TRƯỜNG ĐẠI HỌC TIỀN GIANG
KHOA KHOA HỌC TỰ NHIÊN

HUỲNH HUY VIỆT
VÕ THỊ TRÚC GIANG
VÕ DUY MINH

XÁC SUẤT THỐNG KÊ A

TÀI LIỆU GIẢNG DẠY DÙNG CHO SINH VIÊN ĐẠI HỌC NGÀNH QUẢN TRỊ KINH DOANH, KẾ TOÁN, XÂY DỰNG, CNTT

LƯU HÀNH NỘI BỘ
NĂM 2016

TRƯỜNG ĐẠI HỌC TIỀN GIANG
KHOA KHOA HỌC TỰ NHIÊN

HUỲNH HUY VIỆT
VÕ THỊ TRÚC GIANG
VÕ DUY MINH

XÁC SUẤT THỐNG KÊ A

TÀI LIỆU GIẢNG DẠY DÙNG CHO SINH VIÊN ĐẠI HỌC
NGÀNH QUẢN TRỊ KINH DOANH, KẾ TOÁN, XÂY DỰNG, CNTT
Số tín chỉ: 3

LƯU HÀNH NỘI BỘ
NĂM 2016

LỜI NÓI ĐẦU

Lý thuyết Xác suất - Thống kê là một ngành toán nghiên cứu tìm quy luật chi phối và các phương pháp tính toán khả năng xuất hiện của các hiện tượng ngẫu nhiên. Suốt từ lúc ra đời (thế kỷ 17) đến nay, Xác suất - Thống kê (XS-TK) đã trở thành một công cụ quan trọng cả về phương diện lý thuyết và ứng dụng trong y khoa, nông nghiệp, kinh tế, giáo dục, bảo hiểm,…

Nhà Toán học Pháp Laplace ở thế kỷ 19 đã tiên đoán rằng "Môn khoa học này hứa hẹn trở thành một trong những đối tượng quan trọng nhất của tri thức nhân loại, rất nhiều những vấn đề quan trọng nhất của đời sống thực tế thuộc về những bài toán của Lý thuyết Xác suất".

Đặc biệt thống kê rất cần cho các nhà quản lý bởi vì khoa học thống kê cung cấp cho họ phương pháp thu thập, xử lý và diễn giải các thông tin về dân số, kinh tế, giáo dục… để từ đó có thể hoạch định chính sách và ra quyết định đúng đắn. Ngay từ đầu thế kỷ 20, nhà triết học người Anh Well đã dự báo "Trong một tương lai không xa, kiến thức thống kê và tư duy thống kê sẽ trở thành một yếu tố không thể thiếu được trong học vấn phổ thông của một công dân giống như là khả năng biết đọc, biết viết vậy".

Ngày nay, trên thế giới, XS-TK được giới thiệu, giảng dạy và nghiên cứu ngay từ bậc tiểu học, từ giáo viên mầm non đến các bậc học cao nhất như thạc sĩ, tiến sĩ. Đặc biệt kể từ khi máy tính xuất hiện, các vấn đề XS-TK ngày càng trở nên dễ dàng trong tính toán, trong phân tích, dự báo,… một khi có số liệu đúng đắn và mô hình hợp lý.

Tài liệu giảng dạy "Xác suất – Thống kê A" của Bộ môn Toán, Khoa Khoa học Tự nhiên, Trường Đại học Tiền Giang được biên soạn trên tinh thần đó và trên những quy định về nội dung, số tiết,… của Bộ Giáo dục & Đào tạo. Về phương pháp dạy học, tài liệu rất chú trọng đến việc giúp cho sinh viên tự đọc, bước đầu tự hiểu và nắm chắc bài học sau khi nghe giảng viên phân tích, làm rõ ý nghĩa, hướng dẫn vận dụng bài học vào thực tiễn cuộc sống. Tài liệu cũng xây dựng hệ thống bài tập từ đơn giản đến phức tạp theo hướng sử dụng XS-TK như một công cụ hiệu quả nhằm giải quyết các vấn đề thực tế trong nhiều lĩnh vực kinh tế, kỹ thuật, giáo dục, công nghiệp,… Đối với giảng viên, tài liệu có mở ngỏ một số ý để từ đó giảng viên - tùy theo khả năng tiếp thu và khả năng vận dụng của sinh viên - có thể xây dựng thành các seminar cho sinh viên.

Tài liệu gồm 6 chương. Chương 1, 2 trình bày về xác suất: các khái niệm về biến cố, xác suất, biến ngẫu nhiên, công thức tính xác suất và luật phân phối xác suất; các đặc trưng số của biến ngẫu nhiên và các định lý giới hạn quan trọng trong xác suất. Chương 3, 4, 5, 6 trình bày về thống kê: lý thuyết mẫu, các bài toán ước lượng; kiểm định giả thuyết thống kê và lý thuyết tương quan, hồi quy. Cuối tài liệu là các phụ lục về giải tích tổ hợp, giá trị các phân phối nhằm hỗ trợ sinh viên tự học.

Mặc dù đã cố gắng hết sức nhưng không thể tránh khỏi những thiếu sót. Chúng tôi chân thành chờ mong sự góp ý thẳng thắn của các bạn sinh viên, các đồng nghiệp để tài liệu được hoàn chỉnh hơn trong lần tái bản sau.

NHÓM TÁC GIẢ

MỤC LỤC

PHẦN A. XÁC SUẤT

Chương 1. XÁC SUẤT VÀ CÁC CÔNG THỨC TÍNH XÁC SUẤT

§1. Phép thử và sự kiện

1. Khái niệm 1
2. Tính chất của phép toán sự kiện 4

§2. Định nghĩa xác suất

1. Định nghĩa theo cổ điển 5
2. Định nghĩa theo thống kê 5
3. Định nghĩa theo hình học 6
4. Định nghĩa theo tiên đề 6
5. Một số tính chất của xác suất 7

§3. Các công thức tính xác suất

1. Công thức nhân xác suất 7
2. Công thức cộng xác suất 9
3. Công thức xác suất đầy đủ 11
4. Công thức Bayes 12
5. Công thức Bernoulli 12

Chương 2. ĐẠI LƯỢNG NGẪU NHIÊN

§1. Đại lượng ngẫu nhiên - Luật phân phối xác suất của đại lượng ngẫu nhiên

1. Đại lượng ngẫu nhiên 18
2. Luật phân phối xác suất của đại lượng ngẫu nhiên 19
3. Hàm phân phối xác suất của đại lượng ngẫu nhiên 21
4. Phân vị mức xác suất α 23

§2. Các đặc trưng số của đại lượng ngẫu nhiên

1. Kỳ vọng 25
2. Phương sai 26
3. Độ lệch tiêu chuẩn 28
4. Giá trị tin chắc nhất 28
5. Trung vị 29

§3. Một số luật phân phối xác suất đặc biệt

1. Phân phối siêu bội H(N,M,n) 30
2. Phân phối nhị thức B(n,p) 31
3. Phân phối Poisson P (λ) 33
4. Phân phối chuẩn N(μ,σ^2) 34
5. Phân phối chuẩn tắc N(0,1) 36
6. Phân phối Khi bình phương χ^2(n) 40
7. Phân phối Student T(n) 41

§4. Đại lượng ngẫu nhiên hai chiều
1. Khái niệm về đại lượng ngẫu nhiên hai chiều 43
2. Luật phân phối xác suất của đại lượng ngẫu nhiên hai chiều .. . 43
3. Hàm phân phối xác suất của đại lượng ngẫu nhiên hai chiều... 46
4. Kỳ vọng và phương sai của đại lượng ngẫu nhiên hai chiều ... 46
§5. Định lý giới hạn trung tâm - Luật số lớn
1. Định lý giới hạn trung tâm 47
2. Luật số lớn 47

PHẦN B. THỐNG KÊ

Chương 3. LÝ THUYẾT MẪU

1. Các khái niệm cơ bản 53
2. Mẫu cụ thể 57
3. Phân phối của một số thống kê đặc trưng mẫu..... 61
4. Các hình thức thống kê..... 63

Chương 4. BÀI TOÁN ƯỚC LƯỢNG

1. Bài toán ước lượng các đặc trưng số của ĐLNN .....……65
2. Phương pháp ước lượng điểm 66
3. Phương pháp ước lượng khoảng..... 69
4. Ước lượng khoảng cho trung bình..... 69
5. Ước lượng khoảng cho tỷ lệ 75
6. Ước lượng khoảng cho phương sai..... 76

Chương 5. KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

1. Bài toán kiểm định giả thiết thống kê 81
2. Kiểm định giả thiết về trung bình 82
3. Kiểm định giả thiết về sự bằng nhau của hai trung bình 86
4. Kiểm định giả thiết về tỷ lệ 90
5. Kiểm định giả thiết về sự bằng nhau của hai tỉ lệ 91
6. Kiểm định giả thiết về phương sai..... 92

Chương 6. TƯƠNG QUAN VÀ HỒI QUY

1. Mối quan hệ giữa hai biến ngẫu nhiên..... 96
2. Hệ số tương quan 97
3. Hồi quy 99

Các bảng phụ lục 1, 2, 3, 4, 5, 6, 7.
TÀI LIỆU THAM KHẢO.

PHẦN A. XÁC SUẤT

Chương 1. XÁC SUẤT VÀ CÔNG THỨC TÍNH XÁC SUẤT

Mục tiêu

Sau khi học xong chương này, sinh viên cần đạt được:

1. Kiến thức

- Bước đầu thấy được ý nghĩa và tầm quan trọng của XS-TK trong thực tiễn khoa học và thực tiễn cuộc sống.
- Biết được một số khái niệm cơ bản của lý thuyết XS-TK như phép thử, sự kiện, xác suất của một sự kiện,…
- Hiểu được ý nghĩa các công thức tính xác suất: công thức xác suất có điều kiện, công thức nhân, công thức cộng, công thức xác suất đầy đủ, công thức Bayes, công thức Bernoulli, quy tắc hộp kín.

2. Kỹ năng

- Biểu diễn được một sự kiện (bằng ngôn ngữ XS-TK) qua các sự kiện cho trước (bằng ngôn ngữ thông thường).
- Tính được xác suất của một sự kiện.
- Giải được các dạng toán thực tiễn bằng các công cụ cơ bản của lý thuyết XS: các phép toán xác suất cơ bản, quy tắc hộp kín, công thức xác suất đầy đủ, lược đồ Bernoulli.

3. Thái độ

- Bước đầu thấy được sự ứng dụng mạnh mẽ của Toán học vào đời sống.
- Xây dựng ý thức tổ chức kỷ luật và làm việc khoa học khi giải bài toán thực tế.
- Từng bước giới hạn những gian nan buổi đầu khi biến những thực tiễn đa dạng cuộc sống thành ký hiệu, công thức toán học.

§1. PHÉP THỬ VÀ SỰ KIỆN

1. Khái niệm

1.1. Khái niệm phép thử và sự kiện

Ví dụ 1.1:

a) Tung ngẫu nhiên một đồng xu (đồng chất, có một mặt hình và một mặt chữ). Đây là một **phép thử** và quan sát thấy kết cục có thể xảy ra là mặt hình hoặc mặt chữ. Hai kết quả này được gọi là hai **sự kiện sơ cấp**.

b) Tung ngẫu nhiên một con súc sắc (khối lập phương, đồng chất, các mặt có từ 1 đến 6 chấm). Đó là một phép thử và quan sát số chấm k xuất hiện, k = 1, 2, 3, 4, 5, 6. Sáu khả năng có thể này là sáu sự kiện sơ cấp. Xuất hiện mặt có số chấm lẻ cũng là một sự kiện, nhưng không phải là sự kiện sơ cấp của phép thử.

c) Quan sát nhiệt độ ngoài trời. Đó là một phép thử, nhiệt độ ngoài trời t^0 C là một sự kiện.

d) Quan sát tuổi thọ của một bóng đèn. Đó là một phép thử, tuổi thọ t giờ của bóng đèn là một sự kiện.

Phép thử ngẫu nhiên E (Experiment) là một thí nghiệm, một phép đo, một sự quan sát hiện tượng nào đó,… mà kết quả không đoán trước được.

Phép thử E được xác định bởi một nhóm điều kiện S nào đó và mỗi khi làm cho các điều kiện này được thỏa là ta đã thực hiện phép thử E. Nhóm điều kiện S phải rõ ràng ổn định trong quá trình nghiên cứu và có thể lặp lại nhiều lần.

Mỗi một phép thử đều gắn liền với mục đích của nó, đó chính là những kết quả có thể xảy ra mà ta quan tâm khi thực hiện phép thử đó. Những kết quả này được gọi là các kết cục, tập hợp tất cả các kết cục đó được gọi là **không gian mẫu** Ω của phép thử, mỗi kết cục là một phần tử của Ω hay một sự kiện sơ cấp. Chẳng hạn, trong phép thử tung ngẫu nhiên một súc sắc, không gian mẫu có 6 sự kiện sơ cấp, đó là mặt súc sắc có 1 chấm, 2 chấm, 3 chấm, 4 chấm, 5 chấm, 6 chấm.

Sự kiện (Biến cố - Event) là một tập con của Ω. Các sự kiện được ký hiệu là A, B, C,... hoặc A_1, A_2, A_3,...

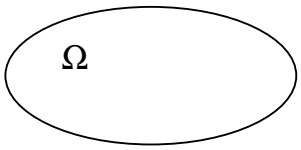
Mỗi sự kiện được xác định bởi một số điều kiện. Sau khi thực hiện phép thử, nếu điều kiện được thỏa mãn, ta nói rằng sự kiện đó xảy ra (xuất hiện), nếu ngược lại, ta nói rằng sự kiện đó không xảy ra (không xuất hiện). Chẳng hạn, xét phép thử rút một cái thăm, quy định thăm có dấu **x** là thăm trúng thưởng, thăm trắng là thăm không trúng thưởng. Gọi A là sự kiện trúng thưởng. Sau khi thực hiện phép thử, nếu rút được thăm có dấu **x** thì ta nói sự kiện A xảy ra, nếu rút được thăm trắng thì ta nói sự kiện A không xảy ra.

Khi thực hiện phép thử, một sự kiện không thể biết trước có xảy ra hay không được gọi là **sự kiện ngẫu nhiên**.

1.2. Một số sự kiện đặc biệt, phép toán và quan hệ giữa các sự kiện trong phép thử

1.2.1. Sự kiện chắc chắn Ω

Sự kiện **chắc chắn** là sự kiện luôn xảy ra sau khi thực hiện phép thử. Sự kiện này cũng chính là không gian mẫu. Ký hiệu: Ω.



Ví dụ 1.2: Gọi Ω là sự kiện súc sắc xuất hiện mặt không quá 6 chấm thì Ω là sự kiện chắc chắn.

1.2.2. Sự kiện không thể Φ

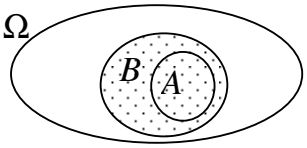
Sự kiện **không thể** là sự kiện luôn luôn không xảy ra sau khi thực hiện phép thử. Ký hiệu: Φ.

Ví dụ 1.3: Gọi Φ là sự kiện súc sắc xuất hiện mặt 7 chấm thì Φ là sự kiện không thể.

1.2.3. Sự kiện thuận lợi

Sự kiện A được gọi là **thuận lợi** cho sự kiện B nếu A xảy ra thì B xảy ra. Ký hiệu: $A \Rightarrow B$ *hay* $A \subseteq B$.

Ví dụ 1.4: Gọi A là sự kiện súc sắc xuất hiện 5 chấm, B là sự kiện súc sắc xuất hiện chấm lẻ. Khi đó $A \Rightarrow B$.



☞ *Một sự kiện có thể thuận lợi cho nhiều sự kiện và một sự kiện có thể có nhiều sự kiện thuận lợi cho nó. Đặc biệt, một sự kiện luôn luôn thuận lợi cho chính nó.*

1.2.4. Sự kiện tương đương

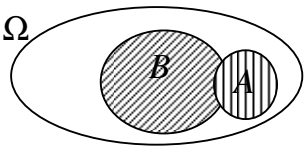
Sự kiện A được gọi là **tương đương** với sự kiện B nếu A xảy ra thì B xảy ra và B xảy ra thì A xảy ra. Ký hiệu: $A \Leftrightarrow B$ hay $A = B$.

Ví dụ 1.5: Gọi A là sự kiện súc sắc xuất hiện 5 chấm, B là sự kiện súc sắc xuất hiện chấm lẻ lớn hơn 3, khi đó $A = B$.

1.2.5. Sự kiện tổng

Sự kiện C được gọi là **tổng** của hai sự kiện A và B nếu C xảy ra khi và chỉ khi A xảy ra hoặc B xảy ra. Ký hiệu: $C = A \cup B$.

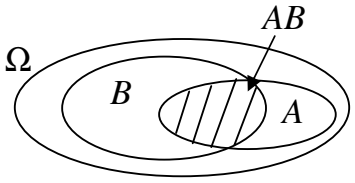
Ví dụ 1.6: Gọi A là sự kiện súc sắc xuất hiện mặt không quá 3 chấm, B là sự kiện súc sắc xuất hiện mặt từ 2 đến 4 chấm. Khi đó $C = A \cup B$ là sự kiện súc sắc xuất hiện mặt không quá 4 chấm.



1.2.6. Sự kiện tích

Sự kiện C được gọi là **tích** của hai sự kiện A và B nếu C xảy ra khi và chỉ khi đồng thời A xảy ra và B xảy ra. Ký hiệu: $C = AB$ hay $C = A \cap B$.

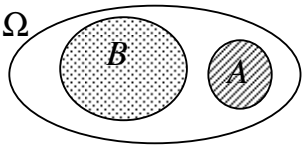
Ví dụ 1.7: Với hai sự kiện A, B trong ví dụ 1.6 và gọi $D = AB$ thì D là sự kiện súc sắc xuất hiện mặt 2 chấm hoặc 3 chấm.



1.2.7. Sự kiện xung khắc

Hai sự kiện A và B được gọi là **xung khắc** nếu A và B không đồng thời xảy ra, hay $AB = \Phi$

Ví dụ 1.8: Gọi A là sự kiện súc sắc xuất hiện mặt không quá 2 chấm, B là sự kiện súc sắc xuất hiện mặt không ít hơn 4 chấm. Khi đó, A xung khắc với B.

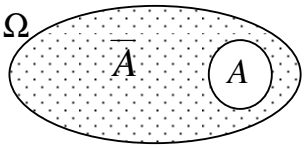


Quy ước: Hai sự kiện A, B xung khắc thì sự kiện tổng có thể viết là $A + B$.

☞ *Hai sự kiện xung khắc với nhau có thể cùng không xảy ra. Trong ví dụ 1.8, khi* súc sắc *xuất hiện mặt 3 chấm thì cả hai sự kiện A và B đều không xảy ra.*

Ví dụ 1.9*: Quan sát gia đình có 2 con. A là sự kiện có 2 con trai, B là sự kiện có 1 con trai. A, B có xung khắc không?*

1.2.8. Sự kiện đối lập



Sự kiện B được gọi là **đối lập** với sự kiện A nếu A xảy ra thì B không xảy ra, nếu A không xảy ra thì B xảy ra. Ký hiệu: $B = \overline{A}$.

Như vậy, A và B đối lập khi và chỉ khi $A + B = \Omega$ và $AB = \Phi$.

Hai hệ thức đặc trưng của cặp sự kiện đối lập A, $\overline{A}$: $A + \overline{A} = \Omega$ và $A\overline{A} = \Phi$.

Ví dụ 1.10: Gọi A là sự kiện súc sắc xuất hiện mặt không quá 2 chấm, B là sự kiện súc sắc xuất hiện mặt không ít hơn 3 chấm. Khi đó $B = \overline{A}$.

2. Tính chất của phép toán sự kiện

Từ khái niệm về sự kiện cho thấy sự tương ứng giữa sự kiện và tập hợp, cụ thể, các phép toán sự kiện $A \cup B$, AB, $\overline{A}$ tương ứng với các phép toán *hợp, giao, phép lấy phần bù* của tập hợp.

Các phép toán sự kiện có các tính chất sau:

i) Kết hợp: $A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$,
$A(BC) = (AB)C = ABC$

ii) Giao hoán: $A \cup B = B \cup A$, $AB = BA$

iii) Phân phối: $A(B \cup C) = AB \cup AC$

iv) Lũy đẳng: $A \cup A = A$, $AA = A$

v) $A \cup \Omega = \Omega$, $A\Omega = A$

vi) $A \cup \Phi = A$, $A\Phi = \Phi$

vii) Nếu $B = \overline{A}$ thì $A = \overline{B}$ hay $\overline{\overline{A}} = A$

viii) Luật đối ngẫu De Morgan: $\overline{A \cup B} = \overline{A}\,\overline{B}$; $\overline{AB} = \overline{A} \cup \overline{B}$

Ví dụ 1.11: Hai người cùng bắn vào bia, mỗi người bắn một viên đạn.
Gọi A_k là sự kiện người thứ k bắn trúng bia, $k = 1, 2$.
Khi đó ta có biểu diễn các sự kiện sau qua A_1, A_2:

a) Chỉ có người thứ 1 bắn trúng bia : $A_1\overline{A_2}$

b) Có đúng một người bắn trúng : $A_1\overline{A_2} + \overline{A_1}A_2$

c) Có ít nhất một người bắn trúng : $A_1 \cup A_2$

d) Cả hai đều bắn trúng : A_1A_2

e) Không có ai bắn trúng : $\overline{A_1}\,\overline{A_2} = \overline{A_1 \cup A_2}$

f) Có không quá một người bắn trúng: $\overline{A_1A_2} = \overline{A_1} \cup \overline{A_2}$

§2. ĐỊNH NGHĨA XÁC SUẤT

1. Định nghĩa xác suất theo cổ điển

Xét một phép thử với không gian các sự kiện sơ cấp gồm n trường hợp (n sự kiện sơ cấp) đồng khả năng (khả năng xảy ra như nhau), trong đó có m trường hợp thuận lợi cho sự kiện A. Xác suất (Probability) của sự kiện A, ký hiệu $P(A)$, được xác định bởi hệ thức:

$$P(A) = \frac{m}{n} \tag{1.1}$$

Ví dụ 1.12: Một công ty cần tuyển 3 nhân viên. Có 7 người nộp đơn gồm 4 nam và 3 nữ. Giả sử khả năng trúng tuyển của 7 người là như nhau. Tính xác suất công ty tuyển được 2 nam và 1 nữ.

Giải

Đặt B là sự kiện công ty tuyển được 2 nam và 1 nữ.

Số trường hợp đồng khả năng: $n = C_7^3$.

Số trường hợp 2 nam trúng tuyển là C_4^2, số trường hợp 1 nữ trúng tuyển là C_3^1.

Suy ra số trường hợp thuận lợi cho B: $m = C_4^2 C_3^1$.

Vậy xác suất của sự kiện B là $P(B) = \frac{C_4^2 C_3^1}{C_7^3} = \frac{18}{35}$. □

Ví dụ 1.12 được phát biểu tổng quát thành ví dụ 1.13 sau đây.

Ví dụ 1.13: (Qui tắc hộp kín) Hộp có N phần tử, trong đó có M phần tử A ($M \le N$). Lấy ngẫu nhiên đồng thời n ($n \le M$) phần tử. Tính xác suất của sự kiện trong n phần tử có k phần tử A ($k \le n$, $k \le M$).

Lập luận tương tự như ví dụ 1.12, ta có: $P = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$.

☞ *Khi* $n = k = 1$, $P = \frac{C_M^1 C_{N-M}^0}{C_N^1} = \frac{M}{N}$, *có nghĩa là khi lấy một phần tử thì* ***xác suất*** *để được phần tử A bằng* ***tỷ lệ*** *phần tử A có trong hộp.*

2. Định nghĩa xác suất theo thống kê

Giả sử phép thử được thực hiện n lần **độc lập** (nghĩa là phép thử sau không phụ thuộc vào kết quả phép thử trước đó), trong đó sự kiện A xuất hiện m lần. Khi đó, m được gọi là **tần số** xuất hiện của sự kiện A, tỉ số $\overline{f} = \frac{m}{n}$ được gọi là **tần suất** (**tỷ lệ**) của sự kiện A.

Khi số phép thử n khá lớn, tần suất $\overline{f}$ đạt giá trị ổn định thì giá trị đó được xem như là **xác suất** của sự kiện A.

$$P(A)=\lim_{n\to\infty}\frac{m}{n} \qquad (1.2)$$

Ví dụ 1.14: Quan sát thí nghiệm tung một đồng xu cân đối và đồng chất n lần

Người thí nghiệm	Số lần thí nghiệm (n)	Số lần xuất hiện mặt sấp (m)	Tần suất $\frac{m}{n}$
Buffon	4.040	2.048	0,5080
Pearson	12.000	6.019	0,5016
Pearson	24.000	12.012	0,5005

Giá trị ổn định trong trường hợp này là 0,5, đó là xác suất xuất hiện mặt sấp khi ta tung một đồng xu cân đối và đồng chất.

☞ *Nhược điểm của định nghĩa này là có những phép thử không thể lặp lại nhiều lần, nên khó xác định được xác suất các sự kiện liên quan tới phép thử đó.*

Ví dụ 1.15: *Có 100 tấm bìa hình vuông như nhau được đánh số từ 1 đến 100. Lấy ngẫu nhiên một tấm bìa. Tìm xác suất:*

a) Được tấm bìa có số mà không có chữ số 5.

b) Được tấm bìa có số chia hết cho 2 hoặc 5 hoặc cả 2 và 5.

3. Định nghĩa xác suất theo hình học

Xét một phép thử với không gian các sự kiện sơ cấp đồng khả năng Ω được biểu diễn bởi miền hình học Ω có độ đo (độ dài, diện tích, thể tích,…) hữu hạn và khác 0. Mỗi sự kiện sơ cấp được xem như một điểm của miền, mỗi sự kiện ngẫu nhiên A được xem như một miền con nào đó của Ω. Khi đó, xác suất của sự kiện A được xác định bởi:

$$P(A)=\frac{m}{n}$$

trong đó, m là độ đo của miền A và n là độ đo của miền Ω.

4. Định nghĩa xác suất theo hệ tiên đề

Giả sử Ω là không gian các sự kiện sơ cấp, $\mathcal{A}$ là họ các tập con của Ω thỏa các điều kiện sau:

i) $\mathcal{A}$ chứa Ω

ii) Nếu $A, B \in \mathcal{A}$ thì $\overline{A}$, $A \cup B$, AB thuộc $\mathcal{A}$

iii) Nếu $A_1, A_2, ..., A_n, ...$ là các phần tử rời nhau của $\mathcal{A}$ thì tổng và tích vô hạn của chúng cũng thuộc $\mathcal{A}$

☞ *Họ $\mathcal{A}$ thỏa các tiên đề i) và ii) được gọi là đại số.*

Họ $\mathcal{A}$ thỏa các tiên đề i), ii) và iii) được gọi là σ - đại số.

Định nghĩa: Ta gọi xác suất trên $(\Omega,\mathcal{A})$ là một hàm số xác định trên $\mathcal{A}$ có giá trị trong [0,1] và thỏa mãn hai tiên đề sau:

i) $P(\Omega) = 1$.

ii) Nếu $\{A_n\}$ dãy các sự kiện xung khắc từng đôi thì $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$.

Khi đó $(\Omega,\mathcal{A}, P)$ được gọi là không gian xác suất.

5. Một số tính chất của xác suất

5.1. Ý nghĩa của xác suất

Xác suất cho biết khả năng xảy ra của một sự kiện trong một phép thử. Sự kiện có xác suất càng lớn thì càng dễ xảy ra và ngược lại.

Ví dụ 1.16: Xét phép thử tung ngẫu nhiên một súc sắc cân đối, đồng chất. Gọi A là sự kiện súc sắc xuất hiện số chấm lẻ, B là sự kiện súc sắc xuất hiện 5 hoặc 6 chấm. Khi đó $P(A) = \frac{1}{2} > P(B) = \frac{1}{3}$ nên sự kiện A dễ xảy ra hơn so với sự kiện B. Tuy nhiên, vẫn có trường hợp sự kiện B xảy ra nhưng sự kiện A không xảy ra, đó là trường hợp súc sắc xuất hiện 6 chấm.

5.2. Tính chất

i) $0 \le P(A) \le 1$, $P(\Omega) = 1$, $P(\Phi) = 0$.

ii) Nếu $A \subseteq B$ thì $P(A) \le P(B)$. Suy ra, nếu $A = B$ thì $P(A) = P(B)$.

iii) $P(A) + P(\overline{A}) = 1$. (1.3)

§3. CÔNG THỨC TÍNH XÁC SUẤT

1. Công thức nhân xác suất

1.1. Xác suất có điều kiện

Khi thực hiện nhóm điều kiện S của thí nghiệm E, ta có không gian mẫu Ω với các sự kiện A, B,... mà ta biết được xác suất $P(A)$, $P(B)$,...

Các xác suất $P(A)$, $P(B)$ ... gọi là các xác suất không điều kiện.

Giả sử sự kiện A xảy ra, ta xem A như là một điều kiện bổ sung vào thí nghiệm E, khi đó xác suất sự kiện B trong điều kiện A xảy ra có thể thay đổi, xác suất sự kiện B lúc này, gọi là xác suất của sự kiện B với điều kiện A.

1.1.1. Định nghĩa

Cho hai sự kiện A và B, xác suất có điều kiện của sự kiện B khi A đã xảy ra là xác suất của B được tính sau khi A đã xảy ra, ký hiệu $P(B|A)$.

Ví dụ 1.17: Hộp có 10 viên bi trong đó có 4 viên màu đỏ, 6 viên màu trắng. Lần lượt lấy không hoàn lại 2 viên bi (mỗi lần lấy một viên). Giả sử lần thứ nhất lấy được bi màu đỏ, tính xác suất để lần thứ hai cũng lấy được bi màu đỏ.

Giải Đặt $Đ_i$ là sự kiện lấy được bi đỏ ở lần lấy thứ i, $i = 1, 2$. Xác suất để lần thứ hai lấy được bi màu đỏ biết rằng lần thứ nhất đã lấy bi màu đỏ:

$$P(Đ_2 \mid Đ_1) = \frac{3}{9} = \frac{1}{3}.\square$$

Ví dụ 1.18*: *Một cuộc điều tra về sở thích mua sắm quần áo của dân cư trong một vùng được bảng số liệu sau:

Sở thích / *Giới tính*	*Thích*	*Không thích*
Nam	*136*	*104*
Nữ	*224*	*36*

Giả sử chọn một người nữ trong vùng, tính xác suất người đó không thích mua sắm.

1.1.2. Định lý

a. Cho hai sự kiện A, B với $P(A) > 0$, ta có:

$$P\left(B|A\right) = \frac{P\left(AB\right)}{P\left(A\right)} \tag{1.4}$$

b. Khi cố định A với $P(A) > 0$ thì xác suất có điều kiện $P\left(B|A\right)$ thỏa tất cả các tính chất của xác suất thông thường.

Ví dụ 1.19: Một lớp thi hai môn: Toán và Lý, có 50% sinh viên thi hỏng môn Toán, 30% sinh viên thi hỏng môn Lý và 20% sinh viên thi hỏng cả hai môn. Chọn ngẫu nhiên một sinh viên của lớp, nếu biết sinh viên này thi hỏng môn Toán thì khả năng người đó thi hỏng môn Lý là bao nhiêu?

Giải

Đặt T, L lần lượt là sự kiện sinh viên thi hỏng môn Toán, Lý.

Ta có $P(TL) = 20\%$, $P(T) = 50\%$, $P(L) = 30\%$

Nên xác suất cần tìm $P\left(L|T\right) = \frac{P\left(LT\right)}{P\left(T\right)} = \frac{2}{5}.\square$

☞ *Trong **Ví dụ 1.19**, nếu biết sinh viên này thi hỏng môn Lý thì khả năng người đó thi hỏng môn Toán là bao nhiêu?*

1.2. Công thức nhân

i) $P\left(AB\right) = P\left(A\right)P\left(B|A\right)$

ii) $P\left(ABC\right) = P\left(A\right)P\left(B|A\right)P\left(C|AB\right)$ (1.5)

iii) $P\left(A_1A_2...A_n\right) = P\left(A_1\right)P\left(A_2|A_1\right)...P\left(A_n|A_1A_2...A_{n-1}\right)$

Ví dụ 1.20: Với giả thiết như **Ví dụ 1.17**, ta có xác suất để cả hai lần đều rút được bi màu đỏ là $P(Đ_1Đ_2) = P(Đ_1).P(Đ_2 \mid Đ_1) = \frac{4}{10}.\frac{3}{9} = \frac{2}{15}.\square$

1.3. Các sự kiện độc lập

+ Nếu $P(B|A) = P(B)$ hay $P(A|B) = P(A)$ thì hai sự kiện A và B được gọi là **độc lập với nhau**.

+ Các sự kiện $\{A_1, A_2,..., A_n\}$, n > 2 được gọi là **độc lập từng đôi** nếu A_i độc lập với A_j với mọi $i \neq j$.

+ Các sự kiện $\{A_1, A_2,..., A_n\}$, n > 2 được gọi là **độc lập toàn phần** nếu mỗi sự kiện độc lập với tích của tùy ý k sự kiện còn lại ($1 \le k \le n-1$).

Hệ quả

i) A, B độc lập : $P(AB) = P(A)P(B)$

ii) $\{A, B, C\}$ độc lập toàn phần : $P(ABC) = P(A)P(B)P(C)$ (1.6)

iii) $\{A_1, A_2,..., A_n\}$ độc lập toàn phần : $P(A_1A_2...A_n) = P(A_1)P(A_2)...P(A_n)$

☞ *Từ đây, để đơn giản, ta sử dụng "độc lập" thay cho "độc lập toàn phần" đối với một nhóm $n \ge 3$ sự kiện.*

Ví dụ 1.21: Tung đồng thời hai súc sắc cân đối, đồng chất. Gọi A_1, A_2 lần lượt là các sự kiện súc sắc thứ nhất, thứ hai xuất hiện mặt 6 chấm thì A_1 độc lập với A_2 nên xác suất để cả hai súc sắc cùng xuất hiện mặt 6 chấm là

$$P(A_1A_2) = P(A_1)P(A_2) = \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{36}\cdot \square$$

Ví dụ 1.22: *Một xí nghiệp có ba phân xưởng cùng sản xuất một loại sản phẩm, với tỷ lệ phế phẩm của mỗi phân xưởng lần lượt là 2%, 3%, 5%. Chọn ngẫu nhiên từ mỗi phân xưởng ra 1 sản phẩm để kiểm tra. Tính xác suất để chọn được 3 phế phẩm.*

☞ *Có thể hiểu trực quan từ định nghĩa là sự kiện A độc lập với sự kiện B nếu A xảy ra hay không xảy ra đều không làm thay đổi xác suất của B. Như vậy, nếu A, B độc lập thì $\overline{A}$, B độc lập; $A, \overline{B}$ độc lập và $\overline{A}, \overline{B}$ cũng độc lập.*

2. Công thức cộng xác suất

i) $P(A \cup B) = P(A) + P(B) - P(AB)$

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

iii) $P(A_1 \cup A_2 \cup ... \cup A_n) =$

$$= \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) - ... + (-1)^{n-1} P(A_1A_2...A_n) \quad (1.7)$$

Ví dụ 1.23: Một nồi hơi có gắn hai van bảo hiểm hoạt động độc lập nhau. Xác suất hoạt động tốt của hai van lần lượt là 0,9; 0,8. Nồi hơi hoạt động an toàn khi có ít nhất một van bản hiểm tốt. Tính xác suất nồi hơi hoạt động an toàn.

Giải

Gọi A_1, A_2 lần lượt là sự kiện van thứ nhất, thứ hai hoạt động tốt.

Ta có $P(A_1) = 0,9$; $P(A_2) = 0,8$. Do đó xác suất nồi hơi hoạt động an toàn là $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1A_2) = 0,9 + 0,8 - 0,9.0,8 = 0,98$.□

Hệ quả

i) $AB = \Phi$ (xung khắc) $\Rightarrow P(A + B) = P(A) + P(B)$

ii) $\begin{cases} AB = \Phi \\ BC = \Phi \\ AC = \Phi \end{cases}$ (xung khắc từng đôi) $\Rightarrow P(A + B + C) = P(A) + P(B) + P(C)$

iii) $A_iA_j = \Phi \ \forall i \neq j$ (xung khắc từng đôi) $\Rightarrow P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$

(1.8)

Ví dụ 1.24: Chọn ngẫu nhiên 13 sản phẩm từ một hộp có 4 sản phẩm xấu và 48 sản phẩm tốt. Tính xác suất các sự kiện sau

A: chọn được 4 sản phẩm xấu,

B: chọn được ít nhất một sản phẩm xấu,

C: không có sản phẩm xấu nào.

Giải

Các xác suất đều có thể được tính theo quy tắc hộp kín

$$P(A) = \frac{C_4^4 C_{48}^9}{C_{52}^{13}} = 0,0026$$

Gọi B_i (i = 1, 2, 3, 4) là sự kiện chọn được i sản phẩm xấu. Các sự kiện $\{B_1, B_2, B_3, B_4\}$ xung khắc từng đôi và $B = B_1 + B_2 + B_3 + B_4$, khi đó

$$P(B) = P(B_1) + P(B_2) + P(B_3) + P(B_4) = \frac{C_4^1 C_{48}^{12}}{C_{52}^{13}} + \cdots + \frac{C_4^4 C_{48}^9}{C_{52}^{13}} = 0,69$$

$$P(C) = \frac{C_4^0 C_{48}^{13}}{C_{52}^{13}} = 0,31$$

Tuy nhiên, bài toán trên có thể tính theo thứ tự sau:

$P(A) = \ldots = 0,0026$; $P(C) = ... = 0,31$

Do $B = \overline{C}$ nên $P(B) = P(\overline{C}) = 1 - P(C) = 1 - 0,31 = 0,69$.□

Ví dụ 1.25: *Có hai người chơi bóng rổ, mỗi người ném một quả. Xác suất ném vào rổ của mỗi người lần lượt là 0,6; 0,7. Tính xác suất:*

a) Có ít nhất một người ném vào rổ.

b) Chỉ có một người ném vào rổ.

3. Công thức xác suất đầy đủ

3.1. Hệ sự kiện đầy đủ, xung khắc từng đôi

Hệ các sự kiện $\{A_1, A_2,...\}$ được gọi là **đầy đủ** nếu $A_1 + A_2 +... = \Omega$, **xung khắc từng đôi** nếu $A_iA_j = \Phi, \forall i \neq j$.

Một ví dụ đơn giản về **hệ đầy đủ và xung khắc từng đôi** là hệ $\{A, \overline{A}\}$ bởi vì $A + \overline{A} = \Omega$, $A.\overline{A} = \Phi$.

Ví dụ 1.26:

i) Tung một xúc xắc. Gọi A_i là sự kiện súc sắc xuất hiện mặt i chấm, $i = \overline{1,6}$. Các sự kiện $\{A_1, A_2,..., A_6\}$ là hệ đầy đủ và xung khắc từng đôi. Các sự kiện $\{A_1, A_3, A_5\}$ là xung khắc từng đôi nhưng không đầy đủ.

ii) Lấy ngẫu nhiên 2 sản phẩm từ 10 sản phẩm gồm 2 sản phẩm xấu và 8 sản phẩm tốt. Đặt B_i là sự kiện lấy được i sản phẩm xấu, $i = 0, 1, 2$. Các sự kiện $\{B_0, B_1, B_2\}$ là hệ đầy đủ và xung khắc từng đôi.

iii) Lần lượt lấy không hoàn lại 2 sản phẩm (mỗi lần lấy 1 sản phẩm) từ 10 sản phẩm gồm 2 sản phẩm xấu và 8 sản phẩm tốt. Đặt C_i là sự kiện lấy được sản phẩm xấu ở lần lấy thứ i, $i = 1, 2$. Các sự kiện $\{C_1, C_2\}$ không là hệ đầy đủ và xung khắc từng đôi vì hệ này không thỏa điều kiện xung khắc từng đôi.

3.2. Công thức xác suất đầy đủ

Giả sử $\{A_1, A_2,..., A_n\}$ là hệ đầy đủ và xung khắc từng đôi, F là một sự kiện ngẫu nhiên có mối liên hệ với hệ sự kiện này và đã biết các xác suất $P(A_i)$, $P(F|A_i)$, $i = 1,...,n$. Khi đó, xác suất của F được tính bởi công thức xác suất đầy đủ sau:

$$P(F) = P(A_1)P(F|A_1) + \cdots + P(A_n)P(F|A_n) \qquad (1.9)$$

Ví dụ 1.27: Giả sử bóng đèn bán ở thị trường do ba công ty I, II, III sản xuất. Số bóng đèn của công ty I, II, III chiếm 30%, 50%, 20% tổng số bóng đèn bán ở thị trường. Trong đó, tỷ lệ bóng đèn hỏng tương ứng của các công ty lần lượt là 1%, 3%, 5%. Mua ngẫu nhiên một bóng đèn, tính xác suất bóng đèn này bị hỏng.

Giải

Gọi A_1, A_2, A_3 lần lượt là sự kiện mua được bóng đèn do công ty I, II, III sản xuất, F là sự kiện mua phải bóng đèn hỏng.

Hệ $\{A_1, A_2, A_3\}$ đầy đủ và xung khắc từng đôi với $P(A_1) = 30\%$; $P(A_2) = 50\%$; $P(A_3) = 20\%$ và $P(F|A_1) = 1\%$; $P(F|A_2) = 3\%$; $P(F|A_3) = 5\%$.

Theo công thức xác suất đầy đủ, xác suất bóng đèn mua được bị hỏng là

$$P(F) = (30\%)(1\%) + (50\%)(3\%) + (20\%)(5\%) = 2,8\%.\square$$

4. Công thức Bayes

Cho hệ n sự kiện đầy đủ và xung khắc từng đôi $\{A_1, A_2,\ldots, A_n\}$ với các xác suất $P(A_i)$, $i=\overline{1,n}$ đã biết. Sự kiện F xảy ra phụ thuộc vào hệ sự kiện trên với xác suất $P\left(F\middle|A_i\right)$, $i=\overline{1,n}$.

Ta có $P\left(A_iF\right)=P\left(FA_i\right)$ hay $P\left(A_i\right)P\left(F\middle|A_i\right)=P\left(F\right)P\left(A_i\middle|F\right)$.

Suy ra, công thức Bayes:

$$P\left(A_i\middle|F\right)=\frac{P\left(A_i\right)P\left(F\middle|A_i\right)}{P\left(F\right)},\ i=1,2,...,n \qquad (1.10)$$

Ví dụ 1.28: Tỷ lệ người dân hút thuốc lá là 30%. Biết rằng tỷ lệ người bị viêm họng trong số người hút thuốc là 60%, còn người bị viêm họng trong số người không hút thuốc là 35%. Chọn ngẫu nhiên một người, biết rằng người đó viêm họng, tính xác suất người đó hút thuốc.

Giải

Gọi A là sự kiện chọn được người hút thuốc lá,

B là sự kiện chọn được người bị viêm họng.

Ta có $P(A) = 30\%$, $P(B|A) = 60\%$, $P\left(B\middle|\overline{A}\right)=40\%$ và $\{A, \overline{A}\}$ là hệ sự kiện đầy đủ. Theo công thức xác suất đầy đủ

$$P\left(B\right)=30\%.60\%+70\%.40\%=46\%$$

Áp dụng công thức Bayes, xác suất để chọn được một người hút thuốc lá trong số những người bị viêm họng là

$$P\left(A\middle|B\right)=\frac{30\%.60\%}{46\%}=39{,}1\%.\ \square$$

Ví dụ 1.29*: Trong kho có 20 thùng hàng, gồm 10 thùng loại I, 6 thùng loại II và 4 thùng loại III. Thùng loại I có 80% sản phẩm loại A, thùng loại II có 60% sản phẩm loại A và thùng loại III có 50% sản phẩm loại A. Lấy ngẫu nhiên một thùng, từ thùng đó lấy ngẫu nhiên ra một sản phẩm.*

a) Tính xác suất để được sản phẩm loại A.

b) Nếu đã lấy ra được sản phẩm loại A, theo bạn bao nhiêu % thùng đã mở thuộc loại I?

5. Công thức Bernoulli

5.1. Lược đồ Bernoulli

Một lược đồ Bernoulli gồm có:

+ Dãy n phép thử độc lập,

+ Trong mỗi phép thử, sự kiện A chỉ xảy ra hoặc không xảy ra với xác suất $P(A) = p$ $(0 < p < 1)$ không đổi.

Ví dụ 1.30:

i) Tung một đồng xu 10 lần. Gọi A là sự kiện đồng xu xuất hiện mặt sấp thì $P(A) = \frac{1}{2}$. Ta có một lược đồ Bernoulli với số phép thử độc lập $n = 10$ và xác suất không đổi $p = \frac{1}{2}$.

ii) Có 15 sản phẩm, trong đó có 6 sản phẩm loại I. Thực hiện 5 phép thử bằng cách kiểm tra 5 sản phẩm trong số đó theo phương pháp chọn lặp (lấy có hoàn lại). Gọi A là sự kiện lấy được sản phẩm loại I trong mỗi lần lấy, ta được một lược đồ Bernoulli với số phép thử độc lập $n = 5$ và xác suất không đổi $p = P(A) = \frac{6}{15}$.

5.2. Công thức Bernoulli

Với một lược đồ Bernoulli gồm n phép thử độc lập và một xác suất $p = P(A)$ không đổi, xác suất để sự kiện A xảy ra k lần được tính bởi công thức Bernoulli

$$P(n;k) = C_n^k p^k q^{n-k} \tag{1.11}$$

với $q = 1 - p$ và $k = 0,1,...,n.$

Ví dụ 1.31: Tỷ lệ sản phẩm loại I trong một lô sản phẩm là 60%.

1) Lấy ngẫu nhiên 5 sản phẩm, tính xác suất trong 5 sản phẩm này có
   a) 2 sản phẩm loại I
   b) ít nhất 1 sản phẩm loại I
2) Cần lấy tối thiểu bao nhiêu sản phẩm để xác suất có ít nhất 1 sản phẩm loại I không nhỏ hơn 0,99?

Giải

Đặt A là sự kiện lấy được chính phẩm, thì $P(A) = 60\% = 0,6$.

Ta có lược đồ Bernoulli với số phép thử độc lập là số sản phẩm lấy ra và xác suất không đổi $p = P(A) = 0,6$.

1) $n = 5$: $P(5;k) = C_5^k (0,6)^k (0,4)^{5-k}$, $k = \overline{0,5}$

   a) $P(5;2) = C_5^2 (0,6)^2 (0,4)^3 = 0,23$

   b) $P(5;k \geq 1) = 1 - P(5;0) = 0,922$

2) Gọi n là số sản phẩm cần lấy. Ta có: $P(n;k \geq 1) = 1 - P(n;0) = 1 - (0,4)^n$

Theo đề bài $P(n;k \geq 1) \geq 0,99$, tức là $1 - (0,4)^n \geq 0,99$. Tính được $n \geq 6$. □

Ví dụ 1.32: *Tín hiệu thông tin được phát 3 lần với xác suất thu được của mỗi lần là 0,4.*

a) Tìm xác suất để nguồn thu nhận được thông tin đó.

b) Nếu muốn xác suất thu được thông tin lên đến 0,9 thì phải phát ít nhất bao nhiêu lần?

ÔN TẬP CUỐI CHƯƠNG

1.- Tung ngẫu nhiên một súc sắc. Gọi A là sự kiện súc sắc xuất hiện mặt chấm chẵn, B là sự kiện súc sắc xuất hiện mặt chấm lẻ. Phương án nào sau đây **sai**?

A. A, B đầy đủ. B. A, B độc lập.
C. A, B đối lập. D. A, B xung khắc.

2.- Lần lượt lấy (mỗi lần 1 sản phẩm) không hoàn lại 2 sản phẩm từ lô hàng có 5 phế phẩm và 8 chính phẩm. Đặt A_i là sự kiện rút được phế phẩm ở lần thứ i ($i = 1, 2$). Sự kiện A_1A_2 là sự kiện rút được

A. 2 phế phẩm. B. ít nhất 1 phế phẩm.
C. 1 hoặc 2 chính phẩm. D. nhiều nhất 2 chính phẩm.

3.- Trong 10 sản phẩm có 2 phế phẩm. Lấy ra ngẫu nhiên 2 sản phẩm, xác suất để cả 2 sản phẩm lấy ra đều là phế phẩm là

A. 0,022. B. 0,04. C. 0,2. D. 0,622.

4.- Có hai lô hàng, lô I có 3 phế phẩm và 7 chính phẩm, lô II có 5 phế phẩm và 4 chính phẩm. Chọn ngẫu nhiên một lô hàng, rồi từ đó lấy ra một sản phẩm. Xác suất lấy được phế phẩm là

A. 0,3. B. 0,43. C. 0,56. D. 0,86.

5.- Khả năng trị khỏi bệnh B của một bác sĩ là 0,7. Xác suất bác sĩ này điều trị cho 10 bệnh nhân có 7 người khỏi bệnh là

A. 0,08. B. 0,27. C. 0,7. D. 1.

6.- Có 30 đề thi trong đó có 10 đề khó, 20 đề trung bình. Tìm xác suất để:

a) Một sinh viên bốc một đề thì gặp được đề trung bình.
b) Một sinh viên bốc hai đề thì gặp được ít ra một đề trung bình.

7.- Cho $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{2}$, $P(A \cup B) = \frac{3}{4}$. Tính $P(AB)$, $P(\overline{A}.\overline{B})$, $P(\overline{A} \cup \overline{B})$, $P(\overline{A}.B)$, $P(A.\overline{B})$.

8.- Một nhóm gồm 5 người ngồi trên một ghế dài. Tính xác suất để 2 người xác định trước luôn ngồi cạnh nhau.

9.- Gieo 6 lần một đồng xu cân đối đồng chất. Tính xác suất để có đúng 4 lần xuất hiện mặt ngửa.

10.- Một xưởng có 3 máy hoạt động độc lập. Trong một ca làm việc, xác suất cần sửa chữa của mỗi máy lần lượt là 0,15; 0,1; 0,12. Tính xác suất sao cho trong một ca làm việc có ít nhất 1 máy cần sửa chữa.

11.- Hộp thứ nhất đựng 3 bi trắng, 7 bi đỏ và 15 bi xanh. Hộp thứ hai đựng 10 bi trắng, 6 bi đỏ và 9 bi xanh. Lấy ngẫu nhiên từ mỗi hộp ra một bi. Tính xác suất lấy được:

a) 2 bi xanh b) 1 bi xanh và 1 bi đỏ c) hai bi cùng màu

12.- Để được nhập kho, sản phẩm của nhà máy phải qua 3 vòng kiểm tra chất lượng độc lập nhau. Xác suất phát hiện ra phế phẩm ở các vòng lần lượt theo thứ tự là 0,8; 0,9; 0,99. Tính xác suất phế phẩm được nhập kho.

13.- Có ba người, mỗi người bắn 1 viên vào mục tiêu với xác suất trúng của mỗi người tương ứng là 0,6; 0,8 và 0,7. Tìm xác suất

a) chỉ có anh thứ 2 bắn trúng.
b) có đúng một người bắn trúng.
c) có ít nhất một người bắn trúng.
d) có đúng hai người bắn trúng.
e) cả ba người bắn trúng.
f) không có ai bắn trúng.

14.- Một lớp học có 100 sinh viên, trong đó có 50 sinh viên giỏi Tiếng Anh, 45 sinh viên giỏi Tiếng Pháp, 10 sinh viên giỏi cả 2 ngoại ngữ. Chọn ngẫu nhiên một sinh viên trong lớp. Tính xác suất sinh viên này

a) giỏi ít nhất một ngoại ngữ.
b) không giỏi ngoại ngữ nào hết.
c) chỉ giỏi Tiếng Anh.
d) chỉ giỏi đúng một ngoại ngữ.

15.- Một sinh viên thi hai môn. Xác suất sinh viên này đạt yêu cầu môn thứ nhất là 80%. Nếu đạt môn thứ nhất thì xác suất đạt yêu cầu môn thứ hai là 60%, nếu không đạt môn thứ nhất thì xác suất đạt yêu cầu môn thứ hai 30%. Hãy tính các xác suất:

a) sinh viên này đạt yêu cầu cả hai môn
b) sinh viên này đạt yêu cầu môn thứ hai
c) sinh viên này đạt yêu cầu ít nhất một môn
d) sinh viên này không đạt yêu cầu cả hai môn

16.- Tại một siêu thị, hệ thống phun nước tự động được lắp liên kết với một hệ thống báo cháy. Khả năng hệ thống phun nước, hệ thống báo cháy bị hỏng lần lượt là 0,1; 0,2. Khả năng để 2 hệ thống này cùng hỏng là 0,04. Tính xác suất:

a) có ít nhất 1 hệ thống hoạt động bình thường.
b) cả 2 hệ thống đều hoạt động bình thường.

17.- Một siêu thị lắp 4 chuông báo cháy hoạt động độc lập nhau. Xác suất để khi có cháy mỗi chuông kêu là 0,95. Tính xác suất để có chuông kêu khi có cháy.

18.- Một lô hạt giống được phân thành ba loại. Loại 1 chiếm 2/3 số hạt cả lô, loại 2 chiếm 1/4, còn lại là loại 3. Loại 1 có tỷ lệ nẩy mầm 80%, loại 2 có tỷ lệ nẩy mầm 60% và loại 3 có tỷ lệ nẩy mầm 40%. Hỏi tỷ lệ nẩy mầm chung của lô hạt giống là bao nhiêu?

19.- Ba sinh viên cùng làm bài thi, xác suất làm được bài của sinh viên A là 0,8; của sinh viên B là 0,7; của sinh viên C là 0,6.

a) Tìm xác suất có 2 sinh viên làm được bài.
b) Nếu có 2 sinh viên làm được bài, tính xác suất để sinh viên A không làm được bài.

20.- Một sinh viên muốn hoàn thành khóa học phải qua 3 kì thi với nguyên tắc cứ đỗ được kỳ thi này mới được thi kì sau. Xác suất sinh viên đó đỗ kì đầu là

0,9. Nếu đỗ kì đầu thì xác suất đỗ được kì thi thứ hai là 0,8; tương tự nếu đỗ kì thi thứ hai thì xác suất đỗ được kì thi thứ ba là 0,7.

a) Tính xác suất để sinh viên đó đỗ cả 3 kì thi.

b) Giả sử sinh viên không đỗ được 3 kì thi, tính xác suất người đó bị trượt ở kì thi thứ hai.

21.- Có 10 hộp bi trong đó:

+ 4 hộp loại 1, mỗi hộp gồm 3 bi trắng và 5 bi đỏ,
+ 3 hộp loại 2, mỗi hộp gồm 4 bi trắng và 6 bi đỏ,
+ 3 hộp loại 3, mỗi hộp gồm 2 bi trắng và 5 bi đỏ.

Lấy ngẫu nhiên một hộp và từ đó lấy ngẫu nhiên một bi thì được bi trắng. Tính xác suất để viên bi đó rút ra từ hộp loại 2.

22.- Trong một cơ quan điều tra người ta dùng máy dò tìm tội phạm, kinh nghiệm cho biết cứ 10 người bị tình nghi thì 7 người là tội phạm. Máy báo đúng người có tội với xác suất 0,85. Máy báo sai người vô tội với xác suất 0,1. Một người được máy phân tích. Hãy tính xác suất:

a) người này là tội phạm.

b) máy báo người này là tội phạm.

c) người này thực sự có tội, biết rằng máy đã báo có tội.

d) máy báo đúng.

23.- Tỷ lệ ôtô tải, ôtô con và xe máy đi qua đường X có trạm xăng là 5 : 2 : 13. Xác suất để ôtô tải, ôtô con và xe máy đi qua đường này vào bơm xăng lần lượt là 0,1; 0,2 và 0,15. Biết rằng có 1 xe đi qua đường X vào bơm xăng, tính xác suất để đó là ôtô con.

24.- Tỷ lệ mắc bệnh basedow ở một địa phương là 10%. Trong đợt khám sức khoẻ cho 100 người, tính xác suất có

a) 6 người bị basedow.

b) 95 người không bị basedow.

c) ít nhất một người bị basedow.

25.- Một công ty sử dụng hai hình thức quảng cáo là đài phát thanh và tivi. Giả sử có 25% khách hàng biết được thông tin quảng cáo qua tivi, 34% khách hàng biết được thông tin quảng cáo qua đài phát thanh và 10% khách hàng biết được thông tin quảng cáo qua cả hai hình thức trên. Tìm xác suất để chọn ngẫu nhiên một khách hàng thì người đó biết được thông tin quảng cáo của công ty.

26.- Một hộp có 3 phiếu gồm 2 phiếu trúng và 1 phiếu trắng. Cho 3 người lần lượt rút thăm, ai được phiếu trúng sẽ có phần thưởng. Hỏi cách rút thăm như thế có công bằng không?

27.- Một thủ kho có một chùm chìa khóa gồm 9 chiếc trông giống hệt nhau, trong đó chỉ có một chiếc mở được kho. Anh ta thử ngẫu nhiên từng chìa khóa một, chiếc nào được thử thì không thử lại. Tính xác suất anh thủ kho mở được khóa ở lần thử thứ 4.

28.- Trước khi đưa sản phẩm ra thị trường người ta đã phỏng vấn ngẫu nhiên 200 khách hàng về sản phẩm đó, thấy có 34 người trả lời "sẽ mua", 97 người trả lời "có thể mua" và 69 người trả lời "không mua". Kinh nghiệm cho thấy tỷ lệ khách hàng thực sự sẽ mua sản phẩm tương ứng với những cách trả lời trên là 70%, 30% và 1%.

a) Hãy đánh giá thị trường tiềm năng của sản phẩm đó.

b) Trong số khách hàng thực sự mua sản phẩm thì có bao nhiêu % trả lời "sẽ mua"?

29.- Hai cửa hàng A và B cung cấp các hộp đĩa CD cho một trung tâm tin học với tỷ lệ $\frac{3}{2}$. Tỷ lệ đĩa bị lỗi của các cửa hàng tương ứng là 1% và 2%. Một sinh viên đến thực tập tại trung tâm chọn ngẫu nhiên một hộp đĩa gồm 20 chiếc và từ đó rút ra ngẫu nhiên một đĩa.

a) Tính xác suất để sinh viên đó rút phải đĩa bị lỗi.

b) Sau khi khởi động máy, sinh viên đó nhận thấy đĩa quả thật đĩa bị lỗi. Tính xác suất để đĩa này do cửa hàng A cung cấp.

30.- Ở một trường học, người ta nhận thấy rằng xác suất để 1 sinh viên khi đi học bị bệnh và phải nằm điều trị tại phòng y tế của trường là 0,04%. Biết rằng trong một buổi học, trung bình có 700 sinh viên. Tính xác suất để trong một buổi học có 3 sinh viên phải nằm điều trị tại phòng y tế và theo bạn, phòng y tế cần trang bị bao nhiêu giường điều trị?

31.- Một khách sạn có 2 hệ thống: báo cháy và báo khói; hai hệ thống này hoạt động độc lập. Xác suất để hệ thống báo cháy và báo khói hỏng tương ứng là 0,07; 0,04. Khách sạn phòng cháy an toàn khi có hệ thống không hỏng. Tính xác suất:

a) Khách sạn phòng cháy an toàn.

b) Khách sạn phòng cháy không an toàn.

c) Giả sử khách sạn phòng cháy an toàn, tính xác suất khi đó hệ báo cháy không hỏng.

Chương 2. ĐẠI LƯỢNG NGẪU NHIÊN

Mục tiêu

Sau khi học xong chương này, sinh viên cần đạt được:

1. Kiến thức

- Biết khái niệm về đại lượng ngẫu nhiên và phân loại.
- Hiểu quy luật của một đại lượng ngẫu nhiên (luật phân phối xác suất, hàm phân phối xác suất).
- Hiểu được bản chất, ý nghĩa các đặc trưng số của đại lượng ngẫu nhiên như kỳ vọng, phương sai, độ lệch tiêu chuẩn, giá trị tin chắc nhất, trung vị, phân vị mức xác suất.
- Hiểu được dấu hiệu bản chất, luật phân phối xác suất, các công thức của phân phối xác suất quan trọng như siêu bội, nhị thức, Poisson, phân phối chuẩn và chuẩn tắc.
- Hiểu được ý nghĩa của 3 định lý xấp xỉ.
- Làm quen khái niệm đại lượng ngẫu nhiên hai chiều.

2. Kỹ năng

- Tìm được luật phân phối xác suất của các đại lượng ngẫu nhiên.
- Tính được xác suất của một đại lượng ngẫu nhiên khi biết luật phân phối xác suất của nó.
- Tính được các giá trị đặc trưng số bằng công thức và bằng máy tính bỏ túi.
- Vận dụng thành thạo các định lý xấp xỉ để tính được các giá trị xác suất của một đại lượng ngẫu nhiên.

3. Thái độ

- Có ý thức vận dụng kiến thức đã học vào việc giải một bài toán thực tiễn.
- Coi trọng tính quy luật trong khoa học và trong cuộc sống, từ đó phải nghiêm túc trong khoa học và trong cuộc sống.
- Xây dựng ý thức chịu khó, kiên nhẫn vì thấy rằng vốn dĩ quy luật cuộc sống (đại lượng ngẫu nhiên) là phức tạp và có mối quan hệ chằng chịt.

§1. ĐẠI LƯỢNG NGẪU NHIÊN LUẬT PHÂN PHỐI XÁC SUẤT CỦA ĐẠI LƯỢNG NGẪU NHIÊN

1. Đại lượng ngẫu nhiên (Random variable)

Trong thực tế, ngoài các sự kiện chúng ta còn gặp các biến số liên kết với các sự kiện đó (thực chất là mối liên hệ giữa định tính và định lượng) nên ta còn xét đến các đại lượng ngẫu nhiên (ĐLNN) liên kết với các sự kiện ngẫu nhiên, chẳng hạn sự kiện trời mưa gắn với đại lượng là thời gian mưa, sự kiện một doanh nghiệp hoạt động có hiệu quả gắn với đại lượng là lượng lãi ròng thu được,…

Ví dụ 2.1:

a) Tung ngẫu nhiên một súc sắc và gọi X là số chấm xuất hiện, ta có X là ĐLNN nhận một trong các giá trị 1, 2,..., 6. Khi đó X được gọi là ĐLNN rời rạc (hữu hạn).

b) Gọi Y là số khách hàng vào một ngân hàng trong ngày, ta có X là ĐLNN nhận một trong các giá trị 0, 1, 2,... Khi đó X được gọi là ĐLNN rời rạc (vô hạn).

c) Gọi Z là trọng lượng của một gói mì ăn liền, ta có Z là ĐLNN nhận một giá trị trong khoảng (0; 200) (g). Khi đó Z được gọi là ĐLNN liên tục.

Đại lượng ngẫu nhiên (Biến số ngẫu nhiên, biến ngẫu nhiên) là biến số **biến thiên** phụ thuộc vào phép thử ngẫu nhiên với một xác suất tương ứng xác định.

Các ĐLNN thường được ký hiệu $X, Y, Z, X_1, X_2, X_3,\ldots$

Có 2 loại ĐLNN:

+ ĐLNN r**ời rạc**: Chỉ nhận hữu hạn hoặc vô hạn đếm được các giá trị.

+ ĐLNN l**iên tục**: Nhận vô hạn không đếm được các giá trị. Nói cách khác, các giá trị nhận được của ĐLNN X liên tục lấp đầy một khoảng số thực $(a;b)$.

2. Luật phân phối xác suất của ĐLNN

Để xác định một đại lượng ngẫu nhiên, trước hết ta phải biết đại lượng đó **có thể nhận những giá trị nào**. Chẳng hạn với phép thử gieo 2 đồng xu thì số mặt ngửa xuất hiện là 0, 1, 2. Ngoài ra, ta phải biết nó nhận các giá trị trên với **xác suất tương ứng là bao nhiêu**.

2.1. Luật phân phối xác suất của ĐLNN rời rạc

Luật phân phối xác suất của ĐLNN X rời rạc (còn gọi là bảng phân phối xác suất) được thể hiện bởi bảng sau (giả thiết X nhận hữu hạn giá trị):

X	x_1	x_2	. . .	x_i	. . .	x_n
P	p_1	p_2	. . .	p_i	. . .	p_n

trong đó: x_i ($i = 1,\ldots, n$) là các giá trị (phân biệt) của X

$$p_i = P[X = x_i] \text{ thỏa } p_i > 0, i = 1,...,n \text{ và } \sum_{i=1}^{n} p_i = 1. \qquad (2.1)$$

Ví dụ 2.2: Một người vào cửa hàng thấy có 5 máy thu thanh giống nhau. Anh ta đề nghị cửa hàng cho anh thử lần lượt các máy đến khi chọn được máy tốt thì mua, nếu cả 5 lần đều xấu thì thôi. Gọi X là số lần thử. Biết xác suất một máy tốt là 0,8 và các máy xấu tốt độc lập với nhau. Tìm luật phân phối xác suất của X và tính xác suất người mua phải thử từ 3 đến 5 lần.

Giải

Gọi A_i $(i = 1,2,3,4,5)$ là sự kiện máy thứ i tốt.

Ta có $P(A_i) = p = 0,8,\ q = 0,2$

Suy ra: $P[X = 1] = P(A_1) = p = 0,8;$

$$P[X=2]=P(\overline{A}_1A_2)=qp=0{,}16;$$
$$P[X=3]=P(\overline{A}_1\overline{A}_2A_3)=q^2p=0{,}032;$$
$$P[X=4]=P(\overline{A}_1\overline{A}_2\overline{A}_3A_4)=q^3p=0{,}0064;$$
$$P[X=5]=P(\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4)=q^4=0{,}0016;$$

Luật phân phối xác suất của X là

X	1	2	3	4	5
P	0,8	0,16	0,032	0,0064	0,0016

Xác suất người mua phải thử từ 3 đến 5 lần

$$P[3\le X\le 5]=P[X=3]+P[X=4]+P[X=5]=0{,}04.\square$$

2.2. Luật phân phối xác suất của ĐLNN liên tục

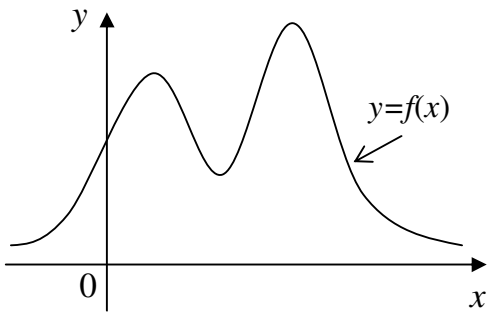
Luật phân phối xác suất của ĐLNN liên tục được biểu thị bởi hàm số $y=f(x)$ xác định trên $\mathbb{R}$ thỏa mãn:

i) $f(x)\ge 0,\ \forall x\in\mathbb{R}$

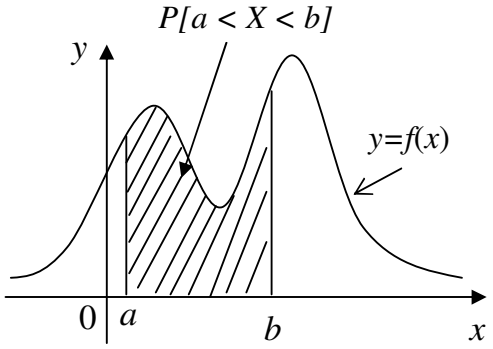
ii) $\int_{-\infty}^{+\infty} f(x)dx=1$ (2.2)

Hàm $y=f(x)$ được gọi là **hàm mật độ xác suất** của ĐLNN X.

Điều kiện ii) cho thấy tổng diện tích tạo bởi đồ thị $y=f(x)$ với trục hoành bằng 1.



Hình 2.1



Hình 2.2

Tính chất: Với ĐLNN X liên tục, ta có

i) $P[X=C]=0$, C là hằng số

ii) $P[a<X<b]=\int_a^b f(x)dx$ (2.3)

☞ *Do tính chất i) nên trong tính chất ii) giá trị xác suất không đổi khi thêm* $X=a$ *hoặc* $X=b$.

Ví dụ 2.3: Cho ĐLNN X liên tục có hàm mật độ xác suất

$$y = f(x) = \begin{cases} k(3x - x^2) & \text{khi} \quad x \in [0;3] \\ 0 & \text{khi} \quad x \notin [0;3] \end{cases}, \ k \text{ là hằng số}$$

a) Xác định hằng số k

b) Tính $P[1 < X < 2]$

Giải

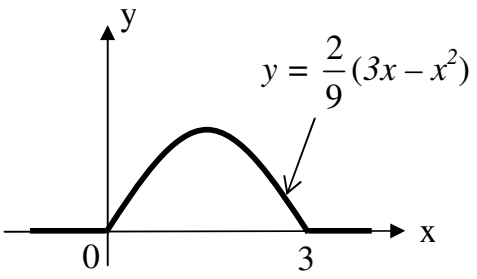
a) Do $f(x)$ là hàm mật độ xác suất của ĐLNN X nên theo (2.2)

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{3} f(x)dx + \int_{3}^{+\infty} f(x)dx$$

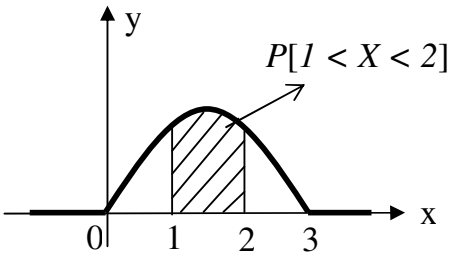
$$= \int_{-\infty}^{0} 0dx + \int_{0}^{3} k(3x - x^2)dx + \int_{3}^{+\infty} 0dx = \frac{9k}{2}.$$

Vậy $k = \frac{2}{9}$.

b) Theo (2.3), $P[1 < X < 2] = \int_{1}^{2} f(x)dx = \int_{1}^{2} \frac{2}{9}(3x - x^2)dx = \frac{13}{27}$.



Hình 2.3



Hình 2.4

☞ *Tính P[X < 6], P[–2 < X < 6].*

3. Hàm phân phối xác suất của ĐLNN

Trong các ứng dụng thực tế, thông thường chỉ sử dụng luật phân phối xác suất. Nhưng trong việc nghiên cứu lý thuyết người ta còn đưa ra khái niệm ***hàm phân phối xác suất***.

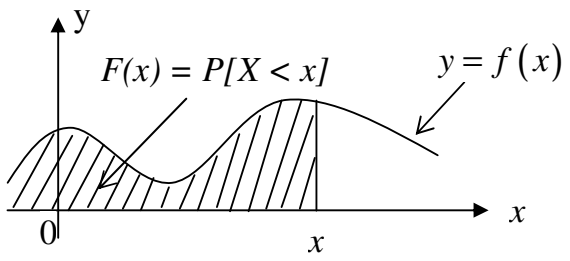
Hàm phân phối xác suất của ĐLNN X, ký hiệu $F(x)$, là hàm được xác định bởi công thức:

$$F(x) = P[X < x] \tag{2.4}$$

Nhận xét:

- Nếu X rời rạc có $P[X = x_i] = p_i$ thì $F(x) = \sum_{x_i < x} p_i$. (2.5)

- Nếu X liên tục có hàm mật độ xác suất $y = f(x)$ thì $F(x) = \int_{-\infty}^{x} f(t)dt.$ (2.6)



Hình 2.5

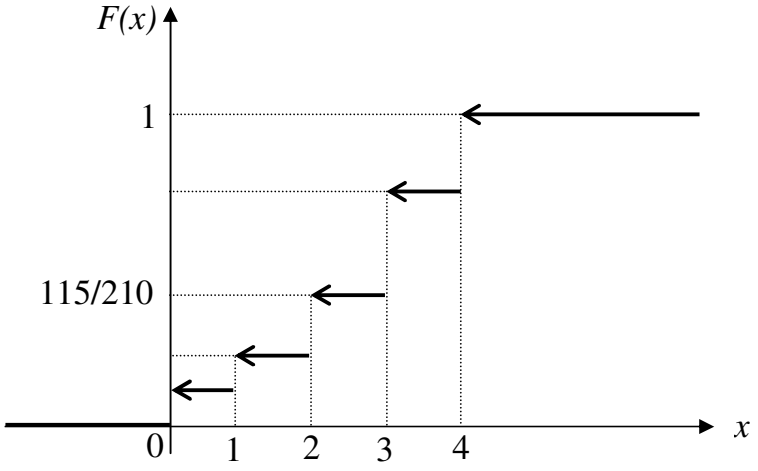
Ví dụ 2.4: Tìm hàm phân phối xác suất của ĐLNN X có luật phân phối

X	0	1	2	3	4
P	$\frac{1}{210}$	$\frac{24}{210}$	$\frac{90}{210}$	$\frac{80}{210}$	$\frac{15}{210}$

Giải: Do X là ĐLNN rời rạc nên $F(x) = \sum_{x_i < x} p_i$, ta có:

- $x \le 0$ $\quad F(x) = 0$
- $0 < x \le 1$ $\quad F(x) = \frac{1}{210}$
- $1 < x \le 2$ $\quad F(x) = \frac{25}{210}$
- $2 < x \le 3$ $\quad F(x) = \frac{115}{210}$
- $3 < x \le 4$ $\quad F(x) = \frac{195}{210}$
- $x > 4$ $\quad F(x) = 1$

Đồ thị hàm $F(x)$



Hình 2.6

☞ *Cho biết giá trị của F(1,3); F(-1); F(5).*

Ví dụ 2.5: Cho ĐLNN X liên tục có hàm mật độ xác suất

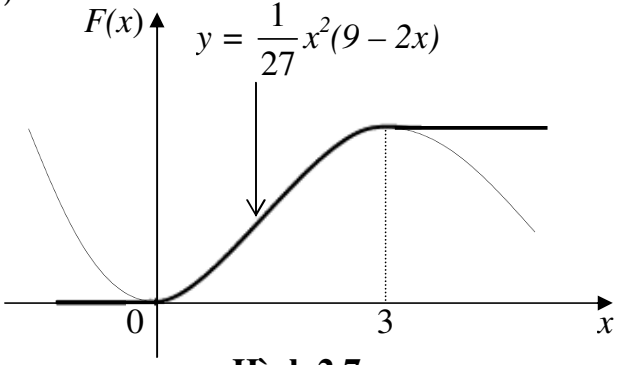
$$y = f(x) = \begin{cases} \frac{2}{9}(3x - x^2) & \text{khi} \quad x \in [0;3] \\ 0 & \text{khi} \quad x \notin [0;3] \end{cases}$$

Xác định hàm phân phối $F(x)$ của X.

Giải : X là ĐLNN liên tục nên $F(x) = \int_{-\infty}^{x} f(t)dt$, do đó:

- $x < 0$ $\qquad F(x) = 0$
- $0 \le x \le 3$ $\qquad F(x) = \int_{0}^{x} \frac{2}{9}(3t - t^2)dt = \frac{1}{27}x^2(9 - 2x)$
- $x > 3$ $\qquad F(x) = 1$

Đồ thị hàm $F(x)$



Hình 2.7

Sau đây là một số tính chất của hàm phân phối xác suất $F(x)$ được suy ra dễ dàng từ định nghĩa:

+ $0 \le F(x) \le 1, \ \forall x.$

+ $F(x)$ là hàm không giảm và liên tục bên trái.

+ Nếu X là ĐLNN rời rạc thì $F(x)$ có dạng hình bậc thang.

+ Nếu X là ĐLNN liên tục có hàm mật độ xác suất $f(x)$ thì $F'(x) = f(x)$.

+ $P[a < X < b] = F(b) - F(a)$, công thức này vẫn đúng khi thêm $X = a$ hoặc $X = b$ (2.7)

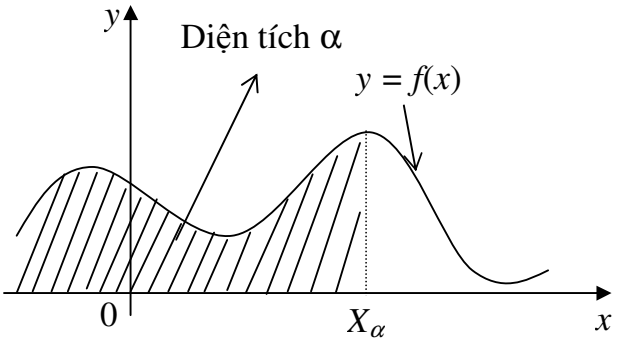
☞ *Ở ví dụ 2.4, tính P[1,3 < X < 3].*
☞ *Ở ví dụ 2.5, tính P[2 < X < 6].*

4. Phân vị mức xác suất α

Cho ĐLNN X liên tục có hàm mật độ xác suất $f(x)$, hàm phân phối xác suất $F(x)$. Với số α cho trước (0 < α < 1), **phân vị** của ĐLNN X với **mức xác suất α** là số X_α sao cho

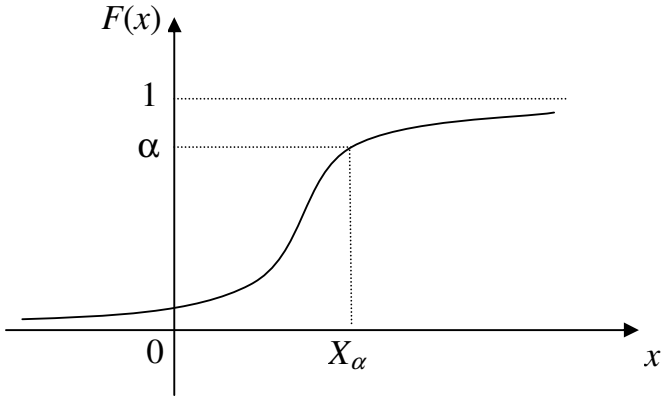
$$P[X < X_\alpha] = \alpha \qquad (2.8)$$

Hệ thức trên 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.



Hình 2.8

Mặt khác, từ hệ thức $F(X_\alpha) = \alpha$ suy ra $X_\alpha = F^{-1}(\alpha)$. Như vậy, X_α là giá trị ngược của hàm phân phối xác suất $F(x)$ tại α cho trước.



Hình 2.9

Ví dụ 2.6: Cho ĐLNN *X* liên tục trong ví dụ 2.5. Với $\alpha = \frac{7}{27}$, hãy xác định X_α (phân vị của *X*, mức xác suất $\frac{7}{27}$).

Đặt $u = X_\alpha$, sử dụng công thức $F(X_\alpha) = \alpha$ ta có $F(u) = \frac{7}{27}$.

Kết hợp ví dụ 2.5, suy ra $\frac{1}{27}u^2(9-2u) = \frac{7}{27};\ 0 < u < 3 \Rightarrow 2u^3 - 9u^2 + 7 = 0$

$\Rightarrow u = \frac{7 \pm \sqrt{105}}{4}$ (loại) hoặc $u = 1$. Vậy $X_\alpha = 1$ với $\alpha = \frac{7}{27}$. □

☞ *Giải thích kết quả* $X_{\frac{7}{27}} = 1$ *bằng hình học.*

§2. CÁC ĐẶC TRƯNG SỐ CỦA ĐẠI LƯỢNG NGẪU NHIÊN

Từ luật phân phối xác suất của ĐLNN người ta thường rút ra một vài số đặc trưng cho mặt này hay mặt khác của biến đó và đôi khi giúp ta so sánh giữa các ĐLNN khác nhau. Các đặc trưng số thường gặp: kỳ vọng, phương sai, độ lệch tiêu chuẩn, giá trị tin chắc nhất, trung vị.

1. Kỳ vọng

1.1. Định nghĩa

Kỳ vọng của ĐLNN *X*, ký hiệu *EX* (Expectation) hoặc *MX* (Mean) được xác định như sau:

- *X* rời rạc có $P\left[X = x_i\right] = p_i$ thì

$$EX = \sum_{i=1}^{n} x_i p_i \tag{2.9}$$

- *X* liên tục có hàm mật độ xác suất $y = f\left(x\right)$ thì

$$EX = \int_{-\infty}^{+\infty} x\, f(x)\, dx \tag{2.10}$$

Ví dụ 2.7:

Với *X* cho trong ví dụ 2.4, ta có $EX = 0\times\frac{1}{210} + 1\times\frac{24}{210} + \cdots + 4\times\frac{15}{210} = 2{,}4.$□

Ví dụ 2.8: Với *X* cho trong ví dụ 2.5, ta có:

$$EX = \int_{-\infty}^{+\infty} xf(x)dx = \frac{2}{9}\int_{0}^{3} x(3x - x^2)dx = \frac{3}{2}\cdot$$ □

1.2. Ý nghĩa

Kỳ vọng là **giá trị trung bình theo xác suất** hay là **giá trị trung tâm** của ĐLNN *X*. Đây là trung tâm điểm của phân phối mà các giá trị cụ thể của *X* sẽ tập trung hay phân tán quanh đó.

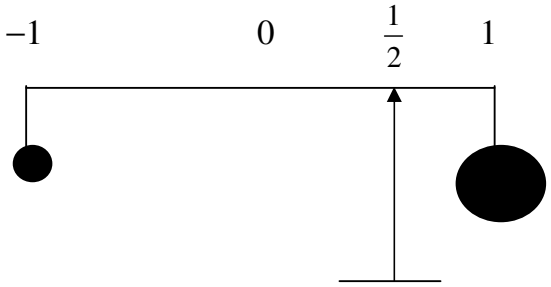
☞ *Phân tích ý nghĩa thực tế của ĐLNN X và giá trị EX* ?

***Lưu ý**: Có hai giá trị trung bình thường được sử dụng để ước lượng xu hướng tập trung của ĐLNN: trung bình cộng và trung bình theo xác suất. Tuy nhiên, trung bình cộng sẽ tốt trong trường hợp các trị số đang xét có cùng khả năng, nhưng phản ảnh không tốt trong trường hợp các trị số được nhận không cùng khả năng, khi đó trung bình theo nghĩa xác suất sẽ tốt hơn.*

Chẳng hạn như:

X	*–1*	*1*
P	1/4	3/4

Ta có $EX = (-1)\frac{1}{4} + (1)\frac{3}{4} = \frac{1}{2}$, *trong khi trung bình cộng là* $\frac{-1+1}{2} = 0$



Hình 2.10

1.3. Tính chất

Cho *X*, *Y* là các ĐLNN và *c* là hằng số, kỳ vọng có các tính chất sau:

i) $Ec = c$.

ii) $E(cX) = c(EX)$.

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

iv) Nếu *X*, *Y* độc lập (tức là, hai sự kiện $\{X < x\}, \{Y < y\}$ độc lập với mọi *x*, *y*) thì $E(XY) = EX.EY$.

2. Phương sai

2.1. Định nghĩa

Phương sai của ĐLNN *X*, ký hiệu *DX* (Dispersion) hoặc *VarX* (Variance), là giá trị được xác định bởi công thức

$$DX = E\left[X - EX\right]^2 \tag{2.11}$$

Bằng cách khai triển công thức (2.11), ta có công thức tính phương sai sau

$$DX = E\left(X^2\right) - \left(EX\right)^2 \tag{2.12}$$

Nhận xét

- Với *X* là ĐLNN rời rạc có $P\left[X = x_i\right] = p_i$: $DX = \sum_{i=1}^{n} x_i^2 p_i - \left(\sum_{i=1}^{n} x_i p_i\right)^2$ (2.13)
- Với *X* là ĐLNN liên tục có hàm mật độ xác suất $y = f(x)$:

$$DX = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left(\int_{-\infty}^{+\infty} xf(x)dx\right)^2 \tag{2.14}$$

Ví dụ 2.9: Với *X* cho trong ví dụ 2.4, kết hợp ví dụ 2.7, ta được:

$$DX = \left(0^2.\frac{1}{210} + 1^2.\frac{24}{210} + ... + 4^2.\frac{15}{210}\right) - \left(2,4\right)^2 = 0,64\,.\square$$

Ví dụ 2.10: Với X cho trong ví dụ 2.3, kết hợp với ví dụ 2.8, ta được:

$$DX = \frac{2}{9}\int_0^3 x^2(3x - x^2)dx - \left(\frac{3}{2}\right)^2 = 0,45.\square$$

2.2. Ý nghĩa

Phương sai là số không âm dùng để đo **mức độ phân tán** của giá trị ĐLNN quanh kỳ vọng. Phương sai càng nhỏ thì độ phân tán của X quanh giá trị đại diện EX càng thấp; và ngược lại, phương sai càng lớn thì độ phân tán của X quanh giá trị đại diện EX càng cao.

Đặc trưng của phương sai trong một số lĩnh vực:

+ Kỹ thuật: phương sai đặc trưng cho độ sai số của thiết bị,
+ Kinh doanh: phương sai đặc trưng cho độ rủi ro của các quyết định,
+ Công nghiệp: phương sai đặc trưng cho độ chính xác của sản xuất,
+ Chăn nuôi: phương sai đặc trưng cho độ đồng đều của các gia súc,
+ Trồng trọt: phương sai đặc trưng cho mức độ ổn định của năng suất,...

Ví dụ 2.11: Năng suất (sản phẩm/phút) của hai máy tương ứng là các ĐLNN X, Y có phân phối xác suất như sau:

X	1	2	3	4
P	0,3	0,1	0,5	0,1

Y	2	3	4	5
P	0,1	0,4	0,4	0,1

Nếu phải chọn mua một trong hai máy này ta nên chọn máy nào?

Giải

Ta tính được: $EX = 2,4$; $EY = 3,5$;

$$EX^2 = 1^2.0,3 + 2^2.0,1 + 3^2.0,5 + 4^2.0,1 = 6,8;$$
$$EY^2 = 2^2.0,1 + 3^2.0,4 + 4^2.0,4 + 5^2.0,1 = 12,9;$$
$$DX = EX^2 - (EX)^2 = 1,04;\ DY = EY^2 - (EY)^2 = 0,65$$

So sánh các số đặc trưng ta thấy:

$EY > EX$: năng suất trung bình của Y cao hơn của X

$DY < DX$: năng suất của Y ổn định hơn của X

Vậy ta nên chọn mua máy Y.$\square$

2.3. Tính chất

Cho X, Y là các ĐLNN và c là hằng số, phương sai có các tính chất sau:

i) $Dc = 0$.

ii) $D(cX) = c^2DX$.

iii) Nếu X, Y độc lập thì $D(X \pm Y) = DX + DY$.

☞ *Tính $D(X + c)$.*

3. Độ lệch tiêu chuẩn

3.1. Định nghĩa

Độ lệch tiêu chuẩn (Standard deviation) của ĐLNN X, ký hiệu σ_X, được xác định như sau:

$$\sigma_X = \sqrt{DX} \tag{2.15}$$

3.2. Ý nghĩa

σ_X có tính chất giống như DX nhưng có ưu điểm là có cùng đơn vị đo lường với EX (và cũng là đơn vị đo lường của X).

4. Giá trị tin chắc nhất (Mode, yếu vị)

Ký hiệu: Mod$X$.

- Nếu X là ĐLNN rời rạc có $P[X = x_i] = p_i$ thì $\text{Mod}X = x_i$ tương ứng với p_i lớn nhất.
- Nếu X là ĐLNN liên tục có hàm mật độ $f(x)$ thì $\text{Mod}X = x_0$ sao cho $f(x)$ đạt giá trị lớn nhất tại $x = x_0$.

Ví dụ 2.12:

a) Nếu X có bảng phân phối xác suất sau thì $\text{Mod}X = 0$

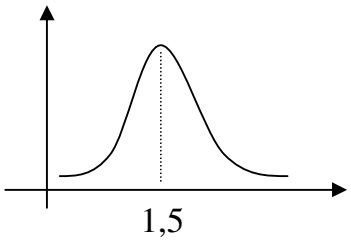
X	0	1	2
P	0,72	0,26	0,02

b) Nếu X có bảng phân phối xác suất:

X	-1	0	1	2
P	0,4	0,18	0,4	0,02

thì $\text{Mod}X = -1$ và $\text{Mod}X = 1$.

c)



Hình 2.11

Với X liên tục có đồ thị của hàm mật độ xác suất $f(x)$ như hình 2.11 thì Mod$X$ = 1,5.□

☞ *ModX có thể nhận nhiều giá trị khác nhau.*

5. Trung vị

Số m được gọi là trung vị (Median) của ĐLNN X nếu

$$P(X<m)\le\frac{1}{2} \text{ và } P(X>m)\le\frac{1}{2} \qquad (2.16)$$

Ký hiệu: Med$X$.

Chú ý:

1. Nếu a, b (a < b) là các trung vị của X thì đoạn [a,b] là tập hợp tất cả các trung vị của ĐLNN X.

2. Nếu X có hàm phân phối xác suất $F(x)$ liên tục thì $F(m)=\frac{1}{2}$.

Ví dụ 2.13: Nếu X có bảng phân phối xác suất:

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

thì Med$X$ = 1 vì

$P(X<1)=P(X=-1)+P(X=0)=0,3<\frac{1}{2}$; $P(X>1)=P(X=2)=0,3<\frac{1}{2}$.□

Ví dụ 2.14: Nếu X có bảng phân phối xác suất:

X	0	1	2	3
P	0,125	0,375	0,375	0,125

thì Med$X\in$ [1; 2] vì

$P(X<1)=P(X=0)=0,125<\frac{1}{2}$ và $P(X>1)=\frac{1}{2}$. Med$X$ = 1.

Tương tự ta có Med$X$ = 2 và không có Med$X$ ngoài đoạn [1; 2].
Vậy các giá trị của Med$X$ thuộc đoạn [1; 2]. □

Ví dụ 2.15: Cho ĐLNN X có hàm mật độ $f(x)=\begin{cases}2e^{-2x} & khi\ x\ge 0\\ 0 & khi\ x<0\end{cases}$.

Tìm Med$X$.

<u>Giải</u>

Gọi m là Med$X$, ta sẽ tìm m thỏa $P(X<m)=\frac{1}{2}$ $\Rightarrow\int_{-\infty}^{m}f(x)\,dx=\frac{1}{2}$

$\Rightarrow\int_{0}^{m}2e^{-2x}\,dx=\frac{1}{2}\Rightarrow m=\frac{\ln 2}{2}$. □

☞ Med$X$ *có thể nhận nhiều giá trị khác nhau.*

Trong ứng dụng, trung vị là đặc trưng vị trí tốt nhất, đôi khi tốt hơn cả kỳ vọng, nhất là khi trong số liệu có nhiều sai sót. Trung vị còn được gọi là phân vị 50% của phân phối.

§3. MỘT SỐ LUẬT PHÂN PHỐI XÁC SUẤT ĐẶC BIỆT

1. Phân phối siêu bội *H(N; M; n)*

1.1. Bài toán

Hộp có N viên bi, trong đó có M viên bi màu đỏ ($M \le N$). Rút đồng thời n viên bi ($n \le M$) và gọi X là số viên bi màu đỏ được rút ra, ta nói rằng X có **luật phân phối siêu bội (siêu hình học) với ba tham số *N, M, n***.

1.2. Định nghĩa

Phân phối siêu bội (Hypergeometric distribution) là phân phối của ĐLNN rời rạc X nhận $n+1$ các giá trị 0, 1, 2, ..., n với xác suất tương ứng của X được tính theo công thức (quy tắc hộp kín):

$$P[X=k]=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n};\quad k=\overline{0,n} \tag{2.17}$$

Ký hiệu: $X \sim H(N; M; n)$ hay $X \in H(N; M; n)$.

1.3. Các đặc trưng số

ĐLNN X có phân phối siêu bội $X \sim H(N; M; n)$ thì kỳ vọng và phương sai được tính theo công thức

$$EX = n\cdot\frac{M}{N} \text{ và } DX = n\cdot\frac{M}{N}\cdot\frac{N-M}{N}\cdot\frac{N-n}{N-1} \tag{2.18}$$

Ví dụ 2.16: Một lô bóng đèn gồm 10 bóng, trong đó có 6 bóng loại A. Lấy ngẫu nhiên ra 4 bóng để kiểm tra. Gọi X là số bóng đèn loại A trong 4 bóng lấy ra.

a) Xác định luật phân phối xác suất của X.

b) Tính số bóng đèn loại A trung bình và phương sai của số bóng đèn loại A trong 4 bóng lấy ra.

c) Tính xác suất lấy được 3 bóng loại A trong 4 bóng đèn lấy ra.

Giải

a) $X \sim H(N=10; M=6; n=4)$.

b) Do X có phân phối siêu bội và theo công thức (2.18) ta có:

Số bóng đèn loại A trung bình là $EX = 4\cdot\frac{6}{10} = 2{,}4$.

Phương sai của số bóng đèn loại A là $DX = 4\cdot\frac{6}{10}\cdot\frac{4}{10}\cdot\frac{10-4}{10-1} = 0{,}64$.

c) $P[X=3]=\frac{C_6^3 C_4^1}{C_{10}^4}=\frac{8}{21}$. □

☞ *Tính xác suất lấy được ít nhất 1 bóng loại A trong 4 bóng đèn lấy ra.*

2. Phân phối nhị thức *B(n; p)*

2.1. Bài toán

Xét một lược đồ Bernoulli gồm

- n phép thử độc lập với nhau;
- Sự kiện A có xác suất xảy ra trong mỗi phép thử không đổi $P(A) = p$ ($P(\overline{A}) = 1 - p = q$).

Gọi X là số lần xảy ra sự kiện A trong n phép thử, khi đó X có **luật phân phối nhị thức với hai tham số *n; p***.

2.2. Định nghĩa

Phân phối nhị thức (Binomial distribution) là phân phối của ĐLNN rời rạc X nhận $n+1$ giá trị 0, 1, 2,..., n với xác suất tương ứng:

$$P[X = k] = C_n^k p^k q^{n-k}; \quad k = \overline{0,n},\ q = 1 - p \tag{2.19}$$

Ký hiệu: $X \sim B(n; p)$.

2.3. Các đặc trưng số

ĐLNN X có phân phối nhị thức $X \sim B(n; p)$ thì kỳ vọng, phương sai và giá trị tin chắc nhất của X được tính theo công thức

$$\begin{aligned} &EX = np; DX = npq \\ &(n+1)p - 1 \le \text{Mod}X \le (n+1)p \end{aligned} \tag{2.20}$$

Ví dụ 2.17: Xác suất sinh bé trai ở một vùng là 0,45. Khảo sát 50 bé sơ sinh của vùng. Gọi X là số bé trai trong 50 trẻ được khảo sát.

a) Xác định phân phối xác suất của X.
b) Tính số bé trai trung bình trong 50 bé sơ sinh.
c) Tính số bé trai có khả năng nhất trong 50 bé sơ sinh.
d) Tính xác suất để có từ 2 đến 3 bé trai trong 50 bé sơ sinh.

<u>Giải</u>

a) Ta có

+ Phép thử: khảo sát bé sơ sinh của vùng $\Rightarrow$ Số phép thử $n = 50$.

+ Gọi A là sự kiện sinh bé trai $\Rightarrow p = P(A) = 0,45;\ q = 1 - p = 0,55$.

Vậy $X \sim B(n = 50;\ p = 0,45)$.

b) Số bé trai trung bình là $EX = 50 \times 0,45 = 22,5$.

c) Số bé trai có khả năng tin chắc nhất là Mod$X$

Theo công thức (2.20), ta có

$$(50+1)0,45 - 1 \le \text{Mod}X \le (50+1)0,45 \text{ hay } 21,95 \le \text{Mod}X \le 22,95$$

Vậy Mod$X$ = 22.

d) $P(2 \le X \le 3) = C_{50}^{2}(0,45)^{2}(0,55)^{48} + C_{50}^{3}(0,45)^{3}(0,55)^{47} = 1,2 \times 10^{-9}$. □

☞ *Tính xác suất được ít nhất 48 bé trai trong 50 bé sơ sinh.*

2.4. Xấp xỉ phân phối siêu bội *H(N;M;n)* bằng phân phối nhị thức *B(n;p)*

Định lý 2.1: Cho $X \sim H(N; M; n)$, nếu N **rất lớn** và n **rất nhỏ so với** N thì phân phối siêu bội xấp xỉ phân phối nhị thức, nghĩa là $X \sim B(n; p)$ với $p = \frac{M}{N}$.

Khi đó $$P[X = k] = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \approx C_n^k p^k (1-p)^{n-k}; k = \overline{0,n} \quad (2.21)$$

☞ *Sự xấp xỉ này khá tốt khi* ***n < 0,05N***.

Ví dụ 2.18: Một lô bóng đèn gồm có 10 000 bóng, trong đó có 2 000 bóng đèn loại A. Lấy ngẫu nhiên 10 bóng ra kiểm tra. Gọi X là số bóng đèn loại A trong 10 bóng lấy ra.

a) Tìm luật phân phối xác suất của X.

b) Tính xác suất trong 10 bóng lấy ra có 3 bóng loại A.

Giải

a) Ta có $X \sim H(N = 10000; M = 2000; n = 10)$.

b) Do $N = 10000$ khá lớn và $n = 10$ rất nhỏ so với N nên theo định lý 2.1, $X \sim B(10; 0,2)$ với $p = \frac{2000}{10000} = 0,2$.

Vậy xác suất có 3 bóng loại A trong 10 bóng đèn được tính theo công thức Bernoulli: $P(10;3) \approx C_{10}^{3}(0,2)^{3}(0,8)^{7} = 0,2013$. □

Ví dụ 2.19: Lô hàng có 10 000 sản phẩm trong đó có 2 000 phế phẩm. Lấy ra 10 sản phẩm để kiểm tra theo hai phương thức:

+ Phương thức 1: Lấy đồng thời.

+ Phương thức 2: Lấy lần lượt có hoàn lại.

Gọi X là số phế phẩm có trong n sản phẩm lấy ra.

Nếu lấy theo phương thức 1 thì $X \sim H(10\,000; 2\,000; 10)$

Nếu lấy theo phương thức 2 thì $X \sim B(10; p)$ với $p = \frac{M}{N} = \frac{2000}{10000} = 0,2$. □

☞ *Khi số phần tử n lấy ra để xem xét rất nhỏ so với kích thước tập hợpN lớn thì sự khác biệt giữa phép thử có hoàn lại và không hoàn lại là không đáng kể. Khi đó, việc lấy mẫu không hoàn lại xấp xỉ với việc lấy mẫu có hoàn lại (phân phối siêu bội xấp xỉ với phân phối nhị thức).*

3. Phân phối Poisson $P(\lambda)$

3.1. Định nghĩa

ĐLNN rời rạc X nhận các giá trị 0, 1, 2, … với xác suất tương ứng

$$P[X=k] = e^{-\lambda}\frac{\lambda^k}{k!}; k=0,1,2,... \quad (2.22)$$

với tham số $\lambda > 0$, được gọi là có **luật phân phối Poisson với tham số λ.**

Ký hiệu: $X \sim P(\lambda)$.

Phân phối này đã có nhiều ứng dụng trong nhiều quá trình có liên quan đến số quan sát đối với một đơn vị thời gian hoặc không gian, chẳng hạn:

+ Số cuộc điện thoại nhận được ở một trạm điện thoại trong 1 phút
+ Số khách hàng đến ngân hàng với mỗi chu kỳ thời gian 30 phút
+ Số máy hỏng trong một ngày
+ Số lỗi in trong một trang sách,...

Nói chung là đầu vào của một hệ phục vụ như nhà hàng, siêu thị, trạm điện thoại, ... là các đại lượng ngẫu nhiên theo luật Poisson.

3.2. Các đặc trưng số

Nếu ĐLNN X có phân phối Poisson $X \sim P(\lambda)$ thì kỳ vọng, phương sai và giá trị tin chắc nhất của X được xác định như sau:

$$\begin{aligned} &EX = \lambda; DX = \lambda \\ &\lambda - 1 \le \text{Mod}X \le \lambda \end{aligned} \quad (2.23)$$

Ví dụ 2.20: Số khách vào một đại lý bưu điện là ngẫu nhiên và độc lập. Quan sát thấy trong 5 phút có 15 khách vào đại lý bưu điện đó. Hãy tính:

a) Xác suất có đúng 4 khách vào đại lý bưu điện đó trong 1 phút.

b) Số khách có khả năng nhất vào đại lý bưu điện đó trong 1 phút.

Giải

a) Gọi X là số người ghé vào đại lý bưu điện đó trong 1 phút. Ta có X là ĐLNN có phân phối Poisson $P(\lambda)$ với $\lambda = \frac{15}{5} = 3$.

Do đó, xác suất cần tìm $P[X=4] = \frac{e^{-3}3^4}{4!} = 0{,}168$.

b) $3-1 \le \text{Mod}X \le 3 \Rightarrow 2 \le \text{Mod}X \le 3 \Rightarrow \text{Mod}X = 2$ và $\text{Mod}X = 3$.□

☞ *Tính xác suất có đúng 16 người vào đại lý bưu điện đó trong 5 phút.*

3.3. Xấp xỉ phân phối nhị thức $B(n; p)$ bằng phân phối Poisson $P(\lambda)$

Định lý 2.2: Cho $X \sim B(n; p)$, nếu số phép thử n khá lớn và xác suất p rất bé thì phân phối nhị thức xấp xỉ phân phối Poisson, $X \sim P(\lambda)$ với $\lambda = np$. Khi đó,

$$P[X=k] = C_n^k p^k (1-p)^{n-k} \approx e^{-\lambda}\frac{\lambda^k}{k}; k=0,1,... \quad (2.24)$$

Ví dụ 2.21: Một tổng đài điện thoại có 5 000 máy con hoạt động độc lập. Trong thời gian một phút, xác suất để mỗi máy con liên lạc với tổng đài là 0,0006.

a) Tính xác suất trong một phút:

i) có 5 máy con liên lạc với tổng đài.

ii) có không quá 2 máy con liên lạc với tổng đài.

iii) có ít nhất một máy con liên lạc với tổng đài.

b) Cho biết số máy con trung bình liên lạc với tổng đài trong một phút.

Giải

Gọi X là số máy con liên lạc với tổng đài trong thời gian một phút.

Ta có $X \sim B(5\,000;\ 0{,}0006)$ và $n = 5\,000$ khá lớn; $p = 0{,}0006$ rất bé với $np = 3$.

Theo định lý 2.2 thì $X \sim P(\lambda = 3)$ và $P[X = k] \approx e^{-\lambda}\dfrac{\lambda^k}{k!}$

a) Do đó

i) $P[X = 5] \approx e^{-3}\dfrac{3^5}{5!} = 0{,}101.$

ii) $P[X \le 2] = P[0 \le X \le 2] \approx \sum_{k=0}^{2} e^{-3}\dfrac{3^k}{k!} = 0{,}423.$

iii) $P[X \ge 1] = 1 - P[X = 0] \approx 1 - e^{-3} = 1 - 0{,}05 = 0{,}95.$

b) Số máy con trung bình liên lạc với tổng đài trong một phút là $EX = np = 3$.

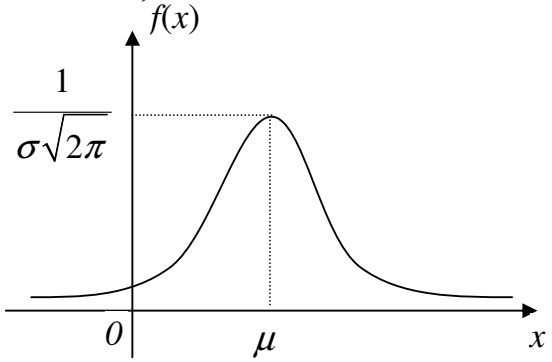
4. Phân phối chuẩn $N(\mu, \sigma^2)$

ĐLNN liên tục X, được gọi là có **luật phân phối chuẩn** (Normal distribution) với tham số $\mu \in \mathbb{R}$ và $\sigma > 0$ nếu X có hàm mật độ xác suất

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad x \in \mathbb{R} \tag{2.25}$$

Ký hiệu: $X \sim N(\mu; \sigma^2)$.

Đồ thị hàm mật độ $f(x)$ có dạng hình ‘‘chuông’’, đối xứng qua đường $x = \mu$ và đạt cực đại tại $\left(\mu; \dfrac{1}{\sigma\sqrt{2\pi}}\right)$.



Hình 2.12

Kỳ vọng và phương sai của $X \sim N(\mu;\sigma^2)$ là $EX = \mu$; $DX = \sigma^2$ (2.26)

Ví dụ 2.22: Cho *X ~ N(5; 9)*. Ta có:

Hàm mật độ xác suất của *X* là $f(x) = \frac{1}{3\sqrt{2\pi}} e^{\frac{-(x-5)^2}{18}}$.

Kỳ vọng *EX* = 5; phương sai *DX* = 9; độ lệch chuẩn σ_X = 3, giá trị tin chắc nhất Mod*X* = 5. □

Phân phối chuẩn chiếm vị trí quan trọng trong lý thuyết xác suất và là vị trí trung tâm trong các kết luận thống kê sau này.

Phân phối chuẩn có ý nghĩa rất lớn trong thực tế. Rất nhiều ĐLNN có luật phân phối chuẩn hoặc gần chuẩn. Những ĐLNN 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 nặng, chiều cao, mức độ thông minh,...) 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 các sản phẩm do một hệ thống máy sản xuất ra.

+ Điểm thi của các thí sinh.

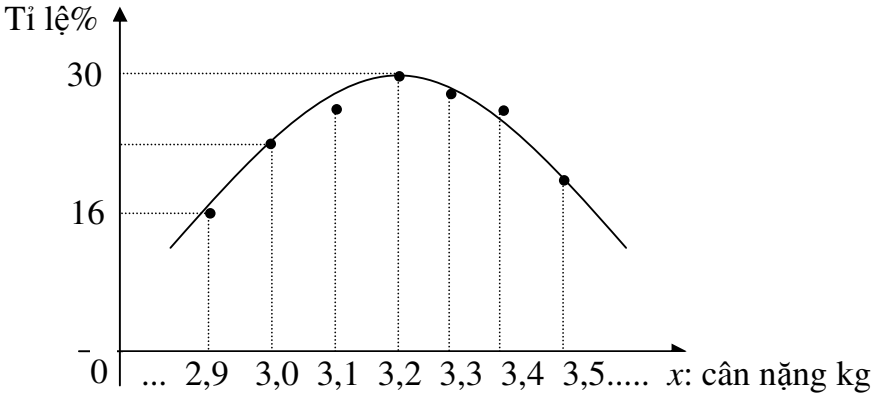
+ Lực chịu đựng của một thanh sắt; các sai số trong đo đạc; sai số quan sát; độ bền dẻo của máy móc; trung bình cộng của một số lớn các ĐLNN độc lập,…

Trong thương mại, kinh tế và khoa học xã hội nhiều phân phối không giống phân phối chuẩn, nhưng phân phối của trung bình cộng đối với mỗi trường hợp lại có thể xem như phân phối chuẩn khi *n* đủ lớn.

Ngoài ra, phân phối chuẩn còn là **tiệm cận** của phân phối nhị thức.

Ví dụ sau minh họa một ĐLNN trong thực tế có luật phân phối chuẩn.

Ví dụ 2.23: Gọi *X* là cân nặng của trẻ sơ sinh ở khu vực dân cư lớn. Khảo sát biến số *X* sau một khoảng thời gian ta có kết quả trong hình.



Hình 2.13

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 khoảng được chia mịn hơn, đường tần suất có dạng trơn dần, xấp xỉ một phần đường hình chuông của phân phối chuẩn với đỉnh tại $X = 3,2$ kg. Do đó có thể cho rằng trung bình cân nặng trẻ sơ sinh khu vực đó là $EX = 3,2$ kg.

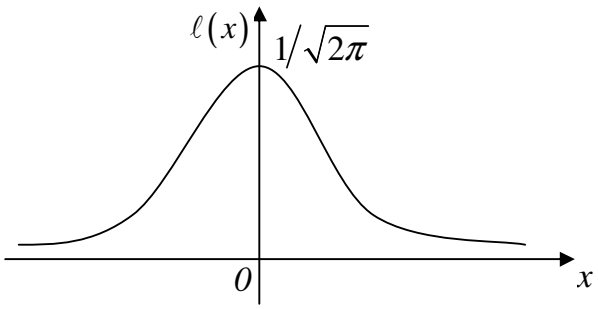
5. Phân phối chuẩn tắc *N(0; 1)*

Là trường hợp đặc biệt của phân phối chuẩn khi $\mu = 0$; $\sigma = 1$.

ĐLNN $U \sim N(0;1)$ được gọi là có **luật phân phối chuẩn tắc** (Standard Normal distribution).

Hàm mật độ xác suất của U có dạng $\ell(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$; $x \in \mathbb{R}$. (2.27)

Hàm $\ell(x)$ được gọi là **hàm Laplace,** là hàm chẵn $\ell(-x) = \ell(x)$, có đồ thị là đường cong hình chuông đối xứng qua trục tung. Các giá trị của hàm $\ell(x)$ được cho trong bảng **Phụ lục 1.**

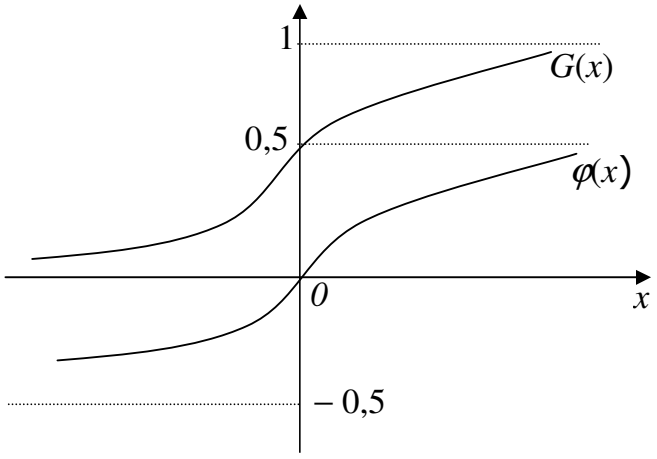


Hình 2.14

Dễ dàng thấy rằng $EU = 0$ và $DU = 1$.

Hàm phân phối xác suất $G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ (2.28)

còn được gọi là **hàm Gauss,** có đồ thị đối xứng qua điểm (0; 0,5).

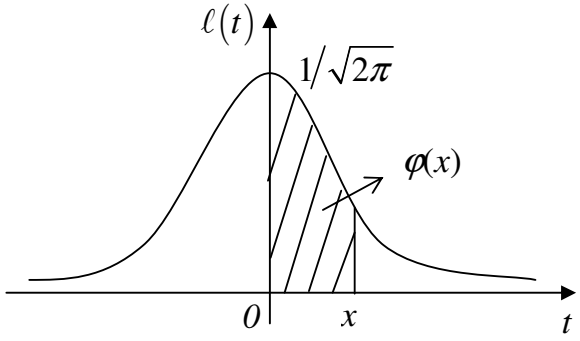


Hình 2.15

Đặt $\varphi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$ (2.29)

Rõ ràng, $G(x) = \varphi(x) + 0{,}5$ (2.30)

Do đó đồ thị hàm $\varphi(x)$ là đồ thị hàm $G(x)$ tịnh tiến xuống 0,5 đơn vị. Hàm $\varphi(x)$ được gọi là **tích phân Laplace.** Với $x > 0$, $\varphi(x)$ là diện tích hình thang cong giới hạn bởi $t = 0$; $t = x$



Hình 2.16

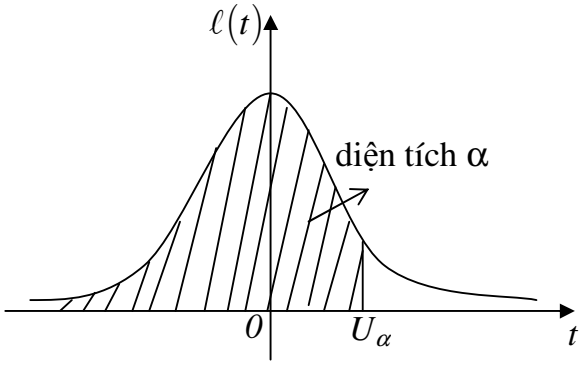
Hàm $\varphi(x)$ là hàm lẻ $\varphi(-x) = -\varphi(x)$. Các giá trị của hàm $\varphi(x)$ được cho trong bảng **Phụ lục 2**.

Về mặt lý thuyết ĐLNN $U \sim N(0;1)$ nhận giá trị trên toàn đường thẳng, song trên thực tế U nhận giá trị trong khoảng (–3;3) với xác suất 0,9973; nhận giá trị trong khoảng (–3,5; 3,5) với xác suất 0,9996.

❖ **Phân vị chuẩn tắc**

Phân vị chuẩn tắc mức xác suất α là số U_α sao cho $P[U < U_\alpha] = \alpha$. Tức là $G(U_\alpha) = \alpha$ hay $U_\alpha = G^{-1}(\alpha)$.

Hệ thức trên có thể viết lại $G(U_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{U_\alpha} e^{-t^2/2} dt = \alpha$.



Hình 2.17

Các giá trị phân vị chuẩn tắc U_α được cho trong bảng **Phụ lục 3.**

❖ **Liên hệ giữa phân phối chuẩn $N(\mu;\sigma^2)$ và phân phối chuẩn tắc $N(0;1)$**

Định lý 2.3: Nếu $X \sim N(\mu;\sigma^2)$ thì $U = \dfrac{X-\mu}{\sigma} \sim N(0;1)$

Hệ quả: Cho $X \sim N(\mu;\sigma^2)$, ta có

1) $P(x_1 < X < x_2) = \varphi\left(\dfrac{x_2-\mu}{\sigma}\right) - \varphi\left(\dfrac{x_1-\mu}{\sigma}\right)$

2) $P(X < x) = 0{,}5 + \varphi\left(\dfrac{x-\mu}{\sigma}\right)$ (2.31)

3) $P(X > x) = 0{,}5 - \varphi\left(\dfrac{x-\mu}{\sigma}\right)$

4) $P(|X-\mu| < \varepsilon) = 2\varphi\left(\dfrac{\varepsilon}{\sigma}\right)$

Trong đó, $P(|X-\mu| < 3\sigma) = 2\varphi(3) \approx 99{,}74\%$ (2.32)

Các công thức trên vẫn đúng khi thêm $X = x$.

Công thức (2.32) còn gọi là quy tắc 3σ, trên 99% giá trị của X nằm trong khoảng $(\mu - 3\sigma;\mu + 3\sigma)$. Trong thực hành, ĐLNN X chưa biết phân phối xác suất nhưng thỏa quy tắc 3σ thì có thể xem X có phân phối chuẩn.

Ví dụ 2.24: Độ dài sản phẩm do một hệ thống máy sản xuất ra là ĐLNN X có luật phân phối chuẩn với trung bình qui định 20 cm, độ lệch tiêu chuẩn 0,2 cm.

a) Tính tỉ lệ sản phẩm có độ dài trong khoảng (19,8;20,4) (cm).

b) Tính tỉ lệ sản phẩm tốt, có độ dài sai lệch so với trung bình qui định không quá 0,3 cm.

Giải

Tỉ lệ sản phẩm có độ dài trong khoảng (a; b) bằng xác suất để khi rút ngẫu nhiên một sản phẩm thì sản phẩm có độ dài trong khoảng (a; b).

Theo đề bài, $X \sim N(\mu;\sigma^2)$ với $\mu = 20$ và $\sigma = 0{,}2$. Ta có:

a) $P[19{,}8 < X < 20{,}4] = \varphi\left(\dfrac{20{,}4-20}{0{,}2}\right) - \varphi\left(\dfrac{19{,}8-20}{0{,}2}\right)$

$= \varphi(2) - \varphi(-1) = \varphi(2) + \varphi(1)$

$= 0{,}4772 + 0{,}3413 = 0{,}8185 = 81{,}85\%$

b) $P[|X-20| \le 0{,}3] = 2\varphi\left(\dfrac{0{,}3}{0{,}2}\right) = 2\varphi(1{,}5) = 2(0{,}4332) = 0{,}8664 = 86{,}64\%$ □

❖ ***Xấp xỉ của phân phối nhị thức B(n; p) với phân phối chuẩn*** $N(\mu;\sigma^2)$

Định lý 2.4: (Định lý giới hạn địa phương Moivre – Laplace)

Cho $X \sim B(n; p)$, nếu số phép thử n khá lớn và xác suất p không quá gần 0 và 1 ($0 << p << 1$) thì X xấp xỉ phân phối chuẩn $N(\mu;\sigma^2)$ trong đó $\mu = np$ và $\sigma^2 = npq$.

$$P[X=k] = C_n^k p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{npq}\sqrt{2\pi}} e^{-\frac{(k-np)^2}{2npq}},\ k = 0,1,... \qquad (2.33)$$

☞ *Sự xấp xỉ này tốt nhất là khi np và nq lớn hơn 5.*

Hệ quả: Cho $X \sim B(n; p)$ với n rất lớn, $0 << p << 1$, ta có:

$$P[X=k] \approx \frac{1}{\sqrt{npq}} \ell\left(\frac{k-np}{\sqrt{npq}}\right) \qquad (2.34)$$

trong đó $\ell(u)$ là hàm Laplace.

Ví dụ 2.25: Xác suất bắn trúng bia của một xạ thủ là 0,6. Tính xác suất sau:

a) Xạ thủ bắn 10 viên, trúng bia 6 viên.
b) Xạ thủ bắn 100 viên, trúng bia 60 viên.

Giải

Gọi X là số viên đạn bắn trúng bia.

a) $X \sim B(10; 0,6)$ suy ra $P[X = 6] = C_{10}^6 (0,6)^6 (0,4)^4 = 0,251$.

b) $X \sim B(100; 0,6)$, $n = 100$ khá lớn, $p = 0,6$ không quá gần 0 và 1, $np = 60$, $npq = 24$. Theo 2.34, $P[X=60] \approx \frac{1}{\sqrt{24}} \ell\left(\frac{60-60}{\sqrt{24}}\right) = \frac{1}{\sqrt{24}} \ell(0) = 0,08$.□

Định lý 2.5: (Định lý giới hạn trung tâm Moivre – Laplace)

Cho $X \sim B(n; p)$ với n rất lớn, $0 << p << 1$, ta có

$$P[k_1 \le X \le k_2] = \varphi\left(\frac{k_2-np}{\sqrt{npq}}\right) - \varphi\left(\frac{k_1-np}{\sqrt{npq}}\right) \qquad (2.35)$$

trong đó $\varphi(u)$ là tích phân Laplace.

Công thức (2.35) vẫn đúng khi bỏ $X = k_1$ hoặc $X = k_2$.

Ví dụ 2.26: Xác suất bắn trúng bia của một xạ thủ là 0,6. Xạ thủ bắn 100 viên. Tính các xác suất sau:

a) Xạ thủ bắn trúng bia từ 55 viên đến 70 viên.
b) Xạ thủ bắn trúng bia từ 65 viên đến 90 viên.

Giải

Ta có, $X \sim B(100; 0,6)$. Mà $n = 100$ khá lớn, $p = 0,6$ không quá gần 0, không quá gần 1, theo định lý 2.5, với $np = 60$, $npq = 24$

a) $P[55 \le X \le 70] = \varphi\left(\frac{70-60}{\sqrt{24}}\right) - \varphi\left(\frac{55-60}{\sqrt{24}}\right)$

$= \varphi(2,04) - \varphi(-1,02) = \varphi(2,04) + \varphi(1,02)$

$= 0,4793 + 0,3461 = 0,8254 = 82,54\%$.

b) $P[65 \le X \le 90] = \varphi\left(\frac{90-60}{\sqrt{24}}\right) - \varphi\left(\frac{65-60}{\sqrt{24}}\right) = \varphi(6,12) - \varphi(1,02) = 15,39\%$.□

☞ *Tính xác suất xạ thủ bắn trúng bia ít nhất 80 viên.*

6. Phân phối Khi bình phương χ^2

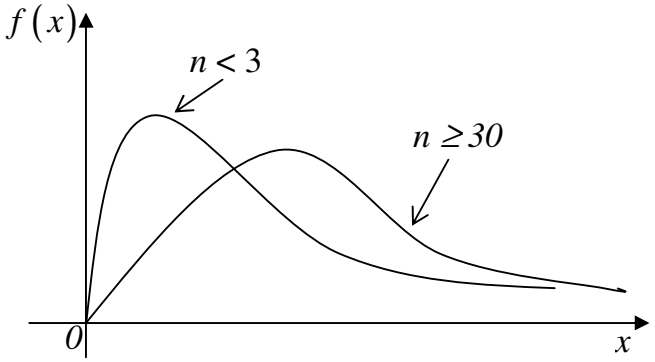
Cho các ĐLNN U_i, $i = \overline{1,n}$ độc lập với nhau và cùng có luật phân phối chuẩn tắc. ĐLNN $\chi^2 = \sum_{i=1}^{n} U_i^2$ được gọi là có luật phân phối **Khi bình phương *n* bậc tự do**. Ký hiệu: $\chi^2 \sim \chi^2(n)$.

Hàm mật độ xác suất của $\chi^2(n)$ bậc tự do là

$$f(x) = \begin{cases} 0 & khi \quad x \le 0 \\ \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}} x^{\frac{n}{2}-1} & khi \quad x > 0 \end{cases}$$

trong đó $\Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx$ là hàm Gamma thỏa $\Gamma(1) = 1$, $\Gamma(n+1) = n\Gamma(n)$.

ĐLNN χ^2 nhận giá trị không âm. Đồ thị hàm mật độ xác suất $f(x)$ là đường cong không đối xứng. Khi bậc tự do $n \ge 30$, đồ thị hàm mật độ gần đối xứng (dạng hình chuông), do đó, **phân phối χ^2 tiệm cận phân phối chuẩn**.



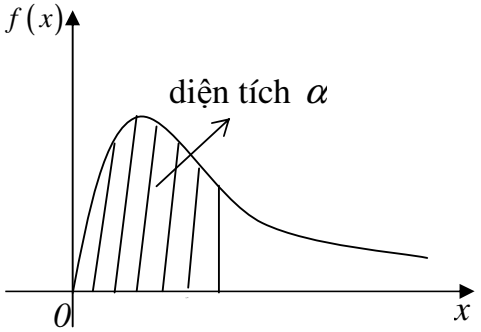
Hình 2.18

Các đặc trưng số của $\chi^2(n)$ là $\begin{cases} E\chi^2 = n \\ D\chi^2 = 2n \end{cases}$

Phân vị khi bình phương n bậc tự do với mức xác suất α là số $\chi^2_\alpha(n)$ sao cho

$$P[\chi^2 < \chi^2_\alpha(n)] = \alpha$$

Giá trị $\chi^2_\alpha(n)$ là vị trí cạnh phải sao cho diện tích của hình bằng α cho trước



Hình 2.19

Các giá trị phân vị $\chi^2_\alpha(n)$ được cho trong bảng **Phụ lục 4.**

Phân phối χ^2 do **Karl Pearson** đưa ra năm 1900.

7. Phân phối Student *T*

Cho các ĐLNN $U \sim N(0;1)$; $\chi^2 \sim \chi^2(n)$, trong đó U và χ^2 độc lập với nhau.

Đại lượng ngẫu nhiên $T = \dfrac{U}{\sqrt{\dfrac{\chi^2}{n}}} = \dfrac{U\sqrt{n}}{\sqrt{\chi^2}}$ được gọi là có **luật phân phối Student *n* bậc tự do**. Ký hiệu: **T ~ *T(n)***.

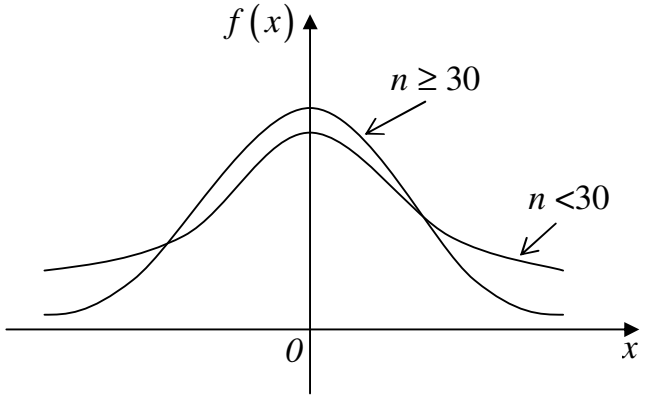
Hàm mật độ xác suất của T với n bậc tự do là

$$f(x) = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\dfrac{n}{2}\right)}\left(1+\frac{x^2}{2}\right)^{-\frac{n+1}{2}}$$

trong đó, $\Gamma(n)$ là hàm Gamma.

Đồ thị hàm mật độ $f(x)$ là đường cong đối xứng qua trục tung. Phân phối Student có cùng dạng và tính đối xứng như phân phối chuẩn nhưng nó phản ánh tính biến đổi của phân phối sâu sắc hơn. Chẳng hạn, các ĐLNN về giá và thời gian thường giới hạn một cách nghiêm ngặt kích thước mẫu, chính vì thế phân phối chuẩn không thể dùng để xấp xỉ phân phối của ĐLNN khi mẫu có kích thước nhỏ, trong trường hợp này ta dùng phân phối Student. Khi bậc tự do

$n \geq 30$, đồ thị hàm mật độ **tiệm cận** đồ thị hàm Laplace. Như vậy, khi $n \geq 30$, **phân phối Student xấp xỉ phân phối chuẩn tắc.**

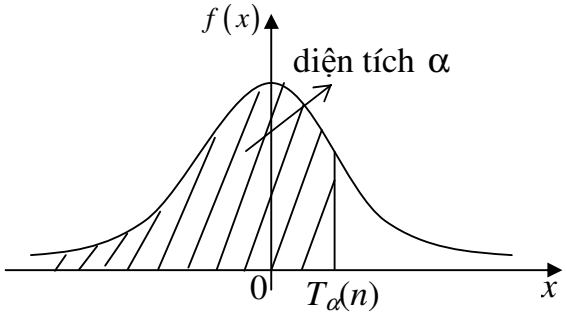


Hình 2.20

Các đặc trưng số của *T(n)* là $ET = 0$ và $DT = \frac{n}{n-2}$.

Phân vị Student *n* bậc tự do với mức xác suất α là số $T_\alpha(n)$ sao cho

$$P[T < T_\alpha(n)] = \alpha$$



Hình 2.21

Giá trị $T_\alpha(n)$ là vị trí ở cạnh phải sao cho diện tích bằng α cho trước. Các giá trị $T_\alpha(n)$ được cho trong bảng **Phụ lục 5**.

Theo **Lincohn L.Chao,** phân phối này do **William S.Gosset** đưa ra năm 1908, Gosset khi đó đang làm thuê cho **Guinness Brewery** ở **Dublin** và hãng này không cho phép Gosset dùng tên thật của mình để công bố các kết quả lý thuyết của ông. Vì thế ông đã viết dưới bài báo của mình một bút danh "Student" và tên phân phối Student bắt nguồn như vậy.

§4. ĐẠI LƯỢNG NGẪU NHIÊN HAI CHIỀU

Ở các phần trên ta đã xét các ĐLNN mà các giá trị có thể của chúng được biểu diễn bằng một số. Các ĐLNN như vậy được gọi là ĐLNN một chiều. Trong thực tế ta còn gặp trường hợp nhiều ĐLNN một chiều được xét đồng thời. Một bộ nhiều ĐLNN một chiều được xét như thế được gọi là một ĐLNN nhiều chiều (vectơ ngẫu nhiên). Như vậy các giá trị có thể có của ĐLNN nhiều chiều được xác định bằng 2, 3,..., *n* số; được gọi tương ứng là ĐLNN 2 chiều, 3 chiều,..., *n* chiều.

Để đơn giản ta chỉ xét các ĐLNN hai chiều.

1. Khái niệm về ĐLNN hai chiều (vectơ ngẫu nhiên hai chiều)

Một cặp hai ĐLNN (X,Y) được xét đồng thời được gọi là ĐLNN 2 chiều.

Người ta cũng chia ĐLNN (X,Y) thành hai loại: rời rạc nếu *X* và *Y* rời rạc; liên tục nếu *X* và *Y* liên tục.

Ví dụ 2.27: Một máy sản xuất một loại sản phẩm. Nếu kích thước của sản phẩm được đo bằng chiều dài *X* và chiều rộng *Y* thì ta có kích thước là ĐLNN hai chiều (X,Y). Đây là ĐLNN liên tục.

2. Luật phân phối xác suất của ĐLNN hai chiều

2.1. Trường hợp rời rạc

2.1.1. Bảng phân phối xác suất đồng thời (Mật độ đa biến)

Luật phân phối xác suất của ĐLNN 2 chiều rời rạc (X,Y) là bảng phân phối xác suất có dạng sau:

X \ Y	y_1	y_2	...	y_j	...	y_m
x_1	$p(x_1,y_1)$	$p(x_1,y_2)$	...	$p(x_1,y_j)$	...	$p(x_1,y_m)$
x_2	$p(x_2,y_1)$	$p(x_2,y_2)$	...	$p(x_2,y_j)$	...	$p(x_2,y_m)$
...	...	...	...	...	...	...
x_i	$p(x_i,y_1)$	$p(x_i,y_2)$	...	$p(x_i,y_j)$	...	$p(x_i,y_m)$
...	...	...	...	...	...	...
x_n	$p(x_n,y_1)$	$p(x_n,y_2)$	...	$p(x_n,y_j)$	...	$p(x_n,y_m)$

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*,

$$p(x_i, y_j) = \mathrm{P}\left(X = x_i, Y = y_j\right) \; ; i = \overline{1,n}, j = \overline{1,m} \tag{2.36}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) = 1 \, , \; p(x_i,y_j) \geq 0. \tag{2.37}$$

2.1.2. Bảng phân phối xác suất biên (Mật độ lề)

Biết được bảng phân phối xác suất đồng thời của ĐLNN 2 chiều bao giờ ta cũng tìm được bảng phân phối xác suất biên của mỗi thành phần bằng cách cộng dòng hay cộng cột của bảng.

Bảng phân phối xác suất biên của thành phần X có dạng:

X	x_1	x_2	...	x_i	...	x_n
P	$p(x_1)$	$p(x_2)$	...	$p(x_i)$	...	$p(x_n)$

trong đó $p(x_i) = \sum_{j=1}^{m} p(x_i, y_j)$, $\sum_{i=1}^{n} p(x_i) = 1$. (2.38)

Bảng phân phối xác suất biên của thành phần Y có dạng:

Y	y_1	y_2	...	y_j	...	y_m
P	$p(y_1)$	$p(y_2)$	...	$p(y_j)$	...	$p(y_m)$

trong đó $p(y_j) = \sum_{i=1}^{n} p(x_i, y_j)$, $\sum_{j=1}^{m} p(y_j) = 1$. (2.39)

Ví dụ 2.28: Tìm bảng phân phối xác suất biên của các thành phần của ĐLNN 2 chiều có bảng phân phối xác suất như sau:

X \ Y	0	1
-1	0,1	0,06
2	0,3	0,18
3	0,2	0,16

Giải

Cộng các xác suất theo hàng ta thu được các xác suất tương ứng với các giá trị của thành phần X. Cộng các xác suất theo cột ta thu được các xác suất tương ứng với các giá trị của thành phần Y. Các bảng phân phối xác suất biên cần tìm:

X	-1	2	3
P	0,16	0,48	0,36

Y	0	1
P	0,6	0,4

2.1.3. Bảng phân phối xác suất có điều kiện (Mật độ có điều kiện)

Cũng từ bảng phân phối xác suất đồng thời ta tìm được bảng phân phối xác suất có điều kiện của mỗi thành phần như sau:

Bảng phân phối xác suất có điều kiện của X với điều kiện $\{Y = y_j\}$

$X \mid Y = y_j$	x_1	x_2	...	x_i	...	x_n
P	$p(x_1 \mid y_j)$	$p(x_2 \mid y_j)$	...	$p(x_i \mid y_j)$	...	$p(x_n \mid y_j)$

trong đó: $p\left(x_i \middle| y_j\right) = \dfrac{p\left(x_i, y_j\right)}{p\left(y_j\right)}; \quad i = \overline{1,n}; j = \overline{1,m}$ (2.40)

Bảng phân phối xác suất có điều kiện của Y với điều kiện $\{X = x_i\}$

$Y \mid X = x_i$	y_1	y_2	...	y_j	...	y_m
P	$p(y_1 \mid x_i)$	$p(y_2 \mid x_i)$	...	$p(y_j \mid x_i)$	...	$p(y_m \mid x_i)$

trong đó: $p\left(y_j \middle| x_i\right) = \dfrac{p\left(x_i, y_j\right)}{p\left(x_i\right)}; \quad i = \overline{1,n}; j = \overline{1,m}$ (2.41)

Ví dụ 2.29: Phân phối xác suất có điều kiện của X với điều kiện $Y = 1$ trong ví dụ 2.28 là

$X \mid Y = 1$	-1	2	3
P	0,15	0,45	0,40

2.1.4. X và Y độc lập khi và chỉ khi $p\left(x_i, y_j\right) = p\left(x_i\right) p\left(y_j\right) \ \forall i, j$ (2.42)

2.2. Trường hợp liên tục

2.2.1. Hàm mật độ xác suất đồng thời

Hàm mật độ xác suất của ĐLNN 2 chiều (X,Y) là hàm $f(x,y)$ không âm, liên tục trên $\mathbb{R}^2$ và thỏa:

i) $f(x,y) \geq 0 \ \forall (x,y)$ ii) $\iint\limits_{\mathbb{R}^2} f(x,y)\,dxdy = 1$ (2.43)

Ta cũng có: $P(X \in A, Y \in B) = \int\limits_A dx \int\limits_B f(x,y)dy$ (2.44)

2.2.2. Hàm mật độ xác suất biên (mật độ lề)

Ta có $P(X \in A) = P(X \in A, Y \in \mathbb{R}) = \int\limits_A \left(\int\limits_{\mathbb{R}} f(x,y)dy \right) dx$ (2.45)

Đặt $f_X(x) = \int_{\mathbb{R}} f(x,y)dy$ thì $f_X(x)$ được gọi là hàm mật độ biên của *X*.

Tương tự, ta có hàm mật độ biên của *Y*: $f_Y(y) = \int_{\mathbb{R}} f(x,y)dx$ (2.46)

3. Hàm phân phối xác suất của ĐLNN hai chiều

Hàm phân phối xác suất của ĐLNN hai chiều (X,Y), ký hiệu *F(x,y)*, là hàm được xác định như sau: $F(x,y) = P(X < x, Y < y)$. (2.47)

Hàm mật độ xác suất của ĐLNN hai chiều liên tục (X,Y) còn là đạo hàm riêng hỗn hợp bậc hai của hàm phân phối xác suất:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} \quad (2.48)$$

Ví dụ 2.30: Tìm hàm mật độ xác suất của ĐLNN hai chiều liên tục (X,Y) nếu biết hàm phân phối xác suất của nó là

$$F(x,y) = \sin x.\sin y \quad (0 \le x \le \frac{\pi}{2}; 0 \le y \le \frac{\pi}{2})$$

Giải

Ta có: $\frac{\partial F(x,y)}{\partial x} = \cos x.\sin y$

Suy ra: $f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \cos x.\cos y$ với $0 \le x \le \frac{\pi}{2}; 0 \le y \le \frac{\pi}{2}$. □

4. Kỳ vọng và phương sai của ĐLNN hai chiều (X,Y)

4.1. (X,Y) rời rạc

$$+ EX = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i p(x_i,y_j);\ DX = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i^2 p(x_i,y_j) - (EX)^2 \quad (2.49)$$

$$+ EY = \sum_{j=1}^{m}\sum_{i=1}^{n} y_j p(x_i,y_j);\ DY = \sum_{j=1}^{m}\sum_{i=1}^{n} y_j^2 p(x_i,y_j) - (EY)^2 \quad (2.50)$$

$$+ E(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_i p(x_i,y_j) \quad (2.51)$$

4.2. (X,Y) liên tục

$$+ EX = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)dxdy;\ DX = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2 f(x,y)dxdy - (EX)^2 \quad (2.52)$$

$$+ EY = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)dxdy;\ DY = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^2 f(x,y)dxdy - (EY)^2 \quad (2.53)$$

$$+ E(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy \quad (2.54)$$

§5. ĐỊNH LÝ GIỚI HẠN TRUNG TÂM
LUẬT SỐ LỚN

1. Định lý giới hạn trung tâm

Định lý giới hạn trung tâm cho kết luận về phân phối của tổng các đại lượng ngẫu nhiên độc lập xấp xỉ phân phối chuẩn. Kết quả này đóng vai trò quan trọng trong thống kê khi xác định phân phối của mẫu.

Định lý 2.6: Nếu dãy $X_1,...,X_n,...$ các ĐLNN độc lập cùng phân phối với $EX_n = \mu,\ DX_n = \sigma^2, \forall n$ thì $S_n = \dfrac{\sum_{k=1}^{n} X - n\mu}{\sigma\sqrt{n}}$ xấp xỉ phân phối chuẩn tắc $N(0;1)$ khi n đủ lớn.

2. Luật số lớn

Lý thuyết xác suất nghiên cứu khả năng xuất hiện của các sự kiện ngẫu nhiên. Khi tung một đồng xu ta sẽ không biết trước mặt sấp hay mặt ngửa xuất hiện nhưng nếu tung nhiều lần thì ta thấy rằng số lần mặt sấp và mặt ngửa xuất hiện xấp xỉ bằng nhau. Như vậy khi thực hiện nhiều lần phép thử ngẫu nhiên ta sẽ tìm được quy luật xuất hiện của sự kiện ngẫu nhiên, đó là nội dung của luật số lớn. Đây cũng là cơ sở để định nghĩa xác suất của sự kiện thông qua tần suất của sự kiện đó.

2.1. Bất đẳng thức Chebyshev

2.1.1. Bất đẳng thức Markov

Giả sử X là ĐLNN nhận các giá trị không âm, ta có:

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon},\ \forall \varepsilon > 0 \qquad (2.55)$$

2.1.2. Bất đẳng thức Chebyshev

Cho X là ĐLNN có kỳ vọng, phương sai hữu hạn, ta có: $\forall \varepsilon > 0$

$$P(|X - EX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2} \qquad (2.56)$$

$$\text{hay } P(|X - EX| < \varepsilon) \geq 1 - \frac{DX}{\varepsilon^2} \qquad (2.57)$$

Ý nghĩa: Bất đẳng thức Markov và Chebyshev cho ta phương tiện xác định được giới hạn của xác suất.

Ví dụ 2.31: Giả sử số sản phẩm được sản xuất của một nhà máy trong một tuần là một ĐLNN với kỳ vọng $\mu = 50$.

a) Có thể kết luận gì về khả năng số sản phẩm của tuần này vượt quá 75.

b) Nếu phương sai của số sản phẩm trong tuần này là $\sigma^2 = 25$ thì có thể kết luận gì về khả năng số sản phẩm tuần này ở giữa 40 và 60.

Giải

Gọi X là số sản phẩm của một nhà máy trong một tuần.

a) Theo bất đẳng thức Markov: $P(X \geq 75) \leq \frac{E(X)}{75} = \frac{50}{75} = \frac{2}{3}$

b) Theo bất đẳng thức Chebyshev:

$$P(|X - 50| \geq 10) \leq \frac{DX}{\varepsilon^2} = \frac{25}{100} = \frac{1}{4}$$

Do đó:

$$P(40 < X < 60) = P(|X - 50| < 10) > 1 - \frac{1}{4} = \frac{3}{4}. \square$$

2.2. Luật số lớn

2.2.1. Luật số lớn cho lược đồ Bernoulli

Định lý 2.7 (Định lý Bernoulli): Xét một lược đồ Bernoulli gồm n phép thử độc lập và sự kiện A có xác suất xảy ra trong mỗi phép thử không đổi $p = P(A)$. Gọi m là tần số xuất hiện của sự kiện A trong n phép thử. Với $\varepsilon > 0$ tùy ý, ta có:

$$\lim_{n \to +\infty} P\left(\left| \frac{m}{n} - p \right| < \varepsilon \right) = 1$$

Định lý cho thấy rằng tần suất xuất hiện sự kiện A trong n phép thử độc lập dần về xác suất xuất hiện sự kiện A trong mỗi phép thử khi số phép thử tăng lên vô hạn. Đây chính là cơ sở lý thuyết để định nghĩa xác suất theo thống kê.

2.2.2 Luật số lớn tổng quát

Nếu dãy $X_1, X_2, ..., X_n, ...$ các ĐLNN độc lập có kỳ vọng hữu hạn và phương sai bị chặn đều thì dãy đó thỏa mãn luật số lớn, nghĩa là, với $\varepsilon > 0$ tùy ý, ta có:

$$\lim_{n \to +\infty} P\left[\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{n} \sum_{i=1}^{n} EX_i \right| < \varepsilon \right] = 1$$

Ý nghĩa: Mặc dù từng ĐLNN 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 ĐLNN 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 ĐLNN.

Định lý 2.8 (Định lý Markov): Các ĐLNN $X_1, X_2, ..., X_n$ có phương sai thỏa

$$\lim_{n \to +\infty} \frac{1}{n^2} D\left(\sum_{i=1}^{n} X_i \right) = 0$$

thì thỏa mãn luật số lớn.

Định lý 2.9 (Định lý Chebyshev): Cho các ĐLNN X_1, X_2,..., X_n độc lập từng đôi và có phương sai bị chặn $DX_i < +\infty$, $i = \overline{1,n}$ thì thỏa mãn luật số lớn.

Định lý 2.10: Các ĐLNN X_1, X_2,..., X_n cùng có luật phân phối xác suất như nhau và có phương sai hữu hạn thì thỏa mãn luật số lớn.

ÔN TẬP CUỐI CHƯƠNG

1.- Một thúng cam có 10 trái trong đó có 4 trái hư. Lấy ngẫu nhiên ra 3 trái. Gọi X là số trái cam hư trong 3 trái cam lấy ra. Phương án nào sau đây đúng?

A. $X \sim H(10; 4; 3)$. B. $X \sim B(3; 0,4)$.

C. $X \sim P(1,2)$. D. $X \sim N(1,2; 0,72)$.

2.- Một công ty may mặc có 150 máy hoạt động độc lập với nhau. Xác suất để một máy chạy an toàn trong một giờ làm việc là 85%. Xác suất để trong một giờ làm việc của nhà máy có ít nhất 120 máy chạy an toàn là

A. 0,0436. B. 0,4564. C. 0,9564. D. 1,000.

3.- Cho ĐLNN X có luật phân phối xác suất là

X	-1	1	2	3
P	0,3	0,3	0,1	0,3

Mod$X$ bằng

A. -1. B. 1. C. 3. D. -1, 1 và 3.

4.- Cho ĐLNN liên tục X có hàm mật độ

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-5)^2}{2}}, x \in \mathbb{R}$$

Phương án nào sau đây là **sai**?

A. $EX = 5$ B. $DX = 1$. C. $\sigma_X = \sqrt{2}$. D. Mod$X = 5$.

5.- Tiến hành khảo sát số khách trên một ôtô buýt tại một tuyến giao thông người ta thu được bảng số liệu sau

Số người trên 1 chuyến	25	30	35	40	45
Tỷ lệ	0,15	0,2	0,25	0,1	0,3

Số khách trung bình trên một chuyến xe là

A. 35. B. 36. C. 45. D. 34.

6.- Một xí nghiệp có 3 ô tô vận tải hoạt động. Xác suất trong ngày làm việc có ô tô bị hỏng tương ứng là 0,1; 0,15 và 0,2. Gọi X là số ô tô bị hỏng trong thời gian làm việc. Lập bảng phân phối xác suất, tính kỳ vọng và phương sai của X.

7.- Có 3 hộp, mỗi hộp đựng 10 sản phẩm. Số phế phẩm có trong mỗi hộp tương ứng là 1; 2 và 3.

a) Lấy ngẫu nhiên từ mỗi hộp ra một 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 đã chọn lấy ngẫu nhiên không hoàn lại ra 3 sản phẩm. Tìm luật phân phối xác suất của số phế phẩm có trong 3 sản phẩm lấy ra.

8.- Cho đại lượng ngẫu nhiên rời rạc X có bảng phân phối xác suất như sau:

X	-1	1	3	5
P	$2a$	a	$4a$	$3a$

a) Xác định a.

b) Xác định $P(X=4)$.

c) Xác định $P\left[(-1<X\le 2)\cup(X=3)\right]$.

9.- Trọng lượng của một con gà 6 tháng tuổi là một ĐLNN X (đơn vị: kg) có hàm mật độ

$$f(x)=\begin{cases} k\left(x^2-1\right), & x\in[1,3] \\ 0, & x\notin[1,3] \end{cases}$$, k là tham số

a) Tìm k. b) Tính $P(1<X<2)$.

10.- Năng suất của 3 máy tương ứng là các ĐLNN X, Y, Z (đơn vị tính là sản phẩm/phút). Cho biết quy luật phân phối xác suất của X, Y, Z như sau:

X	1	2	3	4
P	0,1	0,2	0,5	0,2

Y	2	3	4
P	0,4	0,3	0,3

Z	2	3	4	5
P	0,1	0,4	0,4	0,1

Giả sử bạn cần mua một trong 3 loại máy này thì bạn chọn mua loại máy nào? Vì sao? (Giả sử chất lượng và giá bán của 3 loại máy này là như nhau).

11.- Một người nuôi 100 con gà mái. Xác suất để một con gà bất kỳ đẻ trứng trong ngày là 0,6. Tính xác suất để có không ít hơn 60 con đẻ trứng trong một ngày.

12.- Tại một bến cảng, trung bình mỗi ngày có 5 tàu cập bến. Tính xác suất để trong một ngày mà ta xét có:

a) Không tàu nào cập bến. b) Từ 5 đến 7 tàu cập bến.

c) Đúng 5 tàu cập bến. d) Ít nhất 2 tàu cập bến.

13.- Theo tài liệu thống kê về tai nạn giao thông ở một khu vực thì người ta thấy tỷ lệ xe máy bị tai nạn là 0,0055 (vụ/tổng số xe/năm). Một công ty bảo hiểm đề nghị tất cả các chủ xe phải mua bảo hiểm xe máy với số tiền 30.000đ/xe và số tiền bảo hiểm trung bình cho một vụ tai nạn là 3.000.000đ. Hỏi lợi nhuận công ty kỳ vọng thu được đối với mỗi hợp đồng bảo hiểm là bao nhiêu biết rằng chi phí cho quản lý và các chi phí khác chiếm 30% số tiền bảo hiểm.

14.- Qua kinh nghiệm, một cửa hàng bán bánh trung thu biết được số bánh trung thu có thể bán được trong dịp tết trung thu là ĐLNN có qui luật phân phối xác suất như sau:

Số bánh bán được (X)	500	600	700	800	900	1000
Xác suất	0,05	0,15	0,45	0,30	0,03	0,02

a) Tìm số bánh bán được trung bình và độ lệch chuẩn của số bánh bán được.

b) Nếu cửa hàng đặt mua 700 chiếc thì xác suất bán hết bánh là bao nhiêu? Xác suất còn thừa bánh là bao nhiêu?

c) Để có thể chắc chắn đến 95% là sẽ đủ bánh bán thì cửa hàng cần đặt mua bao nhiêu bánh?

15.- Lãi suất thu được trong một năm (tính theo %) khi đầu tư vào công ty A, công ty B tương ứng là các ĐLNN X và Y (X, Y độc lập). Cho biết quy luật phân phối xác suất của X và Y như sau:

X	4	6	8	10	12
P	0,05	0,1	0,3	0,4	0,15

Y	–4	2	8	10	12	16
P	0,1	0,2	0,2	0,25	0,15	0,1

a) Đầu tư vào công ty nào có lãi suất kỳ vọng cao hơn.

b) Đầu tư vào công ty nào có mức độ rủi ro ít hơn? Vì sao?

c) Nếu muốn đầu tư vào cả hai công ty thì nên đầu tư theo tỉ lệ như thế nào để cho mức độ rủi ro về lãi suất thấp nhất?

16.- Một khối tín hiệu gồm 150 bit được dẫn theo một kênh truyền với xác suất bị lỗi của mỗi bit là 10^{-3}. Biết rằng khả năng bị lỗi của mỗi bit độc lập nhau. Tính xác suất để khối tín hiệu đó có ít nhất 3 bit bị lỗi.

17.- Xác suất trúng số là 1%. Mỗi tuần mua một vé số. Hỏi phải mua vé số liên tiếp trong tối thiểu bao nhiêu tuần để có không ít hơn 95% hy vọng trúng số ít nhất 1 lần?

18.- Giả sử trung bình một ngày Ngân hàng Nhà nước hủy 10 triệu tiền cũ, phát hành 11 triệu tiền mới. Đơn vị phát hành hủy tính chẵn bằng triệu. Việc phát hành và hủy tiền độc lập nhau. Tính xác suất để trong 1 ngày nào đó số tiền bị hủy và số tiền phát hành đều là 10 triệu.

19.- Trong một nước có 50% hy vọng trong năm sẽ không bị thiên tai. Tính xác suất trong năm nay sẽ có:

a) Đúng 2 thiên tai.

b) Ít nhất 2 thiên tai.

c) Trung bình trong 10 năm có bao nhiêu thiên tai.

20.- Nhà máy dệt muốn tuyển dụng người biết rành về một loại sợi. Có một người nói chỉ cần nhìn thoáng qua cũng có thể phân biệt "thật hay giả với xác

suất đúng là 80%". Nhà máy thử thách người này 7 lần. Mỗi lần nhà máy đem ra hai loại sợi trong đó có một loại sợi thật. Nếu nói đúng ít nhất 6 lần thì người này được tuyển dụng.

a) Nếu người này nói thật, tính xác suất được tuyển dụng.
b) Nếu người này không biết gì, tính xác suất được tuyển dụng.

21.- Đề thi trắc nghiệm gồm 6 câu hỏi, mỗi câu hỏi có 4 câu trả lời (trong đó chỉ có một câu trả lời đúng). Điều kiện thi đạt là trả lời đúng ít nhất 4 câu hỏi. Tính xác suất một thí sinh không học bài

a) Thi đạt.
b) Thi đạt nếu đã biết trả lời đúng 2 câu đầu.
c) Thi đạt nếu đã biết trả lời đúng ít nhất 2 câu.

22.- Trọng lượng của các gói mì ăn liền do một máy đóng bao sản xuất là biến số ngẫu nhiên theo qui luật phân phối chuẩn với kì vọng là 250gr và độ lệch chuẩn 5gr. Gói mì đạt tiêu chuẩn nếu có trọng lượng từ 245gr trở lên. Tính xác suất để máy sản xuất ra gói mì đạt tiêu chuẩn.

23.- Một máy dệt có 6000 ống sợi, xác suất để một ống sợi bị đứt trong một giờ máy làm việc là 0,0005. Gọi X là số ống sợi bị đứt trong một giờ máy làm việc. Tìm luật phân phối xác suất của X.

24.- Thống kê dân số của một vùng theo 2 chỉ tiêu: giới tính X và học vấn Y được kết quả cho trong bảng:

X \ Y	Thất học (0)	Phổ thông (1)	Đại học (2)
Nam (0)	0,10	0,25	0,16
Nữ (1)	0,15	0,22	0,12

a) Lập bảng phân phối xác suất của học vấn, của giới tính.
b) Học vấn có độc lập với giới tính không?
c) Tìm xác suất để chọn ngẫu nhiên một người thì được người không bị thất học.

25.- Người thợ chép tranh mỗi tuần chép hai bức tranh độc lập A và B với xác suất hỏng tương ứng là 0,03 và 0,05. Nếu thành công thì người thợ sẽ kiếm lời từ bức tranh A là 1,3 triệu đồng và B là 0,9 triệu đồng, nhưng nếu hỏng thì bị lỗ do bức tranh A là 0,8 triệu đồng và do B là 0,6 triệu đồng. Hỏi trung bình người thợ nhận được bao nhiêu tiền chép tranh mỗi tuần?

26.- Quan sát tại siêu thị A thấy trung bình 5 phút có 18 khách đến mua hàng.

a) Tính xác suất để trong 7 phút có 25 khách đến siêu thị A.
b) Tính xác suất để trong 2 phút có từ 3 đến 5 khách đến siêu thị A.
c) Tính số khách chắc chắn nhất sẽ đến siêu thị A trong 1 giờ.

27.- Tốc độ chuyển dữ liệu từ máy chủ của ký túc xá đến máy tính của sinh viên vào buổi sáng chủ nhật có phân phối chuẩn với trung bình 60Kbits/s và độ lệch chuẩn 4Kbits/s. Tính xác suất để tốc độ chuyển dữ liệu lớn hơn 63Kbits/s.