14	39.1	9.6
2	29.7	11.1	15	46.8	10.9
3	30.8	12.5	16	48.5	9.6
4	58.8	8.4	17	59.3	10.1

5	61.4	9.3	18	70.0	8.1
6	71.3	8.7	19	70.0	6.8
7	74.4	6.4	20	74.4	8.9
8	76.7	8.5	21	72.1	7.7
9	70.7	7.8	22	58.1	8.5
10	57.5	9.1	23	44.6	8.9
11	46.4	8.2	24	33.4	10.4
12	28.9	12.2	25	28.6	11.1
13	28.1	11.9			

Hãy tìm hàm hồi qui tuyến tính mẫu $\overline{y_x} = ax + b$.

Ta lập bảng sau:

x_i	y_i	x_i^2	$x_i y_i$
35.3	11.0	1246.09	388.3
29.7	11.1	882.09	329.67
30.8	12.5	948.64	385
58.8	8.4	3457.44	493.92
61.4	9.3	3769.96	571.02
71.3	8.7	5083.69	620.31
74.4	6.4	5535.36	476.16
76.7	8.5	5882.89	651.95
70.7	7.8	4998.49	551.46
57.5	9.1	3306.25	523.25
46.4	8.2	2152.96	380.48
28.9	12.2	835.21	352.58
28.1	11.9	789.61	334.39
39.1	9.6	1528.81	375.36
46.8	10.9	2190.24	510.12
48.5	9.6	2352.25	465.60
59.3	10.1	3516.49	598.93
70.0	8.1	4900	567
70.0	6.8	4900	476
74.4	8.9	5535.36	662.16
72.1	7.7	5198.41	555.17
58.1	8.5	3375.61	493.85
44.6	8.9	1989.16	396.94

33.4 28.6	10.4 11.1	1115.56 817.96	347.36 317.46
$\sum x_i = 1314.9$	$\sum y_i = 235.7$	$\sum x_i^2 = 76308.53$	$\sum 11824.44$

Ta có:

$$a = \frac{n.\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{n.\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} = \frac{25.(11824.44) - (1314.9)(235.7)}{25.(76308.53) - (1314.9)^2} = -0.08$$

$$b = \overline{y} - a\overline{x} = \frac{235.7}{25} - (-0.08)\frac{1314.9}{25} = 13.64$$

Vậy hàm hồi qui tuyến tính mẫu là $\overline{y_x} = -0.08x + 13.64$

Ví dụ 4: *Xác định hệ số tương quan và hàm hồi qui tuyến mẫu* $\overline{y_x} = ax + b$ *của các đại lượng ngẫu nhiên X và Y cho bởi bảng tương quan thực nghiệm sau:*

Y \ X	1	2	3
10	20		
20		30	1
30		1	48

Ta lập bảng sau:

Y \ X	1	2	3	m_j	$m_j y_j$	$m_j y_i^2$
10	200 20			20	200	2000
20		1200 30	60 1	31	620	12400
30		60 1	4320 48	49	1470	44100
n_i	20	31	49	$n = 100$	$\sum y = 2290$	$\sum y^2 = 58500$
$n_i x_i$	20	62	147	$\sum x = 229$		
$n_i x_i^2$	20	124	441	$\sum x^2 = 585$		$\sum xy = 5840$

Trong đó: $\sum xy = 200 + 1200 + 60 + 60 + 4320 = 5840$.

Phần trên góc trái của ô ghi các tích $n_{ij}x_i y_j$.

Ta có: $\overline{x} = \frac{229}{100} = 2.29$; $\overline{y} = \frac{2290}{100} = 22.9$;

$\overline{x^2} = \frac{585}{100} = 5.85$; $\overline{y^2} = \frac{58500}{100} = 585$; $\overline{xy} = \frac{5840}{100} = 58.4$;

$s_x^2 = \overline{x^2} - \left(\overline{x}\right)^2 = 5.85 - (2.29)^2 \approx 0.6059 \quad \Rightarrow s_x = 0.78$

$s_y = \sqrt{s_y^2} = \sqrt{\overline{y^2} - \left(\overline{y}\right)^2} = \sqrt{585 - (22.9)^2} \approx 7.78$

Do đó: $a = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{s_x^2} = \dfrac{58.4 - (2.29)(22.9)}{0.6059} = 9.835$

$b = \overline{y} - a\overline{x} = 22.9 - (9.835)(2.29) = 0.378$

Hàm hồi qui tuyến tính mẫu là $\overline{y_x} = 9.835x + 0.378$

Hệ số tương quan là: $r_{xy} = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{s_x.s_y} = \dfrac{58.4 - (2.29)(22.9)}{(0.78)(7.78)} \approx 0.982$

TÀI LIỆU THAM KHẢO

TÀI LIỆU THAM KHẢO ĐỂ BIÊN SOẠN NỘI DUNG MÔN HỌC:

1. Đặng Hấn, 1996: Xác suất thống kê – NXB Thống kê.
2. Nguyễn Hữu Khánh: Bài giảng Xác suất thống kê – ĐH Cần Thơ.
3. Đinh Văn Gắng: Xác suất và Thống kê toán – NXB Thống kê.
4. Hoàng Ngọc Nhậm: Xác suất và Thống kê toán – ĐH Kinh tế TP HCM.

5. Đặng Hấn, 1996: Bài tập Xác suất thống kê – NXB Thống kê.
6. Hoàng Hữu Như: Bài tập Xác xuất thống kê – NXB Thống kê.
7. Lê Khánh Luận: Bài tập Xác suất thống kê - Trường ĐH Kinh tế TP HCM.
8. Ninh Quang Hải: Xác suất và Thống kê toán – ĐH Kiến trúc Hà Nội.

TÀI LIỆU THAM KHẢO ĐỀ NGHỊ CHO HỌC VIÊN:

1. Đặng Hấn, 1996: Xác suất thống kê – NXB Thống kê.
2. Nguyễn Hữu Khánh: Bài giảng Xác suất thống kê – ĐH Cần Thơ.
3. Đinh Văn Gắng: Xác suất và Thống kê toán – NXB Thống kê.
4. Hoàng Ngọc Nhậm: Xác suất và Thống kê toán – ĐH Kinh tế TP HCM.
5. Đặng Hấn, 1996: Bài tập Xác suất thống kê – NXB Thống kê.
6. Lê Khánh Luận: Bài tập Xác suất thống kê - Trường ĐH Kinh tế TP HCM.