TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC
School of Applied Mathematics and Informatics

Xác suất thống kê

Giải bài tập đề cương

Nhóm ngành 1 MI2020

Nguyễn Quang Huy 20185454

Mục lục

Lời mở đầu

Xác suất thống kê là một lĩnh vực mà mình thấy rất thú vị và đặc biệt nhức não. Nhiều khi dù mình đọc lời giải rồi mà vẫn không hiểu người ta viết gì, biết mình ra kết quả sai mà không biết mình sai ở đâu ☺ Và bản thân mình là một người sợ, rất sợ môn khoa học của sự không chắc chắn này.

Thật trùng hợp là với mình thì đây là môn đại cương đầu tiên cô giáo kiểm tra và chấm điểm đề cương, và cũng là một học kì rất đặc biệt, khi mà tất cả mọi người đều làm việc ở nhà qua Internet. Chắc là nếu không có các điều kiện này, thì mình không bao giờ làm đề cương và có thể kiên nhẫn để gõ hết lại bài tập …

Trong quá trình hoàn thiện đề cương, có lúc mình bận quá, có lúc gặp biến cố trong học tập và công việc, có lúc lười học chán đời… nên không ít lần mình từng nghĩ sẽ bỏ dở. Nhưng cũng chính nhờ những kí ức không vui, mà mình đã nhận ra rằng cái gì đã khởi đầu tốt đẹp thì nên cố gắng hết sức để nó kết thúc thật mỹ mãn. Và mình đã quyết định hoàn thành những thứ mà mình đã bắt đầu vẫn còn đang dang dở, kết quả, chính là những trang mà bạn đang đọc đây.

Trong tài liệu này mình giải đủ các bài tập đề cương Xác suất thống kê năm 2020 nhóm ngành 1, mã học phần MI2020 các chương 1, 2, 3, 4 và 5. Tuy nhiên, còn nhiều chỗ do mình học chưa kỹ lắm, không ghi chép bài đầy đủ, chữa bài tập trên lớp… nên có thể sẽ có nhiều bài làm sai, nhiều bài làm không hay… Rất mong bạn đọc bỏ qua không ném đá ☺

Xin cảm ơn bạn Nguyễn Minh Hiếu, tác giả của template này đã chia sẻ và cho phép mình sử dụng mẫu LaTeX. Con nhà người ta nghĩ ra cái này cái kia còn mình chỉ đi xin về thôi ☺

Lời cuối cùng, mình muốn gửi lời cảm ơn chân thành và sâu sắc nhất tới cô Nguyễn Thị Thu Thủy, cô giáo dạy Xác suất thống kê của mình. Cô luôn nhiệt tình chỉ bảo, giúp đỡ em hoàn thiện tài liệu này và trong cả suốt quá trình học tập. Em xin cảm ơn cô vì đã dạy em, đã luôn tận tình hướng dẫn, giúp đỡ và quan tâm đến em. Thật may mắn khi em được tiếp xúc với cô. Học với cô, em có thêm nhiều động lực, và em học hỏi được rất rất nhiều từ phong cách làm việc chuyên nghiệp của cô. Một lần nữa, em cảm ơn cô nhiều lắm ạ. Kính chúc cô luôn sức khỏe và vui vẻ ạ.

Hà Nội, ngày 16 tháng 8 năm 2020

Nguyễn Quang Huy

1 Sự kiện ngẫu nhiên và phép tính xác suất

1.1 Quan hệ và phép toán của các sự kiện. Giải tích kết hợp

Bài tập 1.1.

Một hộp có 10 quả cầu cùng kích cỡ được đánh số từ 0 đến 9. Từ hộp người ta lấy ngẫu nhiên 1 quả ra và ghi lại số của quả đó, sau đó trả lại vào trong hộp. Làm như vậy 5 lần ta thu được một dãy số có 5 chữ số.

1. Có bao nhiêu kết quả cho dãy số đó?
2. Có bao nhiêu kết quả cho dãy số đó sao cho các chữ số trong đó là khác nhau?

1. Số kết quả cho dãy đó là 10^5
2. Số kết quả cho dãy có các chữ số khác nhau là $10.9.8.7.6 = 30240$

Bài tập 1.2.

Có 6 bạn Hoa, Trang, Vân, Anh, Thái, Trung ngồi quanh một bàn tròn để uống cà phê, trong đó bạn Trang và Vân không ngồi cạnh nhau.

1. Có bao nhiêu cách xếp 6 bạn này trên bàn tròn nếu tất cả các ghế là không phân biệt?
2. Có bao nhiêu cách xếp 6 bạn này trên bàn tròn nếu tất cả các ghế có phân biệt?

1. Số cách xếp để Trang và Vân không ngồi cạnh nhau là $5! - 2.4! = 72$
2. Số cách xếp nếu các ghế có phân biệt là $6! - 6.2.4! = 432$. Ta thấy rằng $432 = 6.72$

Bài tập 1.3.

Từ một bộ bài tú lơ khơ 52 cây rút ngẫu nhiên và không quan tâm đến thứ tự 4 cây. Có bao nhiêu khả năng xảy ra trường hợp trong 4 cây đó:

1. đều là át;
2. có duy nhất 1 cây át;
3. có ít nhất 1 cây át;
4. có đủ 4 loại rô, cơ, bích, nhép.

1. Chỉ có 1 khả năng do 1 bộ bài chỉ có 4 con át
2. Có 4 cách lấy ra 1 con át, có C_{48}^3 cách chọn 3 lá bài còn lại.
   Như vậy, số cách lấy ra 4 lá để có duy nhất 1 con át là
   $$4 \times C_{48}^3 = 69184$$
3. Số cách chọn ra 4 lá từ bộ bài là C_{52}^3. Số cách để chọn ra 4 lá bài trong đó không có cây át nào là C_{48}^3 (không lấy thứ tự)
   Suy ra số khả năng là $C_{52}^3 - C_{48}^3 = 76145$
4. Số cách lấy 1 lá bài cơ là $C_{13}^1 = 13$. Tương tự với các loại rô, bích, nhép. Suy ra số khả năng là $13^4 = 28561$

Bài tập 1.4.

Có 20 sinh viên. Có bao nhiêu cách chọn ra 4 sinh viên (không xét tới tính thứ tự) tham gia câu lạc bộ Văn và 4 sinh viên tham gia câu lạc bộ Toán trong trường hợp:

1. một sinh viên chỉ tham gia nhiều nhất một câu lạc bộ;
2. một sinh viên có thể tham gia cả hai câu lạc bộ.

1. Chọn 4 học sinh tham gia câu lạc bộ Văn có C_{20}^4 cách.
   Do 1 sinh viên không thể tham gia cùng lúc 2 câu lạc bộ, nên số cách chọn 4 sinh viên tham gia câu lạc bộ Toán là C_{16}^4. Số khả năng là
   $$C_{20}^4 \, C_{16}^4 = 8817900$$
2. Chọn 4 học sinh tham gia câu lạc bộ Văn có C_{20}^4 cách.
   Do 1 sinh viên có thể tham gia cùng lúc 2 câu lạc bộ, nên số cách chọn 4 sinh viên tham gia câu lạc bộ Toán là C_{20}^4. Số khả năng là
   $$C_{20}^4 \, C_{20}^4 = 23474025$$

Bài tập 1.5.

Cho phương trình $x + y + z = 100$. Phương trình đã cho có bao nhiêu nghiệm:

1. nguyên dương;
2. nguyên không âm.

1. Ta đánh dấu trên trục số từ số 1 đến 100 bởi 100 số 1 cách đều nhau 1 đơn vị. Khi đó, ta có 99 khoảng giữa 2 số 1 liên tiếp.

Nếu chia đoạn thẳng $[1, 100]$ này bởi 2 điểm chia nằm trong đoạn thì ta sẽ có 3 phần có độ dài ít nhất là 1.

Có thể thấy rằng ta có song ánh giữa bài toán chia đoạn này với bài toán tìm nghiệm nguyên dương của phương trình $x + y + z = 100$.

Như vậy, số nghiệm của phương trình này bằng số cách chia, và bằng $\binom{99}{2}$

2. Sử dụng ý trên. Đặt $a = x + 1, b = y + 1, c = z + 1$ thì $a, b, c \in \mathbb{Z}^+$ và

$$a + b + c = 103$$

Do đó số nghiệm x, y, z là $\binom{102}{2}$

Bài tập 1.6.

Thực hiện một phép thử tung 2 con xúc xắc, rồi ghi lại số chấm xuất hiện trên mỗi con. Gọi x, y là số chấm xuất hiện tương ứng trên con xúc xắc thứ nhất và thứ hai. Ký hiệu không gian mẫu $W = \{(x, y) \mid 1 \leq x, y \leq 6\}$. Hãy liệt kê các phần tử của các sự kiện sau:

1. A : "tổng số chấm xuất hiện lớn hơn 8";
2. B : "có ít nhất một con xúc xắc ra mặt 2 chấm";
3. C : "con xúc xắc thứ nhất có số chấm lớn hơn 4";
4. $A + B, A + C, B + C, A + B + C$, sau đó thể hiện thông qua sơ đồ $Venn$;
5. AB, AC, BC, ABC, sau đó thể hiện thông qua sơ đồ $Venn$.

1. $A = \{(3, 6), (4, 5), (5, 4), (6, 3), (4, 6), (5, 5), (6, 4), (5, 6), (6, 5), (6, 6)\}$
2. $B = \{(2, 2), (2, 1), (2, 3), (2, 4), (2, 5), (2, 6), (1, 2), (3, 2), (4, 2), (5, 2), (6, 2)\}$
3. $C = \{(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$
4. $A + B, A + C, B + C, A + B + C$
5. $AB = \varnothing$

   $AC = \{(5, 4), (5, 5), (5, 6), (6, 3), (6, 4), (6, 5), (6, 6)\}$

   $BC = \{(5, 2), (6, 2)\}$

   $ABC = \varnothing$

1.2 Định nghĩa xác suất

Bài tập 1.7.

Số lượng nhân viên của công ty A được phân loại theo lứa tuổi và giới tính như sau:

Tuổi \ Giới tính	Nam	Nữ
Dưới 30	120	170
Từ 30 đến 40	260	420
Trên 40	400	230

Tìm xác suất để lấy ngẫu nhiên một người của công ty thì được:

1. một nhân viên trong độ tuổi 30 – 40;
2. một nam nhân viên trên 40 tuổi;
3. một nữ nhân viên từ 40 tuổi trở xuống.

1. Gọi A là "lấy được một nhân viên trong độ tuổi $30 - 40$"
$$P(A) = \frac{260 + 420}{120 + 260 + 400 + 170 + 420 + 230} = \frac{17}{40} = 0.425$$
2. Gọi B là "lấy được nam nhân viên trên 40 tuổi"
$$P(B) = \frac{400}{120 + 260 + 400 + 170 + 420 + 230} = 0.25$$
3. Gọi C là "lấy được nữ nhân viên từ 40 tuổi trở xuống"
$$P(C) = \frac{170 + 420}{120 + 260 + 400 + 170 + 420 + 230} \simeq 0.3688$$

Bài tập 1.8.

Một kiện hàng có 24 sản phẩm, trong số đó có 14 sản phẩm loại I, 8 sản phẩm loại II và 2 sản phẩm loại III. Người ta chọn ngẫu nhiên 4 sản phẩm để kiểm tra. Tính xác suất trong 4 sản phẩm đó:

1. có 3 sản phẩm loại I và 1 sản phẩm loại II;
2. có ít nhất 3 sản phẩm loại I;
3. có ít nhất 1 sản phẩm loại III.

Ta tính xác suất theo định nghĩa cổ điển. Số trường hợp đồng khả năng là C_{24}^4.

1. Số cách lấy 3 sản phẩm loại I là C_{14}^3. Số cách lấy 1 sản phẩm loại II là C_8^1. Số kết cục thuận lợi là $C_{14}^3\,C_8^1$. Suy ra
$$P(A) = \frac{C_{14}^3\,C_8^1}{C_{24}^4} \simeq 0.2740$$

2. Để trong 4 sản phẩm chọn ra có ít nhất 3 sản phẩm loại I, chỉ có 2 khả năng là cả 4 đều loại I, hoặc 3 loại I, 1 loại II, hoặc loại III. Dễ dàng tính được
$$P(B) = \frac{C_{14}^4 + C_{14}^3\,C_{10}^1}{C_{24}^4} \simeq 0.4368$$

3. Ta tính xác suất trong 4 sản phẩm không có sản phẩm loại III: $P(\overline{C}) = \dfrac{C_{22}^4}{C_{24}^4} \simeq 0.6884.$

   Do đó, ta có $P(C) = 1 - P(\overline{C}) \simeq 0.3116$

Bài tập 1.9.

Có 30 tấm thẻ đánh số từ 1 tới 30. Chọn ngẫu nhiên ra 10 tấm thẻ. Tính xác suất để:

1. tất cả tấm thẻ đều mang số chẵn;
2. có đúng 5 số chia hết cho 3;
3. có 5 tấm thẻ mang số lẻ, 5 tấm thẻ mang số chẵn trong đó chỉ có một số chia hết cho 10.

Sử dụng công thức xác suất cổ điển. Số kết cục đồng khả năng khi chọn 10 tấm thẻ là $n = C_{30}^{10}$

1. Gọi A là "tất cả thẻ đều mang số chẵn" thì số kết cục thuận lợi cho A là $m = C_{15}^{10}$.
   Có $P(A) = \dfrac{C_{15}^{10}}{C_{30}^{10}} \simeq 9.995 \times 10^{-5}$

2. Gọi B là "có đúng 5 số chia hết cho 3". Có $P(B) = \dfrac{C_{10}^5\,C_{20}^5}{C_{30}^{10}} \simeq 0.13$

3. Gọi C là sự kiện cần tính xác suất.
   Dễ tính được số kết cục thuận lợi cho C là $C_3^1\,C_{12}^4\,C_{15}^5$. Suy ra
$$P(C) = \frac{C_3^1\,C_{12}^4\,C_{15}^5}{C_{30}^{10}} \simeq 0.1484$$

Bài tập 1.10.

Việt Nam có 64 tỉnh thành, mỗi tỉnh thành có 2 đại biểu quốc hội. Người ta chọn ngẫu nhiên 64 đại biểu quốc hội để thành lập một ủy ban. Tính xác suất để:

1. trong ủy ban có ít nhất một người của thành phố Hà Nội;
2. mỗi tỉnh có đúng một đại biểu trong ủy ban.

1. Gọi A là "có ít nhất 1 người từ Hà Nội". Ta có
$$P(A) = 1 - P(\overline{A}) = 1 - \frac{C_{126}^{64}}{C_{128}^{64}} \simeq 0.7520$$

2. Gọi B là "mỗi tỉnh có một đại diện" ta có $P(B) = \dfrac{2^{64}}{C_{128}^{64}} \approx 7.5 \times 10^{-19}$

Bài tập 1.11.

Một đoàn tàu có 4 toa được đánh số I, II, III, IV đỗ ở sân ga. Có 6 hành khách từ sân ga lên tàu. Mỗi người độc lập với nhau chọn ngẫu nhiên một toa. Tính xác suất để:

1. toa I có 3 người, toa II có 2 người và toa III có 1 người;
2. một toa có 3 người, một toa 2 người, một toa có 1 người;
3. mỗi toa có ít nhất 1 người.

1. Lần lượt chọn 3 người xếp vào toa đầu, 2 người xếp vào toa II và 1 người xếp vào toa III, ta có
$$P(A) = \frac{C_6^3\ C_3^2\ C_1^1}{4^6} = \frac{15}{1024} \simeq 0.0146$$
2. Có chọn ra 3 người xếp vào một toa, rồi chọn ra 2 người xếp vào một toa khác, cuối cùng cho người còn lại vào một toa. Ta có
$$P(B) = \frac{C_6^3 \times 4 \times C_3^2 \times 3 \times C_1^1 \times 2}{4^6} = \frac{45}{128} \simeq 0.3516$$
3. Gọi C "mỗi toa có ít nhất một người", khi đó chỉ có thể xảy ra 2 khả năng.
Khả năng thứ nhất là có 1 toa 3 người, 3 toa còn lại 1 người.
Khả năng thứ 2 là có 2 toa 2 người và 2 toa 1 người. Theo công thức cổ điển ta có
$$P(C) = \frac{C_6^3 \times 4 \times 3! + C_4^2 \times C_6^2 \times C_4^2 \times 2!}{4^6} = \frac{195}{512} \simeq 0.3809$$

Bài tập 1.12.

Gieo hai con xúc xắc cân đối và đồng chất. Một con xúc xắc có số chấm các mặt là 1, 2, 3, 4, 5, 6, con xúc xắc còn lại có số chấm các mặt là 2, 3, 4, 5, 6, 6. Tính xác suất:

1. có đúng 1 con xúc xắc ra mặt 6 chấm;
2. có ít nhất 1 con xúc xắc ra mặt 6 chấm;
3. tổng số chấm xuất hiện bằng 7.

Số kết cục đồng khả năng là $6.6 = 36$

1. $P(A) = \dfrac{1.4 + 5.2}{36} \simeq 0.3889$

2. $P(B) = 1 - \dfrac{5.4}{36} \simeq 0.4444$

3. Để số chấm xuất hiện tổng bằng 7 thì tập kết cục thuận lợi phải là

$$\{(1,6),(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

suy ra $m = 7$. Do đó ta có $P(C) = \dfrac{7}{36} \simeq 0.1944$

Bài tập 1.13.

Trong một thành phố có 5 khách sạn. Có 3 khách du lịch đến thành phố đó, mỗi người chọn ngẫu nhiên một khách sạn. Tìm xác suất để:

1. mỗi người ở một khách sạn khác nhau;

2. có đúng 2 người ở cùng một khách sạn.

Mỗi người có 5 cách chọn khách sạn để ở. Do đó số trường hợp đồng khả năng có thể xảy ra là 5^3

1. Gọi A là "mỗi người ở một khách sạn khác nhau".

   Số kết cục thuận lợi cho A là $5.4.3 = 60$. Từ đó có $P(A) = \dfrac{60}{5^3} = 0.48$

2. Gọi B là "có đúng 2 người ở cùng một khách sạn".
   Có C_3^2 cách để chọn ra 2 người. Có 5 cách để họ chọn khách sạn. Người còn lại ở một trong số 4 cái còn lại. Số kết cục thuận lợi cho B, theo quy tắc nhân, là $C_3^2 \times 5 \times 4$.
   Suy ra $P(B) = \dfrac{C_3^2 \times 5 \times 4}{5^3} = 0.48$

Bài tập 1.14.

Một lớp có 3 tổ sinh viên: tổ I có 12 người, tổ II có 10 người và tổ III có 15 người. Chọn hú họa ra một nhóm sinh viên gồm 4 người.

1. Tính xác suất để trong nhóm có đúng một sinh viên tổ I.

2. Biết trong nhóm có đúng một sinh viên tổ I, tính xác suất để trong nhóm đó có đúng một sinh viên tổ III.

1. Gọi A là "trong nhóm có đúng 1 sinh viên tổ I". Ta có

$$P(A) = \frac{C_{12}^1\, C_{25}^3}{C_{37}^4} = \frac{1840}{4403} \simeq 0.4179$$

2. Gọi B "có đúng 1 sinh viên tổ III". Theo định nghĩa xác suất điều kiện,

$$P(B \mid A) = \frac{P(AB)}{P(A)} = \frac{\dfrac{C_{12}^1\, C_{10}^2\, C_{15}^1}{C_{37}^4}}{\dfrac{1840}{4403}} = \frac{27}{92} \simeq 0.2935$$

Nếu ta tính trực tiếp không qua công thức xác suất điều kiện, thì với giả thiết biết có đúng 1 sinh viên tổ I, số trường hợp đồng khả năng là C_{25}^3.
Số kết cục thuận lợi là $C_{10}^2\, C_{15}^1$, suy ra $P = \dfrac{C_{10}^2\, C_{15}^1}{C_{25}^3} = \dfrac{27}{92}$

Bài tập 1.15.

Ba nữ nhân viên phục vụ A, B và C thay nhau rửa đĩa chén và giả sử ba người này đều “khéo léo” như nhau. Trong một tháng có 4 chén bị vỡ. Tìm xác suất để:

1. chị A đánh vỡ 3 chén và chị B đánh vỡ 1 chén;
2. một trong ba người đánh vỡ 3 chén;
3. một trong ba người đánh vỡ cả 4 chén.

Số kết cục đồng khả năng là 3^4

1. $P(A) = \dfrac{C_4^3\, C_1^1}{3^4} \simeq 0.0494$
2. Chọn một người đánh vỡ 3 chén, và một trong 2 người còn lại đánh vỡ 1 chén.

   Suy ra $P(B) = \dfrac{C_3^1\, C_4^3\, C_2^1\, 1}{3^4} \simeq 0.2963$
3. $P(C) = \dfrac{C_3^1\, 1}{3^4} \simeq 0.0370$

Bài tập 1.16.

Đội A có 3 người và đội B có 3 người tham gia vào một cuộc chạy thi, 6 người có khả năng như nhau và xuất phát cùng nhau. Tính xác suất để 3 người đội A về vị trí nhất, nhì, ba.

Vì chỉ có 3 giải nhất, nhì, ba và mỗi giải chỉ có thể trao cho 1 trong 6 người, nên số kết cục đồng khả năng là $A_6^3 = 20$.
Mặt khác, với mỗi cách trao giải cho 3 người đội A, ta có một hoán vị của "nhất, nhì, ba" nên số kết cục thuận lợi là 3!.
Tóm lại, xác suất cần tính $P = \dfrac{3!}{A_6^3} = 0.05$

Bài tập 1.17.

Phân phối ngẫu nhiên n viên bi vào n chiếc hộp (biết rằng mỗi hộp có thể chứa cả n viên bi). Tính xác suất để:

1. Hộp nào cũng có bi;
2. Có đúng một hộp không có bi.

Số kết cục thuận lợi là n^n

1. Gọi A là "hộp nào cũng có bi". Khi đó, số kết cục thuận lợi là $n!$. Vậy $P(A) = \dfrac{n!}{n^n}$
2. Gọi B là "Có đúng một hộp không có bi". Khi đó, có một hộp có 2 bi, $n-2$ hộp chứa 1 bi và 1 hộp chứa 0 bi.
   Chọn 2 trong n hộp để bi có C_n^2 cách. Chọn 2 trong n bi có C_n^2 cách chọn.
   Xếp 2 bi này vào một trong 2 hộp, có $2!$ cách xếp. Xếp số bi còn lại vào các hộp có $(n-2)!$ cách xếp. Suy ra số kết cục thuận lợi là
   $$2!\, C_n^2\, C_n^2\, (n-2)!$$
   Như vậy
   $$P(B) = \frac{2!\, C_n^2\, C_n^2\, (n-2)!}{n^n} = \frac{(n!)^2}{2\,(n-2)!\, n^n}$$

Bài tập 1.18.

Hai người hẹn gặp nhau ở công viên trong khoảng thời gian từ 5h00 đến 6h00 để cùng đi tập thể dục. Hai người quy ước ai đến không thấy người kia sẽ chỉ chờ trong vòng 10 phút. Giả sử rằng thời điểm hai người đến công viên là ngẫu nhiên trong khoảng từ 5h00 đến 6h00. Tính xác suất để hai người gặp nhau.

Gọi x, y là thời gian người thứ nhất và người thứ hai đến. Ta có tập kết cục đồng khả năng là

$$G = \left\{(x, y) \in \mathbb{R}^2 \mid 0 \leq x, y \leq 60\right\}$$

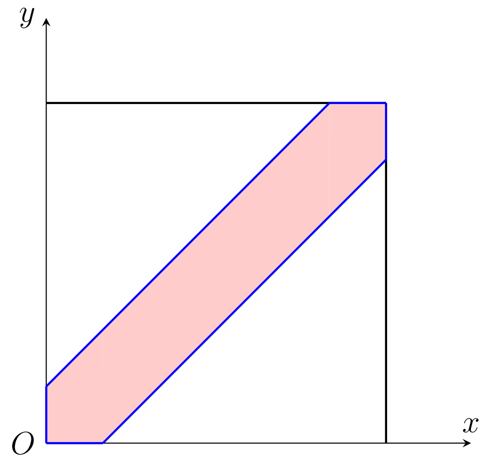
Gọi H "hai người gặp được nhau". Khi đó tập kết cục thuận lợi là

$$H = \{(x, y) \in G\colon\ |x-y| \leq 10\}$$

Suy ra $P = \dfrac{|H|}{|G|} = \dfrac{60^2 - 50^2}{50^2} = \dfrac{11}{36} \simeq 0.3056$

Bài tập 1.19.

Cho đoạn thẳng AB có độ dài 10 cm. Lấy một điểm C bất kỳ trên đoạn thẳng đó. Tính xác suất chênh lệch độ dài giữa hai đoạn thẳng AC và CB không vượt quá 4cm.



Gọi x là độ dài AC, hiển nhiên $CB = 10 - x$. Số kết cục đồng khả năng ở đây là độ dài đoạn thẳng AB, chính là 10 cm.
Gọi A là "chênh lệch độ dài giữa AC và CB không quá 4 cm", khi đó, A biểu thị bởi miền hình học

$$H = \left\{x \in [0, 10] \text{ mà } |x - (10 - x)| \leq 4\right\}$$

Vì H là đoạn thẳng có độ dài $7 - 3 = 4$ (cm) nên ta dễ dàng tính $P(A)$ theo định nghĩa hình học: $P(A) = \dfrac{4}{10} = 0.4$

Bài tập 1.20.

Cho đoạn thẳng AB độ dài 10 cm. Lấy hai điểm C, D bất kỳ trên đoạn AB (C nằm giữa A và D). Tính xác suất độ dài AC, CD, DB tạo thành 3 cạnh một tam giác.

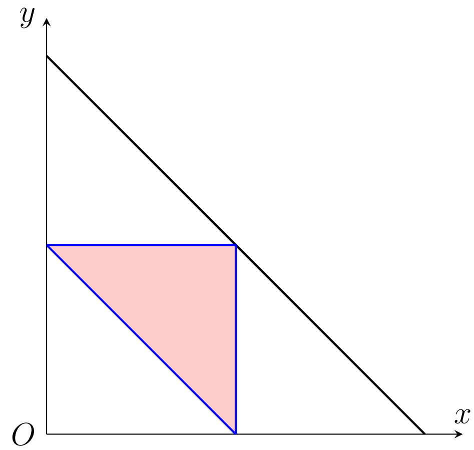
Gọi x, y lần lượt là độ dài các đoạn thẳng AC, CD.
Khi đó ta có $DB = 10 - x - y$, với điều kiện $x \geq 0, y \geq 0, 10 - x - y \geq 0$.
Miền đồng khả năng là

$$G = \left\{(x, y) \in \mathbb{R}^2 \mid x \geq 0, y \geq 0, 10 - x - y \geq 0\right\}$$

Gọi A là "độ dài AC, CD, DB tạo thành 3 cạnh tam giác" thì miền kết cục thuận lợi cho A là

$$H = \{(x, y) \in G \mid x + y > 10 - x - y,\ x + (10 - x - y) > y,\ y + (10 - x - y) > x\}$$

Như vậy, xác suất của sự kiện A là $P(A) = \dfrac{|H|}{|G|} = \dfrac{1}{4} = 0.25$



1.3 Xác suất điều kiện. Công thức cộng, nhân xác suất. Công thức Bernoulli

Bài tập 1.21.

Cho các sự kiện A, B với $P(A) = P(B) = \dfrac{1}{2}; P(A\overline{B}) = \dfrac{1}{8}$. Tìm:

1. $P(\overline{A} + \overline{B})$;
2. $P(\overline{A}B), P(A + \overline{B})$.

1. $P(\overline{A} + \overline{B}) = 1 - P(AB) = 1 - P(A) + P(A\overline{B}) = 0.625$
2. $P(\overline{A}B) = P(B) - P(AB) = P(B) - P(A) + P(A\overline{B}) = 0.125$
   và $P(A + \overline{B}) = 1 - P(\overline{A}B) = 0.875$

Bài tập 1.22.

Cho ba sự kiện A, B, C độc lập từng đôi thỏa mãn $P(A) = P(B) = P(C) = p$ và $P(ABC) = 0$.

1. Tính $P(AB\overline{C}); P(A\overline{B}\,\overline{C}); P(\overline{A}\,\overline{B}\,\overline{C})$.
2. Tìm giá trị p lớn nhất có thể có.

1. $P(AB\overline{C}) = P(AB) - P(ABC) = p^2$
   $P(A\overline{B}\,\overline{C}) = P(A\overline{B}) - P(A\overline{B}C) = p(1-p) - p^2 = p - 2p^2$
   Chú ý rằng vì A, B, C có vai trò như nhau nên $P(AB\overline{C}) = P(A\overline{B}C)$
   Suy ra $P(\overline{A}\,\overline{B}\,\overline{C}) = P(\overline{B}\,\overline{C}) - P(A\overline{B}\,\overline{C}) = (1-p)^2 - p + 2p^2 = 3p^2 - 3p + 1$

2. Các xác suất có thể có là

$$\begin{cases} P(A)=P(B)=P(C)=p \\ P(\overline{A})=P(\overline{B})=P(\overline{C})=1-p \\ P(AB)=P(BC)=P(CA)=p^2 \\ P(A\overline{B})=P(B\overline{A})=P(B\overline{C})=P(C\overline{B})=P(C\overline{A})=P(A\overline{C})=p(1-p) \\ P(\overline{A}\,\overline{B})=P(\overline{B}\,\overline{C})=P(\overline{C}\,\overline{A}))=1-2p+p^2 \\ P(ABC)=0 \\ P(AB\overline{C})=P(BC\overline{A})=P(CA\overline{B})=p^2 \\ P(A\overline{B}\,\overline{C})=P(B\overline{C}\,\overline{A})=P(C\overline{A}\,\overline{B})=p-2p^2 \\ P(\overline{A}\,\overline{B}\,\overline{C})=3p^2-3p+1 \end{cases}$$

Ta có $0 \le p,\ 1-p,\ p^2,\ p-p^2,\ p-2p^2,\ (1-p)^2,\ 3p^2-3p+1 \le 1$ suy ra $p \le \dfrac{1}{2}$

Bài tập 1.23.

Trong cùng một phép thử, A và B là các sự kiện thỏa mãn $P(A)=\dfrac{1}{4}, P(B)=\dfrac{1}{2}$. Tính xác suất để A không xảy ra nhưng B xảy ra trong các trường hợp sau:

1. A và B xung khắc;
2. A suy ra B;
3. $P(AB)=\dfrac{1}{8}$.

1. A và B xung khắc thì $\overline{A}\,B=B$ suy ra $P(B)=0.5$
2. A suy ra B thì $\overline{A}\,B=B\setminus A$ suy ra $P(\overline{A}\,B)=P(B)-P(AB)=P(B)-P(A)=0.25$
3. $P(\overline{A}\,B)=P(B)-P(AB)=0.375$

Bài tập 1.24.

Cho hai sự kiện A và B trong đó $P(A)=0,4$ và $P(B)=0,7$. Xác định giá trị lớn nhất và nhỏ nhất của $P(AB)$ và $P(A+B)$ và điều kiện đạt được các giá trị đó.

Có $0.7 \le P(A+B) \le 1$ vì $P(A)=0.4, P(B)=0.7$.
Dấu bằng đạt được lần lượt tại $A \subset B$ và $P(AB)=0.1$
Suy ra $0.1 \le P(AB) \le 0.4$. Dấu bằng đạt được lần lượt khi $P(A+B)$ đạt max và min

Bài tập 1.25.

Ba người A, B và C lần lượt tung một đồng xu. Giả sử rằng A tung đồng xu đầu tiên, B tung thứ hai và thứ ba C tung. Quá trình lặp đi lặp lại cho đến khi ai thắng bằng việc trở thành người đầu tiên thu được mặt ngửa. Xác định khả năng mà mỗi người sẽ giành chiến thắng.

Gọi A, B, C lần lượt là "A, B, C thắng", và A_i, B_i, C_i lần lượt là "A, B, C tung được mặt ngửa ở lần i", sử dụng tổng của chuỗi, hoặc dùng cấp số nhân, ta có

$$P(A) = P(A_1) + P(\overline{A_1}\,\overline{B_2}\,\overline{C_3}A_4) + \ldots = \frac{1}{2} + \frac{1}{2} \times \frac{1}{2^3} + \frac{1}{2} \times \frac{1}{2^6} + \ldots = \frac{\frac{1}{2}}{1 - \frac{1}{8}} = \frac{4}{7}$$

Tương tự $P(B) = \dfrac{\frac{1}{4}}{1 - \frac{1}{8}} = \dfrac{2}{7}$, $P(C) = \dfrac{\frac{1}{8}}{1 - \frac{1}{8}} = \dfrac{1}{7}$

Bài tập 1.26.

Trong một thùng kín có 6 quả cầu đỏ, 5 quả cầu trắng, 4 quả cầu vàng. Lấy ngẫu nhiên lần lượt từng quả cầu cho đến khi lấy được cầu đỏ thì dừng lại. Tính xác suất để:

1. Lấy được 2 cầu trắng, 1 cầu vàng.
2. Không có quả cầu trắng nào được lấy ra.

Gọi D_i, T_j, V_k là "lấy được quả đỏ, trắng, vàng ở lần thứ i, j, k"

1. Có $A = T_1T_2V_3D_4 + T_1V_2T_3D_4 + V_1T_2T_3D_4$ suy ra

$$P(A) = \frac{5}{15}.\frac{4}{14}.\frac{4}{13}.\frac{6}{12} + \frac{5}{15}.\frac{4}{14}.\frac{4}{13}.\frac{6}{12} + \frac{4}{15}.\frac{5}{14}.\frac{4}{13}.\frac{6}{12} = \frac{4}{91}$$

ở đó $P(T_iT_jV_kD_l) = P(T_i)\ P(T_j \mid T_i)\ P(T_k \mid T_iT_j)\ P(D_l \mid T_iT_jT_k)$

2. Có $B = D_1 + V_1D_2 + V_1V_2D_3 + V_1V_2V_3D_4 + V_1V_2V_3V_4D_5$
Vì các sự kiện trong tổng trên là xung khắc, nên áp dụng công thức cộng và xác suất của một tích ta có

$$P(B) = \frac{6}{15} + \frac{4}{15}.\frac{6}{14} + \frac{4}{15}.\frac{3}{14}.\frac{6}{13} + \frac{4}{15}.\frac{3}{14}.\frac{2}{13}.\frac{6}{12} + \frac{4}{15}.\frac{3}{14}.\frac{2}{13}.\frac{1}{12}\frac{6}{11} = \frac{6}{11}$$

Bài tập 1.27.

Ba xạ thủ A, B, C độc lập với nhau cùng bắn súng vào bia. Xác suất bắn trúng bia của 3 người A, B và C tương ứng là 0,7, 0,6 và 0,9. Tính xác suất để:

1. có duy nhất một xạ thủ bắn trúng bia;
2. có đúng hai xạ thủ bắn trúng bia;
3. có ít nhất một xạ thủ bắn trúng bia;
4. xạ thủ A bắn trúng bia biết rằng có hai xạ thủ bắn trúng bia.

Gọi A, B, C lần lượt là "A, B, C bắn trúng bia". Dễ thấy A, B, C là các sự kiện độc lập. Ta có

1. $P(A_1) = \sum P(\overline{A}\,\overline{B}C) = 0.154$
2. $P(A_2) = \sum P(\overline{A}BC) = 0.456$
3. $P(A_3) = 1 - P(\overline{A}\,\overline{B}\,\overline{C}) = 0.988$
4. Gọi A_4 là "xạ thủ A bắn trúng bia biết rằng có hai xạ thủ bắn trúng bia". Ta có $A_4 = A \mid A_2$. Sử dụng xác suất điều kiện,

$$P(A_4) = P(A \mid A_2) = \frac{P(AB\overline{C}) + P(AC\overline{B})}{P(A_2)} = 0.648$$

Bài tập 1.28.

Trên một bảng quảng cáo, người ta mắc hai hệ thống bóng đèn độc lập. Hệ thống I gồm 4 bóng mắc nối tiếp, hệ thống II gồm 3 bóng mắc song song. Khả năng bị hỏng của mỗi bóng trong 18 giờ thắp sáng liên tục là 0,1. Việc hỏng của mỗi bóng của mỗi hệ thống được xem như độc lập. Tính xác suất để trong 18 giờ thắp sáng liên tục:

1. cả hai hệ thống bị hỏng;
2. chỉ có một hệ thống bị hỏng.

Gọi A_i là "bóng thứ i của hệ thống I hỏng" và B_j là "bóng thứ j của hệ thống II hỏng".
Hệ thống I bị hỏng khi và chỉ khi 1 trong 4 bóng của nó hỏng, ta biểu diễn sự kiện này là

$$A = A_1 + A_2 + A_3 + A_4$$

Có $P(A) = 1 - (1 - 0.1)^4 = 0.3439$
Hệ thống II hỏng khi và chỉ khi tất cả 3 bóng mắc song song đều hỏng, sự kiện này là

$$B = B_1B_2B_3$$

Có $P(B) = 0.1^3 = 0.001$

1. Gọi C là "cả hai hệ thống hỏng". C xảy ra khi và chỉ khi hệ thống I và hệ thống II đều hỏng, nói cách khác,

$$C = AB = (A_1 + A_2 + A_3 + A_4)B_1B_2B_3$$

Suy ra $P(C) = 0.3439 \times 0.001 = 3.439 \times 10^{-4}$

2. Gọi D là "chỉ có một hệ thống hỏng" thì ta có

$$D = A\overline{B} + \overline{A}B = (A_1 + A_2 + A_3 + A_4)(\overline{B_1} + \overline{B_2} + \overline{B_3}) + (\overline{A_1}\,\overline{A_2}\,\overline{A_3}\,\overline{A_4})B_1B_2B_3$$

Suy ra
$$P(D) = 0.3439 \times (1 - 0.001) + (1 - 0.3439) \times 0.001 \simeq 0.3442$$

Bài tập 1.29.

Có 6 khẩu súng cũ và 4 khẩu súng mới, trong đó xác suất trúng khi bắn bằng súng cũ là 0,8, còn súng mới là 0,95. Bắn hú họa bằng một khẩu súng vào một mục tiêu thì thấy trúng. Điều gì có khả năng xảy ra lớn hơn: bắn bằng khẩu súng mới hay bắn bằng khẩu súng cũ?

Gọi M là "bắn bằng khẩu mới" thì $\overline{M}$ là "bắn bằng khẩu cũ".
Có $P(M) = 0.4$ và $P(\overline{M}) = 0.6$.
Gọi T là "bắn trúng" thì theo đề bài, ta có $P(T \mid M) = 0.95$ và $P(T \mid \overline{M}) = 0.8$.
Áp dụng công thức xác suất điều kiện suy ra

$$P(M \mid T) = \frac{P(M)P(T \mid M)}{P(T)} = \frac{0.38}{P(T)},\ P(\overline{M} \mid T) = \frac{P(\overline{M})P(T \mid \overline{M})}{P(T)} = \frac{0.48}{P(T)}$$

Suy ra sự kiện bắn bằng khẩu cũ có khả năng xảy ra cao hơn.
Chú ý: Ở đây ta hoàn toàn có thể tính được $P(T)$ theo công thức đầy đủ, tuy nhiên trong bài toán này là không cần thiết.

Bài tập 1.30.

Theo thống kê xác suất để hai ngày liên tiếp có mưa ở một thành phố vào mùa hè là 0,5; còn không mưa là 0,3. Biết các sự kiện có một ngày mưa, một ngày không mưa là đồng khả năng. Tính xác suất để ngày thứ hai có mưa, biết ngày đầu không mưa.

Gọi A là "ngày đầu mưa" và B là "ngày thứ hai mưa" thì ta có $P(AB) = 0.5, P(\overline{A}\,\overline{B}) = 0.3$.
Vì các sự kiện có một ngày mưa, một ngày không mưa là đồng khả năng nên

$$P(A\,\overline{B}) = P(\overline{A}\,B) = \frac{1 - 0.5 - 0.3}{2} = 0.1$$

Xác suất cần tính là $P(B \mid \overline{A})$, có

$$P(B \mid \overline{A}) = \frac{P(B\,\overline{A})}{P(\overline{A})} = \frac{P(B\,\overline{A})}{P(\overline{A}\,\overline{B}) + P(\overline{A}B)} = \frac{0.1}{0.1 + 0.3} = 0.25$$

Bài tập 1.31.

Một hộp chứa a quả bóng màu đỏ và b quả bóng màu xanh. Một quả bóng được chọn ngẫu nhiên và quan sát màu sắc của nó. Sau đó bóng được trả lại cho vào hộp và k bóng cùng màu cũng được thêm vào hộp. Một quả bóng thứ hai sau đó được chọn một cách ngẫu nhiên, màu sắc của nó được quan sát, và nó được trả lại cho vào hộp với k bóng bổ sung cùng một màu. Quá trình này được lặp đi lặp lại 4 lần. Tính xác suất để ba quả bóng đầu tiên sẽ có màu đỏ và quả bóng thứ tư có màu xanh.

Gọi D_i, X_j lần lượt là "lấy được quả đỏ ở lần i" và "lấy được quả xanh ở lần j". Sự kiện cần tính xác suất là $A = D_1D_2D_3X_4$. Sử dụng công thức xác suất của tích

$$\begin{aligned} P(A) &= P(D_1D_2D_3X_4) = P(D_1)\ P(D_2 \mid D_1)\ P(D_2 \mid D_1D_2)\ P(X_4 \mid D_1D_2D_3) \\ &= \frac{a}{a+b}.\frac{a+k}{a+b+k}.\frac{a+2k}{a+b+2k}.\frac{b}{a+b+3k} \end{aligned}$$

Bài tập 1.32.

Một cửa hàng sách ước lượng rằng: trong tổng số các khách hàng đến cửa hàng có 30% khách cần hỏi nhân viên bán hàng, 20% khách mua sách và 15% khách thực hiện cả hai điều trên. Gặp ngẫu nhiên một khách trong nhà sách. Tính xác suất để người này:

1. không thực hiện cả hai điều trên;
2. không mua sách, biết rằng người này đã hỏi nhân viên bán hàng.

Gọi A là "khách hỏi nhân viên bán hàng" và B là "khách mua sách"

1. $P(\overline{A}\,\overline{B}) = 1 - P(A+B) = 1 - P(A) - P(B) + P(AB) = 0.65$
2. $P(\overline{B} \mid A) = \dfrac{P(\overline{B}A)}{P(A)} = \dfrac{P(A) - P(AB)}{P(A)} = 0.5$

Bài tập 1.33.

Một cuộc khảo sát 1000 người về hoạt động thể dục thấy có 80% số người thích đi bộ và 60% thích đạp xe vào buổi sáng và tất cả mọi người đều tham gia ít nhất một trong hai hoạt động trên. Chọn ngẫu nhiên một người hoạt động thể dục. Nếu gặp được người thích đi xe đạp thì xác suất mà người đó không thích đi bộ là bao nhiêu?

Gọi A là "người thích đi bộ", B là "người thích đi xe đạp"
Theo giả thiết, $P(A) = 0.8, P(B) = 0.6$ và $P(A+B) = 1$. Ta có

$$\begin{aligned} P(\overline{A} \mid B) &= \frac{P(\overline{A}\,B)}{P(B)} = \frac{P(B) - P(AB)}{P(B)} = \frac{P(B) + [P(A+B) - P(A) - P(B)]}{P(B)} \\ &= \frac{P(A+B) - P(A)}{P(B)} = \frac{1 - 0.8}{0.6} \simeq 0.3333 \end{aligned}$$

Bài tập 1.34.

Để thành lập đội tuyển quốc gia về một môn học, người ta tổ chức một cuộc thi tuyển gồm 3 vòng. Vòng thứ nhất lấy 80% thí sinh; vòng thứ hai lấy 70% thí sinh đã qua vòng thứ nhất và vòng thứ ba lấy 45% thí sinh đã qua vòng thứ hai. Để vào được đội tuyển, thí sinh phải vượt qua được cả 3 vòng thi. Tính xác suất để một thí sinh bất kỳ:

1. được vào đội tuyển;
2. bị loại ở vòng thứ ba;
3. bị loại ở vòng thứ hai, biết rằng thí sinh này bị loại.

Gọi A_i là "thí sinh vượt qua vòng thứ i" thì ta có $P(A_1) = 0.8, P(A_2 \mid A_1) = 0.7$ và $P(A_3 \mid A_1A_2) = 0.45$

1. Gọi A là "thí sinh được vào đội tuyển" thì A xảy ra nếu thí sinh vượt qua cả 3 vòng, nghĩa là $A = A_1A_2A_3$

$$P(A) = P(A_1A_2A_3) = P(A_1)\ P(A_2 \mid A_1)\ P(A_3 \mid A_1A_2) = 0.8 \times 0.7 \times 0.45 = 0.252$$

2. Gọi B là "thí sinh bị loại ở vòng thứ 3" thì $B = A_1A_2\overline{A_3}$

$$P(B) = P(A_1)\ P(A_2 \mid A_1)\ P(\overline{A_3} \mid A_1A_2) = 0.8 \times 0.6 \times (1 - 0.45) = 0.308$$

3. Gọi C là sự kiện đang quan tâm: "thí sinh bị loại ở vòng 2, biết thí sinh này bị loại". Ta biểu diễn $C = A_1\overline{A_2} \mid \overline{A}$.

$$\begin{aligned} P(C) &= \frac{P\left[(A_1\overline{A_2})\overline{A}\right]}{P(\overline{A})} = \frac{P(A_1\overline{A_2})}{P(\overline{A})} \quad \left(\text{vì } A_1\overline{A_2} \subset \overline{A}\right) \\ &= \frac{P(A_1)P(\overline{A_2} \mid A_1)}{P(\overline{A})} \\ &= \frac{0.8\,.(1-0.7)}{1-0.252} \simeq 0.3208 \end{aligned}$$

Bài tập 1.35.

Theo thống kê ở các gia đình có hai con thì xác suất để con thứ nhất và con thứ hai đều là trai là 0,27 và hai con đều là gái là 0,23, còn xác suất con thứ nhất và con thứ hai có một trai và một gái là đồng khả năng. Biết sự kiện khi xét một gia đình được chọn ngẫu nhiên có con thứ nhất là gái, tìm xác suất để con thứ hai là trai.

Gọi A là "con thứ nhất là con trai" và B là "con thứ hai là con trai" thì theo đề, $P(AB) = 0.27$, $P(\overline{A}\,\overline{B}) = 0.23$ và $P(A\,\overline{B}) = P(\overline{A}\,B) = 0.25$.
Sự kiện quan tâm là $B \mid \overline{A}$.

Ta có

$$P(B \mid \overline{A}) = \frac{P(B\,\overline{A})}{P(\overline{A})} = \frac{P(B\,\overline{A})}{P(\overline{A}B) + P(\overline{A}\,\overline{B})} = \frac{0.25}{0.25 + 0.23} \simeq 0.5208$$

Bài tập 1.36.

Một tổ có 15 sinh viên trong đó có 5 sinh viên học giỏi môn "Xác suất thống kê". Cần chia làm 5 nhóm, mỗi nhóm 3 sinh viên. Tính xác suất để nhóm nào cũng có một sinh viên học giỏi môn "Xác suất thống kê".

Gọi A_i là "nhóm thứ i có 1 người giỏi Xác suất thống kê" và A là sự kiện nhóm nào cũng có người giỏi Xác suất thống kê, thì dễ dàng nhận thấy

$$A = A_1A_2A_3A_4A_5$$

Ta có

$P(A_1) = \dfrac{C_5^1\,C_{10}^2}{C_{15}^3} = \dfrac{45}{91}$, $P(A_2 \mid A_1) = \dfrac{C_4^1\,C_8^2}{C_{12}^3} = \dfrac{28}{55}$, $P(A_3 \mid A_1A_2) = \dfrac{C_3^1\,C_6^2}{C_9^3} = \dfrac{15}{28}$

$P(A_4 \mid A_1A_2A_3) = \dfrac{C_2^1\,C_4^2}{C_6^3} = \dfrac{3}{5}$, $P(A_5 \mid A_1A_2A_3A_4) = \dfrac{C_1^1\,C_2^2}{C_3^3} = 1$

Áp dụng công thức xác suất của tích ta có

$$\begin{aligned} P(A) &= P(A_1)\,P(A_2 \mid A_1)\,P(A_3 \mid A_1A_2)\,P(A_4 \mid A_1A_2A_3)\,P(A_5 \mid A_1A_2A_3A_4) \\ &= \frac{C_5^1\,C_{10}^2}{C_{15}^3}.\frac{C_4^1\,C_8^2}{C_{12}^3}.\frac{C_3^1\,C_6^2}{C_9^3}.\frac{C_2^1\,C_4^2}{C_6^3}.\frac{C_1^1\,C_2^2}{C_3^3} \\ &\simeq 0.0809 \end{aligned}$$

Bài tập 1.37.

Một hộp có n áo trắng và $2n$ áo xanh. Chia ngẫu nhiên các áo trong hộp thành n nhóm mỗi nhóm 3 áo.

1. Tính xác suất để trong mỗi nhóm đều có áo trắng;
2. Áp dụng cho $n = 5$.

1. Số kết cục đồng khả năng là số cách chia áo sao cho mỗi nhóm có 3 áo:

$$\binom{3n}{3}\binom{3n-3}{3}\cdots\binom{3}{3} = \frac{(3n)!}{(3n-3)!3!}\frac{(3n-3)!}{(3n-6)!3!}\cdots\frac{(3)!}{0!3!} = \frac{(3n)!}{(3!)^n}$$

Nếu đánh số n cái áo trắng thì mỗi cách chia mà mỗi nhóm chỉ có 1 áo trắng cho ta một hoán vị của $1, 2, \ldots, n$. Suy ra số cách chia áo trắng "thuận lợi" là $n!$

Số cách chia $2n$ áo xanh còn lại cho các nhóm là

$$\binom{2n}{2}\binom{2n-2}{2}\cdots\binom{2}{2} = \frac{(2n)!}{(2n-2)!2!}\frac{(2n-2)!}{(2n-4)!2!}\cdots\frac{(2)!}{0!2!} = \frac{(2n)!}{(2!)^n}$$

Như vậy, số kết cục thuận lợi là $n! \times \dfrac{(2n)!}{(2!)^n}$.

Suy ra

$$P = \frac{3^n\ n!\ (2n)!}{(3n)!} = \frac{3^n}{C_{3n}^n}$$

2. Thay $n = 5$ thì $P = \dfrac{3^5}{C_{15}^5} \simeq 0.0809$

Bài tập 1.38.

Hai vận động viên bóng bàn A và B đấu một trận gồm tối đa 5 ván (không có kết quả hòa sau mỗi ván và trận đấu sẽ dừng nếu một người nào đó thắng trước 3 ván). Xác suất để A thắng được ở một ván là 0,7.

1. Tính các xác suất để A thắng sau x ván ($x = 3, 4, 5$).

2. Tính xác suất để trận đấu kết thúc sau 5 ván.

Gọi A là "A thắng được ở một ván" thì $p = P(A) = 0.7$

1. A thắng sau x ván nếu ván thứ x A thắng và trong $x-1$ ván trước đó A thắng 2 ván. Vì ở mỗi ván, A chỉ có thể thắng hoặc thua nên theo công thức *Bernoulli*,

$$P_{x-1}(2) = p\binom{x-1}{2}p^2(1-p)^{x-1-2} = \binom{x-1}{2}0.7^3 \times 0.3^{x-3}$$

Thay $x = 3, P_2(2) = 0.343,\ x = 4, P_3(2) = 0.3087,\ x = 5, P_4(2) = 0.1852$

2. Trận đấu kết thúc sau 5 ván nghĩa là trong 4 ván đầu, A và B mỗi người thắng 2 ván. Áp dụng công thức *Bernoulli*,

$$P = P_4(2) = \binom{4}{2}0.7^2 \times 0.3^2 = 0.2646$$

Bài tập 1.39.

Một bài thi trắc nghiệm (multiple-choice test) gồm 12 câu hỏi, mỗi câu hỏi cho 5 phương án trả lời, trong đó chỉ có 1 phương án đúng. Giả sử một câu trả lời đúng được 4 điểm và mỗi câu trả lời sai bị trừ đi 1 điểm. Một học sinh kém làm bài bằng cách chọn hú họa câu trả lời. Tìm xác suất để:

1. Học sinh đó được 13 điểm.

2. Học sinh đó bị điểm âm.

Giả sử học sinh đó làm đúng x câu, làm sai $12 - x$ câu $(0 \le x \le 12)$. Số điểm học sinh đạt được là $4x - (12 - x) = 5x - 12$. Ta có xác suất học sinh làm đúng mỗi câu là $p = 0.2$.

1. Mỗi kết cục thuận lợi cho sự kiện được 13 điểm là một phần tử của M

$$M = \{x \in \mathbb{N} \mid 5x - 12 = 13, x \leq 12\}$$

Thu được $x = 5$

Bài toán thỏa mãn lược đồ *Bernoulli* với $p = 0.2$, ta có

$$P = P_{12}(5) = \binom{12}{5} 0.2^5 \times 0.8^7 \simeq 0.0531$$

2. Mỗi kết cục thuận lợi là một phần tử của K

$$K = \{x \in \mathbb{N} \mid 5x - 12 < 0, x \leq 12\}$$

Như vậy, xảy ra các trường hợp $x = 0, x = 1, x = 2$. Dễ thấy các trường hợp này xung khắc. Bài toán thỏa mãn lược đồ *Bernoulli* với $p = 0.2$, ta có

$$P = P_{12}(0) + P_{12}(1) + P_{12}(2) = \sum_{k=0}^{2} \binom{12}{k} 0.2^k \times 0.8^{12-k} \simeq 0.5583$$

Bài tập 1.40.

Một nhân viên bán hàng mỗi ngày đi chào hàng ở 10 nơi với xác suất bán được hàng ở mỗi nơi là 0,2. Tìm xác suất để:

1. người đó bán được hàng ở 2 nơi;
2. người đó bán được hàng ở ít nhất 1 nơi.

Bài toán này thỏa mãn lược đồ *Bernoulli*

1. $P(A) = P_{10}(2) = \binom{10}{2} 0.2^2 \times 0.8^8 \simeq 0.3020$

2. $P(B) = 1 - P_{10}(0) = 1 - \binom{10}{0} 0.2^0 \times 0.8^{10} \simeq 1 - 0.1074 \simeq 0.8927$

Bài tập 1.41.

Xác suất trúng đích của một lần bắn là 0,4. Cần phải bắn bao nhiêu phát đạn để xác suất có ít nhất một viên bắn trúng sẽ lớn hơn 0,95?

Giả sử cần bắn n lần. Biết xác suất bắn trúng mỗi lần là $p = 0.4$, xác suất để n lần bắn đều trượt là 0.6^n
Suy ra xác suất để có ít nhất 1 lần trúng là $P = 1 - 0.6^n$
Giải bất phương trình $P \geq 0.95$ thu được $n \geq 6$

Bài tập 1.42.

Hai cầu thủ bóng rổ, mỗi người ném bóng 2 lần vào rổ. Xác suất ném trúng rổ của mỗi cầu thủ theo thứ tự lần lượt là 0,6 và 0,7. Tìm xác suất để

1. số lần ném trúng rổ của hai người bằng nhau;
2. số lần ném trúng rổ của cầu thủ thứ nhất nhiều hơn số lần ném trúng rổ của cầu thủ thứ hai.

Cầu thủ ném bóng vào rổ 2 lần, có thể ném trúng rổ 0, 1 hoặc cả 2 lần. Gọi A_i là "cầu thủ 1 ném trúng rổ i lần" và B_j là "cầu thủ 2 ném trúng rổ j lần"

1. Gọi A là "số lần ném trúng rổ của cả 2 cầu thủ bằng nhau". Có nghĩa là ta quan tâm đến sự kiện 2 cầu thủ cùng ném trúng rổ $0, 1$ hoặc cả 2 lần. Như vậy,

$$A = A_0B_0 + A_1B_1 + A_2B_2$$

   Có $P(A) = 0.4^2 \times 0.3^2 + (2 \times 0.6 \times 0.4) \times (2 \times 0.7 \times 0.3) + 0.6^2 \times 0.7^2 = 0.3924$

2. Gọi B là "số lần ném trúng của cầu thủ 1 nhiều hơn của cầu thủ 2". Ta viết B dưới dạng

$$B = A_2B_1 + A_2B_0 + A_1B_0$$

   Có $P(B) = 0.6^2 \times (2 \times 0.7 \times 0.3) + 0.6^2 \times 0.3^2 + (2 \times 0.6 \times 0.4) \times 0.3^2 = 0.2268$

Bài tập 1.43.

Xác suất sản xuất ra phế phẩm của một máy là 0,005. Tìm xác suất để trong 800 sản phẩm của máy đó có đúng 3 phế phẩm.

$n = 800$ rất lớn và $p = 0.005$ rất nhỏ. Ta có $\lambda = np = 4 < 7$. Áp dụng công thức *Poisson*

$$P_{800}(3) \simeq \frac{4^3}{3!} e^{-4} \simeq 0.1954$$

Bài tập 1.44.

Một công nhân đứng máy 1000 ống sợi. Xác suất mỗi ống bị đứt trong vòng một giờ là 0,005. Tính xác suất để trong vòng một giờ:

1. 40 ống sợi bị đứt;
2. không quá 40 ống sợi bị đứt.

$n = 1000$ rất lớn và $p = 0.005$ rất nhỏ. Ta có $\lambda = np = 5 < 7$. Áp dụng công thức *Poisson*

1. $P_{1000}(40) = \dfrac{5^{40}}{40!} e^{-5} \simeq 7.5107 \times 10^{-23}$

2. $P = \sum_{k=0}^{40} \frac{5^k}{k!} e^{-5} \simeq 1$

Bài tập 1.45.

Xác suất ném trúng rổ của một cầu thủ là 0,8. Tìm xác suất để trong 100 lần cầu thủ đó:

1. ném trúng 75 lần;
2. ném trúng không ít hơn 75 lần

$n = 100$ khá lớn và $p = 0.8$ "tương đối"

1. Gọi A là "100 cầu thủ ném trúng 75 lần", theo công thức $Gauss$ ta có

$$P(A) = P_{100}(75) \simeq \frac{\varphi\left(\dfrac{75 - 0.8.100}{\sqrt{100\,.0.8\,.0.2}}\right)}{\sqrt{100\,.0.8\,.0.2}} = \frac{\varphi(-1.25)}{4} = \frac{\varphi(1.25)}{4} \simeq 0.0456$$

2. Gọi B là "100 cầu thủ ném trúng ít nhất 75 lần", theo công thức $Moirve\,Laplace$ ta có

$$\begin{aligned} P_{100}(75;100) &= \phi\left(\frac{100 - 0.8.100}{\sqrt{100\,.0.8\,.0.2}}\right) - \phi\left(\frac{75 - 0.8.100}{\sqrt{100\,.0.8\,.0.2}}\right) \\ &= \phi(5) - \phi(-1.25) = \phi(5) + \phi(1.25) \\ &\simeq 0.49999 + 0.39435 = 0.8943 \end{aligned}$$

1.4 Công thức xác suất đầy đủ. Công thức Bayes

Bài tập 1.46.

Một phân xưởng có 3 máy tự động: máy I sản xuất 25%, máy II sản xuất 30%, máy III sản xuất 45% số sản phẩm. Tỷ lệ phế phẩm tương ứng của các máy lần lượt là 0,1%, 0,2% và 0,3%. Chọn ngẫu nhiên ra một sản phẩm của phân xưởng.

1. Tìm xác suất nó là phế phẩm.
2. Biết nó là phế phẩm. Tính xác suất để sản phẩm đó do máy I sản xuất.

Gọi A_i là "lấy ra sản phẩm từ lô i" thì A_1, A_2, A_3 tạo thành hệ đầy đủ.

1. Gọi A là "lấy ra sản phẩm là phế phẩm". Áp dụng công thức xác suất đầy đủ, ta có

$$\begin{aligned} P(A) &= P(A_1)P(A \mid A_1) + P(A_2)P(A \mid A_2) + P(A_3)P(A \mid A_3) \\ &= 0.25 \times 0.1\% + 0.3 \times 0.2\% + 0.45 \times 0.3\% = 0.22\% \end{aligned}$$

2. Gọi B là "sản phẩm do máy I sản xuất". Khi đó ta cần tính $P(B \mid A)$

$$P(B \mid A) = \frac{P(B)P(A \mid B)}{P(A)} = \frac{0.25 \times 0.1\%}{0.22\%} \simeq 0.1136$$

Bài tập 1.47.

Có 3 hộp đựng bi: hộp thứ nhất có 3 bi đỏ, 2 bi trắng; hộp thứ hai có 2 bi đỏ, 2 bi trắng; hộp thứ ba không có viên nào. Lấy ngẫu nhiên 1 viên bi từ hộp thứ nhất và 1 viên bi từ hộp thứ hai bỏ vào hộp thứ ba. Sau đó từ hộp thứ ba lấy ngẫu nhiên ra 1 viên bi.

1. Tính xác suất để viên bi đó màu đỏ.
2. Biết rằng viên bi lấy ra từ hộp thứ ba màu đỏ, tính xác suất để lúc đầu ta lấy được viên bi đỏ từ hộp thứ nhất bỏ vào hộp thứ ba.

Gọi A_1, A_2 lần lượt là "lấy bi đỏ từ hợp thứ 1 (thứ 2) bỏ vào hộp thứ ba" thì $A_1A_2, \overline{A_1}A_2, A_1\overline{A_2}, \overline{A_1}\,\overline{A_2}$ tạo thành một hệ đầy đủ. Ta có

$$P(A_1A_2) = 0.3, \quad P(\overline{A_1}A_2) = 0.2$$
$$P(A_1\overline{A_2}) = 0.3, \quad P(\overline{A_1}\,\overline{A_2}) = 0.2$$

1. Gọi A "lấy ra từ hộp 3 một viên bi màu đỏ". Ta có

$$P(A \mid A_1A_2) = 1, \quad P(A \mid \overline{A_1}\,\overline{A_2}) = 0$$
$$P(A \mid \overline{A_1}A_2) = 0.5, \quad P(A \mid A_1\overline{A_2}) = 0.5$$

Áp dụng công thức xác suất đầy đủ ta có

$$\begin{aligned}P(A) &= P(A_1A_2)P(A \mid A_1A_2) + P(\overline{A_1}A_2)P(A \mid \overline{A_1}A_2) + P(A_1\overline{A_2})P(A \mid A_1\overline{A_2}) \\ &\quad + P(\overline{A_1}\,\overline{A_2})P(A \mid \overline{A_1}\,\overline{A_2}) \\ &= 0.3 \times 1 + 0.3 \times 0.5 + 0.2 \times 0.5 + 0.2 \times 0 = 0.55\end{aligned}$$

2. Gọi B là sự kiện cần tính xác suất. Dễ thấy $B = (A_1A_2 + A_1\overline{A_2}) \mid A$.
Theo công thức *Bayes* ta có

$$\begin{aligned}P(B) &= \frac{P\left[(A_1A_2 + A_1\overline{A_2})\,A\right]}{P(A)} \\ &= \frac{P\left[(A_1A_2)A\right] + P\left[(A_1\overline{A_2})A\right]}{P(A)} \\ &= \frac{P(A_1A_2)P(A \mid A_1A_2) + P(A_1\overline{A_2})P(A \mid A_1\overline{A_2})}{P(A)} \\ &= \frac{0.3 \times 1 + 0.2 \times 0.5}{0.55} = \frac{9}{11} \simeq 0.8181\end{aligned}$$

Bài tập 1.48.

Hộp I có 4 viên bi đỏ, 2 viên bi xanh; hộp II có 3 viên bi đỏ, 3 viên bi xanh. Bỏ ngẫu nhiên một viên bi từ hộp I sang hộp II, sau đó lại bỏ ngẫu nhiên một viên bi từ hộp II sang hộp I. Cuối cùng rút ngẫu nhiên từ hộp I ra một viên bi.

1. Tính xác suất để viên bi rút ra sau cùng màu đỏ.
2. Nếu viên rút ra sau cùng màu đỏ, tìm xác suất lúc ban đầu rút được viên bi đỏ ở hộp I cho vào hộp II.

Gọi D_1, X_1 tương ứng là "lấy được viên bi đỏ, xanh từ hộp I sang hộp II", D_2, X_2 tương ứng là "lấy được viên bi đỏ, xanh từ hộp II sang hộp I".
Khi đó hệ $D_1D_2, D_1X_2, X_1D_2, X_1X_2$ tạo thành hệ đầy đủ. Ta có

$$P(D_1D_2) = \frac{4}{6}.\frac{4}{7}, \quad P(D_1X_2) = \frac{4}{6}.\frac{3}{7}$$

$$P(X_1D_2) = \frac{2}{6}.\frac{3}{7}, \quad P(X_1X_2) = \frac{2}{6}.\frac{4}{7}$$

1. Gọi A là "viên bi rút ra sau cùng là màu đỏ". Áp dụng công thức xác suất đầy đủ

$$P(A \mid D_1D_2) = \frac{4}{6}, \quad P(A \mid D_1X_2) = \frac{3}{6}$$

$$P(A \mid X_1D_2) = \frac{5}{6}, \quad P(A \mid X_1X_2) = \frac{4}{6}$$

$$\begin{aligned} P(A) &= P(D_1D_2)P(A \mid D_1D_2) + P(D_1X_2)P(A \mid D_1X_2) + P(X_1D_2)P(A \mid X_1D_2) \\ &\quad + P(X_1X_2)P(A \mid X_1X_2) \\ &= \frac{4}{6}.\frac{4}{7}.\frac{4}{6} + \frac{4}{6}.\frac{3}{7}.\frac{3}{6} + \frac{2}{6}.\frac{3}{7}.\frac{5}{6} + \frac{2}{6}.\frac{4}{7}.\frac{4}{6} \\ &= \frac{9}{14} \simeq 0.6429 \end{aligned}$$

2. Sự kiện cần tính xác suất là $B = (D_1D_2 + D_1X_2) \mid A$

$$\begin{aligned} P(B) &= \frac{P\left[(D_1D_2 + D_1X_2)\, A\right]}{P(A)} \\ &= \frac{P\left[(D_1D_2)A\right] + P\left[(D_1X_2)A\right]}{P(A)} \\ &= \frac{P(D_1D_2)P(A \mid D_1D_2) + P(D_1X_2)P(A \mid D_1X_2)}{P(A)} \\ &= \frac{\frac{4}{6}.\frac{4}{7}.\frac{4}{6} + \frac{4}{6}.\frac{3}{7}.\frac{3}{6}}{\frac{9}{11}} = \frac{50}{81} \simeq 0.6173 \end{aligned}$$

Bài tập 1.49.

Trong một kho rượu, số lượng rượu loại A và loại B bằng nhau. Người ta chọn ngẫu nhiên một chai và đưa cho 5 người nếm thử. Biết xác suất đoán đúng của mỗi người là 0,8. Có 3 người kết luận rượu loại A, 2 người kết luận rượu loại B. Hỏi khi đó xác suất chai rượu đó thuộc loại A là bao nhiêu?

Gọi A là "chai rượu thuộc loại A" thì $A, \overline{A}$ tạo thành hệ đầy đủ và $P(A) = P(\overline{A}) = \dfrac{1}{2}$.
Gọi H là "có 3 người kết luận rượu loại A và 2 người kết luận rượu loại B". Theo công thức đầy đủ

$$\begin{aligned} P(H) &= P(A)P(H \mid A) + P(\overline{A})P(\overline{A} \mid H) \\ &= 0.5 \times \binom{5}{3} 0.8^3 \times 0.2^2 + 0.5 \times \binom{5}{2} 0.8^2 \times 0.2^3 \\ &= 0.128 \end{aligned}$$

Xác suất cần tính là $P(A \mid H) = \dfrac{P(A)P(H \mid A)}{P(H)} = \dfrac{0.5 \times C_5^3 \times 0.8^3 \times 0.2^2}{0.128} = 0.8$

Bài tập 1.50.

Có hai lô sản phẩm: lô I có 7 chính phẩm 3 phế phẩm; lô II có 6 chính phẩm 2 phế phẩm. Lấy ngẫu nhiên 2 sản phẩm từ lô I sang lô II, sau đó từ lô II lấy ngẫu nhiên ra 2 sản phẩm được 2 chính phẩm. Tính xác suất để 2 chính phẩm lấy ra sau cùng là của lô I.

Gọi A'_j là "lấy j chính phẩm từ lô I sang lô II" thì A'_0, A'_1, A'_2 tạo thành hệ đầy đủ, và

$$P(A'_0) = \frac{C_3^2}{C_{10}^2}, \quad P(A'_1) = \frac{C_7^1 C_3^1}{C_{10}^2}, \quad P(A'_2) = \frac{C_7^2}{C_{10}^2}$$

Gọi H là "2 sản phẩm lấy ra sau cùng là chính phẩm", ta tính $P(H)$ theo hệ đầy đủ này

$$P(H) = \frac{C_3^2}{C_{10}^2}.\frac{C_6^2}{C_{10}^2} + \frac{C_7^1 C_3^1}{C_{10}^2}.\frac{C_7^2}{C_{10}^2} + \frac{C_7^2}{C_{10}^2}.\frac{C_8^2}{C_{10}^2} = \frac{358}{675} \simeq 0.5304$$

Gọi A_i là "2 sản phẩm lấy ra sau cùng có i sản phẩm của lô I" thì A_0, A_1, A_2 cũng tạo thành hệ đầy đủ.
Sự kiện cần tính xác suất là $A = A_2 \mid H$. Sử dụng công thức *Bayes* ta có

$$P(A_2 \mid H) = \frac{P(A_2)P(H \mid A_2)}{P(H)} = \frac{\dfrac{C_2^2}{C_{10}^2}.\dfrac{C_7^2}{C_{10}^2}}{0.5304} \simeq 0.0196$$

ở đó $P(A_2) = \dfrac{C_2^2}{C_{10}^2}$ và $P(H \mid A_2) = \dfrac{C_7^2}{C_{10}^2}$

Bài tập 1.51.

Có hai lô sản phẩm: lô I có 7 chính phẩm, 3 phế phẩm; lô II có 8 chính phẩm, 2 phế phẩm. Từ lô I lấy ngẫu nhiên ra 2 sản phẩm, từ lô II lấy ngẫu nhiên ra 3 sản phẩm. Sau đó từ số sản phẩm này lại lấy ngẫu nhiên 2 sản phẩm. Tính xác suất để trong 2 sản phẩm lấy ra sau cùng có ít nhất 1 chính phẩm.

Gọi A_i là "trong 5 sản phẩm cuối có i chính phẩm".
Khi đó hệ $A_0, A_1, A_2, A_3, A_4, A_5$ tạo thành hệ đầy đủ

- A_0 xảy ra thì phải lấy 3 phế phẩm từ lô II, điều này là không thể. Suy ra $P(A_0) = 0$
- A_1 xảy ra nếu lấy 2 phế từ lô I và 1 chính, 1 phế từ lô II.

$$P(A_1) = \frac{C_3^2}{C_{10}^2}.\frac{C_8^1C_2^2}{C_{10}^3} = \frac{1}{225}$$

- A_2 xảy ra nếu lấy 1 chính, 1 phế từ lô I, 1 chính, 2 phế từ lô II **hoặc** 2 phế từ lô I, 2 chính, 1 phế từ lô II

$$P(A_2) = \frac{C_7^1C_3^1}{C_{10}^2}.\frac{C_8^1C_2^2}{C_{10}^3} + \frac{C_3^2}{C_{10}^2}.\frac{C_8^2C_2^1}{C_{10}^3} = \frac{14}{225}$$

- A_3 xảy ra nếu lấy 2 chính từ lô I, 1 chính, 2 phế từ lô II **hoặc** 1 chính, 1 phế từ lô I, 2 chính, 1 phế từ lô II **hoặc** 2 phế từ lô I, 3 chính từ lô II

$$P(A_3) = \frac{C_7^2}{C_{10}^2}.\frac{C_8^1C_2^2}{C_{10}^3} + \frac{C_7^1C_3^1}{C_{10}^2}.\frac{C_8^2C_2^1}{C_{10}^3} + \frac{C_3^2}{C_{10}^2}.\frac{C_8^3}{C_{10}^3} = \frac{7}{25}$$

- A_4 xảy ra nếu lấy 2 chính từ lô I, 2 chính, 2 phế từ lô II **hoặc** 1 chính, 1 phế từ lô I, 3 chính từ lô II

$$P(A_4) = \frac{C_7^2}{C_{10}^2}.\frac{C_8^2C_2^1}{C_{10}^3} + \frac{C_7^1C_3^1}{C_{10}^2}.\frac{C_8^3}{C_{10}^3} = \frac{98}{225}$$

- A_5 xảy ra nếu lấy 2 chính từ lô I, 3 chính từ lô II

$$P(A_5) = \frac{C_7^2}{C_{10}^2}.\frac{C_8^3}{C_{10}^3} = \frac{49}{225}$$

Gọi A là "trong 2 sản phẩm lấy ra có ít nhất 1 chính phẩm", áp dụng công thức xác suất đầy đủ

$$\begin{aligned}
P(\overline{A}) &= \sum_{i=0}^{5} P(A_i)P(\overline{A} \mid A_i) \\
&= \frac{C_5^2}{C_5^2}.0 + \frac{C_4^2}{C_5^2}.\frac{1}{225} + \frac{C_3^2}{C_5^2}.\frac{14}{225} + \frac{C_2^2}{C_5^2}.\frac{7}{25} + 0.\frac{98}{225} + 0.\frac{49}{225} \\
&\simeq 0.4933
\end{aligned}$$

Suy ra $P(A) = 1 - P(\overline{A}) \simeq 0.6507$

Bài tập 1.52.

Có ba kiện hàng (mỗi kiện hàng có 20 sản phẩm) với số sản phẩm tốt tương ứng của mỗi kiện là 18, 16, 12. Lấy ngẫu nhiên một kiện hàng, rồi từ đó lấy ngẫu nhiên một sản phẩm thì được sản phẩm tốt. Trả sản phẩm này lại kiện hàng vừa lấy, sau đó lại lấy ngẫu nhiên một sản phẩm thì được sản phẩm tốt. Tính xác suất để các sản phẩm tốt đó được lấy từ kiện hàng thứ nhất.

Gọi A_i là "sản phẩm lấy từ kiện thứ i" thì A_1, A_2, A_3 tạo thành hệ đầy đủ.

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$$

Gọi A là "các sản phẩm lấy ra đều là tốt", áp dụng công thức xác suất đầy đủ

$$P(A \mid A_1) = \frac{18}{20}.\frac{18}{20}, \quad P(A \mid A_2) = \frac{16}{20}.\frac{16}{20}, \quad P(A \mid A_3) = \frac{12}{20}.\frac{12}{20}$$

Thay vào suy ra

$$P(A) = \frac{1}{3}.\frac{18}{20}.\frac{18}{20} + \frac{1}{3}.\frac{16}{20}.\frac{16}{20} + \frac{1}{3}.\frac{12}{20}.\frac{12}{20} = \frac{181}{300} \simeq 0.6033$$

Sử dụng công thức *Bayes* ta có

$$P(A_1 \mid A) = \frac{P(A_1)P(A \mid A_1)}{P(A)} = \frac{\frac{1}{3}.\frac{18}{20}.\frac{18}{20}}{\frac{181}{300}} = \frac{81}{181} \approx 0.4475$$

Bài tập 1.53.

Tỷ lệ người nghiện thuốc là ở một vùng là 30%. Biết rằng tỷ lệ người bị viêm họng trong số những người nghiện thuốc là 60%, còn tỷ lệ người bị viêm họng trong số những người không nghiện là 40%

1. Lấy ngẫu nhiên một người thấy người ấy bị viêm họng. Tính xác suất người đó nghiện thuốc lá.

2. Nếu người đó không bị viêm họng, tính xác suất người đó nghiện thuốc lá.

Gọi A là "người nghiện thuốc" và B là "người viêm họng" thì từ đề bài

$$P(A) = 0.3, \quad P(B \mid A) = 0.6, \quad P(B \mid \overline{A}) = 0.4$$

1. Sự kiện cần tính xác suất là $C = A \mid B$. Sử dụng công thức *Bayes*

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(\overline{A})P(B \mid \overline{A})} = \frac{0.3 \times 0.6}{0.3 \times 0.6 + 0.7 \times 0.4} \simeq 0.3913$$

2. Gọi $D = A \mid \overline{B}$. Ta có

$$P(D) = \frac{P(A\overline{B})}{P(\overline{B})} = \frac{P(A) - P(AB)}{1 - P(B)} = \frac{P(A) - P(A)P(B \mid A)}{1 - P(B)} \simeq 0.2222$$

Bài tập 1.54.

Một công nhân đi làm ở thành phố khi trở về nhà có 2 cách: hoặc đi theo đường ngầm hoặc đi qua cầu. Biết rằng ông ta đi lối đường ngầm trong $\frac{1}{3}$ các trường hợp, còn lại đi lối cầu. Nếu đi lối đường ngầm 75% trường hợp ông ta về đến nhà trước 6 giờ tối; còn nếu đi lối cầu chỉ có 70% trường hợp (nhưng đi lối cầu thích hơn). Tìm xác suất để công nhân đó đã đi lối cầu biết rằng ông ta về đến nhà sau 6 giờ tối.

Gọi A là "đi đường ngầm" thì $\overline{A}$ là "đi đường cầu" và $P(A) = \frac{1}{3}, P(\overline{A}) = \frac{2}{3}$.
Gọi B là "về nhà sau 6 giờ tối", ta cần tính $P(\overline{A} \mid B)$. Sử dụng công thức *Bayes*

$$P(\overline{A} \mid B) = \frac{P(\overline{A})P(B \mid \overline{A})}{P(B)} = \frac{\frac{2}{3} \times 0.3}{\frac{2}{3} \times 0.3 + \frac{1}{3} \times 0.25} \simeq 0.7059$$

Bài tập 1.55.

Tại một phòng khám chuyên khoa tỷ lệ người đến khám có bệnh là 0,8. Người ta áp dụng phương pháp chẩn đoán mới thì thấy nếu khẳng định có bệnh thì đúng 9 trên 10 trường hợp; còn nếu khẳng định không bệnh thì đúng 5 trên 10 trường hợp. Tính xác suất để

1. chẩn đoán có bệnh;

2. chẩn đoán đúng.

Gọi A là "người đến khám có bệnh" thì $A, \overline{A}$ tạo thành hệ đầy đủ

1. Gọi B là "Chẩn đoán có bệnh". Ta có $P(A \mid B) = 0.9,\ P(A \mid \overline{B}) = 0.5$. Tìm $P(B)$ từ:

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(A) - P(A \mid \overline{B})P(\overline{B})}{P(B)} = \frac{P(A) - P(A \mid \overline{B})\,(1 - P(B))}{P(B)}$$

Thay số vào ta có

$$0.9 = \frac{0.8 - 0.5\,(1 - P(B))}{P(B)}$$

Giải ra tìm được $P(B) = 0.75$

2. Gọi C là "chẩn đoán đúng", thì C xảy ra khi người bị bệnh được chẩn đoán có bệnh hoặc người không bị bệnh được chẩn đoán không bị bệnh. Như vậy

$$C = AB + \overline{A}\,\overline{B}$$

Hiển nhiên 2 sự kiện $AB, \overline{A}\,\overline{B}$ xung khắc, nên

$$\begin{aligned} P(C) &= P(AB + \overline{A}\,\overline{B}) = P(B)P(A \mid B) + P(\overline{B})P(\overline{A} \mid \overline{B}) \\ &= 0.75 \times 0.9 + 0.25 \times 0.5 = 0.8 \end{aligned}$$

Bài tập 1.56.

Một hãng hàng không cho biết rằng 5% số khách đặt trước vé cho các chuyến đã định sẽ hoãn không đi chuyến bay đó. Do đó hãng đã đưa ra một chính sách là sẽ bán 52 ghế cho một chuyến bay mà trong đó mỗi chuyến chỉ trở được 50 khách hàng. Tìm xác suất để tất cả các khách đặt chỗ trước và không hoãn chuyến bay đều có ghế. Biết rằng xác suất bán được 51 vé hoặc 52 vé là như nhau và bằng 10%.

Gọi A là "bán được 52 vé", B là "bán được 51 vé" và C là "bán được nhiều nhất 50 vé". Khi đó A, B, C tạo thành hệ đầy đủ. Ta có

$$P(A) = 0.1, \quad P(B) = 0.1, \quad P(C) = 0.8$$

Gọi H là "khách đặt chỗ trước và không hoãn chuyến đều có ghế".
Sự kiện $H \mid A$ xảy ra nếu có ít nhất 2 khách hủy chuyến, $H \mid B$ xảy ra nếu có ít nhất 1 khách hủy chuyến. Tính trực tiếp xác suất của các sự kiện này đều khá phức tạp
Do đó để cho đơn giản ta tìm $P(\overline{H})$. Ta có

$$P(\overline{H} \mid A) = 0.95^{52} \times 0.05^{0} + 52 \times 0.95^{51} \times 0.05^{1}$$
$$P(\overline{H} \mid B) = 0.95^{51} \times 0.05^{0}, \quad P(\overline{H} \mid C) = 0$$

$$\begin{aligned} P(\overline{H}) &= P(A)P(\overline{H} \mid A) + P(B)P(\overline{H} \mid B) + P(C)P(\overline{H} \mid C) \\ &= 0.1 \times \left(0.95^{52} \times 0.05^{0} + 52 \times 0.95^{51} \times 0.05^{1}\right) + 0.1 \times 0.95^{51} \times 0.05^{0} + 0.8 \times 0 \\ &\simeq 0.033 \end{aligned}$$

Suy ra $P(H) = 1 - P(\overline{H}) \simeq 0.9667$

Bài tập 1.57.

Một trạm chỉ phát hai loại tín hiệu A và B với xác suất tương ứng là 0,84 và 0,16. Do có nhiễu trên đường truyền nên $\dfrac{1}{6}$ tín hiệu A bị méo và được thu như là tín hiệu B, còn $\dfrac{1}{8}$ tín hiệu B bị méo thành tín hiệu A.

1. Tìm xác suất thu được tín hiệu A;
2. Giả sử thu được tín hiệu A, tìm xác suất để thu được đúng tín hiệu lúc phát.

Gọi A, B lần lượt là "phát ra tín hiệu A, B". Khi đó A, B tạo thành hệ đầy đủ.

1. Gọi C là "thu được tín hiệu A". Áp dụng công thức xác suất đầy đủ

$$P(C) = P(A)P(C \mid A) + P(B)P(C \mid B) = 0.84 \times \frac{5}{6} + 0.16 \times \frac{1}{8} = 0.72$$

2. Ta cần tính $P(A \mid C)$. Áp dụng công thức *Bayes*

$$P(A \mid C) = \frac{P(A)P(C \mid A)}{P(C)} = \frac{0.84 \times \frac{5}{6}}{0.72} = \frac{35}{36} \simeq 0.9722$$

Bài tập 1.58.

Một người có ba chỗ ưa thích như nhau để câu cá. Xác suất để câu được cá ở mỗi chỗ tương ứng là 0,6; 0,7 và 0,8. Biết rằng đến một chỗ người đó thả câu 3 lần và chỉ câu được một con cá. Tính xác suất để cá câu được ở chỗ thứ nhất.

Gọi A_1, A_2, A_3 lần lượt là "cá câu được ở chỗ thứ i" thì hệ A_1, A_2, A_3 tạo thành hệ đầy đủ. Dễ thấy

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$$

Gọi H là "thả câu 3 lần và chỉ câu được 1 con cá". Theo công thức đầy đủ, ta có

$$P(H) = P(A_1)P(H \mid A_1) + P(A_2)P(H \mid A_2) + P(A_3)P(H \mid A_3)$$

ở đó

$$P(H \mid A_1) = 3 \times 0.6^1 \times 0.4^2, \quad P(H \mid A_2) = 3 \times 0.7^1 \times 0.3^2$$

$$P(H \mid A_3) = 3 \times 0.8^1 \times 0.2^2$$

Như vậy, $P(H) = 0.191$. Theo công thức *Bayes* suy ra

$$P(A_1 \mid H) = \frac{P(A_1)P(H \mid A_1)}{P(H)} \simeq 0.5026$$

Bài tập 1.59.

Trong học kỳ I năm học 2018 - 2019, sinh viên phải thi 4 học phần. Xác suất để sinh viên thi đạt một học phần trong mỗi lần thi đều là 0,8. Nếu thi không đạt học phần nào phải thi lại học phần đó. Tính xác suất để một sinh viên thi đạt cả 4 học phần trong đó không có học phần nào thi quá 2 lần.

Gọi A_i là "đạt i học phần ở lần thi đầu".

Khi đó, A_0, A_1, A_2, A_3, A_4 tạo thành hệ đầy đủ và $P(A_i) = \binom{4}{i} 0.8^i \times 0.2^{4-i}$

Gọi A là "đạt cả 4 học phần trong đó không có học phần nào thi quá 2 lần". Áp dụng công thức xác suất đầy đủ ta có

$$\begin{aligned}P(A) &= \sum_{i=0}^{4} P(A_i)P(A \mid A_i)\\ &= \binom{4}{0}0.8^0 \times 0.2^4 \times \left(0.8^4\right) + \binom{4}{1}0.8^1 \times 0.2^3 \times \left(0.8^3\right) + \binom{4}{2}0.8^2 \times 0.2^2 \times \left(0.8^2\right)\\ &\qquad + \binom{4}{3}0.8^3 \times 0.2^1 \times (0.8) + \binom{4}{4}0.8^4 \times 0.2^0 \times \left(0.8^0\right)\\ &\simeq 0.8493\end{aligned}$$

Bài tập 1.60.

Ba người thợ cùng may một loại áo với xác suất may được sản phẩm chất lượng cao tương ứng là 0,9; 0,9 và 0,8. Biết một người khi may 8 áo thì có 6 sản phẩm chất lượng cao. Tìm xác suất để người đó may 8 áo nữa thì có 6 áo chất lượng cao.

Gọi A là "trong 8 áo đầu có 6 áo chất lượng cao" và A_i là "8 áo đầu do người thứ i may" thì A_1, A_2, A_3 tạo thành hệ đầy đủ.

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$$

Áp dụng công thức xác suất đầy đủ

$$\begin{aligned}P(A) &= P(A_1)P(A \mid A_1) + P(A_2)P(A \mid A_2) + P(A_3)P(A \mid A_3)\\ &= \frac{1}{3} \times \binom{8}{6}0.9^6 \times 0.1^2 + \frac{1}{3} \times \binom{8}{6}0.9^6 \times 0.1^2 + \frac{1}{3} \times \binom{8}{6}0.8^6 \times 0.2^2\\ &\simeq 0.1971\end{aligned}$$

Gọi B là "trong 8 áo sau có 6 áo chất lượng cao". Vì trong không gian điều kiện A, hệ A_i vẫn là hệ đầy đủ. Áp dụng công thức xác suất đầy đủ có

$$P(B) = P(A_1 \mid A)P(B \mid A_1A) + P(A_2 \mid A)P(B \mid A_2A) + P(A_3 \mid A)P(B \mid A_3A)$$

ở đó

$$P(A_1 \mid A) = \frac{P(A_1)P(A \mid A_1)}{P(A)} \simeq \frac{\frac{1}{3} \times C_8^6 \times 0.9^6 \times 0.1^2}{0.1971} \simeq 0.2516$$

$$P(A_2 \mid A) \simeq 0.2516, \quad P(A_3 \mid A) \simeq 0.4965$$

Thay vào ta tính được

$$\begin{aligned}P(A) &\simeq 0.2516 \times \binom{8}{6}0.9^6 \times 0.1^2 + 0.2516 \times \binom{8}{6}0.9^6 \times 0.1^2\\ &\qquad + 0.4965 \times \binom{8}{6}0.8^6 \times 0.2^2\\ &\simeq 0.2206\end{aligned}$$

2 Biến ngẫu nhiên và luật phân phối xác suất

2.1 Biến ngẫu nhiên rời rạc

Bài tập 2.1.

Một chùm chìa khóa gồm 4 chiếc giống nhau, trong đó chỉ có một chiếc mở được cửa. Người ta thử ngẫu nhiên từng chiếc cho đến khi mở được cửa. Gọi X là số lần thử.

1. Tìm phân phối xác suất của X.
2. Tìm kỳ vọng và phương sai của X.
3. Viết hàm phân phối xác suất của X.

Gọi X là số lần thử thì X là biến ngẫu nhiên rời rạc và nó nhận các giá trị $X = 1, 2, 3, 4$. Gọi X_i là "mở được cửa ở lần thứ i" thì X_1, X_2, X_3, X_4 tạo thành hệ đầy đủ.

i) $X = 1$ nếu mở được cửa ngay lần đầu. Có $P(X = 1) = P(X_1) = \dfrac{1}{4} = 0.25$

ii) $X = 2$ nếu lần đầu không mở được và lần 2 mở được. Có

$$P(X = 2) = P\left(\overline{X_1}X_2\right) = \frac{3}{4}.\frac{1}{3} = \frac{1}{4} = 0.25$$

iii) $X = 3$ là sự kiện $\overline{X_1}\,\overline{X_2}X_3$. Có $P(X = 3) = \dfrac{3}{4}.\dfrac{2}{3}.\dfrac{1}{2} = 0.25$

iv) Tương tự với $X = 4$, có $P(X = 4) = P\left(\overline{X_1}\,\overline{X_2}\,\overline{X_3}X_4\right) = 0.25$

1. Bảng phân phối xác suất của X

X	1	2	3	4
$P(X)$	0.25	0.25	0.25	0.25

2. $E[X] = 1 \times 0.25 + 2 \times 0.25 + 3 \times 0.25 + 4 \times 0.25 = 2.5$

$$\begin{aligned} V[X] &= (1 - 2.5)^2 \times 0.25 + (2 - 2.5)^2 \times 0.25 + (3 - 2.5)^2 \times 0.25 + (4 - 2.5)^2 \times 0.25 \\ &= 1.25 \end{aligned}$$

3. Hàm phân phối của X

$$F_X(x) = \begin{cases} 0, & x \le 1 \\ 0.25, & 1 < x \le 2 \\ 0.5, & 2 < x \le 3 \\ 0.75, & 3 < x \le 4 \\ 1, & x > 4 \end{cases}$$

Bài tập 2.2.

Một xạ thủ có 5 viên đạn. Anh ta phải bắn vào bia với quy định khi nào có 2 viên trúng bia hoặc hết đạn thì dừng. Biết xác suất bắn trúng bia ở mỗi lần bắn là 0,4 và gọi X là số đạn cần bắn.

1. Tìm phân phối xác suất của X.
2. Tìm kỳ vọng, phương sai và viết hàm phân phối xác suất của X.

Gọi X là số đạn cần bắn thì X là biến ngẫu nhiên rời rạc và nhận các giá trị $X = 2, 3, 4, 5$.

i) $X = 2$ có $P(X = 2) = 0.4 \times 0.4 = 0.16$

ii) $X = 3$ xảy ra nếu có 1 trong 2 lần đầu bắn trúng và lần thứ 3 bắn trúng. Bài toán thỏa mãn lược đồ *Bernoulli*, có

$$P(X = 3) = P_2(1) \times 0.4 = 0.192$$

iii) Tương tự $P(X = 4) = P_3(1) \times 0.4 = 0.1728$

iv) $X = 5$ xảy ra nếu cả hết đạn, trượt cả 5 viên hoặc viên cuối trúng và 1 trong 4 lần đầu bắn trúng hoặc chỉ trúng 1 viên duy nhất

$$P(X = 5) = 0.6^5 + P_4(1) \times 0.4 + P_5(1) = 0.4752$$

1. Bảng phân phối xác suất của X

X	2	3	4	5
$P(X)$	0.16	0.192	0.1728	0.4752

2. Theo định nghĩa, ta có $E[X] = 3.9632$ và $V[X] \simeq 1.3059$. Hàm phân phối của X là

$$F_X(x) = \begin{cases} 0, & x \leq 2 \\ 0.16, & 2 < x \leq 3 \\ 0.352, & 3 < x \leq 4 \\ 0.5248, & 4 < x \leq 5 \\ 1, & x > 5 \end{cases}$$

Bài tập 2.3.

Tỷ lệ cử tri ủng hộ ứng cử viên A trong một cuộc bầu cử tổng thống là 40%. Người ta hỏi ý kiến 20 cử tri được chọn một cách ngẫu nhiên. Gọi X là số người bỏ phiếu cho ông A trong 20 người đó.

1. Tìm giá trị trung bình, độ lệch chuẩn của X và $\operatorname{mod} X$.
2. Tìm $P(X = 10)$.

Gọi X là số người bỏ phiếu cho ông A trong 20 người. Khi đó, $X = x$ xảy ra nếu có đúng x người trong $n = 20$ người bầu cho ông A, biết xác suất mỗi người bầu cho ông A là $p = 0.4$ và mọi người bỏ phiếu độc lập với nhau.
Do đó bài toán thỏa mãn lược đồ *Bernoulli*. Như vậy

$$P(X = x) = P_{20}(x) = \binom{20}{x} 0.4^x \times 0.6^{20-x}$$

Hay nói cách khác, X có phân phối nhị thức.

1. $E[X] = np = 8,\ \sigma(X) = \sqrt{V[X]} = \sqrt{np(1-p)} = \sqrt{20 \times 0.4 \times 0.6} \simeq 2.19$ và $\operatorname{mod} X$ chính là số có khả năng nhất trong lược đồ *Bernoulli*

$$\operatorname{mod} X = \lfloor np - q \rfloor + 1 = 8$$

2. $P(X = 10) = P_{20}(10) \simeq 0.1171$

Bài tập 2.4.

Biến ngẫu nhiên rời rạc X chỉ có 2 giá trị x_1 và x_2 ($x_1 < x_2$). Xác suất để X nhận giá trị x_1 là 0,2. Tìm luật phân phối xác suất của X, biết kỳ vọng $E(X) = 2,6$ và độ lệch tiêu chuẩn $\sigma(X) = 0,8$.

Ta có hệ phương trình

$$\begin{cases} 0.2x_1 + 0.8x_2 = E[X] = 2.6 \\ (x_1 - 2.6)^2 \times 0.2 + (x_2 - 2.6)^2 \times 0.8 = \sigma^2(X) = 0.64 \end{cases}$$

Giải ra được $x_1 = 1, x_2 = 3$ và $x_1 = 4.2 > x_2 = 2.2$, loại. Ta thu được bảng phân phối

X	1	3
$P(X)$	0.2	0.8

Bài tập 2.5.

Mỗi khách uống cà phê tại quán cà phê mỗi ngày đều được phát ngẫu nhiên một vé bốc thăm, xác suất khách hàng trúng thăm là 0,1. Nếu khách hàng trúng thăm liên tục trong 5 ngày (từ thứ hai đến thứ sáu) sẽ nhận được 100$, nếu không sẽ không được gì. An uống cà phê liên tục tại quán này 4 tuần liên tiếp. Gọi $X$$ là số tiền An được thưởng khi bốc thăm trong 4 tuần đó. Xác định kỳ vọng và phương sai của X.

Gọi X là số tiền An nhận được khi bốc thăm trong 4 tuần và Y là số tuần An được thưởng thì khi đó

$$X = 100Y$$

và Y là biến ngẫu nhiên rời rạc có phân phối nhị thức với $n = 4$ phép thử độc lập và p là xác suất được thưởng trong 1 tuần bất kì. Dễ tính $p = 0.1^5$
Suy ra $E[X] = 100\ E[Y] = 100 \times 4 \times 0.1^5 = 0.004$ và $V[X] = 10^4\ V[Y] \simeq 0.4$

Bài tập 2.6.

Tung đồng xu 10 lần. Biến ngẫu nhiên X được định nghĩa như sau: $(X = 1)$ nếu sự kiện đúng 3 lần ra mặt sấp xảy ra và $(X = 0)$ trong trường hợp còn lại. Tính kỳ vọng $E(X)$ và phương sai $V(X)$.

X được coi như một kiểu *indicator random variable*.
Gọi A là "đúng 3 lần xảy ra mặt sấp" thì dễ tính được $P(A)$ theo lược đồ *Bernoulli* và có

$$P(X = 1) = P(A) = P_{10}(3) = \binom{10}{3} 0.5^3 \times 0.5^7 \simeq 0.1172$$

Như vậy ta có hàm khối lượng

$$p_X(x) = \begin{cases} 0.1172, & x = 1 \\ 0.8828, & x = 0 \end{cases}$$

Suy ra $E[X] = p = 0.1172$ và $V[X] = p - p^2 \simeq 0.1035$

Bài tập 2.7.

Có 5 sản phẩm trong đó có 4 chính phẩm và 1 phế phẩm. Người ta lấy ra lần lượt hai sản phẩm (lấy không hoàn lại).

1. Gọi X là "số chính phẩm gặp phải". Lập bảng phân phối xác suất của X. Tính $E(X)$ và $V(X)$.
2. Gọi Y là "số phế phẩm gặp phải". Lập hệ thức cho mối quan hệ giữa X và Y.

1. Gọi X là số chính phẩm gặp phải thì nó là biến ngẫu nhiên rời rạc.
   Do chỉ có 1 phế phẩm nên X không thể bằng 0. X nhận giá trị $X = 1;\ X = 2$

   i) $X = 1$ xảy ra nếu ta lấy ra 1 chính, 1 phế. Dễ tính $P(X = 1) = 2 \times \dfrac{4 \times 1}{5 \times 4} = 0.4$

   ii) Tương tự có $P(X = 2) = \dfrac{4 \times 3}{5 \times 4} = 0.6$

   Bảng phân phối xác suất của X

X	1	2
$P(X)$	0.4	0.6

   Suy ra $E[X] = 1.6$ và $V[X] = 0.24$

2. Gọi Y là số phế phẩm gặp lại thì $Y = 2 - X$ vì ta chỉ chọn ra 2 sản phẩm và mỗi sản phẩm chỉ có thể là chính phẩm hoặc phế phẩm

Bài tập 2.8.

Người ta đặt ngẫu nhiên 10 thẻ (trong đó có 5 thẻ màu đỏ và 5 thẻ màu xanh) vào 10 phong bì (5 phong bì có màu đỏ và 5 phong bì có màu xanh), mỗi phong bì một thẻ. Gọi X là số phong bì có chứa một thẻ cùng màu. Tính giá trị:

1. $P(X = 1)$.
2. $E(X)$.

Gọi X là số phong bì có chứa một thẻ cùng màu thì X là biến ngẫu nhiên rời rạc và nó nhận các giá trị $X = 0, 1, \ldots, 10$.

i) $X = 0$ xảy ra nếu 5 phòng bì đỏ chứa 5 thẻ xanh, và 5 phong bì xanh chứa 5 thẻ đỏ. Ta có

$$P(X = 0) = \frac{5!\ 5!}{10!} \simeq 0.004$$

ii) $X = 1$ xảy ra khi có 1 phong bì chứa thẻ cùng màu. Không mất tính tổng quát, giả sử một phong bì đỏ chứa thẻ đỏ.
Khi đó còn lại 4 phong bì đỏ chứa 4 thẻ xanh, 5 phong bì xanh chứa 5 thẻ đỏ.
Điều này là vô lý, do ta chỉ có 10 thẻ. Như vậy, ta có

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

Tương tự, $X = 3, 5, 7, 9$ đều là những sự kiện không thể có

iii) Tương tự $P(X = 2) = \dfrac{C_5^1\ C_5^1\ 1!\ C_5^1\ C_5^1\ 1!\ 4!\ 4!}{10!} \simeq 0.0992$

iv) $P(X = 4) = \dfrac{C_5^2\ C_5^2\ 2!\ C_5^2\ C_5^2\ 2!\ 3!\ 3!}{10!} \simeq 0.3968$

v) $P(X = 6) = \dfrac{C_5^3\ C_5^3\ 3!\ C_5^3\ C_5^3\ 3!\ 2!\ 2!}{10!} \simeq 0.3968$

vi) $P(X = 8) = \dfrac{C_5^4\ C_5^4\ 4!\ C_5^4\ C_5^4\ 4!}{10!} \simeq 0.0992$

vii) $P(X = 10) = \dfrac{5!\ 5!}{10!} \simeq 0.004$

Bảng phân phối xác suất của X

X	0	2	4	6	8	10
$P(X)$	0.004	0.0992	0.3968	0.3968	0.0992	0.004

1. $P(X = 1) = 0$

2. $E[X] = (0 + 10) \times 0.004 + (2 + 8) \times 0.0992 + (4 + 6) \times 0.3968 = 5$

Bài tập 2.9.

Có 2 kiện hàng. Kiện I có 3 sản phẩm tốt và 2 sản phẩm xấu. Kiện II có 2 sản phẩm tốt và 3 sản phẩm xấu. Lấy ngẫu nhiên từ kiện I ra 2 sản phẩm và từ kiện II ra 1 sản phẩm. Lập bảng phân phối xác suất cho biến ngẫu nhiên chỉ số sản phẩm tốt trong 3 sản phẩm lấy ra.

Gọi A_i $(i = 0, 1, 2)$ là "lấy ra i sản phẩm tốt từ kiện I ra" và B_j $(j = 0, 1)$ là "lấy ra j sản phẩm tốt từ kiện II ra" thì A_iB_j tạo thành hệ đầy đủ.
Gọi X là số sản phẩm tốt lấy ra trong 3 sản phẩm thì X là biến ngẫu nhiên rời rạc nhận các giá trị $X = 0, 1, 2, 3$.

i) $X = 0$ chỉ xảy ra khi 2 sản phẩm từ kiện I và 1 sản phẩm từ kiện II là xấu, có nghĩa là $X = 0$ chính là sự kiện A_0B_0. Suy ra $P(A_0B_0) = \dfrac{C_2^2}{C_5^2}.\dfrac{C_3^1}{C_5^1} = 0.06$

ii) Tương tự, $X = 1$ xảy ra nếu lấy ra 2 xấu từ I, 1 tốt từ II hoặc 1 tốt, 1 xấu từ I, 1 xấu từ II, hay $X = 2$ là $A_1B_0 + A_0B_1$, Có

$$P(X = 2) = P(A_1B_0) + P(A_0B_1) = \frac{C_3^1\,C_2^1}{C_5^2}.\frac{C_3^1}{C_5^1} + \frac{C_2^2}{C_5^2}.\frac{C_2^1}{C_5^1} = 0.36 + 0.04 = 0.4$$

iii) $X = 2$ là $A_2B_0 + A_1B_1$. Có

$$P(X = 2) = P(A_2B_0) + P(A_1B_1) = \frac{C_3^2}{C_5^2}.\frac{C_3^1}{C_5^1} + \frac{C_3^1\,C_2^1}{C_5^2}.\frac{C_2^1}{C_5^1} = 0.18 + 0.24 = 0.42$$

iv) $X = 3$ là A_2B_1. Suy ra $P(X = 3) = P(A_2B_1) = \dfrac{C_2^2}{C_5^2}.\dfrac{C_2^1}{C_5^1} = 0.12$

Bảng phân phối xác suất của X

X	0	1	2	3
$P(X)$	0.06	0.4	0.42	0.12

Bài tập 2.10.

Có hai kiện hàng. Kiện thứ nhất có 8 sản phẩm tốt và 2 sản phẩm xấu. Kiện thứ hai có 5 sản phẩm tốt và 3 sản phẩm xấu. Lấy ngẫu nhiên 2 sản phẩm từ kiện I bỏ sang kiện II. Sau đó từ kiện II lấy ngẫu nhiên ra 2 sản phẩm. Lập bảng phân phối xác suất của biến ngẫu nhiên chỉ số sản phẩm tốt có trong 2 sản phẩm lấy ra từ kiện II.

Gọi A_i $(i = 0, 1, 2)$ là "lấy được i sản phẩm tốt từ kiện I sang kiện II" thì A_i tạo thành hệ đầy đủ với

$$P(A_0) = \frac{C_2^2}{C_{10}^2} = \frac{1}{45},\ P(A_1) = \frac{C_8^1\,C_2^1}{C_{10}^2} = \frac{16}{45},\ P(A_2) = \frac{C_8^2}{C_{10}^2} = \frac{28}{45}$$

Gọi X là số sản phẩm tốt trong 2 sản phẩm lấy ra từ kiện II thì X là biến ngẫu nhiên rời rạc và nó nhận các giá trị $X = 0, 1, 2$. Áp dụng công thức xác suất đầy đủ ta có

$$P(X = x) = \sum_{i=0}^{2} P(X = x \mid A_i)$$

suy ra

$$P(X = 0) = \frac{1}{45}.\frac{C_5^2}{C_{10}^2} + \frac{16}{45}.\frac{C_4^2}{C_{10}^2} + \frac{28}{45}.\frac{C_3^2}{C_{10}^2} \simeq 0.0938$$
$$P(X = 1) = \frac{1}{45}.\frac{C_5^1\,C_5^1}{C_{10}^2} + \frac{16}{45}.\frac{C_6^1\,C_4^1}{C_{10}^2} + \frac{28}{45}.\frac{C_7^1\,C_3^1}{C_{10}^2} \simeq 0.4923$$
$$P(X = 2) = \frac{1}{45}.\frac{C_5^2}{C_{10}^2} + \frac{16}{45}.\frac{C_6^2}{C_{10}^2} + \frac{28}{45}.\frac{C_7^2}{C_{10}^2} \simeq 0.4139$$

Bảng phân phối xác suất của X

X	0	1	2
$P(X)$	0.0938	0.4923	0.4138

Bài tập 2.11.

Gieo hai con xúc sắc đồng chất 5 lần, gọi X là số lần xuất hiện hai mặt 6.

1. Tính xác suất của sự kiện số lần xuất hiện hai mặt 6 ít nhất là 2.
2. Tính $E(X), V(X)$.
3. Viết hàm phân phối $F_X(x)$.

Gọi X là số lần xuất hiện hai mặt 6 thì nó là biến ngẫu nhiên liên tục nhận giá trị $X = 0, \ldots, 5$. Dễ thấy X có phân phối nhị thức, do 5 lần gieo là độc lập và xác suất mỗi lần xuất hiện hai mặt 6 là $p = \frac{1}{6}.\frac{1}{6} = \frac{1}{36}$.
Hàm khối lượng xác suất

$$p_X(x) = P_5(x) = \binom{5}{x}\left(\frac{1}{36}\right)^x\left(\frac{35}{36}\right)^{5-x}$$

Áp dụng công thức, thu được bảng phân phối xác suất của X

X	0	1	2	3	4	5
$P(X)$	0.86861	0.12409	0.00709	2.025×10^{-4}	2.89×10^{-6}	1.65×10^{-8}

1. Xác suất cần tính là $1 - (p_X(0) + p_X(1)) \simeq 1 - 0.9927 \simeq 0.0073$
2. Dễ có $E[X] = np = \dfrac{5}{36} \simeq 0.1389$ và $V[X] = np(1-p) \simeq 0.135$
3. Hàm phân phối của X là

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ 0.86861, & 0 < x \leq 1 \\ 0.9927, & 1 < x \leq 2 \\ 0.99979, & 2 < x \leq 3 \\ 0.09999925, & 3 < x \leq 4 \\ 0.099999539, & 4 < x \leq 5 \\ 1, & x > 5 \end{cases}$$

Chú ý: Trong trường hợp này do $P(X = 3, 4, 5)$ rất nhỏ nên $X = 3, 4, 5$ là các sự kiện gần như không bao giờ xảy ra. Ta cũng có thể coi $P(X = 3, 4, 5) \simeq 0$ để tính toán.

Bài tập 2.12.

Một thanh niên nam vào cửa hàng thấy 5 máy thu thanh giống nhau. Anh ta đề nghị cửa hàng cho anh ta thử lần lượt các máy đến khi chọn được máy tốt thì mua, nếu cả 5 lần đều xấu thì thôi. Biết rằng xác suất để một máy xấu là 0,6 và các máy xấu tốt độc lập với nhau. Gọi X là số lần thử. Lập bảng phân phối xác suất của X.

Gọi X là số lần thử thì X là biến ngẫu nhiên rời rạc nhận các giá trị $X = 1, 2, 3, 4, 5$. Ta thấy rằng

i) $X = x,\ (x = 1, 2, 3, 4)$ xảy ra nếu $x - 1$ lần đầu không chọn được máy tốt và lần thứ x chọn được máy tốt.

$$P(X = x) = 0.6^{x-1} \times 0.4 \quad x = 1, 2, 3, 4$$

ii) $X = 5$ xảy ra nếu lần cuối chọn được máy tốt hoặc cả 5 lần đều không chọn được máy tốt.

$$P(X = 5) = 0.6^4 \times 0.4 + 0.6^5 = 0.1296$$

Bảng phân phối xác suất của X

X	1	2	3	4	5
$P(X)$	0.4	0.24	0.144	0.0864	0.1296

Bài tập 2.13.

Có hai hộp bi. Hộp I có 2 bi trắng, 3 bi đỏ. Hộp II có 2 bi trắng, 2 bi đỏ. Lấy ngẫu nhiên 2 bi từ hộp I bỏ sang hộp II, sau đó lại lấy ngẫu nhiên 3 bi từ hộp II bỏ vào hộp I. Lập bảng phân phối xác suất của biến ngẫu nhiên chỉ số bi trắng có mặt ở hộp I, hộp II sau khi đã chuyển xong.

Gọi X và Y lần lượt là số bi trắng ở hộp I, hộp II sau khi đã chuyển xong. Dễ thấy ta có liên hệ $X+Y=4$. Do đó, ta chỉ cần tìm phân phối của X là đủ.
Dễ thấy X là biến ngẫu nhiên rời rạc. Vì $X+Y=4$ và $X,Y\geq 0$ nên chúng nhận các giá trị $1,2,3,4$.

i) $X=1$ xảy ra nếu lấy ra 2 trắng, lấy về 1 trắng, 2 đỏ hoặc lấy ra 1 trắng, 1 đỏ, lấy về 3 đỏ

$$P(X=1)=\frac{C_2^2}{C_5^2}\frac{C_4^1\,C_2^2}{C_6^3}+\frac{C_2^1\,C_3^1}{C_5^2}\frac{C_3^3}{C_6^3}=0.05$$

ii) $X=2$ xảy ra nếu lấy ra 2 trắng, lấy về 2 trắng, 1 đỏ hoặc lấy ra 1 trắng, 1 đỏ, lấy về 1 trắng, 2 đỏ hoặc lấy ra 2 đỏ, lấy về 3 đỏ

$$P(X=2)=\frac{C_2^2}{C_5^2}\frac{C_4^2\,C_2^1}{C_6^3}+\frac{C_2^1\,C_3^1}{C_5^2}\frac{C_3^1\,C_3^2}{C_6^3}+\frac{C_3^2}{C_5^2}\frac{C_4^3}{C_6^3}=0.39$$

iii) $X=3$ xảy ra nếu lấy ra 2 trắng, lấy về 3 trắng hoặc lấy ra 1 trắng, 1 đỏ, lấy về 2 trắng, 1 đỏ hoặc lấy ra 2 đỏ, lấy về 1 trắng, 2 đỏ

$$P(X=3)=\frac{C_2^2}{C_5^2}\frac{C_4^3}{C_6^3}+\frac{C_2^1\,C_3^1}{C_5^2}\frac{C_3^1\,C_3^2}{C_6^3}+\frac{C_3^2}{C_5^2}\frac{C_2^1\,C_4^2}{C_6^3}=0.47$$

iv) $X=4$ xảy ra nếu lấy ra 1 trắng, 1 đỏ, lấy về 3 trắng hoặc lấy ra 2 đỏ, lấy về 2 trắng, 1 đỏ

$$P(X=4)=\frac{C_2^1\,C_3^1}{C_5^2}\frac{C_3^3}{C_6^3}+\frac{C_3^2}{C_5^2}\frac{C_2^2\,C_4^1}{C_6^3}=0.09$$

Bảng phân phối xác suất của X

X	1	2	3	4
$P(X)$	0.05	0.39	0.47	0.09

Chú ý: Nếu gọi A_i là "có i bi trắng trong 2 bi chuyển từ hộp I sang hộp II" và B_j là "có j bi trắng trong 3 bi chuyển từ hộp II sang hộp I" thì A_iB_j tạo thành hệ đầy đủ.
Ta có thể trình bày lại lời giải trên theo cách biểu diễn các sự kiện $X=x$ qua A_i,B_j.
Ở đây, ta không áp dụng trực tiếp công thức đầy đủ mà chỉ tính xác suất điều kiện của các sự kiện có thể xảy ra trong hệ đầy đủ, bỏ qua các sự kiện có xác suất bằng 0.

Bài tập 2.14.

Một người đi làm từ nhà đến cơ quan phải qua 3 ngã tư. Xác suất để người đó gặp đèn đỏ ở các ngã tư tương ứng là 0,2; 0,4 và 0,5. Gọi X là số đèn đỏ mà người đó gặp phải trong một lần đi làm (giả sử 3 đèn giao thông ở ngã tư hoạt động độc lập với nhau).

1. Lập bảng phân phối xác suất của X. Tính kỳ vọng, phương sai của X. Tìm hàm phân phối xác suất của X.

2. Hỏi thời gian trung bình phải ngừng trên đường là bao nhiêu biết rằng mỗi khi gặp đèn đỏ người ấy phải đợi khoảng 3 phút.

Gọi X là số đèn đỏ người đó gặp phải thì X là biến ngẫu nhiên rời rạc nhận các giá trị $X = 0, 1, 2, 3$.
Sử dụng công thức cộng, công thức nhân, có

i) $P(X = 0) = 0.8 \times 0.6 \times 0.5 = 0.24$

ii) $P(X = 1) = 0.2 \times 0.6 \times 0.5 + 0.8 \times 0.4 \times 0.5 + 0.8 \times 0.6 \times 0.5 = 0.46$

iii) $P(X = 2) = 0.2 \times 0.4 \times 0.5 + 0.8 \times 0.4 \times 0.5 + 0.2 \times 0.6 \times 0.5 = 0.26$

iv) $P(X = 3) = 0.2 \times 0.4 \times 0.5 = 0.04$

1. Bảng phân phối xác suất của X

X	0	1	2	3
$P(X)$	0.24	0.46	0.26	0.04

Ta tính được $E[X] = 1.1$ và $V[X] = 0.65$
Hàm phân phối của X là

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ 0.24, & 0 < x \leq 1 \\ 0.7, & 1 < x \leq 2 \\ 0.96, & 2 < x \leq 3 \\ 1, & x > 3 \end{cases}$$

2. Gọi Y là thời gian phải ngừng trên đường thì $Y = 3X$ (phút). Từ đó suy ra

$$E[Y] = 3\,E[X] = 3.3 \quad \text{(phút)}$$

Bài tập 2.15.

Một người chơi trò chơi tung con xúc sắc cân đối đồng chất ba lần. Nếu cả ba lần đều xuất hiện mặt 6 thì thu về 36\$, nếu hai lần xuất hiện mặt 6 thì thu về 2,8\$, nếu một lần xuất hiện mặt 6 thì thu về 0,4\$. Biết rằng khi chơi người đó phải nộp x\$.

1. Tìm x sao cho trò chơi là vô thưởng vô phạt.

2. x bằng bao nhiêu thì trung bình mỗi lần chơi, người chơi mất 1\$?

Gọi X là số tiền người chơi thu về sau 3 lần thì X là biến ngẫu nhiên rời rạc nhận các giá trị $X = 36,\ 2.8,\ 0.4,\ 0$
Ta có

$$P(X = 36) = \frac{1}{6^3},\ P(X = 2.8) = \frac{3.5}{6^3},\ P(X = 0.4) = \frac{3.5^2}{6^3},\ P(X = 0) = \frac{5^3}{6^3}$$

1. Trò chơi là vô thưởng vô phạt nếu $E[X] = x$, hay $x = 0.5055$
2. Điều kiện này có nghĩa là $E[X] = x - 1$, hay $x = 1.5055$

Bài tập 2.16.

Một kiện hàng có 12 sản phẩm, trong đó có 7 sản phẩm loại I và 5 sản phẩm loại II. Khi bán được một sản phẩm loại I thì được lãi 50 ngàn đồng; còn nếu bán được một sản phẩm loại II thì được lãi 20 ngàn đồng. Lấy ngẫu nhiên từ kiện hàng ra 3 sản phẩm.

1. Tìm quy luật phân phối xác suất của số tiền lãi thu được do bán 3 sản phẩm đó; tính kỳ vọng, phương sai của số tiền lãi thu được do bán 3 sản phẩm đó.
2. Viết hàm phân phối, vẽ đồ thị hàm phân phối của số tiền lãi thu được khi bán 3 sản phẩm đó.

Gọi X là số sản phẩm loại I trong 3 sản phẩm lấy ra, thì X là biến ngẫu nhiên rời rạc nhận các giá trị $X = 0, 1, 2, 3$.
Gọi Y (ngàn đồng) là số tiền lãi thu được do bán 3 sản phẩm thì Y là biến ngẫu nhiên rời rạc có thỏa mãn

$$Y = 50X + 20(3 - X) = 30X + 60$$

Dễ tính được

$$P(X = 0) = \frac{C_5^3}{C_{12}^3} = \frac{1}{22},\ P(X = 1) = \frac{C_7^1\,C_5^2}{C_{12}^3} = \frac{7}{22},$$
$$P(X = 2) = \frac{C_7^2\,C_5^1}{C_{12}^3} = \frac{21}{44},\ P(X = 3) = \frac{C_7^3}{C_{12}^3} = \frac{7}{44}$$

1. Bảng phân phối xác suất của Y

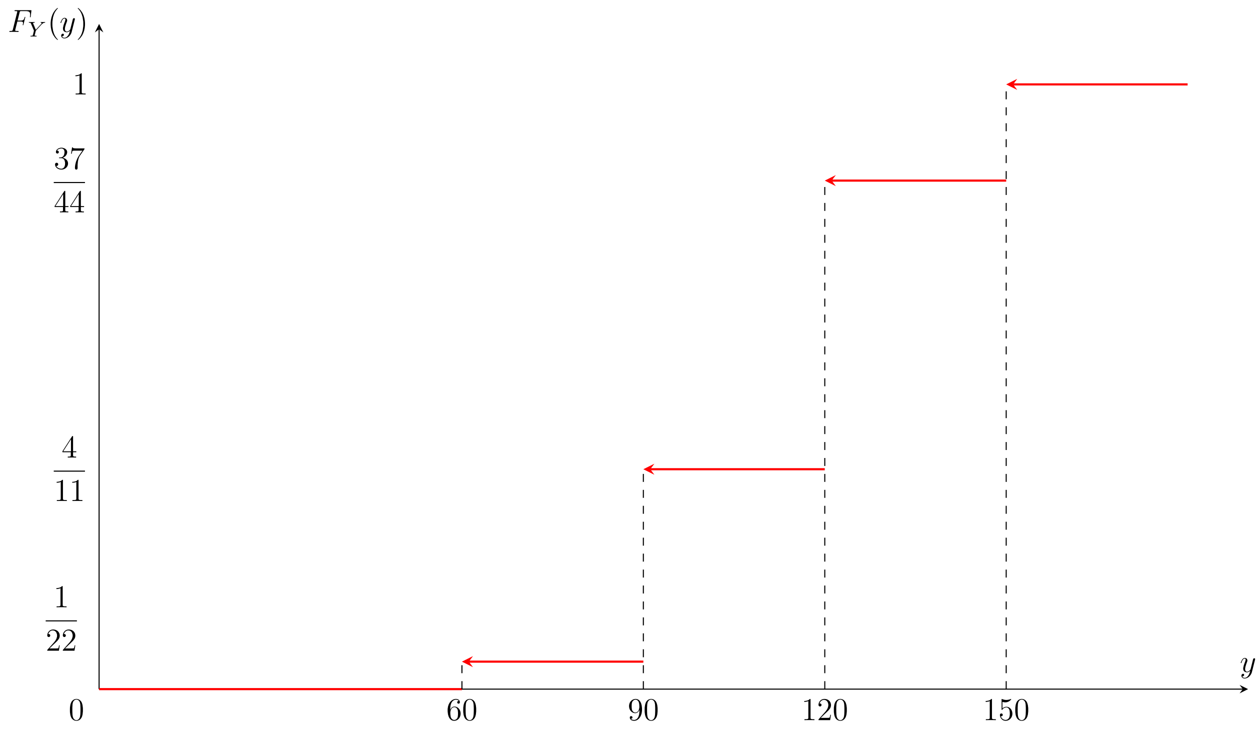
Y	60	90	120	150
$P(Y)$	$1/22$	$7/22$	$21/44$	$7/44$

   Tính được $E[Y] = 112.5$ và $V[Y] \simeq 536.93$

2. Hàm phân phối của Y là

$$F_Y(y) = \begin{cases} 0, & y \le 60 \\ \frac{1}{22}, & 60 < y \le 90 \\ \frac{4}{11}, & 90 < y \le 120 \\ \frac{37}{44}, & 120 < y \le 150 \\ 1, & y > 150 \end{cases}$$

   Đồ thị của hàm phân phối $F_Y(y)$



Hình 1: Hàm phân phối xác suất của biến ngẫu nhiên Y

Bài tập 2.17.

Một hộp đựng 15 quả bóng bàn trong đó có 10 quả còn mới. Lần đầu ta lấy ra 3 quả để thi đấu, sau đó lại trả 3 quả đó vào hộp. Lần thứ hai lại lấy ra 3 quả. Gọi X là biến ngẫu nhiên chỉ số quả bóng mới trong 3 quả lấy ra. Lập bảng phân phối xác suất, tính kì vọng, phương sai của X

Gọi A_i $(i = 0, 1, 2, 3)$ là "số quả mới lấy ra ở lần đầu" thì A_i tạo thành hệ đầy đủ với

$$P(A_0) = \frac{C_5^3}{C_{15}^3} = \frac{2}{91},\ P(A_1) = \frac{C_{10}^1\ C_5^2}{C_{15}^3} = \frac{20}{91},\ P(A_2) = \frac{C_{10}^2\ C_5^1}{C_{15}^3} = \frac{45}{91},\ P(A_3) = \frac{C_{10}^3}{C_{15}^3} = \frac{24}{91}$$

Gọi X là số bóng mới trong 3 quả lấy ra (lần sau) thì X là biến ngẫu nhiên rời rạc nhận các giá trị $0, 1, 2, 3$. Áp dụng công thức xác suất đầy đủ ta có

i) $P(X = 0) = \frac{2}{91}\frac{C_5^3}{C_{15}^3} + \frac{20}{91}\frac{C_6^3}{C_{15}^3} + \frac{45}{91}\frac{C_7^3}{C_{15}^3} + \frac{24}{91}\frac{C_8^3}{C_{15}^3} \simeq 0.0806$

ii) $P(X = 1) = \frac{2}{91}\frac{C_{10}^1\ C_5^2}{C_{15}^3} + \frac{20}{91}\frac{C_9^1\ C_6^2}{C_{15}^3} + \frac{45}{91}\frac{C_8^1\ C_7^2}{C_{15}^3} + \frac{24}{91}\frac{C_7^1\ C_8^2}{C_{15}^3} \simeq 0.3663$

iii) $P(X = 2) = \frac{2}{91}\frac{C_{10}^2\ C_5^1}{C_{15}^3} + \frac{20}{91}\frac{C_9^2\ C_6^1}{C_{15}^3} + \frac{45}{91}\frac{C_8^2\ C_7^1}{C_{15}^3} + \frac{24}{91}\frac{C_7^2\ C_8^1}{C_{15}^3} \simeq 0.4256$

iv) $P(X = 3) = \frac{2}{91}\frac{C_{10}^3}{C_{15}^3} + \frac{20}{91}\frac{C_9^3}{C_{15}^3} + \frac{45}{91}\frac{C_8^3}{C_{15}^3} + \frac{24}{91}\frac{C_7^3}{C_{15}^3} \simeq 0.1275$

Bảng phân phối xác suất của X

X	0	1	2	3
$P(X)$	0.0806	0.3663	0.2456	0.1275

Dễ tính được $E[X] = 1.6$ và $V[X] = 0.6562$

Bài tập 2.18.

Một cơ sở thí nghiệm có 3 phòng thí nghiệm như nhau. Xác suất thực hiện thành công một thí nghiệm của các phòng lần lượt là 0,6; 0,7 và 0,8. Một sinh viên chọn một phòng thí nghiệm bất kỳ và tiến hành 3 thí nghiệm độc lập. Gọi X là số thí nghiệm thành công.

1. Lập bảng phân phối xác suất của X, tính kỳ vọng $E(X)$ và phương sai $V(X)$.
2. Theo anh (chị) thì khả năng chắc chắn sẽ thành công mấy thí nghiệm?

Gọi A_i $(i = 1, 2, 3)$ là "sinh viên chọn phòng thí nghiệm thứ i" thì A_i là hệ đầy đủ với $P(A_1) = P(A_2) = P(A_3) = \dfrac{1}{3}$.
Gọi X là số thí nghiệm thành công thì X là biến ngẫu nhiên rời rạc nhận các giá trị $0, 1, 2, 3$.
Áp dụng công thức xác suất đầy đủ và công thức *Bernoulli* ta có

i) $P(X = 0) = \dfrac{1}{3}\left(0.4^3 + 0.3^3 + 0.2^3\right) = 0.033$

ii) $P(X = 1) = \dfrac{1}{3}\dbinom{3}{1}\left[0.6 \times 0.4^2 + 0.7 \times 0.3^2 + 0.8 \times 0.2^2\right] = 0.191$

iii) $P(X = 2) = \dfrac{1}{3}\dbinom{3}{2}\left[0.6^2 \times 0.4 + 0.7^2 \times 0.3 + 0.8^2 \times 0.2\right] = 0.419$

iv) $P(X = 3) = \dfrac{1}{3}\left(0.6^3 + 0.7^3 + 0.8^3\right) = 0.357$

1. Bảng phân phối xác suất của X

X	0	1	2	3
$P(X)$	0.033	0.191	0.4334	0.357

   Từ đó có được $E[X] = 2.1288$ và $V[X] \simeq 0.6711$

2. Số thí nghiệm chắc chắn nhất về khả năng thành công chính là điểm mà tại đó xác suất là lớn nhất: $\text{mod}\, X = 2$

2.2 Biến ngẫu nhiên liên tục

Bài tập 2.19.

Biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f_X(x) = \begin{cases} k\sin 3x, & x \in \left(0, \dfrac{\pi}{3}\right) \\ 0, & x \notin \left(0, \dfrac{\pi}{3}\right) \end{cases}$$

1. Xác định k và hàm phân phối $F_X(x)$.
2. Tính $P\left(\dfrac{\pi}{6} \le x \le \dfrac{\pi}{3}\right)$

1. Ta giải hệ phương trình

$$\begin{cases} k\sin 3x \ge 0,\ \forall x \in \left(0, \dfrac{\pi}{3}\right) \\ \displaystyle\int_{-\infty}^{+\infty} k\sin 3x\, dx = 1 \end{cases} \Rightarrow \begin{cases} k \ge 0, \\ \displaystyle\int_0^{\frac{\pi}{3}} k\sin 3x\, dx = 1 \end{cases} \Rightarrow k = \frac{3}{2}$$

Thử lại. Hàm phân phối $F_X(x)$

$$F_X(x) = \begin{cases} \displaystyle\int_{-\infty}^{x} 0\, dt, & x \le 0 \\ \displaystyle\int_0^x \frac{3}{2}\sin 3t\, dt, & 0 < x \le \dfrac{\pi}{3} \\ \displaystyle\int_0^{\frac{\pi}{3}} \frac{3}{2}\sin 3t\, dt, & x > \dfrac{\pi}{3} \end{cases}$$

Rút gọn ta được

$$F_X(x) = \begin{cases} 0, & x \le 0 \\ \dfrac{1}{2} - \dfrac{\cos 3x}{2}, & 0 < x \le \dfrac{\pi}{3} \\ 1, & x > \dfrac{\pi}{3} \end{cases}$$

2. $P\left(\dfrac{\pi}{6} \le x \le \dfrac{\pi}{3}\right) = F_X\left(\dfrac{\pi}{3}\right) - F_X\left(\dfrac{\pi}{6}\right) = \dfrac{1}{2}$

Bài tập 2.20.

Biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f_X(x) = \frac{c}{e^x + e^{-x}}$$

Xác định hằng số c và sau đó tính kỳ vọng của X.

Giải hệ phương trình

$$\begin{cases} \dfrac{c}{e^x + e^{-x}} \geq 0, \ \forall x \\ \displaystyle\int_{-\infty}^{+\infty} \frac{c}{e^x + e^{-x}}\, dx = 1 \end{cases} \Rightarrow \begin{cases} c \geq 0, \\ \displaystyle\int_{-\infty}^{+\infty} \frac{c\, e^x}{e^{2x} + 1}\, dx = 1 \end{cases} \Rightarrow \begin{cases} c \geq 0, \\ \dfrac{\pi}{2}\, c = 1 \end{cases} \Rightarrow c = \frac{2}{\pi}$$

Thử lại. Kỳ vọng của X

$$E[X] = \frac{2}{\pi} \int_{-\infty}^{+\infty} \frac{x}{e^x + e^{-x}}\, dx = 0,$$

vì $\dfrac{x}{e^x + e^{-x}}$ là hàm lẻ

Bài tập 2.21.

Biến ngẫu nhiên liên tục X có hàm mật độ là $f_X(x) = ae^{-|x|}, \ (-\infty < x < \infty)$

1. Xác định a.
2. Tìm hàm phân phối của biến ngẫu nhiên X, biến ngẫu nhiên $Y = X^2$.
3. Tìm $E(X), V(X)$.
4. Tính xác suất để sau ba lần lặp lại phép thử một cách độc lập có 2 lần X nhận giá trị trong khoảng $(0; \ln 3)$.

1. Ta giải hệ phương trình

$$\begin{cases} ae^{-|x|} \geq 0, \ \forall x \\ \displaystyle\int_{-\infty}^{+\infty} ae^{-|x|}\, dx = 1 \end{cases} \Rightarrow \begin{cases} a \geq 0, \\ 2\displaystyle\int_{0}^{+\infty} ae^{-x}\, dx = 1 \end{cases} \Rightarrow a = \frac{1}{2}$$

   Thử lại.

2. Hàm phân phối $F_X(x)$

$$F_X(x) = \begin{cases} \dfrac{1}{2}\displaystyle\int_{-\infty}^{x} e^t\, dt, & x \leq 0 \\ \dfrac{1}{2}\displaystyle\int_{-\infty}^{0} e^t\, dt + \dfrac{1}{2}\displaystyle\int_{0}^{x} e^{-t}\, dt, & 0 < x < +\infty \end{cases}$$

   Rút gọn ta được

$$F_X(x) = \begin{cases} \dfrac{e^x}{2}, & x \leq 0 \\ 1 - \dfrac{e^{-x}}{2}, & x > 0 \end{cases}$$

Có $Y = X^2$

$$F_Y(y) = P(Y < y) = P(X^2 < y) = \begin{cases} 0, & y \leq 0 \\ P\left(-\sqrt{y} < X < \sqrt{y}\right), & y > 0 \end{cases}$$

Suy ra hàm phân phối của Y

$$F_Y(y) = \begin{cases} 0, & y \leq 0 \\ 1 - \dfrac{e^{-\sqrt{y}}}{2}, & y > 0 \end{cases}$$

3. Dễ có $E[X] = 0$ vì $xf_X(x) = \dfrac{1}{2}\,xe^{-|x|}$ là hàm lẻ

$$V[X] = \frac{1}{2}\int_{-\infty}^{+\infty} x^2 e^{-|x|}\,dx = 2.\frac{1}{2}\int_0^{+\infty} x^2 e^{-|x|}\,dx = \int_0^{+\infty} x^2 e^{-x}\,dx = \Gamma(3) = 2! = 2$$

4. Xác suất để X nhận giá trị trong khoảng $(0, \ln 3)$ là $p = F_X(\ln 3) - F_X(0) = \dfrac{1}{3}$.

Bài toán thỏa mãn lược đồ *Bernoulli* với $n = 3$ và $p = \dfrac{1}{3}$. Suy ra xác suất cần tìm là

$$P_3(2) = \binom{3}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^1 = \frac{2}{9} \simeq 0.2222$$

Bài tập 2.22.

Nhu cầu hàng năm về loại hàng A là biến ngẫu nhiên liên tục X có hàm mật độ xác suất như sau (đơn vị: ngàn sản phẩm):

$$f_X(x) = \begin{cases} k(30 - x), & x \in (0, 30) \\ 0, & x \notin (0, 30) \end{cases}$$

1. Tìm k.

2. Tìm hàm phân phối $F_X(x)$.

3. Tìm nhu cầu trung bình hàng năm về loại hàng đó.

1. Ta giải hệ phương trình

$$\begin{cases} k(30 - x) \geq 0, \ \forall x \in (0, 30) \\ \displaystyle\int_{-\infty}^{+\infty} k(30 - x)\,dx = 1 \end{cases} \Rightarrow \begin{cases} k \geq 0, \\ \displaystyle\int_0^{30} k(30 - x)\,dx = 1 \end{cases} \Rightarrow k = \frac{1}{450}$$

Thử lại.

2. Hàm phân phối $F_X(x)$

$$F_X(x) = \begin{cases} \int_{-\infty}^{x} 0\,dt, & x \leq 0 \\ \dfrac{1}{450}\int_{0}^{x}(30-t)\,dt, & 0 < x \leq 30 \\ \dfrac{1}{450}\int_{0}^{30}(30-t)\,dt, & x > 30 \end{cases}$$

Rút gọn ta được

$$F_X(x) = \begin{cases} 0 \\ \dfrac{x}{15} - \dfrac{x^2}{900}, & 0 < x \leq 30 \\ 1 \end{cases}$$

3. Nhu cầu trung bình hàng năm

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_{0}^{30} \frac{1}{450}\, x(30-x)\,dx = 10$$

Bài tập 2.23.

Cho biến ngẫu nhiên liên tục X có hàm phân phối xác suất

$$F_X(x) = \begin{cases} 0, & x \leq 0, \\ \dfrac{1}{2} - k\cos x, & 0 < x \leq \pi \\ 1, & x > \pi \end{cases}$$

1. Tìm k.

2. Tìm $P\left(0 < X < \dfrac{\pi}{2}\right)$.

3. Tìm $E(X)$.

1. Vì $F_X(x)$ là hàm phân phối của biến ngẫu nhiên liên tục X nên nó là hàm liên tục.

Giải hệ phương trình $\begin{cases} F_X(0^-) = F_X(0^+) \\ F_X(\pi^-) = F_X(\pi^+) \end{cases} \Rightarrow \begin{cases} 0 = \dfrac{1}{2} - k, \\ \dfrac{1}{2} + k = 1 \end{cases} \Rightarrow k = \dfrac{1}{2}$

Thử lại.

2. $P\left(0 < X < \dfrac{\pi}{2}\right) = F_X\left(\dfrac{\pi}{2}\right) - F_X(0) = \dfrac{1}{2}$

3. Tìm được hàm mật độ

$$f_X(x) = \begin{cases} \dfrac{1}{2}\sin x, & x \in (0, \pi) \\ 0, & x \notin (0, \pi) \end{cases}$$

nên X có kỳ vọng là

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)\, dx = \frac{1}{2}\int_0^{\pi} x \sin x\, dx = \frac{\pi}{2}$$

Bài tập 2.24.

Cho biến ngẫu nhiên liên tục X có hàm phân phối xác suất

$$F_X(x) = \begin{cases} 0, & x \leq -a, \\ A + B \arcsin \dfrac{x}{a}, & x \in (-a, a) \\ 1, & x \geq a \end{cases}$$

1. Tìm A và B.

2. Tìm hàm mật độ xác suất $f_X(x)$

1. Vì $F_X(x)$ là hàm phân phối của biến ngẫu nhiên liên tục X nên nó là hàm liên tục.

Giải hệ phương trình $\begin{cases} F_X(-a^-) = F_X(-a^+) \\ F_X(a^-) = F_X(a^+) \end{cases} \Rightarrow \begin{cases} 0 = A - \dfrac{\pi}{2} B, \\ A + \dfrac{\pi}{2} B = 1 \end{cases} \Rightarrow \begin{cases} A = \dfrac{1}{2} \\ B = \dfrac{1}{\pi} \end{cases}$

Thử lại.

2. Từ biểu thức $f_X(x) = F'_X(x)$, ta tìm được hàm mật độ

$$f_X(x) = \begin{cases} \dfrac{1}{\pi\sqrt{a^2 - x^2}}, & x \in (-a, a) \\ 0, & x \notin (-a, a) \end{cases}$$

Bài tập 2.25.

Hàm phân phối xác suất của biến ngẫu nhiên liên tục X có dạng

$$F_X(x) = a + b\arctan x, \ (-\infty < x < \infty).$$

1. Tìm hệ số a và b.
2. Tìm hàm mật độ xác suất $f_X(x)$.
3. Tìm xác suất để khi tiến hành 3 phép thử độc lập có 2 lần X nhận giá trị trong khoảng $(-1, 1)$.

1. Vì $F_X(x)$ là hàm phân phối nên ta phải có

$$\begin{cases} F_X(-\infty) = 0 \\ F_X(+\infty) = 1 \end{cases} \Rightarrow \begin{cases} a - \dfrac{\pi}{2}\ b = 0 \\ a + \dfrac{\pi}{2}\ b = 1 \end{cases} \Rightarrow \begin{cases} a = \dfrac{1}{2} \\ b = \dfrac{1}{\pi} \end{cases}$$

Thử lại.

2. Hàm mật độ xác suất $f_X(x) = F'_X(x) = \dfrac{1}{\pi(1+x^2)}$

3. Xác suất X nhận giá trị trong khoảng $(-1, 1)$ là

$$p = P(-1 < X < 1) = F_X(1) - F_X(-1) = \frac{1}{2}$$

Bài toán thỏa mãn lược đồ *Bernoulli* với $n = 3$ và $p = 0.5$. Xác suất cần tính là

$$P_3(2) = \binom{3}{2}\ 0.5^2 \times 0.5^1 = 0.375$$

Bài tập 2.26.

Biến ngẫu nhiên X liên tục trên toàn trục số và có hàm phân phối xác suất $F_X(x) = \dfrac{1}{2} + \dfrac{1}{\pi}\arctan\dfrac{x}{2}$. Tìm giá trị có thể có của x_1 thỏa mãn điều kiện $P(X > x_1) = \dfrac{1}{4}$.

Theo định nghĩa, ta dễ dàng tìm được x_1:

$$P(X > x_1) = \frac{1}{4} \Leftrightarrow 1 - F_X(x_1) = \frac{1}{4} \Leftrightarrow 1 - \left(\frac{1}{2} + \frac{1}{\pi}\arctan\frac{x}{2}\right) = \frac{1}{4}$$

Giải ra được $x_1 = 2$

Bài tập 2.27.

Thu nhập của dân cư tại một vùng là biến ngẫu nhiên liên tục có hàm phân phối xác suất như sau:

$$F_X(x) = \begin{cases} 1 - \left(\dfrac{x_0}{x}\right)^{\alpha}, & x \leq x_0,\ \alpha > 0 \\ 0, & x < x_0 \end{cases}$$

Hãy xác định mức thu nhập sao cho lấy ngẫu nhiên một người ở vùng đó thì thu nhập của người này vượt quá mức trên với xác suất 0,5.

Gọi X là thu nhập của dân cư tại một vùng thì X là biến ngẫu nhiên liên tục có hàm phân phối

$$F_X(x) = \begin{cases} 1 - \left(\dfrac{x_0}{x}\right)^{\alpha}, & x \leq x_0,\ \alpha > 0 \\ 0, & x < x_0 \end{cases}$$

Ta cần tìm x thỏa mãn điều kiện $P(X > x) = 0.5$. Tương tự như bài trên:

$$P(X > x) = 0.5 \Leftrightarrow 1 - F_X(x) = 0.5 \Leftrightarrow F_X(x) = 0.5 \Leftrightarrow 1 - \left(\frac{x_0}{x}\right)^{\alpha} = 0.5$$

Giải ra được $x = \dfrac{x_0}{\log_{\alpha} 0.5} > x_0$

Bài tập 2.28.

Thời gian phục vụ mỗi khách hàng tại một cửa hàng ăn nhanh là biến ngẫu nhiên X tuân theo quy luật lũy thừa với hàm mật độ xác suất

$$f_X(x) = \begin{cases} 5e^{-5x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

với x được tính bằng phút/khách hàng.

1. Tìm xác suất để thời gian phục vụ một khách hàng nào đó sẽ nằm trong khoảng $(0,4;1)$ (phút).

2. Tính thời gian trung bình để phục vụ một khách hàng.

1. Xác suất để thời gian phục vụ khách hàng nào đó trong khoảng $(0.4, 1)$ là

$$P(0.4 < X < 1) = \int_{0.4}^{1} 5e^{-5x}\,dx \simeq 0.1286$$

2. Thời gian trung bình phục vụ mỗi khách hàng

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_0^{+\infty} 5xe^{-5x}\,dx \overset{t=5x}{=} \frac{1}{5}\int_0^{+\infty} te^{-t}\,dt = \frac{1}{5}\,\Gamma(2) = \frac{1}{5}$$

Bài tập 2.29.

Biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f_X(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

1. Tính $P(X \ge 5)$.
2. Xác định hàm phân phối xác suất của biến ngẫu nhiên $Y = -2X + 5$.

1. $P(X \ge 5) = \int_5^{+\infty} e^{-x}\,dx = e^{-5} \simeq 6.72 \times 10^{-3}$
2. Hàm phân phối $F_X(x)$

$$F_X(x) = \begin{cases} \displaystyle\int_{-\infty}^{x} 0\,dt, & x \le 0 \\ \displaystyle\int_0^{x} e^{-t}\,dt, & x > 0 \end{cases}$$

Rút gọn ta được

$$F_X(x) = \begin{cases} 0, & x \le 0 \\ 1 - e^{-x}, & x > 0 \end{cases}$$

3. Có $Y = -2X + 5$

$$F_Y(y) = P(Y < y) = P(-2X + 5 < y) = P\left(X > \frac{5-y}{2}\right) = 1 - F_X\left(\frac{5-y}{2}\right)$$

Suy ra hàm phân phối của Y

$$F_Y(y) = \begin{cases} 1, & y \ge 5 \\ e^{\frac{y-5}{2}}, & y < 5 \end{cases}$$

Bài tập 2.30.

Cho hàm mật độ xác suất

$$f_X(x) = \begin{cases} 3e^{-3x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

của biến ngẫu nhiên liên tục X và định nghĩa $Y = [X]$ là số nguyên lớn nhất không vượt quá X (nghĩa là $[x] = 0$ nếu $0 \le x < 1$, $[x] = 1$ nếu $1 \le x < 2 \ldots$).

1. Tính $P(Y = 0)$.
2. Tính $E(Y)$.

Xét $Y = \lfloor X \rfloor$ là biến ngẫu nhiên rời rạc nhận các giá trị $0, 1, 2, \ldots$

1. $Y = 0$ xảy ra khi và chỉ khi $0 \le X < 1$. Suy ra

$$P(Y=0) = P(0 \le x < 1) = \int_0^1 3e^{-3x}\,dx \simeq 0.9502$$

2. Một cách tổng quát, tương tự như trên, ta có

$$P(Y=y) = P(y \le x < y+1) = \int_y^{y+1} 3e^{-3x}\,dx = \left(1-e^{-3}\right)e^{-3y}, \quad y = 0, 1, 2, \ldots$$

Suy ra

$$E[Y] = \left(1-e^{-3}\right)\sum_{y=0}^{\infty} ye^{-3y}$$

Ta có

$$\begin{aligned}
\sum_{k=1}^{\infty} kx^{k-1} &= \sum_{k=1}^{\infty}\left(x^k\right)' = \left(\sum_{k=1}^{\infty} x^k\right)' = \left(\sum_{k=0}^{\infty} x^k - 1\right)' \\
&= \left(\left(\frac{1}{1-x}\right) - 1\right)' = \left(\frac{x}{1-x}\right)' \\
&= \frac{1}{(1-x)^2}
\end{aligned}$$

Suy ra

$$\begin{aligned}
E[Y] &= \left(1-e^{-3}\right)\sum_{y=0}^{\infty} ye^{-3y} = \left(1-e^{-3}\right)\sum_{y=1}^{\infty} ye^{-3y} \\
&= e^{-3}\left(1-e^{-3}\right)\sum_{y=1}^{\infty} ye^{-3(y-1)} \\
&= \frac{e^{-3}\left(1-e^{-3}\right)}{\left(1-e^{-3}\right)^2} = \frac{1}{e^3-1} \qquad \left(\text{thay } x = e^{-3}\right)
\end{aligned}$$

Chú ý: Ta cũng có thể tính kì vọng, phương sai của Y với phép biển đổi *Moment Generating Function*. Phép biến đổi tương ứng với Y là

$$M(s) = \left(1-e^{-3}\right)\sum_{y=0}^{\infty} e^{sy}\,e^{-3y} = \left(1-e^{-3}\right)\sum_{y=0}^{\infty}\left(e^{s-3}\right)^y = \frac{1-e^{-3}}{1-e^{s-3}}$$

Mà ta có

$$E[Y] = \frac{d}{ds}M(s)\Big|_{s=0}$$

Suy ra

$$E[Y] = \frac{e^{-3}}{1-e^{-3}} = \frac{1}{e^3-1}$$

Phương sai của Y

$$E[Y^2] = \frac{d^2}{ds^2} M(s)\bigg|_{s=0} = \frac{1+e^3}{(e^3-1)^2}$$

Suy ra

$$V[Y] = E[Y^2] - E^2[Y] = \frac{e^3}{(e^3-1)^2}$$

2.3 Một số luật phân phối xác suất thông dụng

Bài tập 2.31.

Bắn 5 viên đạn vào một mục tiêu. Xác suất trúng đích của mỗi lần bắn như nhau và bằng 0,2. Muốn phá hủy mục tiêu phải có ít nhất 3 viên trúng mục tiêu. Tìm xác suất mục tiêu bị phá hủy.

Gọi X là số viên đạn bắn trúng mục tiêu thì X có phân phối nhị thức với $n = 5$ và $p = 0.2 : X \sim \mathcal{B}(5, 0.2)$. Gọi A là "mục tiêu bị phá hủy" thì $A \equiv (X \geq 3)$. Do vậy, ta có

$$\begin{aligned} P(A) = P(X \geq 3) &= \binom{5}{3} 0.2^3 \times 0.8^2 + \binom{5}{4} 0.2^4 \times 0.8^1 + \binom{5}{5} 0.2^5 \times 0.8^0 \\ &= 0.05792 \end{aligned}$$

Bài tập 2.32.

Xác suất để một sinh viên chậm giờ thi là 0,02. Tìm số sinh viên chậm giờ thi có khả năng xảy ra nhiều nhất trong 855 sinh viên dự thi.

Gọi X là số thí sinh chậm giờ thì X có phân phối nhị thức $X \sim \mathcal{B}(855, 0.02)$. Ta cần tìm $\operatorname{mod} X$, chính là số có khả năng nhất trong lược đồ *Bernoulli*

$$\operatorname{mod} X = \lfloor np - q \rfloor + 1 = 17$$

Bài tập 2.33.

Có 10 máy sản xuất sản phẩm (độc lập nhau), mỗi máy sản xuất ra 2% phế phẩm.

1. Từ mỗi máy sản xuất lấy ngẫu nhiên ra một sản phẩm. Hỏi xác suất lấy được nhiều nhất 2 phế phẩm trong 10 sản phẩm này là bao nhiêu?
2. Trung bình có bao nhiêu sản phẩm được sản xuất bởi máy đầu tiên trước khi nó tạo ra phế phẩm đầu tiên (giả sử các sản phẩm sản xuất ra là độc lập)?

Gọi X là số phế phẩm trong 10 sản phẩm thì X có phân phối nhị thức $X \sim \mathcal{B}(10, 0.02)$

1. Sự kiện quan tâm là $X \leq 2$. Dễ tính được

$$P(X \leq 2) = \binom{10}{0} 0.02^0 \times 0.98^{10} + \binom{10}{1} 0.02^1 \times 0.98^9 + \binom{10}{2} 0.02^2 \times 0.98^8$$
$$\simeq 0.9991$$

2. Gọi Z là số sản phẩm sản xuất ra trước khi tạo ra phế phẩm đầu tiên.
Có $P(Z = z) = 0.98^z \times 0.02, \ (z = 0, 1, 2, \dots)$ nên theo định nghĩa,

$$E[Z] = \sum_{z=0}^{+\infty} z \times 0.98^z \times 0.02$$

Chú ý rằng

$$\sum_{n=0}^{+\infty} nx^n = x \sum_{n=0}^{+\infty} nx^{n-1} = x \sum_{n=0}^{+\infty} (x^n)' = x \left(\sum_{n=0}^{+\infty} x^n \right)'$$
$$= x \left(\frac{1}{1-x} \right)' = \frac{x}{(1-x)^2}$$

Thay $x = 0.98$ suy ra $E[Z] = 49$

Chú ý: Hàm khối lượng của Z như trên có dạng của biến ngẫu nhiên có phân phối hình học. Ta có thể dùng kì vọng có điều kiện để tính kì vọng, phương sai của biến ngẫu nhiên có phân phối hình học.
Gọi Y là số sản phẩm được tạo ra bởi máy đầu tiên đến khi nó tạo ra phế phẩm đầu tiên thì Y quan hệ với Z theo biểu thức

$$Z = Y - 1$$

Vì các sản phẩm sản xuất ra là độc lập nên Y có phân phối hình học $Y \sim \mathcal{G}(0.02)$, và

$$p_Y(y) = P(Y = y) = (1-p)^{y-1}\, p = 0.98^{y-1} \times 0.02, \ y = 1, 2, \dots$$

Ta có thể coi trường hợp không bao giờ sản xuất ra phế phẩm có xác suất là 0 vì nếu gọi A là "không bao giờ sản xuất ra phế phẩm" thì ta có

$$P(A) = (1-p)^k \xrightarrow{k \to \infty} 0$$

Trước hết, ta chỉ ra *tính chất không nhớ* của Y

$$p_Y(y) = p_{Y-n\,|\,Y>n}(y)$$

Thật vậy, gọi P_i là "sản xuất được phế phẩm ở lần thứ i", ta có

$$\begin{aligned} p_{Y-n\,|\,Y>n}(y) &= P\left(Y - n = y \mid Y > n\right) \\ &= P\left(\overline{P_{n+1}}\,\overline{P_{n+2}} \dots \overline{P_{n+y-1}}\, P_{n+y} \mid \overline{P_1}\,\overline{P_2} \dots \overline{P_n}\right) \\ &= P\left(\overline{P_{n+1}}\,\overline{P_{n+2}} \dots \overline{P_{n+y-1}}\, P_{n+y}\right) \quad \text{(vì các sản phẩm độc lập)} \\ &= (1-p)^{y-1}\, p \\ &= P(Y = y) = p_Y(y) \end{aligned}$$

Sử dụng tính chất trên và theo *Expected value rule*, ta có

$$\begin{aligned} E[Y] = 1 + E[Y-1] &= 1 + p\,E\left[Y-1 \mid Y=1\right] + (1-p)\,E\left[Y-1 \mid Y>1\right] \\ &= 1 + 0 + (1-p)\,E[Y] \end{aligned}$$

Giải ra được

$$E[Y] = \frac{1}{p} = 50$$

Từ đây ta cũng có $E[Z] = 49$.
Tương tự, ta có thể tìm phương sai của Y. Sử dụng tính không nhớ, ta có

$$\begin{aligned} E[Y \mid Y > 1] &= 1 + E[Y] \\ E[Y^2 \mid Y > 1] &= E\left[(Y+1)^2\right] \end{aligned}$$

Suy ra

$$\begin{aligned} E[Y^2] &= P(Y=1)\ E[Y^2 \mid Y=1] + P(Y>1)\ E[Y^2 \mid Y>1] \\ &= p.1 + (1-p)\left(E[Y^2] + 2E[Y] + 1\right) \end{aligned}$$

Giải ra được

$$E[Y^2] = \frac{2}{p^2} - \frac{1}{p}$$

Như vậy

$$V[Y] = E[Y^2] - E^2[Y] = \frac{1-p}{p^2}$$

Suy ra $V[Z] = V[Y] = \dfrac{1-p}{p^2}$

Bài tập 2.34.

Một ga ra cho thuê ôtô thấy rằng số người đến thuê ôtô vào thứ bảy cuối tuần là một biến ngẫu nhiên có phân bố *Poisson* với tham số $\lambda = 2$. Giả sử gara có 4 chiếc ôtô.

1. Tìm xác suất để tất cả 4 ôtô đều được thuê vào thứ 7.
2. Tìm xác suất gara không đáp ứng được yêu cầu (thiếu xe cho thuê) vào thứ 7.
3. Trung bình có bao nhiêu ôtô được thuê vào ngày thứ 7?

Gọi X là số người đến thuê ô tô vào thứ 7 cuối tuần thì X có phân phối $Poisson$, $X \sim \mathcal{P}(2)$

1. Cả 4 ô tô đều được thuê vào thứ 7 chỉ khi có ít nhất 4 người đến thuê. Vậy

$$P(A) = P(X \geq 4) = 1 - e^{-2}\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!}\right) \simeq 0.1429$$

2. Gara thiếu xe cho thuê nếu có từ 5 người trở lên đến thuê

$$P(B) = P(A) - P(X = 4) \simeq 0.0526$$

3. Gọi Y là số ô tô được thuê trong ngày thứ 7 thì Y là biến ngẫu nhiên rời rạc nhận các giá trị $0, 1, 2, 3, 4$

i) $Y = 0$ xảy ra khi không có người đến thuê, nghĩa là $X = 0$, suy ra

$$P(Y = 0) = P(X = 0) = \frac{e^{-2}\ 2^0}{0!} \simeq 0.1353$$

Tương tự, $Y = k$ chính là sự kiện $X = k$, với $k = 1, 2, 3$.
Có $P(Y = 1) \simeq 0.2707,\ P(Y = 2) \simeq 0.2707,\ P(Y = 3) \simeq 0.1804$

ii) $Y = 4$ xảy ra khi có từ 4 người trở lên đến thuê, chính là sự kiện $A \equiv X \geq 4$

$$P(Y = 4) = P(A) \simeq 0.1429$$

Bảng phân phối xác suất của Y

Y	0	1	2	3	4
$P(Y)$	0.1353	0.2707	0.2707	0.1804	0.1429

Suy ra $E[Y] = 1.9249$

Bài tập 2.35.

Số khách hàng đến một cửa hàng bán lẻ là một biến ngẫu nhiên có phân phối *Poisson* với trung bình 6 khách hàng đến trong vòng một giờ.

1. Nếu có đúng 5 khách hàng đến trong khoảng thời gian từ 10:00 đến 11:00 thì xác suất để có ít nhất 8 khách hàng đến trong khoảng thời gian từ 10:00 đến 11:30 là bao nhiêu?

2. Nếu có ít hơn 6 khách hàng đến trong khoảng thời gian từ 10:00 đến 12:00 thì cửa hàng được xem như là không có lợi nhuận. Tìm xác suất để cửa hàng có đúng 1 ngày có lãi trong một tuần (giả sử cửa hàng mở cửa 6 ngày trong tuần).

1. Gọi X là số khách hàng đến cửa hàng bán lẻ trong vòng nửa giờ, ta có $X \sim \mathcal{P}(3)$.
Gọi A là sự kiện quan tâm, A xảy ra khi và chỉ khi có ít nhất 3 người đến trong khoảng 11h00 đến 11h30, như vậy

$$P(A) = 1 - P(X < 3) = 1 - \sum_{k=0}^{2} \frac{e^{-3}\ 3^k}{k!} \simeq 0.5768$$

2. Gọi Y là "cửa hàng không có lợi nhuận", sử dụng biến ngẫu nhiên có phân phối *Poisson*, tính được

$$p = P(Y) = \sum_{k=0}^{5} \frac{e^{-12}\ 12^k}{k!} \simeq 0.0203$$

Gọi Z là số ngày có lãi trong tuần thì Z có phân phối nhị thức $Z \sim \mathcal{B}(6, 0.0203)$

$$p_Z(5) = \binom{6}{5}\ 0.0203^5 \times 0.9797^1 \simeq 2.02 \times 10^{-8}$$

Bài tập 2.36.

Gọi biến ngẫu nhiên Y là tỷ lệ người trong 1000 người Mỹ xác nhận rằng có uống nhiều hơn 5 cốc bia mỗi ngày. Giả sử rằng tỷ lệ đúng là 10% trên toàn bộ dân số Mỹ. Tính $E(Y), V(Y)$.

Gọi X là số người trong 1000 người Mỹ xác nhận có uống nhiều hơn 5 cốc bia mỗi ngày, thì X có phân phối nhị thức $X \sim \mathcal{B}(1000, 0.1)$ và $Y = \dfrac{X}{1000}$, suy ra

$$E[Y] = \frac{E[X]}{1000} = 0.1, \quad V[Y] = \frac{V[X]}{1000^2} = 9 \times 10^{-5}$$

Bài tập 2.37.

Giả sử X là biến ngẫu nhiên tuân theo luật phân phối chuẩn với trung bình là 3 và phương sai là 0,16.

1. Hãy tính $P(X > 3), P(X > 3,784)$.
2. Tìm c sao cho $P(3 - c < X < 3 + c) = 0,9$.

Ta có biến ngẫu nhiên có phân phối chuẩn $X \sim \mathcal{N}(3, 0.4^2)$

1. Áp dụng công thức
$$P(X > 3) = 0.5 - \phi\left(\frac{3-3}{0.4}\right) = 0.5$$
Tương tự, $P(X > 3.784) = 0.5 - 0.4750 = 0.025$
2. Sử dụng công thức $P\left(|X - \mu| < t\sigma\right) = 2\phi(t)$, ta có
$$2\phi(t) = 0.9, \quad \text{với} \quad t\sigma = c$$
Suy ra $t \simeq 1.65$ và $c \simeq 0.66$

Bài tập 2.38.

Cho biên độ dao động của một vật là biến ngẫu nhiên liên tục có hàm phân phối xác suất là

$$F_X(x) = \begin{cases} 1 - \ e^{-\frac{x^2}{2\sigma^2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

trong đó s là tham số đã biết. Tính xác suất để biên độ giao động đó lớn hơn trị trung bình của nó.

Hàm mật độ của X

$$f_X(x) = \begin{cases} \dfrac{x}{\sigma^2}\ e^{-\frac{x^2}{2\sigma^2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Suy ra

$$\begin{aligned}E[X] = \int_{-\infty}^{+\infty} x f_X(x)\,dx &= \int_0^{+\infty} \frac{x^2}{\sigma^2}\, e^{-\frac{x^2}{2\sigma^2}}\,dx \\ &= \sqrt{2}\sigma \int_0^{+\infty} \sqrt{y}\, e^{-y}\,dy \quad \left(\text{đặt } \frac{x^2}{2\sigma^2} = y\right) \\ &= \sqrt{2}\sigma\ \Gamma\left(\frac{3}{2}\right) \\ &= \sqrt{2}\sigma\ \frac{1}{2}.\Gamma\left(\frac{1}{2}\right) \\ &= \sigma\sqrt{\frac{\pi}{2}}\end{aligned}$$

Từ đó ta có

$$P\left(X > E[X]\right) = 1 - F_X\left(\sigma\sqrt{\frac{\pi}{2}}\right) = e^{-\frac{\pi}{4}}$$

Bài tập 2.39.

Lãi suất (%) đầu tư vào một dự án trong năm 2019 được coi như một biến ngẫu nhiên tuân theo quy luật chuẩn. Theo đánh giá của ủy ban đầu tư thì với xác suất 0,1587 cho lãi suất lớn hơn 20% và với xác suất 0,0228 cho lãi suất lớn hơn 25%. Vậy khả năng đầu tư mà không bị lỗ là bao nhiêu?

Gọi X là lãi suất đầu tư vào dự án thì X là biến ngẫu nhiên có phân phối chuẩn $X \sim \mathcal{N}(\mu, \sigma^2)$. Từ giả thiết, ta có hệ

$$\begin{cases} 0.5 - \phi\left(\dfrac{20-\mu}{\sigma}\right) = 0.1587 \\ 0.5 - \phi\left(\dfrac{25-\mu}{\sigma}\right) = 0.0228 \end{cases} \Rightarrow \begin{cases} \dfrac{20-\mu}{\sigma} = 1 \\ \dfrac{25-\mu}{\sigma} = 2 \end{cases} \Rightarrow \begin{cases} \mu = 15 \\ \sigma = 5 \end{cases}$$

Khả năng đầu tư mà không bị lỗ là

$$p = P(X > 0) = 0.5 - \phi\left(\frac{0-15}{5}\right) = 0.99865$$

Bài tập 2.40.

Tung một đồng xu vô hạn lần, xác suất thu được mặt ngửa mỗi lần là p.

1. Gọi X là số lần tung đến khi xuất hiện mặt ngửa lần đầu tiên (tại lần tung thứ X). Tính $E(X)$.
2. Tính xác suất xuất hiện đúng 6 lần ngửa trong 10 lần tung.
3. Tính xác suất để lần xuất hiện mặt ngửa thứ 6 rơi vào lần tung thứ 10.

1. Bài tập này giống bài tập 2.33 ý b. X là biến ngẫu nhiên có phân phối hình học với kì vọng $E[X] = \frac{1}{p}$. Ở đây đề bài hỏi là cho đến khi, không phải là trước khi.

2. Gọi Y là số lần tung được mặt ngửa trong 10 lần thì Y là biến ngẫu nhiên có phân phối nhị thức
$$P(Y=6) = \binom{10}{6} p^6 (1-p)^4$$

3. Gọi A là sự kiện quan tâm, A xảy ra khi và chỉ khi lần thứ 10 tung được mặt ngửa, và tung được đúng 5 mặt ngửa trong 9 lần đầu tiên. Bài toán thỏa mãn lược đồ *Bernoulli*, hoặc ta cũng có thể dùng phân phối nhị thức *Binomial Distribution*
$$P(A) = \left[\binom{9}{5} p^5 (1-p)^4\right] p = \binom{9}{5} p^6 (1-p)^4$$

Bài tập 2.41.

Xét một phần tư hình tròn tâm $O(0,0)$ bán kính bằng a, ký hiệu là OAB, với tọa độ tương ứng là $A(a,0)$ và $B(0,a)$.

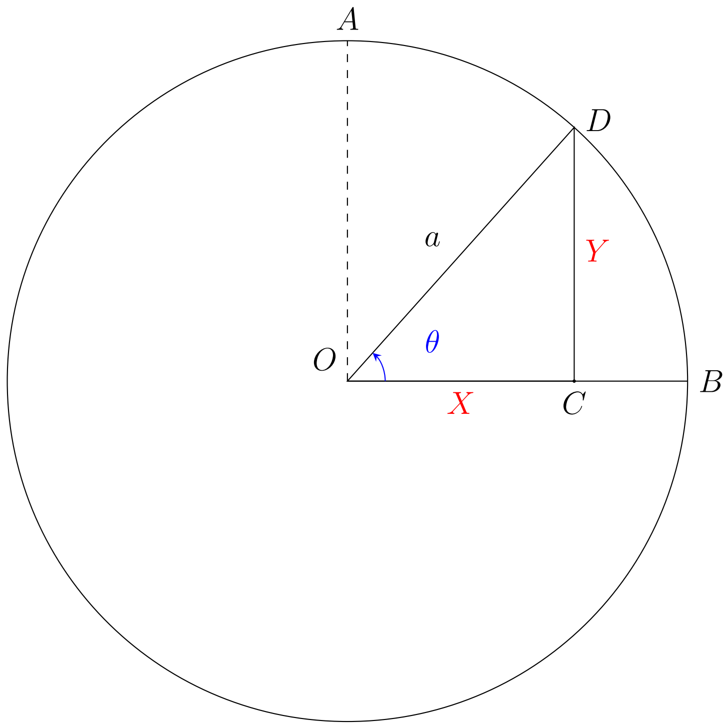
1. Trên đoạn OA lấy ngẫu nhiên một điểm C. Tìm phân phối xác suất của độ dài đoạn OC.

2. Dựng một đường thẳng đi qua C, vuông góc với OA và cắt cung tròn tại điểm D. Tính kỳ vọng và phương sai của độ dài đoạn CD.

1. Gọi X là độ dài đoạn OC thì X là biến ngẫu nhiên có phân phối đều, $X \sim \mathcal{U}[0,a]$ Hàm mật độ xác suất của X
$$f_X(x) = \begin{cases} \frac{1}{a}, & x \in [0,a] \\ 0, & x \notin [0,a] \end{cases}$$

2. Gọi Y là độ dài đoạn CD thì Y là biến ngẫu nhiên thỏa mãn điều kiện
$$Y = \begin{cases} \sqrt{a^2 - x^2}, & x \in [0,a] \\ 0, & x \notin [0,a] \end{cases}$$

Ta sử dụng các kết quả
$$E\left[g(X)\right] = \int_{-\infty}^{+\infty} g(x)\, f_X(x)\, dx$$
và
$$V\left[g(X)\right] = E\left[g^2(X)\right] - \left(E\left[g(X)\right]\right)^2$$



Kỳ vọng của Y

$$\begin{aligned}E[Y] &= \int_{-\infty}^{+\infty} g(X)\, f_X(x)\, dx = \frac{1}{a}\int_0^a \sqrt{a^2-x^2}\, dx \\ &= \frac{1}{a}\int_0^{\frac{\pi}{2}} a^2\cos^2 t\, dt \qquad (\text{đặt } x = a\sin t) \\ &= \frac{a\pi}{4}\end{aligned}$$

Ta có

$$E\left[Y^2\right] = E\left[a^2 - X^2\right] = \frac{1}{a}\int_0^a (a^2 - x^2)\, dx = \frac{2a^2}{3}$$

Suy ra phương sai của Y

$$V[Y] = \frac{2a^2}{3} - \left(\frac{a\pi}{4}\right)^2 = \left(\frac{2}{3} - \frac{\pi^2}{16}\right) a^2$$

Như vậy, trong bài toán này, ta có thể tìm được $E[Y]$ và $V[Y]$ mà không cần phải tìm phân phối của Y.

Chú ý: Ta cũng có thể tìm được kì vọng, phương sai của Y thông qua góc Θ tạo bởi Ox và OD nhờ liên hệ $Y = a\sin\theta$, ở đó Θ là biến ngẫu nhiên có phân phối đều $\Theta \sim \mathcal{U}\left[0, \frac{\pi}{2}\right]$

Bài tập 2.42.

Lấy ngẫu nhiên một điểm M trên nửa đường tròn tâm O, đường kính $AB = 2a$. Biết rằng xác suất điểm M rơi vào cung CD bất kì của nửa đường tròn AMB chỉ phụ thuộc vào độ dài cung CD.

1. Tìm hàm phân phối xác suất của biến ngẫu nhiên Y chỉ diện tích tam giác AMB.

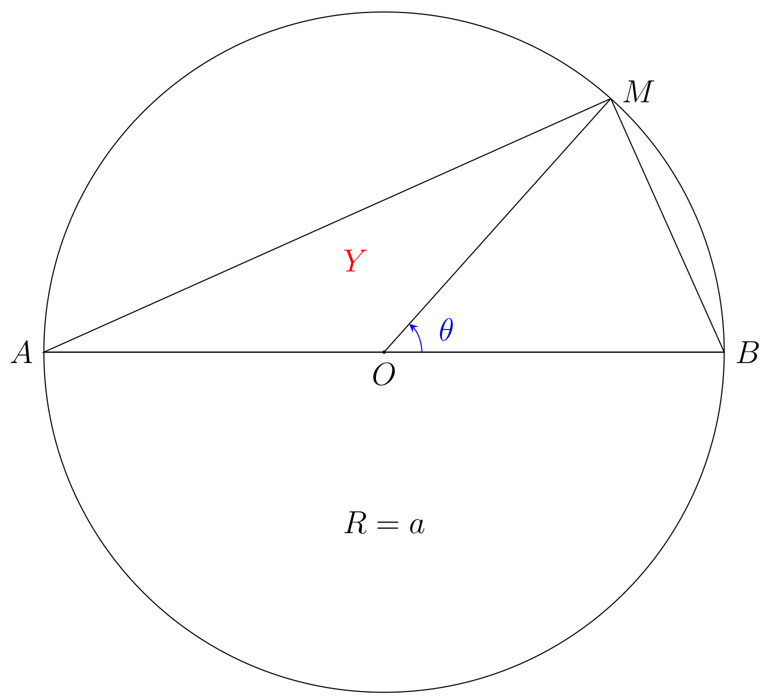
2. Tìm giá trị trung bình của diện tích tam giác ấy.

1. Gọi Θ là góc tạo bởi Ox và OM, dễ thấy Θ là biến ngẫu nhiên có phân phối đều $\Theta \sim \mathcal{U}[0, \pi]$.
   Hàm mật độ xác suất của Θ

$$f_\Theta(\theta) = \begin{cases} \dfrac{1}{\pi}, & \theta \in [0, \pi] \\ 0, & \theta \notin [0, \pi] \end{cases}$$

   Hàm phân phối xác suất của Θ

$$F_\Theta(\theta) = \begin{cases} 0, & \theta \leq 0 \\ \dfrac{\theta}{\pi}, & 0 < \theta \leq \pi \\ 1, & \theta > \pi \end{cases}$$



   Gọi X là diện tích tam giác AMB thì X là biến ngẫu nhiên liên tục thỏa mãn hệ thức

$$X = \begin{cases} a^2 \sin\theta, & x \in [0, a^2] \\ 0, & x \notin [0, a^2] \end{cases}$$

Hàm phân phối xác suất của X

$$\begin{aligned} F_X(x) &= P(X < x) = P(a^2 \sin\theta < x) \\ &= \begin{cases} 0, & x \le 0 \\ P\left(0 < \theta < \arcsin \dfrac{x}{a^2}\right) + P\left(\pi - \arcsin \dfrac{x}{a^2} < \theta < \pi\right), & 0 < x \le a^2 \\ 1, & x > a^2 \end{cases} \end{aligned}$$

Rút gọn đi ta được

$$F_X(x) = \begin{cases} 0, & x \le 0 \\ \dfrac{2}{\pi} \arcsin \dfrac{x}{a^2}, & 0 < x \le a^2 \\ 1, & x > a^2 \end{cases}$$

2. Hàm mật độ xác suất của X

$$f_X(x) = \begin{cases} \dfrac{2}{\pi\sqrt{a^4 - x^2}}, & x \in [0, a^2] \\ 0, & x \notin [0, a^2] \end{cases}$$

Suy ra kỳ vọng của X là

$$E[X] = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx = \frac{2}{\pi} \int_0^{a^2} \frac{x}{\sqrt{a^4 - x^2}}\, dx = \frac{2}{\pi}\, a^2$$

Bài tập 2.43.

Từ điểm $A(0, -a)$ $(a > 0)$ trong nửa mặt phẳng tọa độ xOy phần $x \ge 0$, người ta kẻ ngẫu nhiên một tia At hợp với tia Oy một góc φ. Biết φ là biến ngẫu nhiên có phân phối đều trong khoảng $\left(0, \dfrac{\pi}{4}\right)$. Tia At cắt Ox tại điểm M.

1. Tìm hàm phân phối xác suất của biến ngẫu nhiên X chỉ diện tích tam giác AOM.
2. Tìm giá trị trung bình của diện tích trên.

Theo giả thiết, ta có Ω là biến ngẫu nhiên có phân phối đều $\Omega \sim \mathcal{U}\left[0, \dfrac{\pi}{4}\right]$.

Hàm mật độ xác suất của Ω

$$f_\Omega(\omega) = \begin{cases} \dfrac{4}{\pi}, & \omega \in \left[0, \dfrac{\pi}{4}\right] \\ 0, & \omega \notin \left[0, \dfrac{\pi}{4}\right] \end{cases}$$

Hàm phân phối xác suất của Ω

$$F_\Omega(\omega) = \begin{cases} 0, & \omega \le 0 \\ \dfrac{4\omega}{\pi}, & 0 < \omega \le \dfrac{\pi}{4} \\ 1, & \omega > \dfrac{\pi}{4} \end{cases}$$

1. Gọi X là diện tích tam giác AOM thì X là biến ngẫu nhiên liên tục thỏa mãn hệ thức

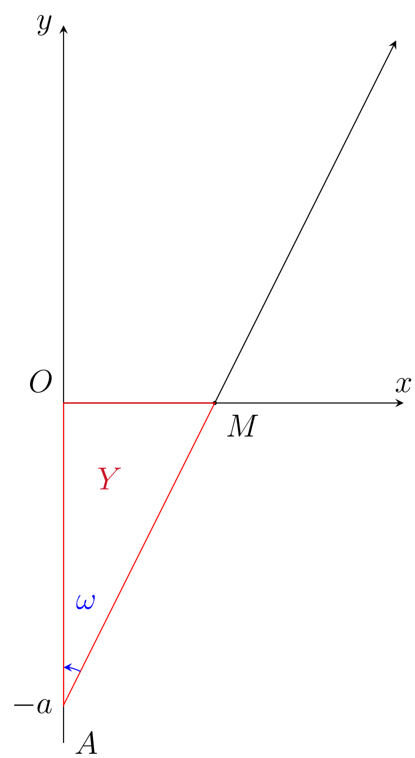
$$X = \begin{cases} \frac{1}{2}a^2 \tan\omega, & x \in [0, a^2] \\ 0, & x \notin [0, a^2] \end{cases}$$

Hàm phân phối xác suất của X

$$F_X(x) = P(X < x) = P\left(\frac{1}{2}a^2 \tan\omega < x\right)$$
$$= \begin{cases} 0, & x \le 0 \\ P\left(0 < \omega < \arctan\frac{2x}{a^2}\right) + P\left(\pi - \arctan\frac{2x}{a^2} < \omega < \pi\right), & 0 < x \le a^2 \\ 1, & x > a^2 \end{cases}$$

Rút gọn ta được

$$F_X(x) = \begin{cases} 0, & x \le 0 \\ \frac{8}{\pi}\arctan\frac{2x}{a^2}, & 0 < x \le a^2 \\ 1, & x > a^2 \end{cases}$$



2. Hàm mật độ xác suất của X

$$f_X(x) = \begin{cases} \frac{16a^2}{\pi(a^4 + 4x^2)}, & x \in [0, a^2] \\ 0, & x \notin [0, a^2] \end{cases}$$

Suy ra kỳ vọng của X là

$$E[X] = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx = \frac{16a^2}{\pi} \int_0^{a^2} \frac{x}{a^4 + 4x^2}\, dx = \frac{2a^2 \ln 5}{\pi}$$

Bài tập 2.44.

Một công ty kinh doanh mặt hàng A dự định sẽ áp dụng một trong hai phương án kinh doanh: Phương án 1: Gọi X_1 (triệu đồng/tháng) là lợi nhuận thu được. X_1 có phân phối chuẩn $\mathcal{N}(140; 2500)$. Phương án 2: Gọi X_2 (triệu đồng/tháng) là lợi nhuận thu được. X_2 có phân phối chuẩn $\mathcal{N}(200; 3600)$. Biết rằng công ty tồn tại và phát triển thì lợi nhuận thu được từ mặt hàng A phải đạt ít nhất 80 triệu đồng/tháng. Hỏi nên áp dụng phương án nào để rủi ro thấp hơn.

Nếu sử dụng phương án 1 thì khả năng công ty tồn tại và phát triển là

$$p_1 = P(X_1 > 80) = 0.5 - \phi\left(\frac{80 - 140}{\sqrt{2500}}\right) = 0.5 + 0.38493 = 0.88493$$

Nếu sử dụng phương án 2 thì khả năng công ty tồn tại và phát triển là

$$p_2 = P(X_2 > 80) = 0.5 - \phi\left(\frac{80 - 200}{\sqrt{3600}}\right) = 0.5 + 0.47725 = 0.97725$$

Vì $p_2 > p_1$ nên công ty sử dụng phương án 2 sẽ có khả năng rủi ro thấp hơn

Bài tập 2.45.

Trọng lượng của một loại trái cây tuân theo luật phân phối chuẩn với trọng lượng trung bình là 250g, độ lệch chuẩn là 5g. Trái cây loại I là trái cây có trọng lượng không nhỏ hơn 260g.

1. Một người lấy 1 trái từ trong sọt trái cây ra. Tính xác suất người này lấy được trái cây loại I.
2. Nếu lấy được trái loại I thì người này sẽ mua sọt đó. Người ngày kiểm tra 100 sọt. Tính xác suất người này mua được 6 sọt.

Gọi X là trọng lượng của loại trái cây, ta có $X \sim \mathcal{N}(250, 5^2)$

1. Xác suất người này lấy được trái cây loại I là

$$p = P(X \geq 260) = 0.5 - \phi\left(\frac{260 - 250}{5}\right) = 0.02275$$

2. Gọi Y là số sọt người đó mua được, dễ thấy Y là biến ngẫu nhiên có phân phối nhị thức $Y \sim \mathcal{B}(100, 0.02275)$.
   Xác suất để người này mua được 6 sọt là

$$p = p_Y(6) = \binom{100}{6} 0.02275^6 \times 0.7725^{94} \simeq 0.0190$$

Bài tập 2.46.

Một dây chuyền tự động khi hoạt động bình thường có thể sản xuất ra phế phẩm với xác suất $p = 0,001$ và được điều chỉnh ngay lập tức khi phát hiện có phế phẩm. Tính số trung bình các sản phẩm được sản xuất giữa 2 lần điều chỉnh.

Gọi X là số sản phẩm được sản xuất giữa 2 lần điều chỉnh. Ta thấy X là biến ngẫu nhiên rời rạc nhận các giá trị $X = 1, 2, 3, \dots$. Hàm khối lượng xác suất

$$p_X(x) = P(X = x) = (1-p)^{x-1}\, p = 0.999^{x-1} \times 0.001, \quad x = 1, 2, 3, \dots$$

Như vậy, X có phân phối hình học, như ở bài tập 2.33 ý b, ta đã chỉ ra

$$E[X] = \frac{1}{p} = 1000$$

Chú ý: Qua bài toán này, ta nhận xét rằng nếu $Y_1, Y_2, \dots$ có phân phối hình học thì biến ngẫu nhiên

$$X_k = Y_k - Y_{k-1}$$

cũng có phân phối hình học.
Ta cũng có thể tìm được $E[X]$ theo hướng này mà không cần công thức
Gọi Y_k là tổng số sản phẩm được sản suất cho đến khi điều chỉnh lần thứ k và X_k là số sản phẩm sản xuất được giữa hai lần điều chỉnh thứ k và $k+1$. Khi đó, ta thấy rằng X chính là $X_1, X_2, \dots$
Dễ có X_1 chính là số sản phẩm sản xuất được khi gặp phế phẩm đầu tiên, nên có phân phối hình học $X_1 \sim \mathcal{G}(\lambda)$. Giả sử khi đã gặp phế phẩm đầu tiên, thì ta biết rằng "tương lai" vẫn tiếp tục là *Bernouli process*, tương tự như process ban đầu: Số sản phẩm sản xuất được X_2 đến khi gặp phế phẩm đầu tiên cũng là phân phối hình học.
Hơn thế nữa, các sản phẩm sản xuất trong quá khứ hoàn toàn độc lập với các sản phẩm trong tương lai. Vì X_2 xác định số sản phẩm trong tương lai, nên nó hoàn toàn độc lập với X_1. Cứ tiếp tục như vậy, ta kết luận các biến ngẫu nhiên $X_1, X_2, \dots$ là độc lập và đều có phân phối hình học. Do đó

$$E[X] = \frac{1}{p}$$

Bài tập 2.47.

Trong một kỳ thi điểm số trung bình của các sinh viên là 80 và độ lệch chuẩn là 10. Giả sử điểm thi của sinh viên tuân theo luật phân phối chuẩn.

1. Nếu giáo viên muốn 25% số sinh viên đạt điểm A (nhóm điểm cao nhất) thì điểm số thấp nhất để đạt điểm A là bao nhiêu?

2. Chọn ngẫu nhiên 50 sinh viên, tính xác suất trong đó có nhiều hơn 10 sinh viên đạt điểm A (điểm A lấy ở câu (a)).

Gọi X là điểm thi của sinh viên, ta có $X \sim \mathcal{N}(80, 10^2)$

1. Ta cần tìm x thỏa mãn $P(X > x) = 0.25$. Sử dụng hàm *Laplace*

$$P(X > x) = 0.25 \Rightarrow 0.5 - \phi\left(\frac{x-80}{10}\right) = 0.25 \Rightarrow \frac{x-80}{10} = 0.09871$$

   Giải ra được $x = 80.9871$

2. Gọi Y là số sinh viên đạt điểm A (lấy ở câu trên) trong 50 sinh viên thì Y là biến ngẫu nhiên có phân phối nhị thức $Y \sim \mathcal{B}(50, 0.25)$. Xác suất cần tính là

$$P(Y > 10) = 1 - \sum_{i=0}^{10} \binom{50}{i} 0.25^i \times 0.75^{50-i} \simeq 0.7378$$

Bài tập 2.48.

Đường kính của một loại chi tiết do một máy sản xuất tuân theo luật phân phối chuẩn, với kỳ vọng là 20mm và độ lêch chuẩn là 0,2mm. Tính xác suất để lấy ngẫu nhiên một chi tiết có đường kính trong khoảng 19,9mm đến 20,3mm.

Có $X \sim \mathcal{N}(20, 0.2^2)$. Xác suất cần tính là

$$P(19.9 < X < 20.3) = \phi\left(\frac{20.3-20}{0.2}\right) - \phi\left(\frac{19.9-20}{0.2}\right) = 0.43319 + 0.19146 = 0.62465$$

Bài tập 2.49.

Chiều cao của nam giới khi trưởng thành là biến ngẫu nhiên tuân theo luật phân phối chuẩn với chiều cao trung bình là 160cm và độ lệch chuẩn là 6cm. Tìm xác suất để đo ngẫu nhiên 4 người thì có ít nhất một người có chiều cao nằm trong khoảng (158–162) cm.

Giả sử X là chiều cao của một người. Ta có $X \sim \mathcal{N}(160, 6^2)$.
Xác suất để một người có chiều cao trong khoảng $158 - 162$ cm là

$$P(158 < X < 162) = \phi\left(\frac{162-160}{6}\right) - \phi\left(\frac{158-160}{6}\right) = 0.12930 + 0.12930 = 0.2586$$

Gọi Y là số người có chiều cao trong khoảng trên, thì Y có phân phối nhị thức. Ta cần tính

$$P(Y \geq 1) = 1 - P(0) = 1 - \binom{4}{0} 0.2586^0 \times 0.7414^4 \simeq 0.6979$$

Bài tập 2.50.

Dùng hai phương pháp để tính sai số của một biến ngẫu nhiên. Phương pháp 1: Cho sai số đó bằng $2X$ với X là biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(0; 25)$. Phương pháp 2: Cho sai số đó bằng tổng hai biến ngẫu nhiên độc lập $Y = Y_1 + Y_2$ trong đó $E(Y_1) = E(Y_2) = 0$ và $\sigma(Y_1) = \sigma(Y_2) = 5$. Hỏi phương pháp nào được ưa dùng hơn?

Phương pháp thứ nhất là một biến ngẫu nhiên có phân phối chuẩn

$$2X \sim \mathcal{N}\left(2 \times 0,\, 4 \times 5^2\right)$$

Phương pháp thứ hai là một biến ngẫu nhiên có phân phối chuẩn

$$X_1 + X_2 \sim \mathcal{N}\left(0 + 0,\, 5^2 + 5^2\right)$$

Vì $V\left[X_1 + X_2\right] < V[2X]$ nên phương pháp thứ hai sẽ được ưa chuộng hơn, do nó phân tán quanh kỳ vọng "hẹp hơn" phương pháp thứ nhất. Trong khi, kỳ vọng của cả hai phương pháp đều là 0, chính là sai số mà ta "mong muốn đạt được".

3 Biến ngẫu nhiên nhiều chiều

3.1 Biến ngẫu nhiên rời rạc

Bài tập 3.1.

Cho biến ngẫu nhiên X và Y có bảng phân bố xác suất đồng thời như sau

$X \backslash Y$	1	2	3
1	0.12	0.15	0.03
2	0.28	0.35	0.07

1. Chứng minh rằng X và Y độc lập.
2. Lập bảng phân phối xác suất của X và Y.
3. Tìm quy luật phân phối của biến ngẫu nhiên $Z = XY$.
4. Tính $E(Z)$ bằng 2 cách và kiểm tra $E(Z) = E(X).E(Y)$.

1. Bảng phân phối xác suất của X và Y là

X	1	2
$P(X)$	0.3	0.7

Y	1	2	3
$P(Y)$	0.4	0.5	0.1

2. Từ giả thiết và ý trên ta kiểm tra được

$$P(X = X_i,\ Y = Y_j) = P(X = X_i)\ P(Y = Y_j), \quad \forall i = 1, 2;\ j = 1, 2, 3$$

nên X, Y là hai biến ngẫu nhiên độc lập

3. Vì X nhận các giá trị $X = 1, 2$ và Y nhận các giá trị $Y = 1, 2, 3$ nên Z là biến ngẫu nhiên rời rạc nhận các giá trị $Z = 1, 2, 3, 4, 6$. Ta có

$$P(Z = 1) = P(XY = 1) = P(X = 1, Y = 1) = 0.12$$

$$P(Z = 2) = P(XY = 2) = P(X = 1, Y = 2) + P(X = 2, Y = 1) = 0.43$$

Tương tự, ta lập được bảng phân phối của $Z = XY$ là

Z	1	2	3	4	6
$P(Z)$	0.12	0.43	0.03	0.35	0.07

4. Từ bảng phân phối xác suất của Z, ta tính kỳ vọng theo định nghĩa

$$E[Z] = 1 \times 0.12 + 2 \times 0.43 + 3 \times 0.03 + 4 \times 0.35 + 6 \times 0.07 = 2.89$$

Ta cũng có thể tính kỳ vọng dựa trên tính độc lập của X và Y

$$E[Z] = E[X]\ E[Y] = (1 \times 0.3 + 2 \times 0.7) \times (1 \times 0.4 + 2 \times 0.5 + 3 \times 0.1) = 2.89$$

Bài tập 3.2.

Cho biến ngẫu nhiên X và Y có bảng phân bố xác suất đồng thời là

$X \backslash Y$	-1	0	1
-1	$\frac{4}{15}$	$\frac{1}{15}$	$\frac{4}{15}$
0	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{15}$
1	0	$\frac{2}{15}$	0

1. Tìm $E(X), E(Y), \operatorname{cov}(X, Y)$.
2. X và Y có độc lập không?
3. Tìm bảng phân phối xác suất của X, của Y.

1. Ta có

$$E[X] = (-1).\left(\frac{4}{15} + \frac{1}{15} + \frac{4}{15}\right) + 0.\left(\frac{1}{15} + \frac{2}{15} + \frac{1}{15}\right) + 1.\left(0 + \frac{2}{15} + 0\right) = -\frac{7}{15}$$
$$E[Y] = (-1).\left(\frac{4}{15} + \frac{1}{15} + 0\right) + 0.\left(\frac{1}{15} + \frac{2}{15} + \frac{2}{15}\right) + 1.\left(\frac{4}{15} + \frac{1}{15} + 0\right) = 0$$
$$E[XY] = (-1) \times (-1) \times \frac{4}{15} + (-1) \times 1 \times \frac{4}{15} + 1 \times (-1) \times 0 + 1 \times 1 \times 0 = 0$$

Suy ra
$$\operatorname{cov}(X, Y) = E[XY] - E[X]\ E[Y] = 0$$

2. Dễ kiểm tra được $P(X = -1,\ Y = -1) \neq P(X = -1)\ P(Y = -1)$ nên X, Y không độc lập

3. Bảng phân phối xác suất của X và Y

X	-1	0	1
$P(X)$	$\frac{9}{15}$	$\frac{4}{15}$	$\frac{5}{15}$

Y	-1	0	1
$P(Y)$	$\frac{5}{15}$	$\frac{5}{15}$	$\frac{5}{15}$

Bài tập 3.3.

Cho biến ngẫu nhiên X và Y có bảng phân bố xác suất đồng thời là

X \ Y	1	2	3
1	0.17	0.13	0.25
2	0.10	0.30	0.05

1. Lập bảng phân phối xác suất của X và của Y.
2. Lập ma trận Covarian của (X, Y).
3. Tìm hệ số tương quan.
4. X và Y có độc lập không?

1. Bảng phân phối xác suất của X và Y

X	1	2
$P(X)$	0.55	0.45

Y	1	2	3
$P(Y)$	0.27	0.43	0.3

2. Từ các bảng phân phối xác suất của X, Y ta có

$$E[X] = 1.45,\ V[X] = 0.2475$$

$$E[Y] = 2.03,\ V[Y] = 0.5691$$

Tính được $E[XY] = 2.88$ suy ra $\operatorname{cov}(X, Y) = -0.0635$
Ma trận hiệp phương sai

$$\Gamma = \begin{pmatrix} V[X] & \operatorname{cov}(X,Y) \\ \operatorname{cov}(X,Y) & V[Y] \end{pmatrix} = \begin{pmatrix} 0.2475 & -0.0635 \\ -0.0635 & 0.5691 \end{pmatrix}$$

3. Hệ số tương quan $\rho_{XY} = \dfrac{\operatorname{cov}(X,Y)}{\sqrt{V[X]\ V[Y]}} = -0.1692$

4. Dễ kiểm tra được $P(X = 1,\ Y = 1) \neq P(X = 1)\ P(Y = 1)$ nên X, Y không độc lập

Bài tập 3.4.

Thống kê về giá thành sản phẩm Y (triệu đồng) và sản lượng X (tấn) của một ngành sản xuất thu được bảng phân phối xác suất sau:

$X \backslash Y$	30	50	80	100
6	0.05	0.06	0.08	0.11
7	0.06	0.15	0.04	0.08
8	0.07	0.09	0.10	0.11

1. Tìm giá thành sản phẩm trung bình và mức độ phân tán của nó.
2. Tìm sản lượng trung bình khi giá thành bằng 8.
3. X và Y có độc lập không?
4. X và Y có tương quan không?

1. Bảng phân phối xác suất của X

X	6	7	8
$P(X)$	0.3	0.33	0.37

Suy ra $E[X] = 7.07$ và $\sigma_X = \sqrt{V[X]} \simeq 0.8155$

2. Bảng phân phối xác suất có điều kiện là

$Y \mid (X = 8)$	30	50	80	100
$P(Y \mid X = 8)$	$\frac{7}{37}$	$\frac{9}{37}$	$\frac{10}{37}$	$\frac{11}{37}$

Như vậy $E[Y \mid X = 8] = \dfrac{2569}{37} \simeq 69.1892$

3. Bảng phân phối xác suất của Y

Y	30	50	80	100
$P(Y)$	0.18	0.3	0.22	0.3

Dễ kiểm tra được $P(X = 6,\ Y = 30) \neq P(X = 6)\ P(Y = 30)$ nên X, Y không độc lập

4. Ta tính được

$$E[X] = 7.07,\ E[Y] = 68,\ E[XY] = 479.7$$

Suy ra

$$\text{cov}(X, Y) = 479.7 - 68 \times 7.07 = -1.06$$

Như vậy X, Y có quan hệ tương quan

Chú ý: Bài này đề bị nhầm, đã tự sửa lại Y là sản lượng và X là giá thành sản phẩm mới làm được ý thứ 2

Bài tập 3.5.

Cho X_1, X_2, X_3 là các biến ngẫu nhiên độc lập theo luật phân phối *Poisson* với tham số $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3$. Tính xác suất của các sự kiện sau:

1. Số lớn nhất trong các số X_1, X_2, X_3 không nhỏ hơn 1.
2. Số lớn nhất trong các số X_1, X_2, X_3 bằng 1.
3. Số nhỏ nhất trong các số X_1, X_2, X_3 không nhỏ hơn 1.
4. Số nhỏ nhất trong các số X_1, X_2, X_3 bằng 1.

Gọi X, Y lần lượt là số lớn nhất và số nhỏ nhất trong các số X_1, X_2, X_3.

1. Gọi A là "X không nhỏ hơn 1". Ta có

$$\begin{aligned} P(A) &= 1 - P(\overline{A}) \\ &= 1 - P(X_1 < 1,\ X_2 < 1,\ X_3 < 1) \\ &= 1 - P(X_1 = 0)\ P(X_2 = 0)\ P(X_3 = 0) \\ &= 1 - \frac{e^{-1}\ 1^0}{0!} . \frac{e^{-2}\ 2^0}{0!} . \frac{e^{-3}\ 3^0}{0!} \\ &\simeq 0.9975 \end{aligned}$$

2. Gọi B là "X bằng 1" thì dễ thấy rằng $P(B) = P(A) - P(X > 1)$. Tương tự ta tính được

$$\begin{aligned} P(X \leq 1) &= P(X_1 \leq 1,\ X_2 \leq 1,\ X_3 \leq 1) \\ &= P(X_1 \leq 1)\ P(X_2 \leq 1)\ P(X_3 \leq 1) \\ &= \left(\frac{e^{-1}\ 1^0}{0!} + \frac{e^{-1}\ 1^1}{1!}\right) . \left(\frac{e^{-2}\ 2^0}{0!} + \frac{e^{-2}\ 2^1}{1!}\right) . \left(\frac{e^{-3}\ 3^0}{0!} + \frac{e^{-3}\ 3^1}{1!}\right) \\ &\simeq 0.0595 \end{aligned}$$

Suy ra $P(B) \simeq 0.9975 - (1 - 0.0595) \simeq 0.057$

3. Gọi C là "Y không nhỏ hơn 1". Ta có

$$\begin{aligned} P(Y \geq 1) &= P(X_1 \geq 1,\ X_2 \geq 1,\ X_3 \geq 1) \\ &= P(X_1 \geq 1)\ P(X_2 \geq 1)\ P(X_3 \geq 1) \\ &= \left(1 - \frac{e^{-1}\ 1^0}{0!}\right) \cdot \left(1 - \frac{e^{-2}\ 2^0}{0!}\right) \cdot \left(1 - \frac{e^{-3}\ 3^0}{0!}\right) \\ &\simeq 0.5194 \end{aligned}$$

4. Gọi D là "Y bằng 1" thì dễ thấy $P(D) = P(C) - P(Y > 1)$. Tương tự ta tính được

$$\begin{aligned} P(Y > 1) &= P(X_1 > 1,\ X_2 > 1,\ X_3 > 1) \\ &= P(X_1 > 1)\ P(X_2 > 1)\ P(X_3 > 1) \\ &= \left(1 - \frac{e^{-1}\ 1^0}{0!} - \frac{e^{-1}\ 1^1}{1!}\right) \cdot \left(1 - \frac{e^{-2}\ 2^0}{0!} - \frac{e^{-2}\ 2^1}{1!}\right) \cdot \left(1 - \frac{e^{-3}\ 3^0}{0!} - \frac{e^{-3}\ 3^1}{1!}\right) \\ &\simeq 0.1257 \end{aligned}$$

Suy ra $P(D) \simeq 0.5194 - 0.1257 \simeq 0.3937$

Bài tập 3.6.

Cho X và Y là hai biến ngẫu nhiên có bảng phân phối xác suất là:

X	0	1	2	3	4	5
P	0.15	0.3	0.25	0.2	0.08	0.02

Y	0	1	2	3	4	5
P	0.3	0.2	0.2	0.15	0.1	0.05

1. Tính $E(X), E(Y), V(X), V(Y)$.

2. Nếu X và Y độc lập, tính $P(X + Y \leq 2)$ và lập bảng phân phối xác suất của $X + Y$.

1. Dễ dàng tính được

$$E[X] = 1.82,\ \ V[X] = 1.5676$$
$$E[Y] = 1.7,\ \ V[Y] = 2.31$$

2. Ta thấy $X + Y \leq 2$ xảy ra khi và chỉ khi X, Y là một trong các bộ dưới đây

$$\{(0,0),\ (0,1),\ (1,0),\ (2,0),\ (1,1),\ (0,2)\}$$

Mà X, Y độc lập, nên ta tính được

$$\begin{aligned} P(X + Y \leq 2) &= P(X + Y = 0) + P(X + Y = 1) + P(X + Y = 2) \\ &= 0.33 \end{aligned}$$

Dễ thấy $X + Y$ là biến ngẫu nhiên rời rạc nhận các giá trị $1, 2, 3, \ldots, 10$. Bảng phân phối xác suất của $X + Y$

0	1	2	3	4	5	6	7	8	9	10
0.045	0.12	0.165	0.1925	0.174	0.137	0.09	0.0485	0.021	0.006	0.001

Bài tập 3.7.

Lấy ngẫu nhiên 3 viên bi từ một hộp gồm 3 bi đỏ, 5 bi xanh và 4 bi vàng. Gọi X, Y lần lượt là số bi xanh, bi vàng trong 3 bi lấy ra. Lập bảng phân phối xác suất đồng thời cho biến ngẫu nhiên hai chiều (X, Y)

Dễ thấy X, Y là hai biến ngẫu nhiên rời rạc nhận các giá trị $0, 1, 2, 3$.
Bảng phân phối xác suất đồng thời của (X, Y)

$X \backslash Y$	0	1	2	3
0	$\frac{C_3^3}{C_{12}^3}$	$\frac{C_3^2\ C_4^1}{C_{12}^3}$	$\frac{C_3^1\ C_4^2}{C_{12}^3}$	$\frac{C_4^3}{C_{12}^3}$
1	$\frac{C_5^1\ C_3^2}{C_{12}^3}$	$\frac{C_3^1\ C_5^1\ C_4^1}{C_{12}^3}$	$\frac{C_5^1\ C_4^2}{C_{12}^3}$	0
2	$\frac{C_3^1\ C_5^2}{C_{12}^3}$	$\frac{C_5^2\ C_4^1}{C_{12}^3}$	0	0
3	$\frac{C_5^3}{C_{12}^3}$	0	0	0

Rút gọn đi ta được bảng

$X \backslash Y$	0	1	2	3
0	$\frac{1}{220}$	$\frac{3}{55}$	$\frac{9}{110}$	$\frac{1}{55}$
1	$\frac{3}{44}$	$\frac{3}{11}$	$\frac{3}{22}$	0
2	$\frac{3}{22}$	$\frac{2}{11}$	0	0
3	$\frac{1}{22}$	0	0	0

3.2 Biến ngẫu nhiên liên tục

Bài tập 3.8.

Cho X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời là

$$f_{X,Y}(x,y) = \begin{cases} kx, \text{ nếu } 0 < y < x < 1, \\ 0, \text{ nếu trái lại.} \end{cases}$$

1. Tìm hằng số k
2. X và Y có độc lập không?

1. Ta giải hệ phương trình

$$\begin{cases} kx \geq 0, \ 0 < y < x < 1 \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} kx \, dxdy = 1 \end{cases} \Rightarrow \begin{cases} k \geq 0, \\ \int_0^1 dx \int_0^x kx \, dy = 1 \end{cases} \Rightarrow k = 3$$

Thử lại.

2. Hàm mật độ xác suất biên $f_X(x)$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y) \, dy = \begin{cases} \int_0^x 3x \, dy, \ 0 < x < 1, \\ 0, \ \text{trái lại} \end{cases} = \begin{cases} 3x^2, \ 0 < x < 1, \\ 0, \ \text{trái lại} \end{cases}$$

Hàm mật độ xác suất biên $f_Y(y)$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y) \, dx = \begin{cases} \int_y^1 3x \, dx, \ 0 < y < 1, \\ 0, \ \text{trái lại} \end{cases} = \begin{cases} \frac{3}{2} - \frac{3}{2}y^2, \ 0 < x < 1, \\ 0, \ \text{trái lại} \end{cases}$$

Vì $f_{X,Y}(x,y) \neq f_X(x) \ f_Y(y)$ với $0 < y < x < 1$ nên X, Y không độc lập.

Bài tập 3.9.

Cho X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời là

$$f_{X,Y}(x,y) = \begin{cases} k\left(x^2 + \frac{xy}{2}\right), \text{ nếu } 0 < x < 1, \ 0 < y < 2, \\ 0, \text{ nếu trái lại.} \end{cases}$$

1. Tìm hằng số k.
2. Tìm hàm phân phối đồng thời của X và Y.

1. Ta giải hệ phương trình

$$\begin{cases} k\left(x^2+\dfrac{xy}{2}\right)\geq 0, \begin{cases} 0<x<1 \\ 0<y<2 \end{cases} \\ \displaystyle\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} k\left(x^2+\dfrac{xy}{2}\right)\,dxdy=1 \end{cases} \Rightarrow \begin{cases} k\geq 0, \\ \displaystyle\int\limits_0^1 dx\int\limits_0^2 k\left(x^2+\dfrac{xy}{2}\right)\,dy=1 \end{cases} \Rightarrow k=\dfrac{6}{7}$$

Thử lại.

2. Hàm phân phối đồng thời $F_{XY}(x,y)$

$$F_{XY}(x,y)=\begin{cases} \displaystyle\int_{-\infty}^{x}\int_{-\infty}^{y} 0\,dudv, & x\leq 0;\ y\leq 0 \\ \displaystyle\int_0^x\int_0^y \frac{6}{7}\left(u^2+\frac{uv}{2}\right)\,dudv, & 0<x<1;\ 0<y<2 \\ \displaystyle\int_0^x\int_0^2 \frac{6}{7}\left(u^2+\frac{uv}{2}\right)\,dudv, & 0<x<1;\ y\geq 2 \\ \displaystyle\int_0^1\int_0^y \frac{6}{7}\left(u^2+\frac{uv}{2}\right)\,dudv, & x\geq 1;\ 0<y<2 \\ \displaystyle\int_0^1\int_0^2 \frac{6}{7}\left(u^2+\frac{uv}{2}\right)\,dudv, & x>1;\ y>2 \end{cases}$$

Rút gọn ta được

$$F_{XY}(x,y)=\begin{cases} 0, & x\leq 0;\ y\leq 0 \\ \dfrac{6}{7}\left(\dfrac{x^3y}{3}+\dfrac{x^2y^2}{8}\right), & 0<x<1;\ 0<y<2 \\ \dfrac{6}{7}\left(\dfrac{2x^3}{3}+\dfrac{x^2}{2}\right), & 0<x<1;\ y\geq 2 \\ \dfrac{6}{7}\left(\dfrac{y}{3}+\dfrac{y^2}{8}\right), & x\geq 1;\ 0<y<2 \\ 1, & x>1;\ y>2 \end{cases}$$

Bài tập 3.10.

Cho X, Y là hai biến ngẫu nhiên có hàm mật độ xác suất đồng thời

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{6\pi}, \text{ nếu } \dfrac{x^2}{9} + \dfrac{y^2}{4} < 1, \\ 0, \text{ nếu trái lại.} \end{cases}$$

1. Tìm hàm mật độ xác suất biên của X, của Y.
2. Tìm xác suất để (X, Y) nằm trong hình chữ nhật $O(0,0); A(0,1); B(1,2); D(2,0)$.

1. Hàm mật độ xác suất biên $f_X(x)$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy = \begin{cases} \displaystyle\int\limits_{-\frac{2}{3}\sqrt{9-x^2}}^{\frac{2}{3}\sqrt{9-x^2}} \frac{1}{6\pi}\,dy, \; -3 < x < 3, \\ 0, \text{ trái lại} \end{cases}$$

Rút gọn ta được

$$f_X(x) = \begin{cases} \dfrac{2\sqrt{9-x^2}}{9\pi}, \; -3 < x < 3, \\ 0, \text{ trái lại} \end{cases}$$

Hàm mật độ xác suất biên $f_Y(y)$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy = \begin{cases} \displaystyle\int\limits_{-\frac{3}{2}\sqrt{4-y^2}}^{\frac{3}{2}\sqrt{4-y^2}} \frac{1}{6\pi}\,dx, \; -2 < y < 2, \\ 0, \text{ trái lại} \end{cases}$$

Rút gọn ta được

$$f_Y(y) = \begin{cases} \dfrac{\sqrt{4-y^2}}{2\pi}, \; -2 < y < 2, \\ 0, \text{ trái lại} \end{cases}$$

2. Dễ thấy hình chữ nhật $OABD$ nằm trọn trong miền hình elip $\dfrac{x^2}{9} + \dfrac{y^2}{4} \leq 1$.

Xác suất để (X, Y) nằm trong hình chữ nhật là

$$\iint\limits_{0 \leq x \leq 2;\; 0 \leq y \leq 1} f_{XY}(x,y)\,dxdy = \int_0^2 dx \int_0^1 \frac{1}{6\pi}\,dy = \frac{1}{3\pi}$$

Chú ý: Trong đề bài tọa độ điểm B bị nhầm. Để nó là hình chữ nhật, đã tự sửa lại $B(2,1)$

Bài tập 3.11.

Cho biến ngẫu nhiên hai chiều liên tục (X,Y) có hàm mật độ xác suất là

$$f_{X,Y}(x,y) = \begin{cases} kx^2, \text{ nếu } -1 \le x \le 1,\ 0 \le y \le x^2, \\ 0, \text{ nếu trái lại.} \end{cases}$$

1. Tìm k.
2. Tìm hàm mật độ xác suất biên $f_X(x), f_Y(y)$.
3. Tính $P\left(Y \le \dfrac{1}{4}\right)$.

1. Ta giải hệ phương trình

$$\begin{cases} kx^2 \ge 0, \begin{cases} -1 \le x \le 1 \\ 0 \le y \le x^2 \end{cases} \\ \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} kx^2\, dxdy = 1 \end{cases} \Rightarrow \begin{cases} k \ge 0, \\ \int\limits_{-1}^{1} dx \int\limits_{0}^{x^2} kx^2\, dy = 1 \end{cases} \Rightarrow k = \frac{5}{2}$$

Thử lại.

2. Hàm mật độ xác suất biên $f_X(x)$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dy = \begin{cases} \int_0^{x^2} \frac{5}{2}x^2\, dy, & -1 \le x \le 1, \\ 0, & \text{trái lại} \end{cases} = \begin{cases} \frac{5}{2}x^4, & -1 \le x \le 1, \\ 0, & \text{trái lại} \end{cases}$$

Hàm mật độ xác suất biên $f_Y(y)$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dx = \begin{cases} \int\limits_{-1}^{-\sqrt{y}} \frac{5}{2}x^2\, dx + \int\limits_{\sqrt{y}}^{1} \frac{5}{2}x^2\, dx, & 0 \le y \le 1, \\ 0, & \text{trái lại} \end{cases}$$

Rút gọn ta được

$$f_Y(y) = \begin{cases} \frac{5}{3} - \frac{5}{3}\, y\sqrt{y}, & 0 \le y \le 1, \\ 0, & \text{trái lại} \end{cases}$$

3. Ta có thể tính $P\left(Y \le \frac{1}{4}\right)$ dùng hàm mật độ đồng thời hoặc dùng hàm mật độ biên Kí hiệu D là miền trên đó $f_{X,Y}(x,y) \neq 0$.

$$D = \left\{(x,y) \in \mathbb{R}^2 \mid -1 \le x \le 1;\ 0 \le y \le x^2\right\}$$

Suy ra

$$P\left(Y \le \frac{1}{4}\right) = \iint\limits_{D \cap \left\{y \le \frac{1}{4}\right\}} f_{XY}(x,y)\,dxdy$$
$$= 2\int_0^{\frac{1}{2}} dx \int_0^{x^2} \frac{5}{2}x^2\,dy + 2\int_{\frac{1}{2}}^1 dx \int_0^{\frac{1}{4}} \frac{5}{2}x^2\,dy \simeq 0.3959$$

$$P\left(Y \le \frac{1}{4}\right) = \int_{-\infty}^{\frac{1}{4}} f_Y(y)\,dy = \int_0^{\frac{1}{4}} \left(\frac{5}{3} - \frac{5}{3}\,y\sqrt{y}\right) dy \simeq 0.3959$$

Bài tập 3.12.

Cho X và Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời là

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{x}, & \text{nếu } 0 < y < x < 1, \\ 0, & \text{nếu trái lại.} \end{cases}$$

1. Tìm hàm mật độ xác suất biên của X, của Y.

2. Tìm hàm mật độ xác suất có điều kiện $f_1(x \mid y)$, $f_2(y \mid x)$.

1. Hàm mật độ xác suất biên $f_X(x)$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy = \begin{cases} \int_0^x \frac{1}{x}\,dy, & 0 < x < 1, \\ 0, & \text{trái lại} \end{cases} = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{trái lại} \end{cases}$$

Hàm mật độ xác suất biên $f_Y(y)$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx = \begin{cases} \int_y^1 \frac{1}{x}\,dx, & 0 < y < 1, \\ 0, & \text{trái lại} \end{cases} = \begin{cases} -\ln y, & 0 < y < 1, \\ 0, & \text{trái lại} \end{cases}$$

2. Các hàm mật độ điều kiện là

$$f_X(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \begin{cases} -\dfrac{1}{x \ln y}, & 0 < y < x < 1, \\ 0, & \text{trái lại} \end{cases}$$

$$f_Y(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)} = \begin{cases} \dfrac{1}{x}, & 0 < y < x < 1, \\ 0, & \text{trái lại} \end{cases}$$

Bài tập 3.13.

Một linh kiện điện tử có thời gian hoạt động X là biến ngẫu nhiên có phân phối mũ với hàm mật độ xác suất là $f_X(x) = \lambda e^{-\lambda x},\ x > 0,\ \lambda > 0$.

1. Tìm hàm phân phối xác suất của biến ngẫu nhiên chỉ thời gian hoạt động của một mạng gồm 2 linh kiện loại trên được mắc song song/mắc nối tiếp.
2. Tính kỳ vọng, phương sai của thời gian hoạt động của mạng đó.

Gọi X_1, X_2 lần lượt là biến ngẫu nhiên chỉ thời gian hoạt động của 2 linh kiện trong mạch. Gọi Y, Z lần lượt là biến ngẫu nhiên chỉ thời gian hoạt động của mạng lắp song song và mạng lắp nối tiếp. Khi đó ta có

$$Y = \max(X_1, X_2) > 0, \quad Z = \min(X_1, X_2) > 0$$

1. Từ đó tính được

$$\begin{aligned} F_Y(y) &= P(Y < y) = P(X_1 < y)\ P(X_2 < y), \quad (\text{vì } X_1, X_2 \text{ độc lập}) \\ &= \left(\int_{-\infty}^{y} f_X(x)\, dx \right)^2 = \left(\int_0^y \lambda\, e^{-\lambda x}\, dx \right)^2 \\ &= \left(1 - e^{-\lambda y}\right)^2 \end{aligned}$$

Như vậy

$$F_Y(y) = \begin{cases} \left(1 - e^{-\lambda y}\right)^2, & y > 0, \\ 0, & \text{trái lại} \end{cases}$$

Tương tự ta có

$$\begin{aligned} F_Z(z) &= 1 - P(Z \geq z) = 1 - P(X_1 \geq z)\ P(X_2 \geq z), \quad (\text{vì } X_1, X_2 \text{ độc lập}) \\ &= 1 - \left(\int_z^{+\infty} f_X(x)\, dx \right)^2 \\ &= 1 - e^{-2\lambda z} \end{aligned}$$

Như vậy

$$F_Z(z) = \begin{cases} 1 - e^{-2\lambda z}, & z > 0, \\ 0, & \text{trái lại} \end{cases}$$

2. Từ hàm phân phối ta tìm được các hàm mật độ xác suất

$$f_Y(y) = \begin{cases} 2\lambda e^{-\lambda y}\left(1 - e^{-\lambda y}\right), & y > 0, \\ 0, & \text{trái lại} \end{cases} \qquad f_Z(z) = \begin{cases} 2\lambda e^{-2\lambda z}, & z > 0, \\ 0, & \text{trái lại} \end{cases}$$

Kỳ vọng và phương sai của biến ngẫu nhiên Y

$$\begin{aligned} E[Y] &= \int_{-\infty}^{+\infty} y\, f_Y(y)\, dy \\ &= \int_0^{+\infty} 2\lambda y\, e^{-\lambda y}\, dy - \int_0^{+\infty} 2\lambda y\, e^{-2\lambda y}\, dy \\ &= \frac{2}{\lambda}\, \Gamma(2) - \frac{1}{2\lambda}\, \Gamma(2) = \frac{3}{2\lambda} \end{aligned}$$

$$\begin{aligned} E[Y^2] &= \int_{-\infty}^{+\infty} y^2\, f_Y(y)\, dy \\ &= \int_0^{+\infty} 2\lambda y^2\, e^{-\lambda y}\, dy - \int_0^{+\infty} 2\lambda y^2\, e^{-2\lambda y}\, dy \\ &= \frac{2}{\lambda^2}\, \Gamma(3) - \frac{1}{4\lambda^2}\, \Gamma(3) = \frac{7}{2\lambda^2} \end{aligned}$$

Suy ra

$$V[Y] = E[Y^2] - (E[Y])^2 = \frac{5}{4\lambda^2}$$

Kỳ vọng và phương sai của biến ngẫu nhiên Z

$$\begin{aligned} E[Z] &= \int_{-\infty}^{+\infty} z\, f_Z(z)\, dz = \int_0^{+\infty} 2\lambda z\, e^{-2\lambda z}\, dz \\ &= \frac{1}{2\lambda}\, \Gamma(2) = \frac{1}{2\lambda} \end{aligned}$$

$$\begin{aligned} E[Z^2] &= \int_{-\infty}^{+\infty} z^2\, f_Z(z)\, dz = \int_0^{+\infty} 2\lambda z^2\, e^{-2\lambda z}\, dz \\ &= \frac{1}{4\lambda^2}\, \Gamma(3) = \frac{1}{2\lambda^2} \end{aligned}$$

Suy ra

$$V[Z] = E[Z^2] - (E[Z])^2 = \frac{1}{4\lambda^2}$$

Chú ý: Ta có thể làm ý b bằng cách sử dụng các đặc trưng kỳ vọng, phương sai của biến ngẫu nhiên có phân phối mũ. Nếu $X \sim \text{Exp}(\lambda)$ thì

$$E[X] = \frac{1}{\lambda} \text{ và } V[X] = \frac{1}{\lambda^2}$$

Dễ thấy Z là biến ngẫu nhiên có phân phối mũ $Z \sim \text{Exp}(2\lambda)$. Suy ra

$$E[Z] = \frac{1}{2\lambda} \text{ và } V[Z] = \frac{1}{4\lambda^2}$$

Biểu diễn Y dưới dạng $Y = 2Y' - Z$, với $Y' \sim \text{Exp}(\lambda),\ Z \sim \text{Exp}(2\lambda)$. Như vậy Y không phải là biến ngẫu nhiên có phân phối mũ. Dựa vào tính tuyến tính, ta có

$$E[Y] = 2E[Y'] - E[Z] = \frac{2}{\lambda} - \frac{1}{2\lambda} = \frac{3}{2\lambda}$$

Tuy nhiên, việc tính toán $V[Y]$ dùng tính tuyến tính như trên là tương đối khó do Y_1, Y_2 không độc lập và $\text{cov}(Y_1, Y_2)$ cũng không dễ tính. Do đó, cần thiết phải tính $V[Y]$ theo định nghĩa như đã làm.

Một số kết quả:

Cho $X_1 \sim \text{Exp}(\lambda)$ và $X_2 \sim \text{Exp}(\mu)$ là hai biến ngẫu nhiên độc lập có phân phối mũ.

- $Z = \min(X_1, X_2)$ là biến ngẫu nhiên có phân phối mũ với tham số $\lambda + \mu$

$$Z = \min(X_1, X_2) \sim \text{Exp}(\lambda + \mu)$$

- Có $\min(X_1, X_2) + \max(X_1, X_2) = X_1 + X_2$. Suy ra kỳ vọng của $\max(X_1, X_2)$

$$\begin{aligned} E\left[\max(X_1, X_2)\right] &= E\left[X_1 + X_2\right] - E\left[\min(X_1, X_2)\right] \\ &= \frac{1}{\lambda} + \frac{1}{\mu} - \frac{1}{\lambda + \mu} \end{aligned}$$

- Với n biến ngẫu nhiên độc lập có phân phối mũ $X_i \sim \text{Exp}(\lambda_i),\ i = 1, 2, \ldots, n$.

$$\min_i X_i \sim \text{Exp}(\lambda_1 + \lambda_2 + \ldots + \lambda_n),$$
$$E\left[\max_i X_i\right] = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \ldots + \frac{1}{\lambda_n} - \frac{1}{\lambda_1 + \lambda_2 + \ldots + \lambda_n}$$

Bài tập 3.14.

Cho X và Y là hai biến ngẫu nhiên độc lập với nhau có cùng phân phối đều trên $[0, 2]$.

1. Tìm hàm phân phối của các biến ngẫu nhiên $Z = X + Y;\ T = XY;\ U = X - Y$.
2. Tính $P(-1 \leq Y - X \leq 1)$.

1. Vì X và Y độc lập nên ta có hàm mật độ đồng thời của (X, Y) là

$$f_{XY}(x, y) = \begin{cases} \frac{1}{4}, & (x, y) \in D, \\ 0, & \text{trái lại} \end{cases}$$

trong đó

$$D = \{0 \leq x \leq 2;\ 0 \leq y \leq 2\}$$

(a) Xét $Z = X + Y$. Ta có $0 \le Z \le 4$. Hàm phân phối xác suất của Z

$$F_Z(z) = P(Z < z) = \iint\limits_{\{x+y<z\}\cap D} f_{XY}(x,y)\,dxdy = \frac{1}{4}\iint\limits_{\{x+y<z\}\cap \text{D}} dxdy$$

Nếu $z \le 0$ thì $F_Z(z) = 0$.

Nếu $0 < z \le 2$ thì $F_Z(z) = \dfrac{1}{4}\displaystyle\int_0^z \left(\int_0^{z-x} dy\right) dx = \frac{1}{4}.\frac{z^2}{2} = \frac{z^2}{8}$

Nếu $2 < z \le 4$ thì

$$F_Z(z) = \frac{1}{4}\int_{z-2}^{2}\left(\int_{z-x}^{2} dy\right) dx = \frac{1}{4}\int_{z-2}^{2}(2 - z + x)\,dx = \frac{1}{8}\left(z^2 - 8z + 16\right)$$

Nếu $z > 4$ thì $F_Z(z) = \dfrac{1}{4}\displaystyle\iint\limits_D dxdy = 1$. Vậy

$$F_Z(z) = \begin{cases} 0, & z \le 0, \\ \dfrac{z^2}{8}, & 0 < z \le 2, \\ \dfrac{1}{8}\left(z^2 - 8z + 16\right), & 2 < z \le 4, \\ 1, & z > 4 \end{cases}$$

(b) Xét $T = XY$ có $0 \le T \le 4$. Hàm phân phối của T được xác định như sau:
Nếu $t \le 0$ thì $F_T(t) = 0$
Nếu $0 < t \le 4$ thì

$$F_T(t) = \frac{1}{4}\left(\int_0^{\frac{1}{2}} dx \int_0^2 dy + \int_{\frac{1}{2}}^{2} dx \int_0^{\frac{1}{x}} dy\right) = \frac{1}{4}\left(t + t\ln 2 - t\ln\frac{t}{2}\right)$$

Nếu $t > 4$ thì $F_T(t) = 1$. Vậy

$$F_Z(z) = \begin{cases} 0, & t \le 0, \\ \dfrac{1}{4}\left(t + t\ln 2 - t\ln\dfrac{t}{2}\right), & 0 < t \le 4, \\ 1, & t > 4 \end{cases}$$

(c) Xét $U = X - Y$ thì ta có $-2 \le U \le 2$. Từ đó hàm phân phối của U được xác đính như sau:
Nếu $u \le -2$ thì $F_U(u) = 0$
Nếu $-2 < u \le 0$ thì

$$F_U(u) = \frac{1}{4}\int_0^{u+2} dx \int_{x-u}^{2} dy = \frac{1}{8}(2+u)^2$$

Nếu $0 < u \leq 2$ thì $F_U(u) = \dfrac{1}{4}\left[4 - \dfrac{1}{2}(2-u)^2\right] = \dfrac{1}{4}\left(-\dfrac{u^2}{2} + 2u + 2\right)$

Nếu $u > 2$ thì $F_U(u) = 1$. Vậy

$$F_U(u) = \begin{cases} 0, & u \leq -2, \\ \frac{1}{8}(2+u)^2, & -2 < u \leq 0, \\ \frac{1}{8}(-u^2 + 4u + 4), & 0 < u \leq 2, \\ 1, & u > 2 \end{cases}$$

2. Ta có

$$\begin{aligned} P(-1 \leq Y - X \leq 1) = P(X - 1 \leq Y \leq X + 1) \\ &= \int_0^1 dx \int_0^{x+1} \frac{1}{4}\, dy + \int_1^2 dx \int_{x-1}^{2} \frac{1}{4}\, dy \\ &= \frac{3}{8} + \frac{3}{8} = \frac{3}{4} \end{aligned}$$

Chú ý: Ta có thể tìm hàm mật độ của biến ngẫu nhiên tổng của hai biến ngẫu nhiên độc lập từ các hàm mật độ ban đầu bằng *phương pháp tích chập*
Xét lại bài toán trên. Giả sử X, Y là hai biến ngẫu nhiên độc lập. Đặt $Z = X + Y$, ta có công thức sau
Nếu X, Y là biến ngẫu nhiên rời rạc thì

$$p_Z(z) = \sum_x p_X(x)\, p_Y(z - x)$$

Nếu X, Y là biến ngẫu nhiên liên tục thì

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)\, f_Y(z - x)\, dx$$

Ta thấy rằng $f_X(x)\, f_Y(z-x)$ khác 0 $\left(\text{và bằng } \dfrac{1}{2}\right)$ nếu $0 \leq x \leq 2$ và $0 \leq z - x \leq 2$.

Kết hợp hai bất đẳng thức này lại, biểu thức dưới dấu tích phân khác 0 với

$$\max\{0, z-2\} \leq x \leq \min\{2, z\}$$

Suy ra

$$f_Z(z) = \begin{cases} \frac{1}{4}\left(\min\{2, z\} - \max\{0, z-2\}\right), & 0 \leq z \leq 4, \\ 0, & \text{trái lại} \end{cases}$$

Tương tự với biến $T = X - Y$, ta có công thức:
Nếu X, Y là biến ngẫu nhiên rời rạc thì

$$p_T(t) = \sum_x p_X(x)\, p_Y(x - z)$$

Nếu X, Y là biến ngẫu nhiên liên tục thì

$$f_T(t) = \int_{-\infty}^{+\infty} f_X(x)\, f_Y(x-z)\, dx$$

Bài tập 3.15.

Hai người A và B hẹn gặp nhau tại cổng trường trong khoảng từ 7h00 đến 8h00. Gọi X và Y lần lượt là thời gian đến điểm hẹn của người A và B trong khoảng thời gian trên. Giả sử X và Y độc lập và có cùng phân phối đều trên $[7;8]$.

1. Tìm hàm phân phối xác suất đồng thời của X và Y.
2. Với quy ước chỉ đợi nhau trong vòng 10 phút, tìm xác suất để 2 người được gặp nhau.

Hàm mật độ xác suất của X là

$$f_X(x) = \begin{cases} 1, & \text{nếu } x \in [7,8], \\ 0, & \text{trái lại} \end{cases}$$

Tương tự ta có hàm mật độ của Y. Suy ra hàm mật độ xác suất đồng thời của X và Y

$$f_{X,Y}(x,y) = \begin{cases} 1, & \text{nếu } x, y \in [7,8], \\ 0, & \text{trái lại} \end{cases}$$

1. Hàm phân phối đồng thời $F_{XY}(x,y)$

$$F_{XY}(x,y) = \begin{cases} \displaystyle\int_{-\infty}^{x} \int_{-\infty}^{y} 0\, dudv, & x < 7;\ y < 7 \\ \displaystyle\int_{7}^{x} \int_{7}^{y} 1\, dudv, & 7 \le x, y \le 8 \\ \displaystyle\int_{7}^{x} \int_{7}^{8} 1\, dudv, & 7 \le x \le 8;\ y > 8 \\ \displaystyle\int_{7}^{8} \int_{7}^{y} 1\, dudv, & x > 8;\ 7 \le y \le 8 \\ \displaystyle\int_{7}^{8} \int_{7}^{8} 1\, dudv, & x > 8;\ y > 8 \end{cases}$$

Rút gọn ta được

$$F_{XY}(x,y) = \begin{cases} 0, & x < 7;\ y < 7 \\ (x-7)(y-7), & 7 \le x, y \le 8 \\ x-7, & 7 \le x \le 8;\ y > 8 \\ y-7, & x > 8;\ 7 \le y \le 8 \\ 1, & x > 8;\ y > 8 \end{cases}$$

2. Ta quy về 1 giờ. Khi đó, xác suất cần tính là $P\left(|X-Y| \leq \frac{1}{6}\right)$. Ta có

$$\begin{aligned}
P\left(|X-Y| \leq \frac{1}{6}\right) &= \iint\limits_{D\cap\left\{|X-Y|\leq\frac{1}{6}\right\}} 1\,dxdy, \qquad \text{với } D = \{(x,y) \in \mathbb{R} \mid 0 \leq x, y \leq 1\} \\
&= \int_0^{\frac{1}{6}} dx \int_0^{x+\frac{1}{6}} dy + \int_{\frac{1}{6}}^{\frac{5}{6}} dx \int_{x-\frac{1}{6}}^{x+\frac{1}{6}} dy + \int_{\frac{5}{6}}^{1} dx \int_{x-\frac{1}{6}}^{1} dy \\
&= \frac{1}{24} + \frac{2}{9} + \frac{1}{24} = \frac{11}{36}
\end{aligned}$$

Bài tập 3.16.

Cho X và Y là hai biên ngẫu nhiên độc lập, $X \sim \mathcal{N}(5; 1^2)$, $Y \sim \mathcal{N}(3; 0.2^2)$.

1. Tìm $P(X+Y < 5,5)$.
2. Tìm $P(X < Y); P(X > 2Y)$.
3. Tìm $P(X < 1; Y < 1)$.

1. Ta có $X+Y \sim \mathcal{N}(5+3,\ 1^2+0.2^2) \sim \mathcal{N}(8, 1.04)$. Suy ra

$$P(X+Y < 5.5) = 0.5 + \phi\left(\frac{5.5-8}{\sqrt{1.04}}\right) = 0.5 - \phi(2.45) \simeq 0.0072$$

2. Ta có $X-Y \sim \mathcal{N}(5-3,\ 1^2+0.2^2) \sim \mathcal{N}(2, 1.04)$. Suy ra

$$P(X < Y) = 0.5 + \phi\left(\frac{0-2}{\sqrt{1.04}}\right) = 0.5 - \phi(1.96) = 0.025$$

Ta có $X - 2Y \sim \mathcal{N}(5 - 2\times 3,\ 1^2 + 4\times 0.2^2) \sim \mathcal{N}(-1, 1.16)$. Suy ra

$$P(X > 2Y) = 0.5 - \phi\left(\frac{0+1}{\sqrt{1.16}}\right) \simeq 0.8212$$

3. Vì X, Y độc lập nên ta có

$$\begin{aligned}
P(X<1;\ Y<1) &= P(X<1)\ P(Y<1) \\
&= \left[0.5 + \phi\left(\frac{1-5}{1}\right)\right]\left[0.5+\phi\left(\frac{1-3}{0.02}\right)\right] \\
&= 9.6\times 10^{-13}
\end{aligned}$$

Bài tập 3.17.

Trọng lượng của những người chồng tuân theo luật phân phối chuẩn với kỳ vọng 70kg và độ lệch chuẩn 9kg, còn trọng lượng của những người vợ tuân theo luật phân phối chuẩn với kỳ vọng 55kg và độ lệch chuẩn 4kg. Hệ số tương quan trọng lượng giữa vợ và chồng là $\frac{2}{3}$. Tính xác suất vợ nặng hơn chồng.

Gọi X, Y lần lượt là "trọng lượng của chồng" và "trọng lượng của vợ" thì ta có

$$X \sim \mathcal{N}(70, 9^2), \quad Y \sim \mathcal{N}(55, 4^2)$$

Xác suất cần tính là $P(X < Y)$.
Ta biết rằng $X - Y$ là biến ngẫu nhiên có phân phối chuẩn có

$$E[X-Y] = E[X] - E[Y] = 15$$

Để tính phương sai, sử dụng công thức sau

$$V[X+Y] = V[X] + V[Y] + 2\operatorname{cov}(X,Y)$$

trong đó $\operatorname{cov}(X,Y) = \rho_{XY}\sqrt{V[X]\ V[Y]} = \frac{2}{3} \times 9 \times 4 = 24$.
Suy ra

$$V[X-Y] = V[X] + V[Y] - 2\operatorname{cov}(X,Y) = 9^2 + 4^2 - 2 \times 24 = 49$$

Như vậy $X - Y \sim \mathcal{N}(15, 49)$. Khi đó

$$P(X<Y) = 0.5 + \phi\left(\frac{0-15}{\sqrt{49}}\right) = 0.5 - \phi(2.14) \simeq 0.0162$$

Chú ý: Công thức $V[X+Y] = V[X] + V[Y]$ không còn đúng nếu X, Y không độc lập

Bài tập 3.18.

Biến ngẫu nhiên liên tục X có hàm mật độ xác suất $f_X(x)$. Tìm hàm mật độ xác suất $g_Y(y)$ của biến ngẫu nhiên Y nếu:

1. $Y = X + 1,\ -\infty < x < \infty$.
2. $Y = 2X,\ -a < x < a$.

1. Ta có

$$F_Y(y) = P(Y = X+1 < y) = P(X < y-1) = F_X(y-1),\ -\infty < y < \infty$$

Đạo hàm của hàm hợp, suy ra

$$g_Y(y) = F'_Y(y) = f_X(y-1),\ -\infty < y < \infty$$

2. Ta có

$$F_Y(y) = P(Y = 2X < y) = P\left(X < \frac{y}{2}\right) = F_X\left(\frac{y}{2}\right), \; -2a < y < 2a$$

Đạo hàm của hàm hợp, suy ra

$$g_Y(y) = F'_Y(y) = \frac{1}{2}\, f_X\left(\frac{y}{2}\right), \; -2a < y < 2a$$

Bài tập 3.19.

Giả sử tại một trường đại học, một sinh viên đạt được điểm X trong bài kiểm tra năng khiếu toán học và điểm Y trong bài kiểm tra năng khiếu âm nhạc là một số trong khoảng từ 0 đến 1. Giả sử X và Y được phân phối theo hàm mật độ sau

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{2}{5}(2x+3y), \text{ nếu } 0 < x, y < 1, \\ 0, \text{ nếu trái lại.} \end{cases}$$

1. Tính tỷ lệ sinh viên đại học đạt điểm cao hơn 0,8 trong bài kiểm tra năng khiếu toán.
2. Giả sử điểm số của một sinh viên trong bài kiểm tra năng khiếu âm nhạc là 0,3. Tính xác suất để điểm của anh ấy trong bài kiểm tra năng khiếu toán học sẽ lớn hơn 0,8.
3. Giả sử điểm số của một sinh viên trong bài kiểm tra năng khiếu toán là 0,3. Tính xác suất để điểm của anh ấy trong bài kiểm tra năng khiếu âm nhạc sẽ lớn hơn 0,8.

1. Gọi A là "điểm cao hơn 0.8 trong bài thi Toán". Ta có

$$P(A) = P(X > 0.8) = \int_0^1 dy \int_{0.8}^1 \frac{2}{5}(2x+3y)\,dx = \frac{33}{125} \simeq 0.264$$

2. Cho $Y = 0.3$, ta có

$$P(X > 0.8 \mid Y = 0.3) = \int_{0.8}^1 f_{X,Y}(x,3)\,dx = \int_{0.8}^1 \frac{2}{5}(2x+0.9)\,dx = 0.216$$

3. Cho $X = 0.3$, ta có

$$P(Y > 0.8 \mid X = 0.3) = \int_{0.8}^1 f_{X,Y}(3,y)\,dy = \int_{0.8}^1 \frac{2}{5}(0.6+3y)\,dy = 0.264$$

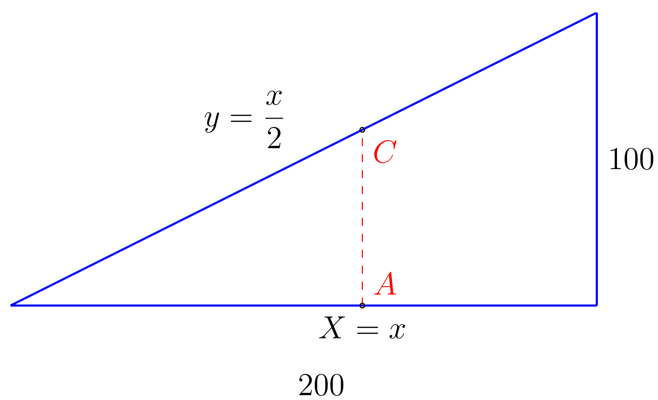
Bài tập 3.20.

Một mảnh đất bằng phẳng có hình tam giác vuông với một bờ phía nam dài 200m, bờ phía đông dài 100m. Ta quan tâm đến điểm mà một hạt giống rơi từ trên cao xuống tiếp đất. Giả sử rằng hạt giống nằm trong ranh giới của mảnh đất với tọa độ X và Y của nó được phân bố đều trên bề mặt của tam giác vuông.

1. Tìm c với c là giá trị của hàm mật độ xác suất của điểm nằm trong ranh giới mảnh đất.
2. Tìm các hàm mật độ xác suất biên của X và Y.
3. Tìm hàm mật độ xác suất của Y biết $X = x$ và tính $P(0,1 \leq Y \leq 0,7 \mid X = 0,5)$.

1. Vì tọa độ X và Y phân bố đều trên bề mặt tam giác vuông nên ta phải có
$$f_{XY}(x,y) = \frac{1}{S}$$
ở đó $S = 10000$ là diện tích tam giác vuông.
Điều này suy ra từ $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)\,dxdy = 1$. Như vậy, $c = \dfrac{1}{S} = \dfrac{1}{10000}$

2. Từ hình vẽ, ta thấy miền tam giác vuông là
$$D = \left\{(x,y) \in \mathbb{R}^2 \mid 0 \leq x \leq 200;\ 0 \leq y \leq \frac{1}{2}x\right\}$$



Hàm mật độ xác suất biên $f_X(x)$

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy = \begin{cases} \displaystyle\int_0^{\frac{x}{2}} \frac{1}{10000}\,dy, & 0 \leq x \leq 200, \\ 0, & \text{trái lại} \end{cases} = \begin{cases} \dfrac{x}{20000}, & 0 \leq x \leq 200, \\ 0, & \text{trái lại} \end{cases}$$

Hàm mật độ xác suất biên $f_Y(y)$

$$f_Y(y) = \int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx = \begin{cases} \int\limits_{2y}^{200} \dfrac{1}{10000}\,dx, & 0 \le y \le 100, \\ 0, & \text{trái lại} \end{cases} = \begin{cases} \dfrac{100-y}{5000}, & 0 \le y \le 100, \\ 0, & \text{trái lại} \end{cases}$$

3. Biết $X = x$, khi đó Y là biến ngẫu nhiên phân phối đều trên AC (xem hình vẽ)
Sử dụng tính chất $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{Y|X}(y \mid x)\,dxdy = 1$ suy ra

$$f_{Y|X}(y \mid x) = \frac{2}{x}, \text{ với } X > 0.$$

Ta có

$$P(0.1 \le Y \le 0.7 \mid X = 0.5) = \int\limits_{D\cap\{0.1\le Y\le 0.7\}} f_{Y|X}(y \mid 0.5)\,dx = \int\limits_{0.1}^{0.25} \frac{2}{0.5}\,dx = 0.6$$

trong đó D là miền được định nghĩa ở ý b.

4 Ước lượng tham số

4.1 Ước lượng khoảng cho kỳ vọng

Bài tập 4.1.

Xác suất để một sinh viên Đại học Bách khoa Hà Nội thi trượt môn Giải tích 2 là p. Một mẫu lớn n sinh viên được lựa chọn ngẫu nhiên và ký hiệu X là số sinh viên đã trượt môn Giải tích 2 trong mẫu.

1. Giải thích tại sao có thể sử dụng $\dfrac{X}{n}$ để ước lượng cho p?
2. Trình bày cách tính xấp xỉ xác suất sự sai khác giữa $\dfrac{X}{n}$ và p nhỏ hơn 0,01? Áp dụng cho n = 500 và p = 0,2.

1. Ta có thể sử dụng $\dfrac{X}{n}$ để ước lượng cho p do tần suất mẫu ngẫu nhiên f là ước lượng không chệch, hiệu quả và vững của xác suất p của tổng thể.

   Ước lượng không chệch

   Lấy mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ trong đó $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập có cùng phân phối *Bernoulli* với tham số p. Tần số xuất hiện sinh viên thi trượt Giải tích 2 trong mẫu là

   $$X = \sum_{i=1}^{n} X_i$$

   Khi đó ta có tần suất mẫu $f = \dfrac{X}{n} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$. Suy ra

   $$E[f] = E\left[\frac{X}{n}\right] = \frac{E[X]}{n} = \frac{np}{n} = p$$

   Như vậy, f là ước lượng không chệch của p.

   Ước lượng hiệu quả

   Định lý. *(Bất đẳng thức Cramér - Rao)* Cho hàm mật độ $f_Y(y,\theta)$ của biến ngẫu nhiên liên tục Y với đạo hàm cấp 1, cấp 2 liên tục. Giả sử tập các giá trị y mà $f_Y(y,\theta) \neq 0$ không phụ thuộc vào θ.

   Xét một mẫu ngẫu nhiên $Y_1, Y_2, \ldots, Y_n$ từ $f_Y(y,\theta)$ và đặt $\hat{\theta} = h(Y_1, Y_2, \ldots, Y_n)$ là một ước lượng không chệch của θ. Khi đó ta có

   $$V\left[\hat{\theta}\right] \geq \left\{nE\left[\left(\frac{\partial \ln f_Y(y,\theta)}{\partial\theta}\right)^2\right]\right\}^{-1} = \left\{-nE\left[\left(\frac{\partial^2 \ln f_Y(y,\theta)}{\partial\theta^2}\right)\right]\right\}^{-1}$$

Kết luận tương tự nếu n quan sát thu được từ biến rời rạc X với $p_X(k, \theta)$ $\square$

Ta có $p_{X_i}(k, p) = p^k(1-p)^{1-k}$ với $k = 0, 1$ và $0 < p < 1$. Suy ra

$$\ln p_{X_i}(X_i, p) = X_i \ln p + (1 - X_i) \ln (1 - p)$$

Hơn thế nữa ta có

$$\frac{\partial \ln p_{X_i}(X_i, p)}{\partial p} = \frac{X_i}{p} - \frac{1 - X_i}{1 - p}$$

và

$$\frac{\partial^2 \ln p_{X_i}(X_i, p)}{\partial p^2} = -\frac{X_i}{p^2} - \frac{1 - X_i}{(1 - p)^2}$$

Lấy kỳ vọng của đạo hàm cấp 2 ta có

$$E\left[\frac{\partial^2 \ln p_{X_i}(X_i, p)}{\partial p^2}\right] = -\frac{p}{p^2} - \frac{1 - p}{(1 - p)^2} = -\frac{1}{p(1 - p)}$$

Suy ra chặn *Cramér - Rao* là

$$\frac{1}{-n\left[-\frac{1}{p(1 - p)}\right]} = \frac{p(1 - p)}{n} = V[f]$$

Như vậy, f là ước lượng hiệu quả của p.

Ước lượng vững

Định lý. *(Bất đẳng thức Chebyshev)* Cho W là biến ngẫu nhiên có kỳ vọng μ và phương sai σ^2. Với mọi $\varepsilon > 0$ ta có

$$P(|W - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

$\square$

Theo bất đẳng thức *Chebyshev* ta có

$$P(|f - p| < \varepsilon) \geq 1 - \frac{V[f]}{\varepsilon^2}$$

Mà $V[f] = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V[X_i] = \frac{p(1-p)}{n}$ nên

$$P(|f - p| < \varepsilon) \geq 1 - \frac{p(1 - p)}{n\varepsilon^2}$$

Với mọi ε, δ và $0 < p < 1$, luôn tìm được n sao cho $\frac{p(1-p)}{n\varepsilon^2} < \delta$. Do đó

$$\lim_{n \to \infty} P(|f - p| < \varepsilon) = 1$$

Như vậy, f là ước lượng vững của p.

2. Vì khi n đủ lớn, $np \geq 5$ và $n(1-p) \geq 5$, thì thống kê $U = \dfrac{f-p}{\sqrt{p(1-p)}}\sqrt{n} \sim \mathcal{N}(0,1)$ nên ta có

$$P\left(|f-p| < 0.01\right) = P\left(\left|f-\mu_f\right| < 0.01\right) = 2\phi\left[\frac{0.01\sqrt{n}}{\sqrt{p(1-p)}}\right]$$

Với $n = 500,\ p = 0.2$, kiểm tra $np = 100 > 5,\ n(1-p) = 400 > 5$. Ta có

$$P\left(\left|\frac{X}{n} - p\right| < 0.01\right) = 2\phi\left(\frac{0.01\sqrt{500}}{\sqrt{0.2 \times 0.8}}\right) = \phi(0.56) = 0.1772$$

Bài tập 4.2.

Tuổi thọ của một loại bóng đèn do một dây chuyền công nghệ sản xuất ra có độ lệch chuẩn là 305 giờ. Người ta lấy ngẫu nhiên ra 45 bóng đèn loại này thấy tuổi thọ trung bình là 2150 giờ. Với độ tin cậy 95% hãy ước lượng tuổi thọ trung bình của loại bóng đèn nói trên.

Gọi X là tuổi thọ của bóng đèn, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với $\sigma = 305$. Tuổi thọ trung bình của bóng đèn là $E[X] = \mu$ chưa biết cần được ước lượng.

Bước 1. Chọn thống kê $U = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n}$. Thống kê $U \sim \mathcal{N}(0,1)$.

Bước 2. Áp dụng khoảng tin cậy đối xứng $\left(\overline{x} - u_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}\right)$

Với $\alpha = 0.05$, $\Phi\left(u_{1-\frac{\alpha}{2}}\right) = 1 - \dfrac{\alpha}{2} = 0.975$, tra bảng giá trị hàm phân phối chuẩn tắc nhận được $u_{1-\frac{\alpha}{2}} = 1.96$

Bước 3. Từ số liệu đã cho ta có $n = 45,\ \sigma = 305$ và tính được $\overline{x} = 2150$, suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(2150 - 1.96 \times \frac{305}{\sqrt{45}},\ 2150 + 1.96 \times \frac{305}{\sqrt{45}}\right) = (2060.8852,\ 2239.1148)$$

Bước 4. Kết luận, với độ tin cậy 95% tuổi thọ trung bình của loại bóng đèn trên từ 2060.8858 giờ đến 2239.1148 giờ.

Bài tập 4.3.

Một kỹ sư cho biết trọng lượng tạp chất trong một sản phẩm tuân theo luật phân phối chuẩn với độ lệch chuẩn bằng 3,8gam. Một mẫu ngẫu nhiên gồm 9 sản phẩm được tiến hành kiểm tra và thấy lượng tạp chất như sau (đơn vị tính là gam):

18,2 13,7 15,9 17,4 21,8 16,6 12,3 18,8 16,2

1. Tìm khoảng tin cậy cho trọng lượng trung bình tạp chất của sản phẩm với độ tin cậy 99%.
2. Không cần tính toán, nếu độ tin cậy 95% thì khoảng ước lượng trung bình sẽ rộng hơn, hẹp hơn hay bằng như trong ý (a)?

1. Gọi X là trọng lượng tạp chất trong một sản phẩm, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với $\sigma = 3.8$. Trọng lượng trung bình của tạp chất trong một sản phẩm là $E[X] = \mu$ chưa biết cần được ước lượng.

   Bước 1. Chọn thống kê $U = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n}$. Thống kê $U \sim \mathcal{N}(0,1)$.

   Bước 2. Áp dụng khoảng tin cậy đối xứng $\left(\overline{x} - u_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}\right)$

   Với $\alpha = 0.01$, $\Phi\left(u_{1-\frac{\alpha}{2}}\right) = 1 - \dfrac{\alpha}{2} = 0.995$, tra bảng giá trị hàm phân phối chuẩn tắc nhận được $u_{1-\frac{\alpha}{2}} = 2.58$

   Bước 3. Từ số liệu đã cho ta có $n = 9$, $\sigma = 3.8$ và có $\overline{x} = 16.76667$, suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

   $$\left(16.7667 - 2.58 \times \frac{3.8}{\sqrt{9}},\ 16.7667 + 2.58 \times \frac{3.8}{\sqrt{9}}\right) = (13.4987,\ 20.0347)$$

   Bước 4. Kết luận, với độ tin cậy 99% trọng lượng trung bình của tạp chất trong một sản phẩm từ 13.4987 gam đến 20.0347 gam.

2. Nếu độ tin cậy $1 - \alpha$ giảm từ 99% xuống 95% thì khoảng ước lượng sẽ hẹp hơn khoảng ước lượng xét trên, do giá trị của $u_{1-\frac{\alpha}{2}}$ giảm từ 2.58 xuống 1.96

Chú ý: Dễ thấy rằng trong bài toán này, độ chính xác tăng $\varepsilon \uparrow$ nếu độ dài khoảng tin cậy giảm $I \downarrow$ nếu độ tin cậy giảm $\gamma \downarrow$

Bài tập 4.4.

Giả sử chiều dài của một chi tiết sản phẩm là biến ngẫu nhiên tuân theo luật phân phối chuẩn với độ lệch chuẩn là 0,2m. Người ta sản xuất thử nghiệm 35 sản phẩm loại này và tính được chiều dài trung bình là 25m. Với độ tin cậy 95% hãy ước lượng khoảng cho chiều dài trung bình của chi tiết sản phẩm đang được thử nghiệm.

Gọi X là chiều dài của một chi tiết sản phẩm, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với $\sigma = 0.2$. Chiều dài trung bình của một chi tiết sản phẩm là $E[X] = \mu$ chưa biết cần được ước lượng.

Bước 1. Chọn thống kê $U = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n}$. Thống kê $U \sim \mathcal{N}(0,1)$.

Bước 2. Áp dụng khoảng tin cậy đối xứng $\left(\overline{x} - u_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}\right)$

Với $\alpha = 0.05$, $\Phi\left(u_{1-\frac{\alpha}{2}}\right) = 1 - \dfrac{\alpha}{2} = 0.975$, tra bảng giá trị hàm phân phối chuẩn tắc nhận được $u_{1-\frac{\alpha}{2}} = 1.96$

Bước 3. Từ số liệu đã cho ta có $n = 35$, $\sigma = 0.2$ và có $\overline{x} = 25$, suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(25 - 1.96 \times \frac{0.2}{\sqrt{35}},\ 25 + 1.96 \times \frac{0.2}{\sqrt{35}}\right) = (24.9337,\ 25.0663)$$

Bước 4. Kết luận, với độ tin cậy 95% chiều dài trung bình của một chi tiết sản phẩm từ 24.9337 m đến 25.0663 m.

Bài tập 4.5.

Để xác định trọng lượng trung bình của các bao gạo được đóng gói bằng máy tự động, người ta chọn ngẫu nhiên ra 20 bao gạo và thấy trung bình mẫu là 49,2kg và độ lệch chuẩn mẫu hiệu chỉnh là 1,8kg. Biết rằng trọng lượng các bao gạo xấp xỉ phân phối chuẩn. Hãy tìm khoảng tin cậy cho trọng lượng trung bình của một bao gạo với độ tin cậy 99%.

Gọi X là trọng lượng của các bao gạo, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Trọng lượng trung bình của các bao gạo là $E[X] = \mu$ chưa biết cần được ước lượng.

Bước 1. Vì phương sai chưa biết và $n = 20 < 30$, chọn thống kê $T = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Thống kê T có phân phối *Student* với $n-1$ bậc tự do.

Bước 2. Sử khoảng tin cậy đối xứng cho $E[X] = \mu$:

$$\left(\overline{x} - t_{1-\frac{\alpha}{2}}^{(n-1)}\frac{s}{\sqrt{n}},\ \overline{x} + t_{1-\frac{\alpha}{2}}^{(n-1)}\frac{s}{\sqrt{n}}\right)$$

trong đó $t_{1-\frac{\alpha}{2}}^{(n-1)} = t_{0.995}^{(19)} = 2.861$, được xác định từ bảng phân phối *Student*.

Bước 3. Từ số liệu của đầu bài, ta có $n = 20$, $\overline{x} = 49.2$, $s = 1.8$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(49.2 - 2.861 \times \frac{1.8}{\sqrt{20}},\ 49.2 + 2.861 \times \frac{1.8}{\sqrt{20}}\right) = (48.0485,\ 50.3515)$$

Bước 4. Kết luận, với độ tin cậy 99% trọng lượng trung bình của các bao gạo từ 48.0485 kg đến 50.3515 kg.

Bài tập 4.6.

Thời gian đợi phục vụ tại một cửa hàng ăn nhanh là biến ngẫu nhiên xấp xỉ phân phối chuẩn. Người ta khảo sát 16 người thì thấy thời gian đợi trung bình là 4 phút và độ lệch chuẩn mẫu hiệu chỉnh là 1,8 phút. Với độ tin cậy 99% hãy tìm khoảng tin cậy cho thời gian chờ đợi trung bình của một khách hàng tại cửa hàng ăn nhanh này.

Gọi X là thời gian đợi phục vụ tại cửa hàng, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Thời gian đợi phục vụ trung bình tại cửa hàng là $E[X] = \mu$ chưa biết cần được ước lượng.

Bước 1. Vì phương sai chưa biết và $n = 16 < 30$, chọn thống kê $T = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Thống kê T có phân phối *Student* với $n - 1$ bậc tự do.

Bước 2. Sử dụng khoảng tin cậy đối xứng cho $E[X] = \mu$:

$$\left(\overline{x} - t_{1-\frac{\alpha}{2}}^{(n-1)} \frac{s}{\sqrt{n}},\ \overline{x} + t_{1-\frac{\alpha}{2}}^{(n-1)} \frac{s}{\sqrt{n}}\right)$$

trong đó $t_{1-\frac{\alpha}{2}}^{(n-1)} = t_{0.995}^{(15)} = 2.947$, được xác định từ bảng phân phối *Student*.

Bước 3. Từ số liệu của đầu bài, ta có $n = 16,\ \overline{x} = 4,\ s = 1.8$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(4 - 2.947 \times \frac{1.8}{\sqrt{16}},\ 4 + 2.947 \times \frac{1.8}{\sqrt{16}}\right) = (2.6739,\ 5.3262)$$

Bước 4. Kết luận, với độ tin cậy 99% thời gian đợi phục vụ trung bình tại cửa hàng từ 2.6739 phút đến 5.3262 phút.

Bài tập 4.7.

Một mẫu ngẫu nhiên gồm 16 thùng hàng được chọn ra từ tất cả các thùng hàng được sản xuất bởi nhà máy trong một tháng. Trọng lượng của 16 thùng hàng lần lượt như sau (đơn vị tính là kg):

18.6	18,4	19,2	19,8	19,4	19,5	18,9	19,4
19.7	20,1	20,2	20,1	18,6	18,4	19,2	19,8

Tìm khoảng tin cậy cho trọng lượng trung bình tổng thể của tất cả các thùng hàng của nhà máy với độ tin cậy 95%, biết rằng trọng lượng thùng hàng được chọn ngẫu nhiên là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Gọi X là trọng lượng của các thùng hàng của nhà máy, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Trọng lượng trung bình của các thùng hàng là $E[X] = \mu$ chưa biết cần được ước lượng.

Bước 1. Vì phương sai chưa biết và $n = 16 < 30$, chọn thống kê $T = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Thống kê T có phân phối *Student* với $n - 1$ bậc tự do.

Bước 2. Sử dụng khoảng tin cậy đối xứng cho $E[X] = \mu$:

$$\left(\overline{x} - t_{1-\frac{\alpha}{2}}^{(n-1)}\frac{s}{\sqrt{n}},\ \overline{x} + t_{1-\frac{\alpha}{2}}^{(n-1)}\frac{s}{\sqrt{n}}\right)$$

trong đó $t_{1-\frac{\alpha}{2}}^{(n-1)} = t_{0.975}^{(15)} = 2.131$, được xác định từ bảng phân phối *Student*.
Bước 3. Từ số liệu của đầu bài, ta tính được $n = 16,\ \overline{x} = 19.3313,\ s = 0.6097$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(19.3313 - 2.131 \times \frac{0.6097}{\sqrt{16}},\ 19.3313 + 2.131 \times \frac{0.6097}{\sqrt{16}}\right) = (19.0065,\ 19.6561)$$

Bước 4. Kết luận, với độ tin cậy 95% trọng lượng trung bình của các thùng hàng từ 19.0065 kg đến 19.6561 kg.

Bài tập 4.8.

Để định mức thời gian gia công một chi tiết máy, người ta theo dõi ngẫu nhiên quá trình gia công 35 chi tiết máy và thu được số liệu:

Thời gian (phút)	16 − 17	17 − 18	18 − 19	19 − 20	20 − 21	21 − 22
Số chi tiết máy	3	4	10	9	5	4

Giả sử thời gian gia công chi tiết máy là biến ngẫu nhiên tuân theo luật phân phối chuẩn. Với độ tin cậy 95% hãy ước lượng khoảng tin cậy cho thời gian gia công trung bình một chi tiết máy nói trên.

Gọi X là thời gian gia công một chi tiết máy, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Thời gian gia công trung bình một chi tiết máy là $E[X] = \mu$ chưa biết cần được ước lượng.
Bước 1. Chọn thống kê $U = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Vì $n = 35 > 30$ nên thống kê $U \sim \mathcal{N}(0, 1)$.
Bước 2. Khoảng tin cậy đối xứng cho $E[X] = \mu$ là

$$\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

trong đó $\alpha = 0.05,\ u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$, được tra từ bảng giá trị phân phối chuẩn tắc.
Bước 3. Từ số liệu đã cho tính được $n = 35,\ \overline{x} = 19.1,\ s = 1.4184$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(19.1 - 1.96 \times \frac{1.4184}{\sqrt{35}},\ 19.1 + 1.96 \times \frac{1.4184}{\sqrt{35}}\right) = (18.6301,\ 19.57)$$

Bước 4. Kết luận, với độ tin cậy 95% thời gian gia công trung bình một chi tiết máy từ 18.6301 phút đến 19.57 phút.

Bài tập 4.9.

Đo áp lực X (tính bằng kg/cm^2) của 18 thùng chứa ta được bảng kết quả sau:

Áp lực (kg/cm^2)	19,6	19,5	19,9	20,0	19,8	20,5	21,0	18,5	19,7
Số thùng	1	2	2	4	2	3	2	1	1

Với độ tin cậy 99% hãy tìm khoảng ước lượng đối xứng của áp lực trung bình của các thùng trên. Biết rằng áp lực là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Gọi X là áp lực của 1 thùng, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Áp lực trung bình của các thùng hàng là $E[X] = \mu$ chưa biết cần được ước lượng.

Bước 1. Vì phương sai chưa biết và $n = 18 < 30$, chọn thống kê $T = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Thống kê T có phân phối *Student* với $n-1$ bậc tự do.

Bước 2. Sử dụng khoảng tin cậy đối xứng cho $E[X] = \mu$:

$$\left(\overline{x} - t^{(n-1)}_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\ \overline{x} + t^{(n-1)}_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

trong đó $t^{(n-1)}_{1-\frac{\alpha}{2}} = t^{(17)}_{0.995} = 2.898$, được xác định từ bảng phân phối *Student*.

Bước 3. Từ số liệu của đầu bài, ta tính được $n = 18$, $\overline{x} = 19.9833$, $s = 0.5864$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(19.38333 - 2.898 \times \frac{0.5864}{\sqrt{18}},\ 19.38333 + 2.898 \times \frac{0.5864}{\sqrt{18}}\right) = (18.9828,\ 19.7838)$$

Bước 4. Kết luận, với độ tin cậy 99% áp lực trung bình của các thùng hàng từ 18.9828 kg/cm^2 đến 19.7838 kg/cm^2.

Bài tập 4.10.

Một bài báo trong Nuclear Engineering International (tháng 2 năm 1988, trang 33) mô tả một số đặc điểm của các thanh nhiên liệu được sử dụng trong một lò phản ứng hạt nhân của một công ty điện lực ở Na Uy. Người ta đo tỷ lệ làm giàu của 12 thanh và có được dữ liệu sau:

2,94 3,00 2,90 2,90 2,75 2,95 2,75 3,00 2,95 2,82 2,81 3,05

Giả sử tỷ lệ làm giàu của các thanh nhiên liệu tuân theo luật phân phối chuẩn. Hãy ước lượng khoảng cho tỷ lệ làm giàu trung bình của các thanh nhiên liệu với độ tin cậy 95%.

Gọi X là tỷ lệ làm giàu của các thanh nhiên liệu, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Tỷ lệ làm giàu trung bình của các thanh nhiên liệu là $E[X] = \mu$ chưa biết cần được ước lượng.

Bước 1. Vì phương sai chưa biết và $n = 12 < 30$, chọn thống kê $T = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Thống kê T có phân phối *Student* với $n-1$ bậc tự do.
Bước 2. Sử dụng khoảng tin cậy đối xứng cho $E[X] = \mu$:

$$\left(\overline{x} - t_{1-\frac{\alpha}{2}}^{(n-1)}\frac{s}{\sqrt{n}},\ \overline{x} + t_{1-\frac{\alpha}{2}}^{(n-1)}\frac{s}{\sqrt{n}}\right)$$

trong đó $t_{1-\frac{\alpha}{2}}^{(n-1)} = t_{0.975}^{(11)} = 2.201$, được xác định từ bảng phân phối *Student*.
Bước 3. Từ số liệu của đầu bài, ta tính được $n = 12,\ \overline{x} = 2.9017,\ s = 0.0993$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(2.9017 - 2.201 \times \frac{0.0993}{\sqrt{12}},\ 2.9017 + 2.201 \times \frac{0.0993}{\sqrt{12}}\right) = (2.8386,\ 2.9648)$$

Bước 4. Kết luận, với độ tin cậy 95% tỷ lệ làm giàu trung bình của các thanh nhiên liệu từ 2.8386 đến 2.9648.

Bài tập 4.11.

Trọng lượng những viên gạch trong một quá trình sản xuất gạch được giả sử là tuân theo luật phân phối chuẩn. Một mẫu ngẫu nhiên gồm 25 viên gạch vừa sản xuất ra trong ngày có trọng lượng trung bình 2,45 kg và độ lệch chuẩn mẫu hiệu chỉnh là 0,15 kg.

1. Tìm khoảng tin cậy của trọng lượng trung bình của tất cả các viên gạch trong ngày với độ tin cậy 99%.

2. Không cần tính toán, với độ tin cậy 95% thì khoảng tin cậy trung bình sẽ rộng hơn, hẹp hơn hay bằng với kết quả ý (a)?

3. Một mẫu ngẫu nhiên gồm 20 viên gạch sẽ được chọn ra trong ngày mai. Không cần tính toán, với độ tin cậy 99% thì khoảng tin cậy cho trọng lượng trung bình của tất cả các viên gạch sản xuất ra trong ngày mai sẽ rộng hơn, hẹp hơn hay bằng như trong ý (a)?

4. Sự thật rằng, độ lệch chuẩn mẫu của các viên gạch sản xuất trong ngày mai là 0,10kg. Không cần tính toán, với độ tin cậy 99% thì khoảng tin cậy cho trọng lượng trung bình của tất cả các viên gạch sản xuất ra trong ngày mai sẽ rộng hơn, hẹp hơn hay bằng như trong ý (a)?

1. Gọi X là trọng lượng của viên gạch trong quá trình sản xuất, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Trọng lượng trung bình của viên gạch là $E[X] = \mu$ chưa biết cần được ước lượng.

   Bước 1. Vì phương sai chưa biết và $n = 25 < 30$, chọn thống kê $T = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Thống kê T có phân phối *Student* với $n-1$ bậc tự do.

Bước 2. Sử khoảng tin cậy đối xứng cho $E[X] = \mu$:

$$\left(\overline{x} - t^{(n-1)}_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}},\ \overline{x} + t^{(n-1)}_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right)$$

trong đó $t^{(n-1)}_{1-\frac{\alpha}{2}} = t^{(24)}_{0.995} = 2.797$, được xác định từ bảng phân phối *Student*.

Bước 3. Từ số liệu của đầu bài, ta có $n = 25,\ \overline{x} = 2.45,\ s = 0.15$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(2.45 - 2.797 \times \frac{0.15}{\sqrt{25}},\ 2.45 + 2.797 \times \frac{0.15}{\sqrt{25}}\right) = (2.3661,\ 2.5339)$$

Bước 4. Kết luận, với độ tin cậy 99% trọng lượng trung bình của viên gạch trong quá trình sản xuất từ 2.3661 kg đến 2.5339 kg.

2. Nếu độ tin cậy $1 - \alpha$ giảm từ 99% xuống 95% thì khoảng ước lượng sẽ hẹp hơn khoảng ước lượng xét trên, do giá trị của $t^{(n-1)}_{1-\frac{\alpha}{2}}$ giảm từ 2.797 xuống 2.064

3. Nếu cỡ mẫu n giảm từ 25 xuống 20 thì khoảng ước lượng sẽ rộng hơn khoảng ước lượng xét ở ý a.

4. Độ lệch chuẩn mẫu hiệu chỉnh mới là $s = \sqrt{\frac{n}{n-1}}\hat{s} = \sqrt{\frac{25}{24}} \times 0.1 = 0.1021$

Nếu độ lệch chuẩn mẫu hiệu chỉnh s giảm từ 0.15 xuống 0.1021 thì khoảng ước lượng sẽ hẹp hơn khoảng ước lượng xét ở ý a.

Bài tập 4.12.

Một hường đại học lớn đang quan tâm về lượng thời gian sinh viên tự nghiên cứu mỗi tuần. Người ta tiến hành khảo sát một mẫu ngẫu nhiên gồm 16 sinh viên, dữ liệu cho thấy thời gian nghiên cứu trung bình của một sinh viên là 15,26 giờ/tuần và độ lệch chuẩn hiệu chỉnh là 6,43 giờ. Giả sử thời gian nghiên cứu của sinh viên của trường đại học trên là tuân theo luật phân phối chuẩn.

1. Tìm khoảng tin cậy cho lượng thời gian tự nghiên cứu trung bình mỗi tuần cho tất cả sinh viên trường đại học này với độ tin cậy 95%.

2. Không cần tính toán, khoảng tin cậy của trung bình tổng thể khi ước lượng sẽ rộng hơn hay hẹp hơn với ba điều kiện sau:

(a) Mẫu gồm 30 sinh viên được chọn ra, với tất cả các điều kiện khác giống như ý (a)?

(b) Độ lệch chuẩn mẫu là 4,15 giờ, tất cả các điều kiện khác giống như ý (a)?

(c) Độ tin cậy 99%, tất cả các điều kiện khác giống như ý (a)?

1. Gọi X là lượng thời gian sinh viên tự nghiên cứu mỗi tuần của một trường đại học, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Lượng thời gian trung bình sinh viên tự nghiên cứu mỗi tuần là $E[X] = \mu$ chưa biết cần được ước lượng.

   Bước 1. Vì phương sai chưa biết và $n = 16 < 30$, chọn thống kê $T = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Thống kê T có phân phối *Student* với $n-1$ bậc tự do.

   Bước 2. Sử khoảng tin cậy đối xứng cho $E[X] = \mu$:

   $$\left(\overline{x} - t^{(n-1)}_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\ \overline{x} + t^{(n-1)}_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

   trong đó $t^{(n-1)}_{1-\frac{\alpha}{2}} = t^{(15)}_{0.975} = 2.131$, được xác định từ bảng phân phối *Student*.

   Bước 3. Từ số liệu của đầu bài, ta có $n = 16,\ \overline{x} = 15.26,\ s = 6.43$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

   $$\left(15.26 - 2.131 \times \frac{6.43}{\sqrt{16}},\ 15.26 + 2.131 \times \frac{6.43}{\sqrt{16}}\right) = (11.8344,\ 18.6856)$$

   Bước 4. Kết luận, với độ tin cậy 95% lượng thời gian trung bình sinh viên tự nghiên cứu mỗi tuần từ 11.8344 giờ đến 18.6856 giờ.

2. Nhận xét rằng

   (a) Nếu cỡ mẫu n tăng từ 16 lên 30 thì bài toán trở thành ước lượng khoảng cho kỳ vọng trong trường hợp chưa biết phương sai với mẫu đủ lớn. Áp dụng khoảng tin cậy đối xứng

   $$\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

   Do đó, khoảng ước lượng sẽ hẹp hơn khoảng ước lượng xét ở ý a, vì $u_{0.975} = 1.96 < t^{(15)}_{0.975} = 2.131$

   (b) Độ lệch chuẩn mẫu hiệu chỉnh mới là $s = \sqrt{\dfrac{n}{n-1}}\hat{s} = \sqrt{\dfrac{16}{15}} \times 4.15 = 4.2861$

   Nếu độ lệch chuẩn mẫu hiệu chỉnh s giảm từ 6.43 xuống 4.2861 thì khoảng ước lượng sẽ hẹp hơn khoảng ước lượng xét ở ý a.

   (c) Nếu độ tin cậy $1-\alpha$ tăng từ 95% lên 99% thì khoảng ước lượng sẽ rộng hơn khoảng ước lượng xét trên, do giá trị của $t^{(n-1)}_{1-\frac{\alpha}{2}}$ tăng từ 2.131 lên 2.947

Bài tập 4.13.

Một kỹ sư nghiên cứu về cường độ nén của bê tông đang được thử nghiệm. Anh ta tiến hành kiểm tra 12 mẫu vật và có được các dữ liệu sau đây:

2216 2234 2225 2301 2278 2255 2249 2204 2286 2263 2275 2295

Giả sử cường độ nén của bê tông đang thử nghiệm tuân theo luật phân phối chuẩn.

1. Hãy ước lượng khoảng với độ tin cậy 95% cho cường độ nén trung bình của bê tông đang được thử nghiệm.
2. Hãy ước lượng khoảng tin cậy phải cho cường độ nén trung bình của bê tông đang được thử nghiệm với độ tin cậy 99%.

1. Gọi X là cường độ nén của bê tông đang được thử nghiệm, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Cường độ nén trung bình của bê tông đang được thử nghiệm là $E[X] = \mu$ chưa biết cần được ước lượng.

   Bước 1. Vì phương sai chưa biết và $n = 12 < 30$, chọn thống kê $T = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Thống kê T có phân phối *Student* với $n-1$ bậc tự do.

   Bước 2. Sử dụng khoảng tin cậy đối xứng cho $E[X] = \mu$:

$$\left(\overline{x} - t_{1-\frac{\alpha}{2}}^{(n-1)}\frac{s}{\sqrt{n}},\ \overline{x} + t_{1-\frac{\alpha}{2}}^{(n-1)}\frac{s}{\sqrt{n}}\right)$$

   trong đó $t_{1-\frac{\alpha}{2}}^{(n-1)} = t_{0.975}^{(11)} = 2.201$, được xác định từ bảng phân phối *Student*.

   Bước 3. Từ số liệu của đầu bài, ta tính được $n = 12,\ \overline{x} = 2256.75,\ s = 31.8123$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(2256.75 - 2.201 \times \frac{31.8123}{\sqrt{12}},\ 2256.75 + 2.201 \times \frac{31.8123}{\sqrt{12}}\right) = (2236.5373,\ 2276.9627)$$

   Bước 4. Kết luận, với độ tin cậy 95% cường độ nén trung bình của bê tông đang được thử nghiệm từ 2236.5373 đến 2276.9627.

2. Gọi X là cường độ nén của bê tông đang được thử nghiệm, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Cường độ nén trung bình của bê tông đang được thử nghiệm là $E[X] = \mu$ chưa biết cần được ước lượng.

   Bước 1. Vì phương sai chưa biết và $n = 12 < 30$, chọn thống kê $T = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Thống kê T có phân phối *Student* với $n-1$ bậc tự do.

   Bước 2. Sử dụng khoảng tin cậy phải cho $E[X] = \mu$:

$$\left(\overline{x} - t_{1-\alpha}^{(n-1)}\frac{s}{\sqrt{n}},\ +\infty\right)$$

trong đó $t_{1-\alpha}^{(n-1)} = t_{0.99}^{(11)} = 2.718$, được xác định từ bảng phân phối *Student*.

Bước 3. Từ số liệu của đầu bài, ta tính được $n = 12,\ \overline{x} = 2256.75,\ s = 31.8123$. Suy ra khoảng tin cậy phải của $E[X] = \mu$ là

$$\left(2256.75 - 2.718 \times \frac{31.8123}{\sqrt{12}},\ +\infty\right) = (2231.7595,\ +\infty)$$

Bước 4. Kết luận, với độ tin cậy 99% cường độ nén trung bình tối thiểu của bê tông đang được thử nghiệm là 2231.7595.

Bài tập 4.14.

Người ta chọn ngẫu nhiên ra 49 sinh viên của một trường đại học và thấy chiều cao trung bình mẫu là 163cm và độ lệch chuẩn mẫu hiệu chỉnh là 12cm. Hãy tìm khoảng ước lượng với độ tin cậy 99% cho chiều cao trung bình của sinh viên của trường đó.

Gọi X là chiều cao của sinh viên của một trường đại học, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Chiều cao trung bình của sinh viên là $E[X] = \mu$ chưa biết cần được ước lượng.

Bước 1. Chọn thống kê $U = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Vì $n = 49 > 30$ nên thống kê $U \sim \mathcal{N}(0, 1)$.

Bước 2. Khoảng tin cậy đối xứng cho $E[X] = \mu$ là

$$\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

trong đó $\alpha = 0.01,\ u_{1-\frac{\alpha}{2}} = u_{0.995} = 2.58$, được tra từ bảng giá trị phân phối chuẩn tắc.

Bước 3. Từ số liệu đã cho có $n = 49,\ \overline{x} = 163,\ s = 12$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(163 - 2.58 \times \frac{12}{\sqrt{49}},\ 163 + 2.58 \times \frac{12}{\sqrt{49}}\right) = (158.5771,\ 167.4229)$$

Bước 4. Kết luận, với độ tin cậy 99% chiều cao trung bình của sinh viên từ 158.5771 cm đến 167.4229 cm.

Bài tập 4.15.

Một trường đại học tiến hành một nghiên cứu xem trung bình một sinh viên tiêu hết bao nhiêu tiền điện thoại trong một tháng. Họ điều tra 60 sinh viên và cho thấy số tiền trung bình mẫu là 95 nghìn và độ lệch chuẩn mẫu hiệu chỉnh là 36 nghìn. Hãy ước lượng khoảng với độ tin cậy 95% cho số tiền điện thoại trung bình trong một tháng của mỗi sinh viên.

Gọi X là số tiền điện thoại trong một tháng của sinh viên, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Số tiền điện thoại trung bình trong một tháng của sinh viên là $E[X] = \mu$ chưa biết cần được ước lượng.

Bước 1. Chọn thống kê $U = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Vì $n = 60 > 30$ nên thống kê $U \sim \mathcal{N}(0, 1)$.
Bước 2. Khoảng tin cậy đối xứng cho $E[X] = \mu$ là

$$\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

trong đó $\alpha = 0.05$, $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$, được tra từ bảng giá trị phân phối chuẩn tắc.
Bước 3. Từ số liệu đã cho có $n = 60$, $\overline{x} = 95$, $s = 36$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(95 - 1.96 \times \frac{36}{\sqrt{60}},\ 95 + 1.96 \times \frac{36}{\sqrt{60}}\right) = (85.8907,\ 104.1093)$$

Bước 4. Kết luận, với độ tin cậy 95% số tiền điện thoại trung bình trong một tháng của sinh viên từ 85.8907 nghìn đến 104.1093 nghìn.

Bài tập 4.16.

Người ta điều tra 35 người nghiện thuốc lá được chọn ngẫu nhiên từ số lượng người nghiện hút thuốc lá của một thành phố thấy số điếu thuốc hút trong 5 ngày của họ là:

31	37	48	40	59	97	98	87	80	68	64	45
48	62	74	76	79	85	83	81	93	82	85	79
34	57	95	49	59	63	48	79	50	55	63	

Hãy tìm khoảng ước lượng cho số điếu thuốc hút trung bình trong 5 ngày của những người nghiện thuốc lá của thành phố đó với độ tin cậy 99%.

Gọi X là số điếu thuốc hút trong 5 ngày của người nghiện thuốc ở một thành phố, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Số điếu thuốc hút trung bình trong 5 ngày của người nghiện thuốc là $E[X] = \mu$ chưa biết cần được ước lượng.
Bước 1. Chọn thống kê $U = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Vì $n = 35 > 30$ nên thống kê $U \sim \mathcal{N}(0, 1)$.
Bước 2. Khoảng tin cậy đối xứng cho $E[X] = \mu$ là

$$\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

trong đó $\alpha = 0.01$, $u_{1-\frac{\alpha}{2}} = u_{0.995} = 2.58$, được tra từ bảng giá trị phân phối chuẩn tắc.
Bước 3. Từ số liệu đã cho tính được $n = 35$, $\overline{x} = 66.6571$, $s = 19.1018$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(35 - 2.58 \times \frac{19.1018}{\sqrt{35}},\ 35 + 2.58 \times \frac{19.1018}{\sqrt{35}}\right) = (26.6697,\ 43.3303)$$

Bước 4. Kết luận, với độ tin cậy 99% số điếu thuốc hút trung bình trong 5 ngày của người nghiện thuốc từ 26.6697 điếu đến 43.3303 điếu.

Bài tập 4.17.

Để nghiên cứu về thời gian xem ti vi của một thanh niên từ 18 đến 35 tuổi trong vòng một tuần, người ta tiến hành khảo sát trên 40 người và cho ta bảng số liệu sau:

39	02	43	35	15	54	23	21	25	07	24	33	17	
23	24	43	11	15	17	15	19	06	43	35	25	37	
15	14	08	11	29	12	13	25	15	28	24	06	16	7

Hãy tìm khoảng ước lượng cho thời gian xem ti vi trung bình của thanh niên trong độ tuổi trên trong vòng một tuần với độ tin cậy 99%.

Gọi X là thời gian xem ti vi trong 1 tuần của thanh niên trong độ tuổi từ 18 đến 35 tuổi, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Thời gian xem ti vi trung bình của thanh niên trong 1 tuần là $E[X] = \mu$ chưa biết cần được ước lượng.

Bước 1. Chọn thống kê $U = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Vì $n = 40 > 30$ nên thống kê $U \sim \mathcal{N}(0, 1)$.

Bước 2. Khoảng tin cậy đối xứng cho $E[X] = \mu$ là

$$\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

trong đó $\alpha = 0.01$, $u_{1-\frac{\alpha}{2}} = u_{0.995} = 2.58$, được tra từ bảng giá trị phân phối chuẩn tắc.

Bước 3. Từ số liệu đã cho tính được $n = 40$, $\overline{x} = 22.2308$, $s = 12.1449$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(22.2308 - 2.58 \times \frac{12.1449}{\sqrt{40}},\ 22.2308 + 2.58 \times \frac{12.1449}{\sqrt{40}}\right) = (17.2765,\ 27.1851)$$

Bước 4. Kết luận, với độ tin cậy 99% thời gian xem ti vi trung bình trong 1 tuần của thanh niên trong độ tuổi từ 18 đến 35 tuổi từ 17.2765 đến 27.1851.

Bài tập 4.18.

Để điều tra tiền điện phải trả trong một tháng của một hộ dân cư ở phường A, người ta kiểm tra ngẫu nhiên 200 hộ gia đình ở phường này và được kết quả sau:

Số tiền	$[80, 180)$	$[180, 280)$	$[280, 380)$	$[380, 480)$	$[480, 580)$	$[580, 680)$	$[680, 780]$
Số hộ	14	25	43	46	39	23	10

Ước lượng khoảng cho số tiền trung bình một hộ dân phải trả ở phường đó với độ tin cậy 95%.

Gọi X là số tiền điện phải trả trong 1 tháng ở phường A, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Số tiền điện phải trả trung bình trong 1 tháng ở phường A là $E[X] = \mu$ chưa biết cần được ước lượng.

Bước 1. Chọn thống kê $U = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Vì $n = 200 > 30$ nên thống kê $U \sim \mathcal{N}(0, 1)$.
Bước 2. Khoảng tin cậy đối xứng cho $E[X] = \mu$ là

$$\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

trong đó $\alpha = 0.05$, $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$, được tra từ bảng giá trị phân phối chuẩn tắc.
Bước 3. Từ số liệu đã cho tính được $n = 200$, $\overline{x} = 420$, $s = 156.597$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(420 - 1.96 \times \frac{156.597}{\sqrt{200}},\ 420 + 1.96 \times \frac{156.597}{\sqrt{200}}\right) = (398.5968,\ 441.7032)$$

Bước 4. Kết luận, với độ tin cậy 95% số tiền điện phải trả trung bình trong 1 tháng ở phường A từ 398.5968 nghìn đến 441.7032 nghìn.

Bài tập 4.19.

Để ước lượng số lượng xăng hao phí trên một tuyến đường của một hãng xe khách, người ta tiến hành chạy thử nghiệm 55 lần liên tiếp trên tuyến đường này và có được số liệu:

Lượng xăng hao phí	$10,5-11$	$11-11,5$	$11,5-12$	$12-12,5$	$12,5-13$	$13-13,5$
Tần số	5	12	15	13	6	4

Hãy ước lượng lượng xăng hao phí trung bình cho một xe với độ tin cậy 95%.

Gọi X là lượng xăng hao phí trên 1 tuyến đường của một hãng xe, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Lượng xăng hao phí trung bình trên 1 tuyến đường của một hãng xe là $E[X] = \mu$ chưa biết cần được ước lượng.
Bước 1. Chọn thống kê $U = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Vì $n = 55 > 30$ nên thống kê $U \sim \mathcal{N}(0, 1)$.
Bước 2. Khoảng tin cậy đối xứng cho $E[X] = \mu$ là

$$\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

trong đó $\alpha = 0.05$, $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$, được tra từ bảng giá trị phân phối chuẩn tắc.
Bước 3. Từ số liệu đã cho tính được $n = 55$, $\overline{x} = 11.8864$, $s = 0.6835$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(11.8864 - 1.96 \times \frac{0.6835}{\sqrt{55}},\ 11.8864 + 1.96 \times \frac{0.6835}{\sqrt{55}}\right) = (11.7058,\ 12.067)$$

Bước 4. Kết luận, với độ tin cậy 95% lượng xăng hao phí trung bình trên 1 tuyến đường của một hãng xe từ 11.7058 đến 12.067.

Bài tập 4.20.

Để xác định giá trung bình đối với một loại hàng hóa trên thị trường, người ta điều tra ngẫu nhiên tại 100 cửa hàng thu được số liệu sau:

Giá (nghìn đồng)	83	85	87	89	91	93	95	97	99	101
Số cửa hàng	5	8	13	14	30	11	8	6	4	1

Với độ tin cậy 95% hãy ước lượng giá trung bình của loại hàng đó tại thời điểm đang xét. Biết rằng giá hàng hóa là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Gọi X là giá của hàng hóa trên thị trường, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Giá trung bình của hàng hóa trên thị trường là $E[X] = \mu$ chưa biết cần được ước lượng.

Bước 1. Chọn thống kê $U = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Vì $n = 100 > 30$ nên thống kê $U \sim \mathcal{N}(0, 1)$.

Bước 2. Khoảng tin cậy đối xứng cho $E[X] = \mu$ là

$$\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

trong đó $\alpha = 0.05$, $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$, được tra từ bảng giá trị phân phối chuẩn tắc.

Bước 3. Từ số liệu đã cho tính được $n = 100$, $\overline{x} = 90.64$, $s = 4.024$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(90.64 - 1.96 \times \frac{4.024}{\sqrt{100}},\ 90.64 + 1.96 \times \frac{4.024}{\sqrt{100}}\right) = (89.8513,\ 91.4287)$$

Bước 4. Kết luận, với độ tin cậy 95% giá trung bình của hàng hóa trên thị trường từ 89.8513 nghìn đến 91.4287 nghìn.

4.2 Ước lượng khoảng cho tỷ lệ hay xác suất

Bài tập 4.21.

Để ước lượng cho tỷ lệ những cây bạch đàn có chiều cao đạt chuẩn phục vụ cho việc khai thác ở một nông trường lâm nghiệp, người ta tiến hành đo ngẫu nhiên chiều cao của 135 cây và thấy có 36 cây cao từ 7,5m trở lên. Hãy ước lượng khoảng cho tỷ lệ các cây bạch đàn có chiều cao trên 7,5m với độ tin cậy 95%.

Gọi p là tỷ lệ những cây bạch đàn có chiều cao trên 7.5 m. Kiểm tra $nf = 135 \times \dfrac{4}{15} = 36 > 5$ và $n(1-f) = 135 \times \dfrac{11}{15} = 99 > 5$.

Bước 1. Chọn thống kê $Z = \dfrac{f - p}{\sqrt{f(1-f)}}\sqrt{n}$. Thống kê $Z \sim \mathcal{N}(0, 1)$.

Bước 2. Khoảng tin cậy đối xứng của sác xuất p là

$$\left(f - u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}},\ f + u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}\right)$$

trong đó $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$ được tra từ bảng giá trị hàm phân phối chuẩn tắc.
Bước 3. Với $n = 135,\ m = 36,\ f = \dfrac{m}{n} \approx 0.2667$, suy ra khoảng tin cậy đối xứng của p là

$$\left(0.2667 - 1.96\sqrt{\frac{0.2667\times 0.7333}{135}},\ 0.2667 + 1.96\sqrt{\frac{0.2667\times 0.7333}{135}}\right) = (0.1921,\ 0.3413)$$

Bước 4. Kết luận, tỷ lệ những cây bạch đàn có chiều cao trên 7.5 m là từ 19.21% đến 34.13% với độ tin cậy 95%.

Bài tập 4.22.

Để ước lượng số cá có trong hồ người ta bắt từ hồ lên 100 con đánh dấu rồi thả lại vào hồ. Sau đó người ta bắt lên 300 con thì thấy có 32 con bị đánh dấu. Hãy ước lượng khoảng cho số cá có trong hồ với độ tin cậy 99%.

Gọi p là tỷ lệ cá bị đánh dấu trong hồ. Kiểm tra $nf = 300 \times \dfrac{8}{75} = 32 > 5$ và $n(1-f) = 300 \times \dfrac{67}{75} = 268 > 5$.
Bước 1. Chọn thống kê $Z = \dfrac{f-p}{\sqrt{f(1-f)}}\sqrt{n}$. Thống kê $Z \sim \mathcal{N}(0,1)$.
Bước 2. Khoảng tin cậy đối xứng của sác xuất p là

$$\left(f - u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}},\ f + u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}\right)$$

trong đó $u_{1-\frac{\alpha}{2}} = u_{0.995} = 2.58$ được tra từ bảng giá trị hàm phân phối chuẩn tắc.
Bước 3. Với $n = 300,\ m = 32,\ f = \dfrac{m}{n} \approx 0.1067$, suy ra khoảng tin cậy đối xứng của p là

$$\left(0.1067 - 2.58\sqrt{\frac{0.1067\times 0.8933}{300}},\ 0.1067 + 2.58\sqrt{\frac{0.1067\times 0.8933}{300}}\right) = (0.0607,\ 0.1527)$$

Bước 4. Suy ra tỷ lệ cá bị đánh dấu trong hồ là từ 6.07% đến 15.27% với độ tin cậy 99%.
Bước 5. Vì người ta bắt lên 100 con đánh dấu rồi thả xuống hồ nên khoảng ước lượng cho số cá trong hồ là

$$\left(\frac{100}{0.1527},\ \frac{100}{0.0607}\right) = (654.8788,\ 1647.4465)$$

Bước 6. Kết luận, số cá trong hồ là từ 654 con đến 1648 con với độ tin cậy 99%

Bài tập 4.23.

Để điều tra thị phần xe máy, người ta chọn ngẫu nhiên ra 450 người mua xe máy trong một tháng ở các địa bàn ở một thành phố thì có 275 người mua xe Honda. Tìm khoảng tin cậy cho tỷ lệ người mua xe Honda với độ tin cậy 95%.

Gọi p là tỷ lệ người mua xe máy Honda. Kiểm tra $nf = 450 \times \dfrac{11}{18} = 275 > 5$ và $n(1-f) = 450 \times \dfrac{7}{18} = 175 > 5$.

Bước 1. Chọn thống kê $Z = \dfrac{f-p}{\sqrt{f(1-f)}}\sqrt{n}$. Thống kê $Z \sim \mathcal{N}(0,1)$.

Bước 2. Khoảng tin cậy đối xứng của sác xuất p là

$$\left(f - u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}},\ f + u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}\right)$$

trong đó $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$ được tra từ bảng giá trị hàm phân phối chuẩn tắc.

Bước 3. Với $n = 450,\ m = 275,\ f = \dfrac{m}{n} \approx 0.6111$, suy ra khoảng tin cậy đối xứng của p là

$$\left(0.6111 - 1.96\sqrt{\frac{0.6111 \times 0.3889}{450}},\ 0.6111 + 1.96\sqrt{\frac{0.6111 \times 0.3889}{450}}\right) = (0.5661,\ 0.6561)$$

Bước 4. Kết luận, tỷ lệ người mua xe máy Honda là từ 56.61% đến 65.61% với độ tin cậy 95%.

Bài tập 4.24.

Kiểm tra ngẫu nhiên 400 sản phẩm do một hệ thống máy mới sản xuất thì thấy có 387 chính phẩm. Hãy ước lượng tỷ lệ chính phẩm tối thiểu của hệ thống máy mới với độ tin cậy 95%.

Gọi p là tỷ lệ chính phẩm của hệ thống máy mới. Kiểm tra $nf = 400 \times \dfrac{387}{400} = 387 > 5$ và $n(1-f) = 400 \times \dfrac{13}{400} = 13 > 5$.

Bước 1. Chọn thống kê $Z = \dfrac{f-p}{\sqrt{f(1-f)}}\sqrt{n}$. Thống kê $Z \sim \mathcal{N}(0,1)$.

Bước 2. Khoảng tin cậy phải của sác xuất p là

$$\left(f - u_{1-\alpha}\sqrt{\frac{f(1-f)}{n}},\ 1\right)$$

trong đó $u_{1-\alpha} = u_{0.95} = 1.65$ được tra từ bảng giá trị hàm phân phối chuẩn tắc.

Bước 3. Với $n = 400,\ m = 387,\ f = \dfrac{m}{n} = 0.9675$, suy ra khoảng tin cậy phải của p là

$$\left(0.9675 - 1.65\sqrt{\frac{0.9675 \times 0.0325}{400}},\ 1\right) = (0.9529,\ 1)$$

Bước 4. Kết luận, tỷ lệ chính phẩm tối thiểu của hệ thống máy mới là 95.29% với độ tin cậy 95%.

Bài tập 4.25.

Thử nghiệm 560 bóng đèn điện tử do một nhà máy sản xuất thì thấy 10 bóng có lỗi kỹ thuật. Hãy tìm ước lượng cho tỷ lệ bóng có lỗi kỹ thuật tối đa với độ tin cậy 95%.

Gọi p là tỷ lệ bóng có lỗi kĩ thuật. Kiểm tra $nf = 560 \times \dfrac{1}{56} = 10 > 5$ và $n(1-f) = 560 \times \dfrac{55}{56} = 550 > 5$.

Bước 1. Chọn thống kê $Z = \dfrac{f-p}{\sqrt{f(1-f)}}\sqrt{n}$. Thống kê $Z \sim \mathcal{N}(0,1)$.

Bước 2. Khoảng tin cậy trái của sác xuất p là

$$\left(0,\ f + u_{1-\alpha}\sqrt{\frac{f(1-f)}{n}}\right)$$

trong đó $u_{1-\alpha} = u_{0.95} = 1.65$ được tra từ bảng giá trị hàm phân phối chuẩn tắc.

Bước 3. Với $n = 560,\ m = 10,\ f = \dfrac{m}{n} \approx 0.0179$, suy ra khoảng tin cậy trái của p là

$$\left(0,\ 0.0179 + 1.65\sqrt{\frac{0.0179 \times 0.9821}{560}}\right) = (0,\ 0.0271)$$

Bước 4. Kết luận, tỷ lệ bóng có lỗi kĩ thuật tối đa là 2.71% với độ tin cậy 95%.

Bài tập 4.26.

Mở thử 200 hộp của kho đồ hộp thấy có 10 hộp bị biến chất. Với độ tin cậy 95% hãy ước lượng tỷ lệ hộp bị biến chất tối đa của kho.

Gọi p là tỷ lệ hộp bị biến chất. Kiểm tra $nf = 200 \times \dfrac{1}{20} = 10 > 5$ và $n(1-f) = 200 \times \dfrac{19}{20} = 190 > 5$.

Bước 1. Chọn thống kê $Z = \dfrac{f-p}{\sqrt{f(1-f)}}\sqrt{n}$. Thống kê $Z \sim \mathcal{N}(0,1)$.

Bước 2. Khoảng tin cậy trái của sác xuất p là

$$\left(0,\ f + u_{1-\alpha}\sqrt{\frac{f(1-f)}{n}}\right)$$

trong đó $u_{1-\alpha} = u_{0.95} = 1.65$ được tra từ bảng giá trị hàm phân phối chuẩn tắc.

Bước 3. Với $n = 200,\ m = 10,\ f = \dfrac{m}{n} = 0.05$, suy ra khoảng tin cậy trái của p là

$$\left(0,\ 0.05 + 1.65\sqrt{\frac{0.05 \times 0.95}{200}}\right) = (0,\ 0.0754)$$

Bước 4. Kết luận, tỷ lệ hộp bị biến chất tối đa là 7.54% với độ tin cậy 95%.

Bài tập 4.27.

Chọn ngẫu nhiên ra 1000 trường hợp điều trị bệnh ung thư phổi, các bác sĩ thống kê thấy có 823 bệnh nhân bị chết trong vòng 10 năm.

1. Ước lượng khoảng cho tỷ lệ tử vong của bệnh nhân điều trị bệnh ung thư phổi với độ tin cậy 99%.
2. Cần phải lấy số lượng mẫu là bao nhiêu để với độ tin cậy 95% các sai số khi dự đoán tỷ lệ bệnh nhân điều trị ung thư phổi tử vong 10 năm là ít hơn 0,03?

1. Gọi p là tỷ lệ tử vong của bệnh nhân điều trị bệnh ung thư phổi. Kiểm tra $nf = 1000 \times \dfrac{823}{1000} = 823 > 5$ và $n(1-f) = 1000 \times \dfrac{177}{1000} = 177 > 5$.

   Bước 1. Chọn thống kê $Z = \dfrac{f-p}{\sqrt{f(1-f)}}\sqrt{n}$. Thống kê $Z \sim \mathcal{N}(0,1)$.

   Bước 2. Khoảng tin cậy đối xứng của sác xuất p là

$$\left(f - u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}},\ f + u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}\right)$$

   trong đó $u_{1-\frac{\alpha}{2}} = u_{0.995} = 2.58$ được tra từ bảng giá trị hàm phân phối chuẩn tắc.

   Bước 3. Với $n = 1000,\ m = 823,\ f = \dfrac{m}{n} = 0.823$, suy ra khoảng tin cậy đối xứng của p là

$$\left(0.823 - 2.58\sqrt{\frac{0.823 \times 0.177}{1000}},\ 0.823 + 2.58\sqrt{\frac{0.823 \times 0.177}{1000}}\right) = (0.7919,\ 0.8541)$$

   Bước 4. Kết luận, tỷ lệ tử vong của bệnh nhân điều trị bệnh ung thư phổi là từ 79.19% đến 85.41% với độ tin cậy 99%.

2. Sai số khi dự đoán tỷ lệ bệnh nhân điều trị ung thư phổi tử vong trong 10 năm với độ tin cậy 95% là $\varepsilon = u_{1-\frac{\alpha}{2}}\sqrt{\dfrac{f(1-f)}{n}}$, trong đó $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$ được tra từ bảng giá trị hàm phân phối chuẩn tắc.

   Thay $f = 0.823$ vào biểu thức $\varepsilon < 0.03$, suy ra

$$n > \frac{u^2_{1-\frac{\alpha}{2}} f(1-f)}{0.03^2} = \frac{1.96^2 \times 0.823 \times 0.177}{0.03^2} \approx 621.7886$$

   Kết luận, cần phải lấy số lượng mẫu nhỏ nhất là $n = 622$ để với độ tin cậy 95% sai số khi dự đoán tỷ lệ bệnh nhân điều trị ung thư phổi tử vong trong 10 năm là ít hơn 0.03.

Bài tập 4.28.

Cần phải lập một mẫu ngẫu nhiên với kích thước là bao nhiêu để tỷ lệ phế phẩm của mẫu là 0,2 và độ dài khoảng tin cậy đối xứng là 0,05 và độ tin cậy của ước lượng là 95%.

Độ dài khoảng tin cậy đối xứng $I = 2\varepsilon = 0.05$ suy ra sai số ước lượng $\varepsilon = 0.025$.
Với $\gamma = 0.95$, ta nhận được $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$ được tra từ bảng giá trị hàm phân phối chuẩn tắc.
Thay tỷ lệ phế phẩm của mẫu $f = 0.2$ vào công thức sai số $\varepsilon = u_{1-\frac{\alpha}{2}}\sqrt{\dfrac{f(1-f)}{n}}$, suy ra cỡ mẫu

$$n = \frac{u^2_{1-\frac{\alpha}{2}}\, f(1-f)}{\varepsilon^2} = \frac{1.96^2 \times 0.2 \times 0.8}{0.025^2} \approx 983.4496$$

Kết luận, cần phải lập một mẫu ngẫu nhiên với kích thước 984 để thỏa mãn yêu cầu đề bài.

Bài tập 4.29.

Làm cách nào để ước lượng số thú hiếm trong một khu rừng với độ tin cậy 95%.

Giả sử ta biết tổng số thú trong khu rừng là N. Trong tổng thể ta chọn một mẫu có cỡ n và quan sát xem trong mẫu có bao nhiêu thú hiếm.
Gọi f là tần suất xuất hiện thú hiếm trong mẫu đã chọn. Chú ý chọn mẫu sao cho điều kiện $nf > 5$ và $n(1-f) > 5$ phải được thỏa mãn.
Từ đây ta có thể ước lượng khoảng tin cậy đối xứng cho tỷ lệ số thú hiếm trong khu rừng với độ tin cậy 95% và dựa vào N để suy ra khoảng ước lượng số thú hiếm trong cả khu rừng.

Bài tập 4.30.

Nghiên cứu về năng suất của loại hoa màu A, người ta kiểm tra năng suất của 64 điểm trồng loại hoa màu này thu được bảng số liệu

Năng suất (tạ/ha)	40 – 45	45 – 50	50 – 55	55 – 60	60 – 65	65 – 70
Số điểm	2	5	15	30	8	4

1. Hãy ước lượng năng suất trung bình của loại hoa màu A với độ tin cậy 95%; Nếu muốn sai số của ước lượng giảm đi 2 lần thì cần kiểm tra bao nhiêu điểm để đảm bảo yêu cầu nêu trên?
2. Biết rằng trên toàn miền Bắc có 10.000 điểm trồng loại hoa màu A. Hãy cho biết có khoảng bao nhiêu điểm đạt năng suất trên 60 tạ/ha? Hãy kết luận với độ tin cậy 99%.
3. Hãy cho biết tỷ lệ những điểm có năng suất trên 60 tạ/ha của loại hoa màu A tối thiểu là bao nhiêu? Hãy kết luận với độ tin cậy 95%?

1. Gọi X là năng suất của loại hoa màu A, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Năng suất trung bình của loại hoa màu A là $E[X] = \mu$ chưa biết cần được ước lượng.

Bước 1. Chọn thống kê $U = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Vì $n = 64 > 30$ nên thống kê $U \sim \mathcal{N}(0, 1)$.

Bước 2. Khoảng tin cậy đối xứng cho $E[X] = \mu$ là

$$\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

trong đó $\alpha = 0.05,\ u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$, được tra từ bảng giá trị phân phối chuẩn tắc.

Bước 3. Từ số liệu đã cho tính được $n = 64,\ \overline{x} = 56.3281,\ s = 5.4$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(56.3281 - 1.96 \times \frac{5.4}{\sqrt{64}},\ 56.3281 + 1.96 \times \frac{5.4}{\sqrt{64}}\right) = (55.0051,\ 57.6511)$$

Bước 4. Kết luận, với độ tin cậy 95% năng suất trung bình của loại hoa màu A từ 55.0051 tạ/ha đến 57.6511 tạ/ha.

Sai số của ước lượng là $\varepsilon = u_{1-\frac{\alpha}{2}}\dfrac{s}{\sqrt{n}} = 1.323$. Để sai số giảm đi 2 lần, tức là $\varepsilon_0 = \dfrac{1.323}{2} = 0.6615$. Ta cần có mẫu có kích thước nhỏ nhất là

$$n = \left[\frac{s^2\ u^2_{1-\frac{\alpha}{2}}}{\varepsilon_0^2}\right] = \left[\frac{5.4^2 \times 1.96^2}{0.6615^2}\right] = 256$$

2. Gọi p là tỷ lệ điểm đạt năng suất trên 60 tạ/ha. Kiểm tra $nf = 64 \times \dfrac{3}{16} = 12 > 5$ và $n(1-f) = 64 \times \dfrac{13}{16} = 48 > 5$.

Bước 1. Chọn thống kê $Z = \dfrac{f - p}{\sqrt{f(1-f)}}\sqrt{n}$. Thống kê $Z \sim \mathcal{N}(0, 1)$.

Bước 2. Khoảng tin cậy đối xứng của sác xuất p là

$$\left(f - u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}},\ f + u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}\right)$$

trong đó $u_{1-\frac{\alpha}{2}} = u_{0.995} = 2.58$ được tra từ bảng giá trị hàm phân phối chuẩn tắc.

Bước 3. Với $n = 64,\ m = 12,\ f = \dfrac{m}{n} = 0.1875$, suy ra khoảng tin cậy đối xứng của p là

$$\left(0.1875 - 2.58\sqrt{\frac{0.1875 \times 0.8125}{64}},\ 0.1875 + 2.58\sqrt{\frac{0.1875 \times 0.8125}{64}}\right) = (0.0616,\ 0.3134)$$

Bước 4. Suy ra tỷ lệ điểm đạt năng suất trên 60 tạ/ha là từ 6.16% đến 31.34% với độ tin cậy 99%.

Bước 5. Vì trên toàn miền Bắc có 10000 điểm trồng hoa màu loại A nên khoảng ước lượng cho số điểm đạt năng suất trên 60 tạ/ha là

$$(10000 \times 0.0616,\ 10000 \times 0.3134) = (616,\ 3134)$$

Bước 6. Kết luận, số điểm đạt năng suất trên 60 tạ/ha là từ 616 điểm đến 3134 điểm với độ tin cậy 99%

3. Gọi p là tỷ lệ điểm đạt năng suất trên 60 tạ/ha. Kiểm tra $nf = 64 \times \dfrac{3}{16} = 12 > 5$ và $n(1-f) = 64 \times \dfrac{13}{16} = 48 > 5$.

Bước 1. Chọn thống kê $Z = \dfrac{f-p}{\sqrt{f(1-f)}}\sqrt{n}$. Thống kê $Z \sim \mathcal{N}(0,1)$.

Bước 2. Khoảng tin cậy phải của sác xuất p là

$$\left(f - u_{1-\alpha}\sqrt{\frac{f(1-f)}{n}},\ 1\right)$$

trong đó $u_{1-\alpha} = u_{0.95} = 1.65$ được tra từ bảng giá trị hàm phân phối chuẩn tắc.

Bước 3. Với $n = 64,\ m = 12,\ f = \dfrac{m}{n} = 0.1875$, suy ra khoảng tin cậy phải của p là

$$\left(0.1875 - 1.65\sqrt{\frac{0.1875 \times 0.8125}{64}},\ 1\right) = (0.107,\ 1)$$

Bước 4. Kết luận, tỷ lệ điểm đạt năng suất trên 60 tạ/ha tối thiểu là 1.07% với độ tin cậy 95%.

5 Kiểm định giả thuyết

5.1 Kiểm định giả thuyết cho một mẫu

5.1.1 Kiểm định giả thuyết cho kỳ vọng

Bài tập 5.1.

Với các thử nghiệm về nhiệt độ nước ở một bình nước sử dụng năng lượng mặt người ta chỉ ra rằng độ lệch tiêu chuẩn là 2^oF. Người ta chọn ra ngẫu nhiên 9 ngày để tiến hành đo đạc thì thấy trung bình mẫu là 98^oF. Giả sử nhiệt độ nước tuân theo luật phân phối chuẩn. Với mức ý nghĩa 5% có thể kết luận rằng nhiệt độ trung bình sử dụng năng lượng mặt trời là bằng 99^oF hay không?

Gọi X là nhiệt độ nước ở bình nước sử dụng năng lượng mặt trời. $X \sim \mathcal{N}(\mu, \sigma^2)$ với $\sigma = 2$. Nhiệt độ nước trung bình sử dụng năng lượng mặt trời là $E[X] = \mu$ chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp đã biết phương sai.
Bước 1. Đặt giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu \neq \mu_0$ với $\mu_0 = 99$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{\overline{X} - \mu_0}{\sigma}\sqrt{n}$ nếu giả thuyết H_0 đúng. $U \sim \mathcal{N}(0, 1)$.
Bước 3. Với $\alpha = 0.05$, $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$, tra từ bảng giá trị hàm phân phối chuẩn tắc. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -1.96) \cup (1.96,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $n = 9,\ \mu_0 = 99,\ \overline{x} = 98,\ \sigma = 2$ suy ra giá trị quan sát

$$u_{qs} = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = \frac{98 - 99}{2}\sqrt{9} = -1.5$$

Bước 5. Vì $u_{qs} = -1.5 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là chưa có cơ sở để bác bỏ kết luận về nhiệt độ nước trung bình sử dụng năng lượng mặt trời là 99^oF với mức ý nghĩa 5%.

Bài tập 5.2.

Người ta tiến hành thử nghiệm một cải tiến kỹ thuật trong bộ chế hòa khí của một loại xe ôtô với hy vọng sẽ tiết kiệm được xăng hơn. Họ thử nghiệm 16 xe ô tô với bộ hòa khí có cải tiến kỹ thuật và thu được kết quả sau về số km chạy được cho một lít xăng:

20,5	20,9	20,3	20,2	20,6	20,6	20,5	21,0
21,1	21,2	20,8	20,7	20,6	20,9	20,3	20,2

Giả thiết số km chạy được cho một lít xăng tuân theo luật phân phối chuẩn. Nếu trước khi cải tiến một lít xăng trung bình chạy được 20,1 km thì có thể kết luận rằng cải tiến trên đã mang lại hiệu quả đáng kể hay không với mức ý nghĩa 5%.

Gọi X là số km chạy được cho một lít xăng. $X \sim \mathcal{N}(\mu, \sigma^2)$ với σ chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 16 < 30$.
Bước 1. Đặt giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu > \mu_0$ với $\mu_0 = 20.1$.
Bước 2. Chọn tiêu chuẩn kiểm định $T = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $T \sim \mathcal{T}^{(n-1)}$.
Bước 3. Với $\alpha = 0.05$, tra bảng phân phối *Student* được $t_{1-\alpha}^{(n-1)} = t_{0.95}^{(15)} = 1.753$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(t_{1-\alpha}^{(n-1)},\ +\infty\right) = (1.753,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $n = 16,\ \overline{x} = 20.65,\ s = 0.3141$ với $\mu_0 = 20.1$ suy ra giá trị quan sát

$$t_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{20.65 - 20.1}{0.3141}\sqrt{16} \simeq 7.0041$$

Bước 5. Vì $t_{qs} = 7.0041 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là có cơ sở để kết luận rằng cải tiến trên đã mang lại hiệu quả đáng kể với mức ý nghĩa 5%.

Bài tập 5.3.

Một nhà máy đưa ra định mức thời gian hoàn thành sản phẩm là 24 phút. Khi khảo sát thời gian hoàn thành sản phẩm của 22 công nhân, ta tính được thời gian trung bình hoàn thành sản phẩm trong mẫu là 25,2 phút, độ lệch chuẩn mẫu hiệu chỉnh 2,6 phút. Với mức ý nghĩa 5% người quản lý nhà máy có cần phải đổi định mức không. Giả sử rằng thời gian hoàn thành một sản phẩm là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Gọi X là thời gian hoàn thành một sản phẩm. $X \sim \mathcal{N}(\mu, \sigma^2)$ với σ chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 22 < 30$.
Bước 1. Đặt giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu \neq \mu_0$ với $\mu_0 = 24$.
Bước 2. Chọn tiêu chuẩn kiểm định $T = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $T \sim \mathcal{T}^{(n-1)}$.
Bước 3. Với $\alpha = 0.05$, tra bảng phân phối *Student* được $t_{1-\frac{\alpha}{2}}^{(n-1)} = t_{0.975}^{(21)} = 2.08$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -t_{1-\frac{\alpha}{2}}^{(n-1)}\right) \cup \left(t_{1-\frac{\alpha}{2}}^{(n-1)},\ +\infty\right) = (-\infty,\ -2.08) \cup (2.08,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $n = 22,\ \overline{x} = 25.2,\ s = 2.6$ với $\mu_0 = 24$ suy ra giá trị quan sát

$$t_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{25.2 - 24}{2.6}\sqrt{22} \simeq 2.1648$$

Bước 5. Vì $t_{qs} = 2.1648 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là có cơ sở để người quản lý nhà máy thay đổi định mức với mức ý nghĩa 5%.

Bài tập 5.4.

Một dây dây chuyền sản xuất dầu gội đầu, mỗi thùng dầu gội có trọng lượng trung bình là 20kg. Một mẫu ngẫu nhiên gồm 10 thùng được chọn ra ngẫu nhiên để cân có trọng lượng (kg) như sau:

21,4 19,7 19,9 20,6 20,8 20,1 19,7 20,3 20,9 20,8

Giả sử rằng trọng lượng của mỗi thùng dầu gội tuân theo luật phân phối chuẩn. Hãy kiểm định giả thuyết ở mức ý nghĩa 5% với giả thuyết cho rằng quá trình sản xuất hoạt động một cách chính xác.

Gọi X là trọng lượng của mỗi thùng dầu gội. $X \sim \mathcal{N}(\mu, \sigma^2)$ với σ chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 10 < 30$.

Bước 1. Đặt giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu \neq \mu_0$ với $\mu_0 = 20$.

Bước 2. Chọn tiêu chuẩn kiểm định $T = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $T \sim \mathcal{T}^{(n-1)}$.

Bước 3. Với $\alpha = 0.05$, tra bảng phân phối *Student* được $t^{(n-1)}_{1-\frac{\alpha}{2}} = t^{(9)}_{0.975} = 2.262$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -t^{(n-1)}_{1-\frac{\alpha}{2}}\right) \cup \left(t^{(n-1)}_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -2.262) \cup (2.262,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $n = 10,\ \overline{x} = 20.42,\ s = 0.5712$ với $\mu_0 = 20$ suy ra giá trị quan sát

$$t_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{20.42 - 20}{0.5712}\sqrt{10} \simeq 2.3252$$

Bước 5. Vì $t_{qs} = 2.3252 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là có cơ sở để cho rằng quá trình sản xuất hoạt động một cách không chính xác với mức ý nghĩa 5%.

Bài tập 5.5.

Gạo được đóng gói bằng máy tự động có trọng lượng đóng bao theo quy định 25kg. Người ta chọn ngẫu ngẫu nhiên 25 bao được đóng bằng máy tự động trên ra kiểm tra trọng lượng của chúng ta được bảng số liệu sau:

Trọng lượng (kg)	$24,6-24,8$	$24,8-25,0$	$25,0-25,2$	$25,2-25,4$	$25,4-25,6$
Tần suất	3	7	8	5	2

Giả sử trọng lượng của các bao gạo tuân theo luật phân phối chuẩn. Hỏi trọng lượng trung bình của các bao gạo được đóng gói tự động giống như yêu cầu hay phải dừng máy để điều chỉnh với mức ý nghĩa 5%?

Gọi X là trọng lượng của các bao gạo. $X \sim \mathcal{N}(\mu, \sigma^2)$ với σ chưa biết. Đây là bài toán kiểm

định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 25 < 30$.
Bước 1. Đặt giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu \neq \mu_0$ với $\mu_0 = 25$.
Bước 2. Chọn tiêu chuẩn kiểm định $T = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $T \sim \mathcal{T}^{(n-1)}$.
Bước 3. Với $\alpha = 0.05$, tra bảng phân phối *Student* được $t^{(n-1)}_{1-\frac{\alpha}{2}} = t^{(24)}_{0.975} = 2.064$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -t^{(n-1)}_{1-\frac{\alpha}{2}}\right) \cup \left(t^{(n-1)}_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -2.064) \cup (2.064,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $n = 25,\ \overline{x} = 25.608,\ s = 0.2286$ với $\mu_0 = 25$ suy ra giá trị quan sát

$$t_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{25.608 - 25}{0.2286}\sqrt{25} \simeq 13.2983$$

Bước 5. Vì $t_{qs} = 13.2983 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là có cơ sở để dừng máy để điều chỉnh với mức ý nghĩa 5%.

Bài tập 5.6.

Định mức thời gian hoàn thành một sản phẩm là 14 phút. Có cần thay đổi định mức không, nếu theo dõi thời gian hoàn thành một sản phẩm ở 25 công nhân ta thu được bảng số liệu sau:

Thời gian sản xuất 1 sản phẩm (phút)	10 – 12	12 – 14	14 – 16	16 – 18	20 – 22
Số công nhân tương ứng	3	6	10	4	2

Yêu cầu kết luận với mức ý nghĩa 5%, biết rằng thời gian hoàn thành một sản phẩm là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Gọi X là thời gian hoàn thành một sản phẩm. $X \sim \mathcal{N}(\mu, \sigma^2)$ với σ chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 25 < 30$.
Bước 1. Đặt giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu \neq \mu_0$ với $\mu_0 = 14$.
Bước 2. Chọn tiêu chuẩn kiểm định $T = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $T \sim \mathcal{T}^{(n-1)}$.
Bước 3. Với $\alpha = 0.05$, tra bảng phân phối *Student* được $t^{(n-1)}_{1-\frac{\alpha}{2}} = t^{(24)}_{0.975} = 2.064$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -t^{(n-1)}_{1-\frac{\alpha}{2}}\right) \cup \left(t^{(n-1)}_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -2.064) \cup (2.064,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $n = 25,\ \overline{x} = 14.84,\ s = 2.5768$ với $\mu_0 = 14$ suy ra giá trị quan sát

$$t_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{14.84 - 14}{2.5768}\sqrt{25} \simeq 1.6299$$

Bước 5. Vì $t_{qs} = 1.6299 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là chưa có cơ sở để thay đổi định mức hoàn thành sản phẩm với mức ý nghĩa 5%.

Bài tập 5.7.

Trọng lượng đóng gói bánh loại 250g một gói trên một máy tự động là biến ngẫu nhiên. Kiểm tra ngẫu nhiên 100 gói thu được kết quả sau:

Trọng lượng (gam)	245	247	248	250	252	253	254
Số gói	8	12	20	32	16	8	4

Có thể coi trọng lượng trung bình của các gói bánh là bằng 250g theo quy định hay không với mức ý nghĩa 5%?

Gọi X là trọng lượng của các gói bánh trên một máy tự động. Ta thấy $E[X] = \mu$ là trọng lượng trung bình của các gói bánh chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 100 > 30$.
Bước 1. Kiểm tra giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu \neq \mu_0$ với $\mu_0 = 250$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $U \sim \mathcal{N}(0, 1)$.
Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -1.96) \cup (1.96,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $\mu_0 = 250, n = 100,\ \overline{x} = 249.56,\ s = 2.3966$ suy ra giá trị quan sát

$$u_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{249.56 - 250}{2.3966}\sqrt{100} \simeq -1.8359$$

Bước 5. Vì $u_{qs} = -1.8359 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là có thể coi trọng lượng trung bình của các gói bánh là 250 g theo quy định với mức ý nghĩa 5%.

Bài tập 5.8.

Kiểm tra lượng điện áp đầu vào của một loại máy tính bảng, người ta tiến hành thử nghiệm 100 lần đo và thu được điện áp trung bình 5,04V với độ lệch chuẩn mẫu hiệu chỉnh 0,064V. Với mức ý nghĩa 5%, hãy kiểm định lượng điện áp trung bình đầu vào của loại máy tính bảng có đúng bằng 5V hay không?

Gọi X là lượng điện áp đầu vào của một máy tính bảng. Ta thấy $E[X] = \mu$ là lượng điện áp đầu vào trung bình của một máy tính bảng chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 100 > 30$.
Bước 1. Kiểm tra giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu \neq \mu_0$ với $\mu_0 = 5$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $U \sim \mathcal{N}(0, 1)$.

Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -1.96) \cup (1.96,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $\mu_0 = 5, n = 100,\ \overline{x} = 5.04,\ s = 0.064$ suy ra giá trị quan sát

$$u_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{5.04 - 5}{0.064}\sqrt{100} \simeq 6.25$$

Bước 5. Vì $u_{qs} = 6.25 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là lượng điện áp đầu vào trung bình của một máy tính bảng không đúng bằng 5 V với mức ý nghĩa 5%.

Bài tập 5.9.

Gọi X là thời gian sản xuất một sản phẩm (phút). Định mức cũ để sản xuất một sản phẩm là 20 phút. Nay do cải tiến kỹ thuật, người ta sản xuất thử 100 sản phẩm và thu được số liệu:

Thời gian sản xuất sản phẩm	16 − 17	17 − 18	18 − 19	19 − 20	20 − 21	21 − 22
Số sản phẩm tương ứng	6	10	24	30	18	12

Với mức ý nghĩa 5% có thể nói rằng việc cải tiến kỹ thuật giảm bớt thời gian sản xuất một sản phẩm hay không? Biết rằng thời gian sản xuất một sản phẩm là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Gọi X là thời gian sản xuất một sản phẩm. Ta thấy $E[X] = \mu$ là thời gian sản xuất trung bình một sản phẩm chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 100 > 30$.

Bước 1. Kiểm tra giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu < \mu_0$ với $\mu_0 = 20$.

Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $U \sim \mathcal{N}(0, 1)$.

Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\alpha} = u_{0.95} = 1.65$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (-\infty,\ -u_{1-\alpha}) = (-\infty,\ -1.65)$$

Bước 4. Từ số liệu của đầu bài ta có $\mu_0 = 20, n = 100,\ \overline{x} = 19.3,\ s = 1.3484$ suy ra giá trị quan sát

$$u_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{19.3 - 20}{1.3484}\sqrt{100} \simeq -5.1913$$

Bước 5. Vì $u_{qs} = -5.1913 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là có cơ sở để nói rằng việc cải tiến kỹ thuật giảm bớt thời gian sản xuất một sản phẩm với mức ý nghĩa 5%.

Bài tập 5.10.

Hàm lượng đường trung bình của một loại trái cây lúc đầu là 5(%). Người ta chăm bón bằng một loại NPK mới và sau một thời gian kiểm tra một số trái cây được kết quả sau:

Hàm lượng	1 – 5	5 – 9	9 – 13	13 – 17	17 – 21	21 – 25	25 – 29	29 – 33	37 – 41
Số trái	51	47	39	36	32	8	7	3	2

Hãy cho kết luận về loại NPK trên với mức ý nghĩa 5%. Giả thiết hàm lượng đường của loại trái là biến ngẫu nhiên tuân theo luật phân phối chuẩn.

Gọi X là hàm lượng đường của một loại trái cây. Ta thấy $E[X] = \mu$ là hàm lượng đường trung bình của một loại trái cây chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 225 > 30$.
Bước 1. Kiểm tra giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu \neq \mu_0$ với $\mu_0 = 5$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $U \sim \mathcal{N}(0, 1)$.
Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -1.96) \cup (1.96,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $\mu_0 = 5, n = 225,\ \overline{x} = 11.5689,\ s = 7.4039$ suy ra giá trị quan sát

$$u_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{11.5689 - 5}{7.4039}\sqrt{225} \simeq 13.3083$$

Bước 5. Vì $u_{qs} = 13.3083 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là loại phân NPK trên làm thay đổi hàm lượng đường của một loại trái cây với mức ý nghĩa 5%.

Bài tập 5.11.

Một nhà phân phối sữa trong một thành phố khẳng định rằng: bằng cách quảng cáo và cách tiếp cận khách hàng mới ở các cửa hàng, mỗi tuần trong các cửa hàng bán trung bình tăng thêm 20 hộp sữa. Người ta tiến hành chọn ra một mẫu ngẫu nhiên gồm 40 cửa hàng để xác định lời khẳng định trên thì thấy trung bình mỗi cửa hàng chỉ bán thêm được 16,4 hộp sữa và độ lệch chuẩn mẫu hiệu chỉnh là 7,2. Kiểm định giả thuyết cho rằng mỗi tuần bán thêm được 20 hộp sữa ở mỗi cửa hàng với mức ý nghĩa 5%.

Gọi X là số hộp sữa bán thêm được mỗi tuần. Ta thấy $E[X] = \mu$ là số hộp sữa trung bình bán thêm được mỗi tuần chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 40 > 30$.
Bước 1. Kiểm tra giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu \neq \mu_0$ với $\mu_0 = 5$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $U \sim \mathcal{N}(0, 1)$.

Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -1.96) \cup (1.96,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $\mu_0 = 20, n = 40,\ \overline{x} = 16.4,\ s = 7.2$ suy ra giá trị quan sát

$$u_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{16.4 - 20}{7.2}\sqrt{40} \simeq -3.1623$$

Bước 5. Vì $u_{qs} = -3.1623 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là không có cơ sở để cho rằng mỗi tuần bán thêm được 20 hộp sữa ở mỗi cửa hàng với mức ý nghĩa 5%.

5.1.2 Kiểm định giả thuyết cho tỷ lệ

Bài tập 5.12.

Người ta quan tâm tới việc lây lan dịch sốt xuất huyết ở một phường. Theo số liệu năm ngoái tỷ lệ mắc bệnh sốt xuất huyết của vùng này là 8%. Người ta tiến hành kiểm tra sức khỏe ngẫu nhiên 200 người ở phường này thì thấy có 17 người mang vi trùng sốt xuất huyết. Tỷ lệ mắc bệnh sốt xuất huyết của phường có tăng lên hay không với mức ý nghĩa 5%.

Gọi p là tỷ lệ mắc bệnh sốt xuất huyết của phường. Đây là bài toán kiểm định giả thuyết về tỷ lệ của tổng thể.
Bước 1. Đặt giả thuyết $H_0\colon p = p_0$, đối thuyết $H_1\colon p > p_0$ với $p_0 = 0.08$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$ nếu giả thuyết H_0 đúng.
Vì $np_0 = 16 > 5$ và $n(1-p_0) = 184 > 5$ khá lớn nên $U \sim \mathcal{N}(0,1)$.
Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\alpha} = u_{0.95} = 1.65$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (u_{1-\alpha},\ +\infty) = (1.65,\ +\infty)$$

Bước 4. Từ số liệu đã cho ta có $n = 200,\ m = 17$ tính được $f = \dfrac{m}{n} = \dfrac{17}{200} = 0.085$, với $p_0 = 0.08$ suy ra giá trị quan sát

$$u_{qs} = \frac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0.085 - 0.08}{\sqrt{0.08 \times 0.92}}\sqrt{200} \approx 0.2606$$

Bước 5. Vì $u_{qs} = 0.2606 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là chưa có cơ sở để khẳng định tỷ lệ mắc bệnh sốt xuất huyết của phường tăng lên với mức ý nghĩa 5%.

Bài tập 5.13.

Một hãng xà phòng A tuyên bố rằng 64% số các bà nội trợ thích sử dụng bột giặt của hãng. Người ta chọn ra một mẫu gồm 100 bà nội trợ và hỏi thì có 58 bà tỏ ra là thích sử dụng bột giặt của hãng A. Với mức ý nghĩa 1%, số liệu trên có chứng tỏ là tuyên bố của hãng xà phòng A là đúng hay không?

Gọi p là tỷ lệ các bà nội trợ thích sử dụng bột giặt của hãng A. Đây là bài toán kiểm định giả thuyết về tỷ lệ của tổng thể.
Bước 1. Đặt giả thuyết $H_0: p = p_0$, đối thuyết $H_1: p \neq p_0$ với $p_0 = 0.64$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$ nếu giả thuyết H_0 đúng.
Vì $np_0 = 64 > 5$ và $n(1-p_0) = 36 > 5$ khá lớn nên $U \sim \mathcal{N}(0,1)$.
Bước 3. Với $\alpha = 0.01$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\frac{\alpha}{2}} = u_{0.995} = 2.58$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -2.58) \cup (2.58,\ +\infty)$$

Bước 4. Từ số liệu đã cho ta có $n = 100,\ m = 58$ tính được $f = \dfrac{m}{n} = \dfrac{29}{50} = 0.58$, với $p_0 = 0.64$ suy ra giá trị quan sát

$$u_{qs} = \frac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0.58 - 0.64}{\sqrt{0.64 \times 0.36}}\sqrt{100} \approx -1.25$$

Bước 5. Vì $u_{qs} = -1.25 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là có cơ sở để chứng tỏ tuyên bố của hãng A là đúng với mức ý nghĩa 1%.

Bài tập 5.14.

Tỷ lệ phế phẩm do một máy tự động sản xuất là 5%. Kiểm tra ngẫu nhiên 300 sản phẩm thấy có 24 phế phẩm. Từ đó có ý kiến cho rằng tỷ lệ phế phẩm do máy đó sản xuất có chiều hướng tăng lên. Hãy kết luận ý kiến nêu trên với mức ý nghĩa 5%.

Gọi p là tỷ lệ phế phẩm do một máy tự động sản xuất. Đây là bài toán kiểm định giả thuyết về tỷ lệ của tổng thể.
Bước 1. Đặt giả thuyết $H_0: p = p_0$, đối thuyết $H_1: p > p_0$ với $p_0 = 0.05$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$ nếu giả thuyết H_0 đúng.
Vì $np_0 = 15 > 5$ và $n(1-p_0) = 285 > 5$ khá lớn nên $U \sim \mathcal{N}(0,1)$.
Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\alpha} = u_{0.95} = 1.65$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (u_{1-\alpha},\ +\infty) = (1.65,\ +\infty)$$

Bước 4. Từ số liệu đã cho ta có $n = 300,\ m = 24$ tính được $f = \dfrac{m}{n} = \dfrac{2}{25} = 0.08$, với $p_0 = 0.05$ suy ra giá trị quan sát

$$u_{qs} = \frac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0.08 - 0.05}{\sqrt{0.05 \times 0.95}}\sqrt{300} \approx 2.3842$$

Bước 5. Vì $u_{qs} = 2.3842 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là có cơ sở để cho rằng tỷ lệ phế phẩm do máy đó sản xuất có chiều hướng tăng lên với mức ý nghĩa 5%.

Bài tập 5.15.

Nếu áp dụng phương pháp công nghệ thứ nhất thì tỷ lệ phế phẩm là 6%, còn nếu áp dụng phương pháp công nghệ thứ hai thì trong 100 sản phẩm có 5 phế phẩm. Vậy có thể kết luận áp dụng phương pháp công nghệ thứ hai thì tỷ lệ phế phẩm thấp hơn tỷ lệ phế phẩm của phương pháp công nghệ thứ nhất không? Yêu cầu kết luận với mức ý nghĩa 5%.

Gọi p là tỷ lệ phế phẩm khi áp dụng phương pháp công nghệ thứ hai. Đây là bài toán kiểm định giả thuyết về tỷ lệ của tổng thể.
Bước 1. Đặt giả thuyết $H_0\colon p = p_0$, đối thuyết $H_1\colon p < p_0$ với $p_0 = 0.06$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$ nếu giả thuyết H_0 đúng.
Vì $np_0 = 6 > 5$ và $n(1-p_0) = 94 > 5$ khá lớn nên $U \sim \mathcal{N}(0, 1)$.
Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\alpha} = u_{0.95} = 1.65$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (-\infty,\ -u_{1-\alpha}) = (-\infty,\ -1.65)$$

Bước 4. Từ số liệu đã cho ta có $n = 100,\ m = 5$ tính được $f = \dfrac{m}{n} = \dfrac{1}{20} = 0.05$, với $p_0 = 0.06$ suy ra giá trị quan sát

$$u_{qs} = \frac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0.05 - 0.06}{\sqrt{0.06 \times 0.94}}\sqrt{100} \approx -0.4211$$

Bước 5. Vì $u_{qs} = -0.4211 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là chưa có cơ sở để kết luận áp dụng phương pháp công nghệ thứ hai thì tỷ lệ phế phẩm thấp hơn áp dụng phương pháp công nghệ thứ nhất với mức ý nghĩa 5%.

Bài tập 5.16.

Tỷ lệ bệnh nhân khỏi bệnh T khi điều trị bằng thuốc A là 85%. Thí nghiệm dùng loại thuốc B để chữa bệnh thì trong số 900 người mắc bệnh T có 810 người được chữa khỏi. Như vậy có thể kết luận thuốc B hiệu quả hơn thuốc A hay không? Yêu cầu kết luận với mức ý nghĩa 5%.

Gọi p là tỷ lệ bệnh nhân khỏi bệnh T khi điều trị bằng thuốc B. Đây là bài toán kiểm định giả thuyết về tỷ lệ của tổng thể.
Bước 1. Đặt giả thuyết $H_0\colon p = p_0$, đối thuyết $H_1\colon p > p_0$ với $p_0 = 0.85$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$ nếu giả thuyết H_0 đúng.
Vì $np_0 = 765 > 5$ và $n(1-p_0) = 135 > 5$ khá lớn nên $U \sim \mathcal{N}(0, 1)$.
Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\alpha} = u_{0.95} = 1.65$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (u_{1-\alpha},\ +\infty) = (1.65,\ +\infty)$$

Bước 4. Từ số liệu đã cho ta có $n = 900,\ m = 810$ tính được $f = \dfrac{m}{n} = \dfrac{9}{10} = 0.9$, với $p_0 = 0.85$ suy ra giá trị quan sát

$$u_{qs} = \frac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0.9 - 0.85}{\sqrt{0.85 \times 0.15}}\sqrt{900} \approx 4.2008$$

Bước 5. Vì $u_{qs} = 4.2008 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là có thể kết luận thuốc B hiệu quả hơn thuốc A với mức ý nghĩa 5%.

5.2 Kiểm định giả thuyết cho hai mẫu

5.2.1 So sánh hai kỳ vọng

Bài tập 5.17.

Hai công thức khác nhau về nhiên liệu động cơ oxy hóa được tiến hành thử nghiệm để đưa ra chỉ số octan. Phương sai của công thức I là $\sigma_1^2 = (1,5)^2$ của công thức II là $\sigma_2^2 = (1,3)^2$. Người ta chọn ngẫu nhiên $n_1 = 15$ mẫu của công thức I và $n_2 = 18$ mẫu của công thức II thì thấy $\overline{x}_1 = 89,7$ và $\overline{x}_2 = 91,5$. Giả sử rằng chỉ số octan của công thức I và II tuân theo luật phân phối chuẩn. Với mức ý nghĩa 5% có thể cho rằng công thức I có chỉ số octan ít hơn so với công thức II hay không?

Gọi X_1, X_2 lần lượt là chỉ số octan của công thức I và II. Ta có $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2),\ X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Đây là bài toán so sánh hai kỳ vọng của hai biến ngẫu nhiên có phân phối chuẩn trường hợp đã biết phương sai $\sigma_1^2 = 1.5^2,\ \sigma_2^2 = 1.3^2$.
Bước 1. Đặt giả thuyết $H_0\colon \mu_1 = \mu_2$, đối thuyết $H_1\colon \mu_1 < \mu_2$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ nếu H_0 đúng. $U \sim \mathcal{N}(0,1)$.
Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\alpha} = u_{0.95} = 1.65$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (-\infty,\ -u_{1-\alpha}) = (-\infty,\ -1.65)$$

Bước 4. Từ số liệu đã cho ta có $n_1 = 15,\ n_2 = 18,\ \overline{x}_1 = 89.7,\ \sigma_1^2 = 1.5^2,\ \overline{x}_2 = 91.5,\ \sigma_2^2 = 1.3^2$ suy ra giá trị quan sát

$$u_{qs} = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{89.7 - 91.5}{\sqrt{\dfrac{1.5^2}{15} + \dfrac{1.3^2}{18}}} \approx -3.6448$$

Bước 5. Vì $u_{qs} = -3.6448 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là có cơ sở để cho rằng công thức I có chỉ số octan ít hơn so với công thức II với mức ý nghĩa 5%.

Bài tập 5.18.

Chọn ngẫu nhiên 100 thiết bị điện tử của nhà máy I thấy tuổi thọ trung bình là 1658 giờ, độ lệch chuẩn mẫu là 123 giờ. Chọn ngẫu nhiên 110 thiết bị điện tử của nhà máy II thấy tuổi thọ trung bình là 1717 giờ, độ lệch chuẩn mẫu là 107 giờ. Với mức ý nghĩa 1%, hãy kiểm định giả thiết có phải thực sự tuổi thọ trung bình thiết bị điện tử của nhà máy II là lớn hơn nhà máy I hay không?

Gọi X, Y lần lượt là tuổi thọ thiết bị điện tử của nhà máy I và II. Ta có $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Đây là bài toán so sánh hai kỳ vọng của hai biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n_1 = 100 > 30$, $n_2 = 110 > 30$.
Bước 1. Đặt giả thuyết $H_0\colon \mu_1 = \mu_2$, đối thuyết $H_1\colon \mu_1 < \mu_2$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ nếu H_0 đúng. $U \sim \mathcal{N}(0,1)$.
Bước 3. Với $\alpha = 0.01$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\alpha} = u_{0.99} = 2.33$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (-\infty,\ -u_{1-\alpha}) = (-\infty,\ -2.33)$$

Bước 4. Từ số liệu đã cho ta có $n_1 = 100$, $n_2 = 110$, $\overline{x} = 1658$, $s_1^2 = \dfrac{100}{99} \times 123^2$, $\overline{y} = 1717$, $s_2^2 = \dfrac{110}{109} \times 107^2$ suy ra giá trị quan sát

$$u_{qs} = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{1658 - 1717}{\sqrt{\dfrac{123^2}{99} + \dfrac{107^2}{109}}} \approx -3.6742$$

Bước 5. Vì $u_{qs} = -3.6742 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là có cơ sở để cho rằng tuổi thọ trung bình của thiết bị điện tử của nhà máy II là lớn hơn nhà máy I với mức ý nghĩa 1%.

Bài tập 5.19.

Hai máy tự động dùng để cắt những thanh thép do cùng một kỹ thuật viên phụ trách và căn chỉnh. Từ mỗi máy lấy ra 35 thanh thép để kiểm tra thu được kết quả sau:

- Máy 1: Trung bình mẫu 11,7m, độ lệch chuẩn mẫu hiệu chỉnh 0,12m.
- Máy 2: Trung bình mẫu 11,6m, độ lệch chuẩn mẫu hiệu chỉnh 0,14m.

Giả sử chiều dài thanh thép do các máy sản xuất tuân theo luật phân phối chuẩn và có phương sai như nhau. Với mức ý nghĩa 5% có thể cho rằng chiều dài của các thanh thép do hai máy sản xuất là khác nhau hay không?

Gọi X, Y lần lượt là chiều dài thanh thép được sản xuất bởi các máy I và II. Ta có $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Đây là bài toán so sánh hai kỳ vọng của hai biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n_1 = n_2 = 35 > 30$.

Bước 1. Đặt giả thuyết $H_0: \mu_1 = \mu_2$, đối thuyết $H_1: \mu_1 \neq \mu_2$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ nếu H_0 đúng. $U \sim \mathcal{N}(0,1)$.
Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -1.96) \cup (1.96,\ +\infty)$$

Bước 4. Từ số liệu đã cho ta có $n_1 = n_2 = 35,\ \overline{x} = 11.7,\ s_1^2 = 0.12^2,\ \overline{y} = 11.6,\ s_2^2 = 0.14^2$ suy ra giá trị quan sát

$$u_{qs} = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{11.7 - 11.6}{\sqrt{\dfrac{0.12^2}{35} + \dfrac{0.14^2}{35}}} \approx 3.2084$$

Bước 5. Vì $u_{qs} = 3.2084 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là có thể cho rằng chiều dài của các thanh thép do hai máy sản xuất là khác nhau với mức ý nghĩa 5%.

Bài tập 5.20.

Hai công ty I và II cùng sản xuất ra một loại sản phẩm và cạnh tranh nhau trên thị trường. Người ta chọn ngẫu nhiên ra $n_1 = 11$ ngày và $n_2 = 18$ ngày để khảo sát số lượng sản phẩm được bán ra trong ngày của hai công ty I và II tương ứng và có được kết quả:

- Công ty I: trung bình mẫu 237, độ lệch chuẩn mẫu hiệu chỉnh 23;
- Công ty II: trung bình mẫu 247, độ lệch chuẩn mẫu hiệu chỉnh 27.

Giả sử số lượng hàng bán ra trong một ngày của hai công ty là tuân theo luật phân phối chuẩn, có cùng phương sai. Phải chăng lượng hàng bán ra của công ty II là nhiều hơn so với công ty I với mức ý nghĩa 1%?

Gọi X, Y lần lượt là số lượng sản phẩm được bán ra trong ngày của công ty I và II. Ta có $X \sim \mathcal{N}(\mu_1, \sigma_1^2),\ Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Đây là bài toán so sánh hai kỳ vọng của hai biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n_1 = 11 < 30,\ n_2 = 18 < 30$ và $\sigma_1^2 = \sigma_2^2$.
Bước 1. Đặt giả thuyết $H_0: \mu_1 = \mu_2$, đối thuyết $H_1: \mu_1 < \mu_2$.
Bước 2. Chọn tiêu chuẩn kiểm định

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

nếu H_0 đúng. Vì X và Y có cùng phương sai nên $T \sim \mathcal{T}^{(n_1+n_2-2)}$.

Bước 3. Với $\alpha = 0.01$, tra bảng phân phối *Student* được $t_{1-\alpha}^{(n_1+n_2-2)} = t_{0.99}^{(27)} = 2.473$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -t_{1-\alpha}^{(n_1+n_2-2)}\right) = (-\infty,\ -2.473)$$

Bước 4. Từ số liệu đã cho ta có $n_1 = 11,\ n_2 = 18,\ \overline{x} = 237,\ s_1^2 = 23^2,\ \overline{y} = 247,\ s_2^2 = 27^2$ suy ra giá trị quan sát

$$t_{qs} = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \approx -1.021$$

Bước 5. Vì $t_{qs} = -1.021 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là chưa thể cho rằng lượng hàng bán ra của công ty II là nhiều hơn so với công ty I với mức ý nghĩa 1%.

Bài tập 5.21.

Người ta nghiên cứu trọng lượng của loại trái cây A ở 2 vùng với hai chế độ canh tác khác nhau. Kiểm tra ngẫu nhiên trong lượng 25 trái ở vùng I, 22 trái ở vùng II ở thời điểm thu hoạch thu được kết quả sau (đơn vị tính là kg):

- Vùng I: 2,0; 2,0; 1,8; 1,9; 1,7; 1,5; 1,9; 2,0; 1,8; 1,6; 1,8; 1,7; 1,6; 1,7; 2,1; 1,5; 1,7; 2,0; 1,8; 1,7; 1,5; 1,6; 1,6; 1,7; 1,7.
- Vùng II: 1,5; 1,4; 1,5; 1,6; 1,1; 1,7; 1,4; 1,7; 1,4; 1,4; 1,7; 1,1; 1,5; 1,2; 2,0; 1,6; 1,2; 1,3; 1,5; 1,7; 1,9; 1,0.

Hỏi có sự khác nhau đáng kể giữa các trọng lượng trung bình của loại trái cây A của hai vùng trên không? Kết luận với mức ý nghĩa 5%.

Gọi X, Y lần lượt là trọng lượng của loại trái cây A ở hai vùng I và II. Ta có $X \sim \mathcal{N}(\mu_1, \sigma_1^2),\ Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Đây là bài toán so sánh hai kỳ vọng của hai biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n_1 = 25 < 30,\ n_2 = 22 < 30$ và $\sigma_1^2 = \sigma_2^2$.
Bước 1. Đặt giả thuyết $H_0\colon \mu_1 = \mu_2$, đối thuyết $H_1\colon \mu_1 \neq \mu_2$.
Bước 2. Chọn tiêu chuẩn kiểm định

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

nếu H_0 đúng. Vì X và Y có cùng phương sai nên $T \sim \mathcal{T}^{(n_1+n_2-2)}$.
Bước 3. Với $\alpha = 0.05$, tra bảng phân phối *Student* được $t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} = t_{0.975}^{(45)} = 1.96$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}\right) \cup \left(t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)},\ +\infty\right) = (-\infty,\ -1.96) \cup (1.96,\ +\infty)$$

Bước 4. Từ số liệu đã cho ta có $n_1 = 25,\ n_2 = 22,\ \overline{x} = 1.756,\ s_1^2 = 0.1734^2,\ \overline{y} = 1.4727,\ s_2^2 = 0.2585^2$ suy ra giá trị quan sát

$$t_{qs} = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \approx 4.4598$$

Bước 5. Vì $t_{qs} = 4.4598 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là có cơ sở để cho rằng có sự khác nhau đáng kể giữa các trọng lượng trung bình của loại trái cây A ở hai vùng trên với mức ý nghĩa 5%.

Bài tập 5.22.

Thời gian tự học trong một tuần của 12 sinh viên lớp A và 15 sinh viên lớp B được thống kê lại như sau (đơn vị tính là giờ):

- Lớp A: 18; 15; 24; 23; 30; 12; 15; 24; 35; 30; 18 ;20
- Lớp B: 19; 18; 24; 25; 30; 36; 28; 25; 30; 12; 14; 28; 22; 28; 20.

Với mức ý nghĩa 5%, xét xem thời gian tự học của sinh viên hai lớp thực chất là như nhau không?

Gọi X, Y lần lượt là thời gian tự học trong một tuần của sinh viên lớp A và lớp B. Ta có $X \sim \mathcal{N}(\mu_1, \sigma_1^2),\ Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Đây là bài toán so sánh hai kỳ vọng của hai biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n_1 = 12 < 30,\ n_2 = 15 < 30$ và $\sigma_1^2 = \sigma_2^2$.

Bước 1. Đặt giả thuyết $H_0\colon \mu_1 = \mu_2$, đối thuyết $H_1\colon \mu_1 \neq \mu_2$.

Bước 2. Chọn tiêu chuẩn kiểm định

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

nếu H_0 đúng. Vì X và Y có cùng phương sai nên $T \sim \mathcal{T}^{(n_1+n_2-2)}$.

Bước 3. Với $\alpha = 0.05$, tra bảng phân phối *Student* được $t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} = t_{0.975}^{(25)} = 2.06$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}\right) \cup \left(t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)},\ +\infty\right) = (-\infty,\ -2.06) \cup (2.06,\ +\infty)$$

Bước 4. Từ số liệu đã cho ta có $n_1 = 12,\ n_2 = 15,\ \overline{x} = 22,\ s_1^2 = 7.0065^2,\ \overline{y} = 23.2667,\ s_2^2 = 6.5625^2$ suy ra giá trị quan sát

$$t_{qs} = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \approx -0.4837$$

Bước 5. Vì $t_{qs} = -0.4837 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là có thể xem thời gian tự học của sinh viên hai lớp thực chất là như nhau với mức ý nghĩa 5%.

Bài tập 5.23.

Người ta muốn so sánh 2 chế độ bón phân cho một loại cây trồng, họ đã chia 10 mảnh ruộng sao cho mỗi mảnh thành 2 nửa có điều kiện trồng trọt tương đối như nhau. Nửa thứ nhất áp dụng phương pháp bón phân I, nửa thứ hai theo phương pháp bón phân II (các chế độ chăm sóc khác nhau). Sau khi thu hoạch ta được số liệu về năng suất như sau (đơn vị tính là kg/sào)

Mảnh	1	2	3	4	5	6	7	8	9	10
Năng suất nửa thứ I	24	14	18	20	21	19	16	18	20	23
Năng suất nửa thứ II	16	20	24	23	25	15	22	24	25	29

Giả sử năng suất của hai chế độ phân bón đều tuân theo luật phân phối chuẩn. Đánh giá xem hai chế độ bón phân có giống nhau không với mức ý nghĩa 1%.

Cách 1. Gọi X, Y lần lượt là năng suất của một loại cây trồng áp dụng phương pháp bón phân thứ I và II. Ta có $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Đây là bài toán so sánh hai kỳ vọng của hai biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n_1 = n_2 = 10 < 30$ và $\sigma_1^2 = \sigma_2^2$.
Bước 1. Đặt giả thuyết $H_0\colon \mu_1 = \mu_2$, đối thuyết $H_1\colon \mu_1 \neq \mu_2$.
Bước 2. Chọn tiêu chuẩn kiểm định

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

nếu H_0 đúng. Vì X và Y có cùng phương sai nên $T \sim \mathcal{T}^{(n_1+n_2-2)}$.
Bước 3. Với $\alpha = 0.01$, tra bảng phân phối *Student* được $t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} = t_{0.995}^{(18)} = 2.878$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}\right) \cup \left(t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)},\ +\infty\right) = (-\infty,\ -2.878) \cup (2.878,\ +\infty)$$

Bước 4. Từ số liệu đã cho ta có $n_1 = n_2 = 10$, $\overline{x} = 19.5$, $s_1^2 = 3.3417^2$, $\overline{y} = 22.3$, $s_2^2 = 4.2701^2$ suy ra giá trị quan sát

$$t_{qs} = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \approx -1.6777$$

Bước 5. Vì $t_{qs} = -1.6777 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là có thể đánh giá hai chế độ bón phân giống nhau với mức ý nghĩa 1%.
Cách 2. Đặt $Z = X - Y$, thiết lập hiệu $z_i = x_i - y_i$, $i = \overline{1, 10}$ với

x_i	24	14	18	20	21	19	16	18	20	23
y_i	16	20	24	23	25	15	22	24	25	29
z_i	8	-6	-6	-3	-4	-6	-6	-6	-5	-6

Ta thấy $Z \sim \mathcal{N}(\mu, \sigma^2)$ với σ chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 10 < 30$.
Bước 1. Đặt giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu \neq \mu_0$ với $\mu_0 = 0$.
Bước 2. Chọn tiêu chuẩn kiểm định $T = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $T \sim \mathcal{T}^{(n-1)}$.
Bước 3. Với $\alpha = 0.01$, tra bảng phân phối *Student* được $t^{(n-1)}_{1-\frac{\alpha}{2}} = t^{(9)}_{0.995} = 3.25$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -t^{(n-1)}_{1-\frac{\alpha}{2}}\right) \cup \left(t^{(n-1)}_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -3.25) \cup (3.25,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $n = 10$, $\overline{z} = -4$, $s = 4.3461$ với $\mu_0 = 0$ suy ra giá trị quan sát

$$t_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{-4 - 0}{4.3461}\sqrt{10} \simeq -2.9105$$

Bước 5. Vì $t_{qs} = -2.9104 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là có thể đánh giá hai chế độ bón phân giống nhau với mức ý nghĩa 1%.

Bài tập 5.24.

Quan sát 12 lọ chất hóa học do hai cân khác nhau cân, ta có số liệu (đơn vị tính là gam):

Cân I	0,5	1	2,5	3	4	5	0,7	0,9	1,5	2,3	3,4	4,5
Cân II	1	1,5	2	2	2,5	3	1,8	1,7	2,2	2,4	4,5	3,1

Giả sử cân nặng của lọ hóa chất tuân theo luật phân phối chuẩn. Kiểm định giả thiết hai cân có cân khác nhau hay không với mức ý nghĩa 5%.

Cách 1. Gọi X, Y lần lượt là cân nặng của các lọ hóa chất khi cân bởi cân I và II. Ta có $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Đây là bài toán so sánh hai kỳ vọng của hai biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n_1 = n_2 = 12 < 30$ và $\sigma_1^2 = \sigma_2^2$.
Bước 1. Đặt giả thuyết $H_0\colon \mu_1 = \mu_2$, đối thuyết $H_1\colon \mu_1 \neq \mu_2$.
Bước 2. Chọn tiêu chuẩn kiểm định

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

nếu H_0 đúng. Vì X và Y có cùng phương sai nên $T \sim \mathcal{T}^{(n_1+n_2-2)}$.

Bước 3. Với $\alpha = 0.05$, tra bảng phân phối *Student* được $t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} = t_{0.975}^{(22)} = 2.074$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}\right) \cup \left(t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)},\ +\infty\right) = (-\infty,\ -2.074) \cup (2.074,\ +\infty)$$

Bước 4. Từ số liệu đã cho ta có $n_1 = n_2 = 12,\ \overline{x} = 2.4417,\ s_1^2 = 1.5553^2,\ \overline{y} = 2.3083,\ s_2^2 = 0.912^2$ suy ra giá trị quan sát

$$t_{qs} = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \approx -0.4674$$

Bước 5. Vì $t_{qs} = -0.4674 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là có thể xem hai cân không khác nhau với mức ý nghĩa 5%.

Cách 2. Đặt $Z = X - Y$, thiết lập hiệu $z_i = x_i - y_i,\ i = \overline{1, 12}$ với

x_i	0.5	1	2.5	3	4	5	0.7	0.9	1.5	2.3	3.4	4.5
y_i	1	1.5	2	2	2.5	3	1.8	1.7	2.2	2.4	4.5	3.1
z_i	-0.5	-0.5	0.4	1	1.5	2	-1.1	-0.8	-0.7	-0.1	-1.1	1.4

Ta thấy $Z \sim \mathcal{N}(\mu, \sigma^2)$ với σ chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 12 < 30$.

Bước 1. Đặt giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu \neq \mu_0$ với $\mu_0 = 0$.

Bước 2. Chọn tiêu chuẩn kiểm định $T = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $T \sim \mathcal{T}^{(n-1)}$.

Bước 3. Với $\alpha = 0.05$, tra bảng phân phối *Student* được $t_{1-\frac{\alpha}{2}}^{(n-1)} = t_{0.975}^{(11)} = 2.201$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -t_{1-\frac{\alpha}{2}}^{(n-1)}\right) \cup \left(t_{1-\frac{\alpha}{2}}^{(n-1)},\ +\infty\right) = (-\infty,\ -2.201) \cup (2.201,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $n = 12,\ \overline{z} = 0.1333,\ s = 1.0999$ với $\mu_0 = 0$ suy ra giá trị quan sát

$$t_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{0.1333 - 0}{1.0999}\sqrt{12} \simeq 0.4198$$

Bước 5. Vì $t_{qs} = 0.4198 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là có thể xem hai cân không khác nhau với mức ý nghĩa 5%.

5.2.2 So sánh hai tỷ lệ

Bài tập 5.25.

Một hãng nước giải khát A muốn đưa vào sản xuất một công thức mới để cải tiến sản phẩm của mình. Người ta tiến hành một cuộc khảo sát với công thức cũ cho 600 người uống thử thì thấy có 132 người thích nó và công thức mới cho 400 người uống thử thì thấy có 91 người thích nó. Hãy kiểm định xem liệu với công thức mới có làm tăng tỉ lệ những người ưa thích nước uống của hãng A hay không với mức ý nghĩa 1%.

Gọi p_1, p_2 lần lượt là tỷ lệ những người ưa thích nước uống của hãng A khi sử dụng công thức cũ và công thức mới tương ứng. Đây là bài toán so sánh tỷ lệ.
Bước 1. Đặt giả thuyết $H_0\colon p_1 = p_2$, đối thuyết $H_1\colon p_1 < p_2$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{f_1 - f_2}{\sqrt{\overline{f}\left(1-\overline{f}\right)\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$ nếu giả thuyết H_0 đúng. Ta thấy $U \sim \mathcal{N}(0,1)$.
Bước 3. Với $\alpha = 0.01$, tra bảng giá trị hàm phân phối chuẩn tắc được $u_{1-\alpha} = u_{0.99} = 2.33$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (-\infty,\ -u_{1-\alpha}) = (-\infty,\ -2.33)$$

Bước 4. Theo đầu bài $n_1 = 600,\ n_2 = 400,\ f_1 = \dfrac{11}{50},\ f_2 = \dfrac{91}{400}$. Từ đó tính được $\overline{f} = \dfrac{n_1 f_2 + n_2 f_2}{n_1 + n_2} = \dfrac{132+91}{600+400} = 0.223$ suy ra

$$u_{qs} = \frac{f_1 - f_2}{\sqrt{\overline{f}\left(1-\overline{f}\right)\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}} \approx -0.2791$$

Bước 5. Vì $u_{qs} = -0.2791 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là công thức mới không làm tăng tỷ lệ những người ưa thích nước uống của hãng A với mức ý nghĩa 1%.

Bài tập 5.26.

Từ kho đồ hộp I, lấy ngẫu nhiên 1000 hộp để kiểm tra thấy có 20 hộp bị hỏng. Từ kho II lấy ngẫu nhiên 900 hộp thấy 30 hộp bị hỏng. Hỏi chất lượng bảo quản của 2 kho có thực sự giống nhau hay không với mức ý nghĩa 5%.

Gọi p_1, p_2 lần lượt là tỷ lệ hộp bị hỏng trong kho đồ hộp I và II tương ứng. Đây là bài toán so sánh tỷ lệ.
Bước 1. Đặt giả thuyết $H_0\colon p_1 = p_2$, đối thuyết $H_1\colon p_1 \neq p_2$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{f_1 - f_2}{\sqrt{\overline{f}\left(1-\overline{f}\right)\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$ nếu giả thuyết H_0 đúng. Ta thấy $U \sim \mathcal{N}(0,1)$.

Bước 3. Với $\alpha = 0.05$, tra bảng giá trị hàm phân phối chuẩn tắc được $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -1.96) \cup (1.96,\ +\infty)$$

Bước 4. Theo đầu bài $n_1 = 1000,\ n_2 = 900,\ f_1 = \dfrac{1}{50},\ f_2 = \dfrac{1}{30}$. Từ đó tính được $\overline{f} = \dfrac{n_1 f_2 + n_2 f_2}{n_1 + n_2} = \dfrac{20 + 30}{1000 + 900} = \dfrac{1}{38}$ suy ra

$$u_{qs} = \frac{f_1 - f_2}{\sqrt{\overline{f}\left(1 - \overline{f}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \approx -1.8129$$

Bước 5. Vì $u_{qs} = -1.8129 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là có thể xem chất lượng bảo quản của 2 kho là thực sự giống nhau với mức ý nghĩa 5%.

Bài tập 5.27.

Bệnh A được điều trị theo hai phương pháp. Sau một thời gian thấy kết quả như sau:

- Trong 102 bệnh nhân điều trị phương pháp I có 82 bệnh nhân khỏi bệnh.
- Trong 98 bệnh nhân điều trị phương pháp II có 69 bệnh nhân khỏi bệnh.

Hỏi có phải phương pháp I điều trị tốt hơn phương pháp II hai hay không với mức ý nghĩa 5%.

Gọi p_1, p_2 lần lượt là tỷ lệ bệnh nhân khỏi bệnh A được điều trị theo phương pháp I và II tương ứng. Đây là bài toán so sánh tỷ lệ.
Bước 1. Đặt giả thuyết $H_0 \colon p_1 = p_2$, đối thuyết $H_1 \colon p_1 > p_2$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{f_1 - f_2}{\sqrt{\overline{f}\left(1 - \overline{f}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ nếu giả thuyết H_0 đúng. Ta thấy $U \sim \mathcal{N}(0, 1)$.
Bước 3. Với $\alpha = 0.05$, tra bảng giá trị hàm phân phối chuẩn tắc được $u_{1-\alpha} = u_{0.95} = 1.65$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (u_{1-\alpha},\ +\infty) = (1.65,\ +\infty)$$

Bước 4. Theo đầu bài $n_1 = 102,\ n_2 = 98,\ f_1 = \dfrac{44}{51},\ f_2 = \dfrac{69}{98}$. Từ đó tính được $\overline{f} = \dfrac{n_1 f_2 + n_2 f_2}{n_1 + n_2} = \dfrac{82 + 69}{102 + 98} = 0.755$ suy ra

$$u_{qs} = \frac{f_1 - f_2}{\sqrt{\overline{f}\left(1 - \overline{f}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \approx 2.6081$$

Bước 5. Vì $u_{qs} = 2.6081 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là có thể nói phương pháp I điều trị tốt hơn phương pháp II với mức ý nghĩa 5%.

Bài tập 5.28.

Để đánh giá hiệu quả của hai dây chuyền sản xuất người ta tiến hành kiểm tra 1000 sản phẩm do dây chuyền I sản xuất có 10 sản phẩm hỏng, kiểm tra 1000 sản phẩm do dây chuyền II sản xuất thấy có 8 sản phẩm hỏng. Với mức ý nghĩa 5%, có kết luận gì về tỷ lệ sản phẩm hỏng từ hai dây chuyền trên.

Gọi p_1, p_2 lần lượt là tỷ lệ sản phẩm hỏng được sản xuất bởi dây chuyền I và II tương ứng.
Bước 1. Đặt giả thuyết $H_0\colon p_1 = p_2$, đối thuyết $H_1\colon p_1 \neq p_2$.
Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{f_1 - f_2}{\sqrt{\overline{f}\left(1-\overline{f}\right)\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$ nếu giả thuyết H_0 đúng. Ta thấy $U \sim \mathcal{N}(0,1)$.
Bước 3. Với $\alpha = 0.05$, tra bảng giá trị hàm phân phối chuẩn tắc được $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -1.96) \cup (1.96,\ +\infty)$$

Bước 4. Theo đầu bài $n_1 = n_2 = 1000$, $f_1 = \dfrac{1}{100}$, $f_2 = \dfrac{1}{125}$. Từ đó tính được $\overline{f} = \dfrac{n_1 f_2 + n_2 f_2}{n_1 + n_2} = \dfrac{10+8}{1000+1000} = 0.009$ suy ra

$$u_{qs} = \frac{f_1 - f_2}{\sqrt{\overline{f}\left(1-\overline{f}\right)\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}} \approx 0.4735$$

Bước 5. Vì $u_{qs} = 0.4735 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là có thể cho rằng tỷ lệ sản phẩm hỏng của dây chuyền I giống dây chuyền II với mức ý nghĩa 5%.

Bài tập 5.29.

Nghiên cứu về năng suất của loại hoa màu A, người ta kiểm tra năng suất của 64 điểm trồng loại hoa màu này thu được bảng số liệu

Năng suất (tạ/ha)	40 − 45	45 − 50	50 − 55	55 − 60	60 − 65	65 − 70
Số điểm	2	5	15	30	8	4

1. Giả sử theo tính toán lý thuyết, năng suất trung bình của loại hoa màu A là 55 tạ/ha. Theo anh chị năng suất trung bình loại hoa màu A có xu hướng tăng không? Hãy kết luận với mức ý nghĩa 1%?

2. Một tài liệu thống kê cho biết tỷ lệ những điểm có năng suất trên 60 tạ/ha của loại hoa màu A là 15%. Hãy cho kết luận về tài liệu nói trên với mức ý nghĩa 5%.

1. Gọi X là năng suất của loại hoa màu A. Ta thấy $E[X] = \mu$ là năng suất trung bình của loại hoa màu A chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 64 > 30$.

Bước 1. Kiểm tra giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu > \mu_0$ với $\mu_0 = 55$.

Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $U \sim \mathcal{N}(0, 1)$.

Bước 3. Với $\alpha = 0.01$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\alpha} = u_{0.99} = 2.33$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (u_{1-\alpha},\ +\infty) = (2.33,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $\mu_0 = 55, n = 64,\ \overline{x} = 56.3281,\ s = 5.4$ suy ra giá trị quan sát

$$u_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{56.3281 - 55}{5.4}\sqrt{64} \simeq 1.9676$$

Bước 5. Vì $u_{qs} = 1.9676 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là năng suất trung bình của loại hoa màu A không có xu hướng tăng lên với mức ý nghĩa 1%.

2. Gọi p là tỷ lệ những điểm có năng suất trên 60 tạ/ha của loại hoa màu A. Đây là bài toán kiểm định giả thuyết về tỷ lệ của tổng thể.

Bước 1. Đặt giả thuyết $H_0\colon p = p_0$, đối thuyết $H_1\colon p \neq p_0$ với $p_0 = 0.15$.

Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$ nếu giả thuyết H_0 đúng.

Vì $np_0 = 9.6 > 5$ và $n(1 - p_0) = 54.4 > 5$ khá lớn nên $U \sim \mathcal{N}(0, 1)$.

Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -1.96) \cup (1.96,\ +\infty)$$

Bước 4. Từ số liệu đã cho ta có $n = 64,\ m = 12$ tính được $f = \dfrac{m}{n} = \dfrac{3}{16} = 0.1875$, với $p_0 = 0.15$ suy ra giá trị quan sát

$$u_{qs} = \frac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0.1875 - 0.15}{\sqrt{0.15 \times 0.85}}\sqrt{64} \approx 0.8402$$

Bước 5. Vì $u_{qs} = 0.8402 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là có thể tin vào kết luận của tài liệu thống kê trên với mức ý nghĩa 5%.

Bài tập 5.30.

Điều tra doanh thu của các hộ gia đình kinh doanh loại mặt hàng A tại địa phương B, người ta điều tra 100 hộ kinh doanh loại mặt hàng này trong một tháng năm 2019 thu được bảng số liệu

Doanh thu (triệu đồng)	20	24	28	32	36	40	44	48	52
Số hộ gia đình	5	10	17	25	20	10	8	3	2

1. Với độ tin cậy 95% hãy ước lượng doanh thu trung bình của các hộ gia đình kinh doanh loại mặt hàng nói trên. Để độ chính xác của ước lượng nhỏ hơn 2 triệu đồng thì cần điều tra ít nhất bao nhiêu hộ?
2. Theo số liệu điều tra năm 2018 thì tỷ lệ những hộ gia đình đạt doanh thu dưới 28 triệu đồng là 20%. Theo anh chị tỷ lệ này năm 2019 có giảm đi hay không? Hãy kết luận với mức ý nghĩa 1%.
3. Hãy ước lượng tỷ lệ những hộ có doanh thu trên 40 triệu đồng với độ tin cậy 99%? Nếu yêu cầu độ tín cậy 95%, độ chính xác của ước lượng là 0,02 thì cần điều tra ngẫu nhiên bao nhiêu hộ gia đình?
4. Một tài liệu báo cáo cho biết doanh thu trung bình của các hộ kinh doanh loại mặt hàng A tại địa phương B là 30 triệu đồng trên tháng. Tài liệu báo cáo này có làm giảm doanh thu trung bình của các hộ gia đình kinh doanh mặt hàng A để giảm thuế hay không? Hãy kết luận với mức ý nghĩa 5%.
5. Theo điều tra cách đây 2 năm thì doanh thu trung bình của các hộ gia đình kinh doanh loại mặt hàng này là 30 triệu đồng/tháng, hãy đánh giá xem doanh thu trung binh sau 2 năm có thay đổi không với mức ý nghĩa 5%.
6. Điều tra doanh thu của 200 hộ gia đình kinh doanh loại mặt hàng A ở địa phương C năm 2019 người ta tính được doanh thu trung bình/tháng là 37 triệu đồng và độ lệch chuẩn mẫu là 1,1 triệu đồng. Doanh thu trung bình loại mặt hàng A ở địa phương C và B có như nhau hay không? Hãy kết luận với độ tín cậy 95%.

1. Gọi X là doanh thu của các hộ gia đình kinh doanh loại mặt hàng A, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ với phương sai σ^2 chưa biết. Doanh thu trung bình của các hộ gia đình là $E[X] = \mu$ chưa biết cần được ước lượng.

   Bước 1. Chọn thống kê $U = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$. Vì $n = 100 > 30$ nên thống kê $U \sim \mathcal{N}(0,1)$.

   Bước 2. Khoảng tin cậy đối xứng cho $E[X] = \mu$ là

   $$\left(\overline{x} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\ \overline{x} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

   trong đó $\alpha = 0.05$, $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$, được tra từ bảng giá trị phân phối chuẩn tắc.

Bước 3. Từ số liệu đã cho tính được $n = 100,\ \overline{x} = 33.36,\ s = 7.1964$. Suy ra khoảng tin cậy đối xứng của $E[X] = \mu$ là

$$\left(33.36 - 1.96 \times \frac{7.1964}{\sqrt{100}},\ 33.36 + 1.96 \times \frac{7.1964}{\sqrt{100}}\right) = (31.9495,\ 34.7705)$$

Bước 4. Kết luận, với độ tin cậy 95% doanh thu trung bình của các hộ gia đình kinh doanh loại mặt hàng A từ 31.9495 triệu đồng đến 34.7705 triệu đồng.

Sai số của ước lượng là $\varepsilon = u_{1-\frac{\alpha}{2}} \dfrac{s}{\sqrt{n}}$. Để độ chính xác của ước lượng nhỏ hơn 2 triệu đồng, hay $\varepsilon < 2$ thì cỡ mẫu phải lấy là

$$n > \frac{u^2_{1-\frac{\alpha}{2}} s^2}{\varepsilon^2} = \frac{1.96^2 \times 7.1964^2}{2^2} \approx 49.7374$$

Vậy cần chọn mẫu nhỏ nhất có cỡ $n = 50$.

2. Gọi p là tỷ lệ những hộ gia đình đạt doanh thu dưới 28 triệu đồng. Đây là bài toán kiểm định giả thuyết về tỷ lệ của tổng thể.

Bước 1. Đặt giả thuyết $H_0\colon p = p_0$, đối thuyết $H_1\colon p < p_0$ với $p_0 = 0.2$.

Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$ nếu giả thuyết H_0 đúng.

Vì $np_0 = 20 > 5$ và $n(1-p_0) = 80 > 5$ khá lớn nên $U \sim \mathcal{N}(0,1)$.

Bước 3. Với $\alpha = 0.01$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\alpha} = u_{0.99} = 2.33$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (-\infty,\ u_{1-\alpha}) = (-\infty,\ -2.33)$$

Bước 4. Từ số liệu đã cho ta có $n = 100,\ m = 15$ tính được $f = \dfrac{m}{n} = \dfrac{3}{20} = 0.15$, với $p_0 = 0.2$ suy ra giá trị quan sát

$$u_{qs} = \frac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0.15 - 0.2}{\sqrt{0.2 \times 0.8}}\sqrt{100} \approx -1.25$$

Bước 5. Vì $u_{qs} = -1.25 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Tức là chưa thể kết luận tỷ lệ những hộ gia đình đạt doanh thu dưới 28 triệu đồng năm 2019 giảm đi với mức ý nghĩa 1%.

3. Gọi p là tỷ lệ những hộ có doanh thu trên 40 triệu đồng. Kiểm tra $nf = 100 \times \dfrac{13}{100} = 13 > 5$ và $n(1-f) = 100 \times \dfrac{87}{100} = 87 > 5$.

Bước 1. Chọn thống kê $Z = \dfrac{f - p}{\sqrt{f(1-f)}}\sqrt{n}$. Thống kê $Z \sim \mathcal{N}(0,1)$.

Bước 2. Khoảng tin cậy đối xứng của sác xuất p là

$$\left(f - u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}},\ f + u_{1-\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}\right)$$

trong đó $u_{1-\frac{\alpha}{2}} = u_{0.995} = 2.58$ được tra từ bảng giá trị hàm phân phối chuẩn tắc.

Bước 3. Với $n = 100,\ m = 13,\ f = \dfrac{m}{n} = 0.13$, suy ra khoảng tin cậy đối xứng của p là

$$\left(0.13 - 2.58\sqrt{\frac{0.13 \times 0.87}{100}},\ 0.13 + 2.58\sqrt{\frac{0.13 \times 0.87}{100}}\right) = (0.0432,\ 0.2168)$$

Bước 4. Kết luận, tỷ lệ những hộ có doanh thu trên 40 triệu đồng là từ 4.32% đến 21.68% với độ tin cậy 99%.

Độ chính xác của ước lượng là $\varepsilon = u_{1-\frac{\alpha}{2}}\sqrt{\dfrac{f(1-f)}{n}}$. Với độ tin cậy $\gamma = 1 - \alpha = 0.95$ và độ chính xác $\varepsilon_0 = 0.02$ cho trước thì kích thước mẫu cần thiết là số tự nhiên n nhỏ nhất thỏa mãn

$$n \geq \frac{u^2_{1-\frac{\alpha}{2}}\ f(1-f)}{\varepsilon_0^2} = \frac{1.96^2 \times 0.13 \times 0.87}{0.02^2} \approx 1086.2124$$

Như vậy mẫu cần tìm có cỡ $n = 1087$.

4. Gọi X là doanh thu của các hộ gia đình kinh doanh loại mặt hàng A. Ta thấy $E[X] = \mu$ là doanh thu trung bình của các hộ gia đình chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 100 > 30$.

Bước 1. Kiểm tra giả thuyết $H_0\colon \mu = \mu_0$, đối thuyết $H_1\colon \mu > \mu_0$ với $\mu_0 = 30$.

Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $U \sim \mathcal{N}(0,1)$.

Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\alpha} = u_{0.95} = 1.65$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (u_{1-\alpha},\ +\infty) = (1.65,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $\mu_0 = 30, n = 100,\ \overline{x} = 33.36,\ s = 7.1964$ suy ra giá trị quan sát

$$u_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{33.36 - 30}{7.1964}\sqrt{100} \simeq 4.669$$

Bước 5. Vì $u_{qs} = 4.669 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là tài liệu báo cáo này làm giảm mức doanh thu trung bình của các hộ gia đình kinh doanh loại mặt hàng A để giảm thuế với mức ý nghĩa 5%.

5. Gọi X là doanh thu của các hộ gia đình kinh doanh loại mặt hàng A. Ta thấy $E[X] = \mu$ là doanh thu trung bình của các hộ gia đình chưa biết. Đây là bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n = 100 > 30$.

Bước 1. Kiểm tra giả thuyết $H_0: \mu = \mu_0$, đối thuyết $H_1: \mu \neq \mu_0$ với $\mu_0 = 30$.

Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ nếu giả thuyết H_0 đúng. $U \sim \mathcal{N}(0, 1)$.

Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -1.96) \cup (1.96,\ +\infty)$$

Bước 4. Từ số liệu của đầu bài ta có $\mu_0 = 30, n = 100,\ \overline{x} = 33.36,\ s = 7.1964$ suy ra giá trị quan sát

$$u_{qs} = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{33.36 - 30}{7.1964}\sqrt{100} \simeq 4.669$$

Bước 5. Vì $u_{qs} = 4.669 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là doanh thu trung bình sau 2 năm có thay đổi với mức ý nghĩa 5%.

6. Gọi X, Y lần lượt là doanh thu loại mặt hàng A năm 2019 ở địa phương B và C. Ta có $X \sim \mathcal{N}(\mu_1, \sigma_1^2),\ Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Đây là bài toán so sánh hai kỳ vọng của hai biến ngẫu nhiên có phân phối chuẩn trường hợp chưa biết phương sai, cỡ mẫu $n_1 = 100 > 30,\ n_2 = 200 > 30$.

Bước 1. Đặt giả thuyết $H_0: \mu_1 = \mu_2$, đối thuyết $H_1: \mu_1 \neq \mu_2$.

Bước 2. Chọn tiêu chuẩn kiểm định $U = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ nếu H_0 đúng. $U \sim \mathcal{N}(0, 1)$.

Bước 3. Với $\alpha = 0.05$, tra bảng giá trị phân phối chuẩn tắc được $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = \left(-\infty,\ -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}},\ +\infty\right) = (-\infty,\ -1.96) \cup (1.96,\ +\infty)$$

Bước 4. Từ số liệu đã cho ta có $n_1 = 100,\ n_2 = 200,\ \overline{x} = 33.36,\ s_1^2 = 7.1964^2,\ \overline{y} = 37,\ s_2^2 = \dfrac{200}{199} \times 1.1^2 = 1.2161$ suy ra giá trị quan sát

$$u_{qs} = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{33.36 - 37}{\sqrt{\dfrac{7.1964^2}{100} + \dfrac{1.2161}{200}}} \approx -5.0286$$

Bước 5. Vì $u_{qs} = -5.0286 \in W_\alpha$ nên có cơ sở để bác bỏ giả thuyết H_0. Tức là doanh thu trung bình mặt hàng A ở địa phương B và C không như nhau với mức ý nghĩa 5%.

Tài liệu tham khảo

1. Bài giảng Xác suất thống kê, PGS. TS Nguyễn Thị Thu Thủy, Viện Toán ứng dụng và Tin học, Đại học Bách khoa Hà Nội, 2020
2. Đề cương bài tập Xác suất thống kê, Viện Toán ứng dụng và Tin học, Đại học Bách khoa Hà Nội, 2020
3. Introduction to Probability, Dimitri P. Bertsekas & John N. Tsitsiklis, MIT Publisher, 2008
4. An Introduction to Mathematical Statistics and Its Applications, Richard J. Larsen & Morris L. Marx, Pearson, 2018