CHƯƠNG 1. BIẾN CỐ VÀ XÁC SUẤT CỦA BIẾN CỐ

1.1. Phép thử và biến cố

Phép thử và biến cố

Phép thử là việc thực hiện một nhóm các điều kiện xác định nào đó (chẳng hạn làm thí nghiệm) để quan sát một hiện tượng có xảy ra hay không. Hiện tượng có xảy ra hay không trong kết cục của phép thử gọi là *biến cố*.

Ví dụ 1: Tung một con xúc xắc sáu mặt xuống đất là một phép thử. Một vài biến cố của phép thử này:

Mặt k chấm xuất hiện (k=1,2,3,4,5,6)

Mặt có số chấm chẵn xuất hiện

Mặt có số chấm lớn hơn 6 xuất hiện.

Mặt có có số chấm nhỏ hơn hoặc bằng 6.

Ví dụ 2: Bắn một phát súng vào bia. Việc bắn súng là phép thử. Còn việc trúng vào một miền nào đó của bia là biến cố.

Ví dụ 3: Đo nhiệt độ ngoài trời. "Nhiệt độ ngoài trời là t độ C" là biến cố

Ví dụ 4: Cho một hộp có 5 bi đỏ, 3 bi vàng. Lấy ngẫu nhiên 1 bi. Biến cố: "Lấy được bi vàng"…

Một phép thử được gọi là *ngẫu nhiên* nếu như ta không biết trước hay dự đoán chính xác kết quả của nó.

Các loại biến cố

- Biến cố chắc chắn

Biến cố chắc chắn là biến cố nhất định sẽ xảy ra khi thực hiện một phép thử. Biến cố chắc chắn được kí hiệu là Ω.

Ví dụ: Khi thực hiện phép thử tung một con xúc sắc. Biến cố " Xuất hiện mặt có số chấm nhỏ hơn hoặc bằng 6" là biến cố chắc chắn.

- Biến cố không thể có

Biến cố không thể có là biến cố nhất định không xảy ra khi thực hiện một phép thử. Biến cố không thể có kí hiệu là $\varnothing$.

Ví dụ: Tung một con xúc sắc. Biến cố " Xuất hiện mặt 7 chấm " là biến cố không thể có.

- Biến cố ngẫu nhiên (BCNN)

Biến cố ngẫu nhiên là biến cố có thể xảy ra hoặc không xảy ra khi thực hiện một phép thử. Các biến cố ngẫu nhiên được kí hiệu là A, B, C….

Ví dụ 1: Tung một con xúc sắc. Biến cố " Xuất hiện mặt 1 chấm " là biến cố ngẫu nhiên

Ví dụ 2: Bóc ngẫu nhiên một tờ lịch trong năm. Phân loại các biến cố sau đây

Biến cố A: "bóc được tờ lịch ghi tháng 2".

Biến cố B: "bóc được tờ lịch ghi tháng 13".

Biến cố C: "bóc được tờ lịch ghi một trong các tháng 1, 2, 3, 4,…,12".

1.2. Mối quan hệ giữa các biến cố

- **Quan hệ kéo theo**

Biến cố *A* gọi là kéo theo biến cố *B*, kí hiệu $A \subset B$, nếu *A* xảy ra thì kéo theo biến cố *B* xảy ra khi thực hiện phép thử.

Ví dụ: Một sinh viên mua một tờ vé số. *A* là biến cố sinh viên trúng độc đắc. *B* là biến cố sinh viên trúng số.

Vậy $A \subset B$ *hay* $B \subset A$?

- **Quan hệ tương đương**

Hai biến cố *A* và *B* được gọi là tương đương, kí hiệu $A = B$ nếu *A* xảy ra thì *B* xảy ra và ngược lại.

$$A = B \Leftrightarrow A \subset B \text{ và } B \subset A.$$

Ví dụ: Tung một con xúc sắc. *A* là biến cố " Xuất hiện mặt chẵn chấm " . *B* là biến cố xuất hiện mặt 2 chấm hoặc 4 chấm hoặc 6 chấm

- **Biến cố tổng**

Biến cố tổng (còn gọi là tổng) của hai biến cố *A* và *B*, ký hiệu $A + B$ hay $A \cup B$ là biến cố xảy ra khi và chỉ khi ít nhất một trong hai biến cố *A*, *B* xảy ra.

Ví dụ 1: Một xạ thủ bắn một viên đạn. *A* là biến cố xạ thủ bắn trúng vòng 10 điểm, *B* là biến cố xạ thủ bắn trúng vòng 9 điểm. *C* là biến cố xạ thủ bắn được ít nhất 9 điểm.

Ta có $C = A + B$

Ví dụ 2: Một sinh viên thi hai môn Toán cao cấp và Kinh tế lượng. Gọi T là biến cố sinh viên đó đậu môn Toán cao cấp, K là biến cố sinh viên đó đậu môn Kinh tế lượng. Khi đó T+K là biến cố Sinh viên đó đậu ít nhất một môn.

Ví dụ 3: Hai xạ thủ bắn vào 1 bia. Mỗi người bắn 1 viên. A là biến cố xạ thủ 1 bắn trúng bia. B là biến cố xạ thủ 2 bắn trúng bia. A+B là biến cố: Bia trúng đạn

Ví dụ 4: Gieo một con xúc xắc. Gọi A_i là biến cố mặt i xuất hiện. Hãy gọi tên biến cố $A_1+A_3+A_5$

Tổng quát: Tổng của n biến cố $A_1, A_2, ..., A_n$ ký hiệu $\sum_{i=1}^{n} A_i = A_1 + A_2 + ... + A_n$ hay $\bigcup_{i=1}^{n} A_i$, là biến cố xảy ra nếu ít nhất có 1 trong các biến cố A_k xảy ra trong phép thử.

Ví dụ: Kiểm tra chất lượng n sản phẩm. A_k là biến cố sản phẩm thứ k xấu. C là biến cố có ít nhất 1 sp xấu. Ta có: $C = A_1 + A_2 + ... + A_n$

- **Biến cố tích**

Biến cố tích (hay còn gọi là tích) của hai biến cố *A* và *B*, kí hiệu $A.B$ hay $A \cap B$ là biến cố xảy ra khi và chỉ khi *A* xảy ra và *B* xảy ra.

Ví dụ: Một sinh viên thi hai môn Toán cao cấp và Kinh tế lượng. Gọi T là biến cố sinh viên đó đậu môn Toán cao cấp, K là biến cố sinh viên đó đậu môn Kinh tế lượng. Khi đó TK là biến cố sinh viên đậu cả hai môn.

Tổng quát: Tích của n biến cố $A_1, A_2, ..., A_n$, ký hiệu $\prod_{i=1}^{n} A_i = A_1.A_2...A_n$ hay $\bigcap_{i=1}^{n} A_i$ là biến cố xảy ra khi và chỉ khi A_1 xảy ra và A_2 xảy ra …và A_n xảy ra trong phép thử.

Ví dụ: Kiểm tra chất lượng n sản phẩm. A_k là bc sản phẩm thứ k xấu. C là bc có tất cả sp xấu. Ta có: $C = A_1.A_2...A_n$

- **Biến cố xung khắc**

Hai biến cố *A* và *B* gọi là hai biến cố *xung khắc* nếu biến cố này xảy ra thì biến cố kia không xảy ra, tức là $A.B = \varnothing$.

Ví dụ 1: Tung một con xúc sắc. Gọi A là biến cố xuất hiện mặt 4 chấm. B là biến cố xuất hiện mặt 1 chấm. C là biến cố xuất hiện mặt chẵn chấm. A và B là biến cố xung khắc với nhau. A và C không xung khắc với nhau.

Ví dụ 2: Trong hộp có 4 bi vàng và 2 bi đỏ. Lấy ngẫu nhiên 1 bi. Gọi A là biến cố lấy được bi vàng. B là biến cố lấy được bi đỏ. A và B xung khắc.

Ví dụ 3: Một sinh viên thi môn XSTK. Gọi A là biến cố sinh viên được 7 điểm. B là biến cố sinh viên được 8 điểm. C là biến cố sinh viên được ít nhất 8 điểm.

- **Biến cố xung khắc từng đôi**

Nhóm các biến cố $A_1, A_2, ..., A_n$ được gọi là xung khắc từng đôi nếu 2 biến cố bất kì trong nhóm xung khắc nhau. Nghĩa là $A_i.A_j = \varnothing (i \neq j)$

- **Biến cố đối lập**

$\overline{A}$ được gọi là biến cố đối lập của A khi và chỉ khi $\overline{A}$ xảy ra thì A không xảy ra và ngược lại. Tức là

$$\begin{cases} A.\overline{A} = \varnothing \\ A + \overline{A} = \Omega \end{cases}$$

Ví dụ 1: Một sinh viên thi môn XSTK. A là biến cố SV đó được 7 điểm môn XSTK. B là biến cố SV đó được ít nhất 9 điểm môn XSTK.

Ví dụ 2: Một lô hàng có 5 sản phẩm trong đó có 3 phế phẩm. Lấy ngẫu nhiên 2 sản phẩm. Gọi A là biến cố có sản phẩm xấu (hay có ít nhất 1 sản phẩm xấu). B là biến cố có 1 sản phẩm xấu. C là biến có tất cả sản phẩm đều xấu $\overline{A}$ biến cố không có sản phẩm xấu.. $\overline{B}$ …

Ví dụ 3: Một sinh viên thi hai môn Toán cao cấp và Kinh tế lượng. Gọi T là biến cố sinh viên đó đậu môn Toán cao cấp, K là biến cố sinh viên đó đậu môn Kinh tế lượng. Khi đó TK là biến cố sinh viên đậu cả hai môn. $C = \overline{T}.K + T.\overline{K}$

- **Biến cố hiệu**

Biến cố hiệu của hai biến cố A và B, kí hiệu $A \setminus B$ là biến cố xảy ra khi và chỉ khi biến cố A xảy ra và biến cố B không xảy ra. Nghĩa là $A \setminus B = A.\overline{B}$

Một số tính chất

$1.\overline{\overline{A}} = A$	$9.A.B = B.A$
$2.A + \Omega = \Omega$	$10.A + (B.C) = (A + B).(A + C)$
$3.A.\Omega = A$	$11.A.(B + C) = A.B + A.C$
$4.A + \varnothing = A$	$12.\overline{A + B} = \overline{A}.\overline{B}$
$5.A.\varnothing = \varnothing$	$13.\overline{A.B} = \overline{A} + \overline{B}$
$6.A + A = A$	$14.\overline{A + B + C} = \overline{A}.\overline{B}.\overline{C}$
$7.A.A = A$	$15.\overline{A.B.C} = \overline{A} + \overline{B} + \overline{C}$
$8.A + B = B + A$	

Biến cố sơ cấp

Biến cố sơ cấp là kết quả đơn giản nhất của phép thử, là biến cố không thể biểu diễn thành tổng của các biến cố khác.

Tập hợp tất các biến cố sơ cấp trong một phép thử được gọi là *không gian các biến cố sơ cấp*. Kí hiệu là Ω.

Ví dụ. Gieo một con xúc sắc xuống đất. Gọi A_i là biến cố "xuất hiện măt i chấm" $i=\overline{1,6}$. C là biến cố xuất hiện mặt chẵn chấm, thì $A_1, A_2, A_3, A_4, A_5, A_6$ là các biến cố sơ cấp và $\Omega = \{A_1, A_2, A_3, A_4, A_5, A_6\}$. C không là biến cố sơ cấp vì $C = A_2 + A_4 + A_6$

Mỗi biến cố ngẫu nhiên *A* đều biểu diễn được thành tổng của các biến cố sơ cấp nào đó. Các biến cố sơ cấp trong tổng này được gọi là các *biến cố thuận lợi* cho *A*.

Ví dụ. Tung một con xúc xắc sáu mặt, cân đối, đồng chất trên mặt phẳng nằm ngang. Gọi *A* là biến cố mặt chẵn chấm xuất hiện. A_i là biến cố mặt i chấm xuất hiện. Khi đó

$$A = A_2 + A_4 + A_6$$

Các biến cố đồng khả năng là các biến cố có cùng khả năng xuất hiện như nhau trong một phép thử.

Ví dụ 1

Tung một con xúc xắc sáu mặt, cân đối, đồng chất trên mặt phẳng nằm ngang. Các biến cố đồng khả năng là biến cố được mặt một chấm, hai chấm,….

Nhóm đầy đủ các biến cố

Nhóm các biến cố $A_1, A_2, \ldots, A_n$ $(n \geq 2)$ của một phép thử được gọi là một nhóm đầy đủ (hay hệ đầy đủ) nếu trong kết quả của phép thử sẽ xảy ra một và chỉ một trong các biến cố đó. Tức là

$$\begin{cases} A_i.A_j = \varnothing, 1 \leq i \neq j \leq n \\ A_1 + A_2 + \cdots + A_n = \Omega \end{cases}$$

Ví dụ : Gieo một con xúc sắc xuống đất. Gọi A_i là biến cố "xuất hiện măt *i* chấm", $i=\overline{1,6}$ thì $\{A_1, A_2, A_3, A_4, A_5, A_6\}$ là một nhóm đầy đủ các biến cố.

Chú ý: *A* và $\overline{A}$ tạo nên một nhóm đầy đủ các biến cố.

1.3. Xác suất biến cố: định nghĩa và tính chất.

Xác suất của một biến cố là một số đặc trưng cho khả năng xuất hiện biến cố đó trong phép thử.

1.3. 1. Định nghĩa cổ điển về xác suất

Định nghĩa

Giả sử trong một phép thử có tất cả n biến cố sơ cấp đồng khả năng, trong đó có m biến cố thuận lợi cho biến cố *A*. Khi đó xác suất của *A*, kí hiệu *P*(*A*) được xác định như sau $P(A) = \dfrac{m}{n}$

Ví dụ. Tung một con xúc xắc sáu mặt, cân đối, đồng chất trên mặt phẳng nằm ngang. Tính xác suất để

a) Mặt 6 chấm xuất hiện ;

b) Mặt có số chấm chẵn xuất hiện.

Giải

Vì con xúc xắc có sáu mặt (cân đối, đồng chất) với số chấm từ 1 đến 6 nên sau khi gieo có đúng 6 trường hợp đồng khả năng. Gọi:

A là biến cố mặt 6 chấm xuất hiện

B là biến cố mặt có số chấm chẵn xuất hiện .

Số trường hợp thuận lợi cho A là 1

Số trường hợp thuận lợi cho B là 3.

Do đó

$a)P(A)=\frac{1}{6}$ $b)P(B)=\frac{3}{6}=\frac{1}{2}$

Ví dụ 2

Một hộp có 7 viên bi đỏ và 3 viên bi trắng (các bi này to như nhau, nặng như nhau, nhẵn nhụi như nhau). Lấy ngẫu nhiên hai bi từ hộp. Tính xác suất lấy được hai bi trắng.

Phép thử : lấy ngẫu nhiên 2 bi.

Số biến cố sơ cấp đồng khả năng là : C_{10}^2

Gọi A là biến cố “lấy đc 2 bi trắng”. Số biến cố thuận lợi cho A là C_3^2

Vậy $P(A)=\frac{C_3^2}{C_{10}^2}$

Ví dụ 3: Một hộp có 10 sản phẩm trong đó có 4 phế phẩm. Lấy ngẫu nhiên 3 sản phẩm. Tính xác suất để lấy được 2 phế phẩm.

Giải

Phép thử: “lấy ngẫu nhiên 3 sp”

Số trường hợp đồng khả năng là C_{10}^3.

Gọi A là biến cố lấy được 2 phế phẩm.

Số trường hợp thuận lợi cho A là. $C_4^2.C_6^1$

Do đó $P(A)=\dfrac{C_4^2.C_6^1}{C_{10}^3}$

Ví dụ 4. Một công ty tuyển 3 nhân viên cho 3 vị trí giám đốc, trợ lý giám đốc và trưởng phòng kinh doanh. Biết có 50 người dự tuyển trong đó có 20 nữ. Tính xác suất để trong 3 người được tuyển có trợ lý giám đốc là nữ.

ĐS. $P(A)=\dfrac{20.A_{49}^2}{A_{50}^3}$

Ý nghĩa xác suất: Xác suất của biến cố *A* là một con số đặc trưng cho khả năng xuất hiện biến cố *A* trong một phép thử. $P(A)$ càng lớn (càng gần 1) thì khả năng xuất hiện *A* càng nhiều, $P(A)$ càng nhỏ (càng gần 0) thì khả năng xuất hiện *A* càng ít.

Các tính chất của xác suất

$P(\Omega)=1$

$P(\varnothing)=0$

$0\le P(A)\le 1$ với *A* là biến cố bất kì.

Nếu $A\subset B$ thì $P(A)\le P(B)$

Nhận xét: Dùng định nghĩa cổ điển về xác suất thì

- Tính được xác suất mà không cần thực hiện phép thử.
- Nó chỉ xét cho phép thử có số kết quả hữu hạn và mọi kết quả là đồng khả năng.

1.3.2. Định nghĩa thống kê về xác suất

Một điều dễ dàng nhận ra rằng định nghĩa cổ điển chỉ áp dụng tính cho các phép thử có số kết cục hữu hạn và mọi kết cục là đồng khả năng . Thế còn với các phép thử khác thì sao?

Ví dụ 1. Tung một đồng xu không cân đối và không đồng chất. Ta không thể nào nói rằng xác suất được mặt sấp và mặt ngửa là như nhau và bằng 0,5. Vậy xác suất để đồng xu ngửa là bao nhiêu?

Ví dụ 2. Một số trường hợp ta không thể hoặc khó có thể xác định được kích thước không gian mẫu cũng như kích thước biến cố.

- Điều tra chiều cao thanh niên Việt Nam ở độ tuổi 18 đến 20. Tìm xác suất để chiều cao lớn hơn 1,60m.
- Tìm xác suất của sản phẩm hư hỏng trong một lô hàng nước giải khát. Ta biết kích thước không gian mẫu nhưng lại không xác định được số sản phẩm hỏng.

Nếu muốn biết chính xác thì phải thử từng sản phẩm, nhưng khi đó xác suất sẽ không ý nghĩa gì nữa.

Khi phép thử không đối xứng, ta xác định xác suất $P(A)$ gần đúng theo cách sau.

Tần suất

Tiến hành n phép thử. Gọi m là số lần xuất hiện biến cố A trong n phép thử này.

Số $\frac{m}{n}$ gọi là *tần suất* xuất hiện biến cố A trong n phép thử. Kí hiệu $f_n(A)$.

Xác suất

Xác suất xuất hiện biến cố A trong một phép thử là một số không đổi mà tần suất $f_n(A)$ dao động rất ít xung quanh nó khi phép thử tăng lên vô hạn lần.

Trong thực tế với phép thử đủ lớn $P(A) \approx f_n(A)$

Ví dụ . Từ ngàn xưa một số người đã tiến hành quan sát tỉ lệ sinh con trai của một số vùng lãnh thổ trong những thời điểm khác nhau. Kết quả các số liệu quan sát được ghi trong bảng sau:

Người thống kê	**Nơi thống kê**	**Tỉ số con trai**
Người Trung hoa cổ đại	Trung Quốc	$\approx \frac{1}{2}$
Laplace	London, Berlin	$\frac{22}{43} \approx 0,5116$
Cramer	Thuỵ Điển	$\frac{45682}{88079} \approx 0,51187$
Darmon	Pháp	$\approx 0,511$
Tổng cục thống kê VN	Việt Nam	$\approx 0,508$

Kết quả ghi trên bảng trên cho ta thấy tỉ lệ sinh con trai (trên tổng số lần sinh) dao động quanh con số 0,511.

Ví dụ 4. Để nghiên cứu khả năng xuất hiện mặt sấp khi tung một đồng xu, người ta tiến hành tung một đồng xu nhiều lần và thu được kết quả sau đây:

Người tiến hành thử	**Số lần tung**	**Số lần được mặt sấp**	**Tần suất xuất hiện mặt sấp**
Button	4040	2048	0,5069
Person	12000	6019	0,5016
Person	24000	12012	0,5005

Từ kết quả các lần thử trên ta thấy khi số phép thử tăng lên thì tần suất xuất hiện mặt sấp dao động ngày càng ít xung quanh giá trị không đổi là 0,5 . Vậy tần suất xuất hiện mặt sấp tiến dần đến 0,5 là xác suất xuất hiện mặt sấp khi tung đồng xu. Vậy tần suất tiến dần đến xác suất khi số phép thử tăng dần đến vô hạn.

Nhận xét: Dùng Định nghĩa thống kê về xác suất thì

- Không đòi hỏi phép thử hữu hạn biến cố và đồng khả năng.
- Đòi hỏi phải lặp lại nhiều lần phép thử. Trong thực tế điều này khó thực hiện.
- Phương pháp định nghĩa xác suất bằng thống kê chỉ cho ta giá trị xấp xỉ và mức độ chính xác của xấp xỉ tùy thuộc vào số lần thực hiện phép thử.
- Phương pháp định nghĩa xác suất bằng thống kê được áp dụng hiệu quả trong việc tìm ra các quy luật diễn biến về thời tiết, về tỷ lệ phế phẩm, nghiên cứu thuốc men, trong nhân chủng học, xã hội học…

1.3.3. Định nghĩa hình học về xác suất

Coi các kết cục đồng khả năng của một phép thử có thể biểu diễn bằng miền hình học D nào đó. Còn các kết cục thuận lợi cho biến cố A được biểu diễn bằng miền hình học G. Khi đó

$$P(A) = \frac{\text{Do do mien G}}{\text{Do do mien D}}$$

Độ đo có thể là số đo biểu thị độ dài, diện tích, thể tích

Ví dụ. Hai người hẹn gặp nhau tại một địa điểm xác định vào khoảng từ 20h đến 21h. Mỗi người đến (và chắc chắn đến) địa điểm đã hẹn trong khoảng thời gian đó một cách độc lập, chờ 20 phút, nếu không gặp người kia thì bỏ đi. Tính khả năng (xác suất) để hai người gặp nhau.

Giải

Gọi A là biến cố hai người gặp nhau; x, y là thời điểm đến của mỗi người. Rõ ràng x, y đều là một điểm ngẫu nhiên trong đoạn [20; 21]. Để G xẩy ra, tức là hai người gặp nhau, ta phải có $|x-y| \le 20$ (phút) = 1/3(giờ).

Xem cặp (x, y) như là một điểm trên mặt phẳng tọa độ. Khi đó ta được hai miền phẳng

D = { (x, y) / 20 ≤ x ≤ 21; 20 ≤ y ≤ 21}: biểu diễn tất cả các trường hợp có thể xảy ra;

$G = \{(x, y) \in D : |x-y| \le 1/3\}$ biểu diễn các trường hợp thuận lợi cho biến cố A xẩy ra.

Do đó

$$P(A)=\frac{S_G}{S_D}=\frac{5/9}{1}=\frac{5}{9}$$

1.4. Nguyên lí xác suất lớn và nguyên lí xác suất nhỏ

Nguyên lý xác suất nhỏ:

Nếu một biến cố có xác suất rất nhỏ thì thực tế có thể cho rằng trong một phép thử biến cố đó sẽ không xảy ra"

Ta gọi đây là nguyên lí xác suất nhỏ hay nguyên lí biến cố hiếm.

Vậy xác suất của một biến cố bằng bao nhiêu thì được xem là nhỏ? Điều này tùy thuộc từng bài toán, từng người, từng hoàn cảnh.

Chẳng hạn mỗi chiếc máy bay trên thực tế có một xác suất rất nhỏ bị tai nạn. Nhưng trên thực tế ta vẫn không từ chối đi máy bay vì tin tưởng rằng trong chuyến bay tai nạn sẽ không xảy ra.

Xác suất để một chiếc máy bay rơi là 0,01 thì không thể coi là nhỏ nhưng xác suất một chuyến xe bị trễ là 0,01 thì có thể hoàn toàn coi là nhỏ.

Xác suất một người đua xe bị chết là 1%. Điều này với các anh hùng xa lộ liều mạng thì có thể xem là nhỏ nhưng đối với những người cẩn thận thì đây là lớn.

Mức xác suất nhỏ này được gọi là *"mức ý nghĩa*". Tùy thuộc vào từng bài toán, mức ý nghĩa này có thể được lấy trong khoảng từ 0,00001 đến 0,1. Nếu α là mức ý nghĩa thì số β =1-α được gọi là *"độ tin cậy"*. Khi dựa trên Nguyên lý xác suất nhỏ ta tuyên bố rằng "Biến cố A có xác suất nhỏ (tức là P(A) ≤ α) sẽ không xảy ra trong thực tế thì độ tin cậy của tuyên bố này là β". Tức là tính đúng đắn của kết luận này đúng trong β.100% trường hợp.

Nguyên lí xác suất lớn

"*Nếu biến cố ngẫu nhiên có xác suất rất lớn (gần bằng 1) thì thực tế có thể cho rằng biến cố đó sẽ xảy ra trong một phép thử"*

Việc quy định một mức xác suất đủ coi là lớn tùy thuộc vào từng bài toán cụ thể

Tùy thuộc vào từng bài toán, xác suất lớn này có thể được lấy 0,95; 0,99….

1.5. Các công thức tính xác suất

1.5.1. Công thức cộng

Cho *A* và *B* là hai biến cố tùy ý**.** Khi đó $P(A+B)=P(A)+P(B)-P(AB)$.

Cho n biến cố $A_1, A_2, \ldots, A_n$. Khi đó

$$\mathrm{P}\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_iA_j) + \sum_{1\le i<j<k\le n} P(A_iA_jA_k) - \cdots + (-1)^{n-1}P(A_1A_2\ldots A_n)$$

"Nói thêm: $P(A+B+C) = P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)$ "

Chú ý:

Nếu *A* và *B* xung khắc thì $P(A+B) = P(A)+P(B)$

Nếu $A_1, A_2, \ldots, A_n$ xung khắc từng đôi thì $\mathrm{P}\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$

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

Nếu $\{A_1, A_2, \ldots, A_n\}$ là một hệ đầy đủ thì: $\mathrm{P}\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) = 1$

Ví dụ

Một hộp có 10 bi trong đó có 4 bi màu đỏ. Lấy ngẫu nhiên 3 bi từ hộp. Tính xác suất để lấy được ít nhất một bi màu đỏ.

Giải

Phép thử: Lấy ngẫu nhiên 3 bi

Cách 1.

Gọi *A* là biến cố " Lấy 3 bi có ít nhất 1 bi đỏ".

A_i là biến cố " lấy 3 bi có *i* bi đỏ", *i* = 0,1,2,3. Ta có $A = A_1 + A_2 + A_3$.

Các biến cố A_0, A_1, A_2, A_3 xung khắc từng đôi , do đó

$$P(A) = P(A_1) + P(A_2) + P(A_3) = \frac{C_4^1.C_6^2}{C_{10}^3} + \frac{C_4^2.C_6^1}{C_{10}^3} + \frac{C_4^3}{C_{10}^3} \approx 0,8333$$

Cách 2: $P(A) = 1 - P(\overline{A}) = 1 - \dfrac{C_6^3}{C_{10}^3} \approx 0,8333$

Ví dụ . Một lớp có 100 sinh viên, trong đó có 40 sinh viên giỏi Toán, 50 sinh viên giỏi Văn và 20 sinh viên giỏi cả Toán và Văn. Sinh viên nào giỏi ít nhất 1 trong 2 môn này sẽ được thưởng. Chọn ngẫu nhiên 1 sinh viên trong lớp. Tìm xác suất để sinh viên được thưởng?

Giải :

Phép thử : chọn ngẫu nhiên 1 sinh viên trong 100 sinh viên.

Gọi T: "Sinh viên được chọn giỏi Toán"

V: " Sinh viên được chọn giỏi Văn"

A: " Sinh viên được chọn giỏi ít nhất 1 trong 2 môn"

Ta có : A = T+V. Theo công thức cộng ta có :

$$P(A)=P(T+V)=P(T)+P(V)-P(TV)=\frac{40}{100}+\frac{50}{100}-\frac{20}{100}=0,7$$

1.5.2. Xác suất có điều kiện- Công thức nhân xác suất

Xác suất có điều kiện

Định nghĩa. Cho hai biến cố *A* và *B*. Xác suất của biến cố *A* được tính trong điều kiện biến cố *B* đã xảy ra rồi gọi là xác suất có điều kiện của *A* với điều kiện *B*, kí hiệu là $P(A|B)$

Ví dụ 1: Trong một bình có 3 quả cầu trắng và 2 quả cầu đen. Lấy ngẫu nhiên lần lượt không hoàn lại hai quả cầu. Tìm xác suất để lần thứ hai lấy được quả cầu trắng (biến cố *A*). Biết rằng lần thứ nhất đã lấy được cầu trắng. (Biến cố *B*)

Giải

Sau khi lấy lần thứ nhất (biến cố *B*) đã xảy ra thì trong bình chỉ còn lại 4 quả cầu, trong đó có hai cầu trắng. Do đó

$$P(A|B)=\frac{2}{4}$$

Ví dụ 2:

Tung một con xúc xắc. Gọi *A* là biến cố mặt 1 chấm xuất hiện. *B* là biến cố mặt lẻ xuất hiện. Ta có P(*A*|*B*)=1/3; P(*B*|*A*)=1

Công thức

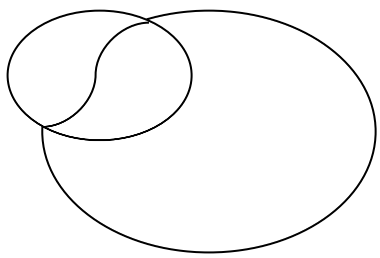
$$P(A|B)=\frac{P(AB)}{P(B)} \quad (P(B)>0)$$

$0\le P(A|B)\le 1$

$P(B|B)=1$

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

$P(A_1+A_2|B)=P(A_1|B)+P(A_2|B)-P(A_1A_2|B)$



Công thức nhân xác suất

Cho *A*, *B* là hai biến cố bất kì. Khi đó

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

Với $A_1, A_2, ..., A_n$ là các biến cố bất kì thì

$$P(A_1.A_2...A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\ldots P(A_n|A_1A_2\ldots A_{n-1})$$

Ví dụ

Cho một lô hàng có 20 sản phẩm trong đó có 5 phế phẩm. Lấy liên tiếp hai sản phẩm. Tính xác suất để cả hai đều là phế phẩm.

Giải:

Gọi A_1, A_2 lần lượt là biến cố sản phẩm thứ nhất, thứ hai là phế phẩm. Ta cần tính $P(A_1A_2)=P(A_1).P(A_2|A_1)=5/20.4/19=1/19$

Hai biến cố độc lập

Hai biến cố *A* và *B* gọi là độc lập với nhau nếu

$$P(A|B) = P(A) \text{ hoặc } P(B|A) = P(B)$$

Tức là việc xảy ra hay không xảy ra của biến cố này đều không ảnh hưởng đến khả năng xảy ra biến cố kia.

Ví dụ:

Trong một bình có 3 quả cầu trắng và 2 quả cầu đen.

a. Lấy ngẫu nhiên có hoàn lại lần lượt từng quả cầu ra hai quả cầu để kiểm tra. Gọi A là biến cố lần 1 lấy được cầu trắng. B là biến cố lần 2 lấy được cầu trắng. A và B độc lập với nhau.

b. Lấy lần lượt hai quả cầu theo phương thức không hoàn lại. A và B cũng là biến cố giống như trên. Khi đó A và B không độc lập với nhau.

Nhận xét

Nếu *A* và *B* độc lập với nhau thì A và $\overline{B}$; $\overline{A}$ và $\overline{B}$; $\overline{A}$ và B cũng độc lập với nhau

$$A, B \text{ độc lập} \Leftrightarrow P(AB) = P(A)P(B)$$

Nhóm các biến cố $A_1, A_2, ..., A_n$ được gọi là *độc lập từng đôi* nếu 2 biến cố bất kì trong nhóm độc lập với nhau.

Nhóm các biến cố $A_1, A_2, ..., A_n$ được gọi là *độc lập toàn phần* với nhau nếu mỗi biến cố độc lập với các biến cố còn lại cũng như với tích của một số bất kỳ các biến cố còn lại.

A_1, A_2, A_3 được gọi là *độc lập toàn phần* nếu A_1 độc lập với A_2, A_1 độc lập với A_3, A_2 độc lập với A_3, A_1 độc lập với A_2A_3, A_2 độc lập với A_1A_3, A_3 độc lập với A_1A_2

Nếu $A_1, A_2, ..., A_n$ độc lập toàn phần với nhau thì

$$P(A_1.A_2...A_n) = P(A_1)P(A_2)\ldots P(A_n)$$

Ví dụ

Một lô hàng có 20 sản phẩm, trong đó có hai sản phẩm xấu. Chọn lần lượt mỗi lần một sản phẩm cho đến khi phát hiện đủ hai sản phẩm xấu thì dừng. Tính xác suất để dừng lại ở lần chọn thứ 3 nếu việc chọn là chọn có hoàn lại.

Giải

Đặt X_i là biến cố chọn được sản phẩm xấu ở lần chọn thứ i $i=\overline{1,20}$.

D là biến cố dừng ở lần chọn thứ 3. Ta có

$D = X_1\overline{X_2}X_3 + \overline{X_1}X_2X_3$

Các biến cố này xung khắc với nhau nên $P(D) = P(X_1\overline{X_2}X_3) + P\left(\overline{X_1}X_2X_3\right)$

Do chọn hoàn lại nên các biến cố X_1, X_2, X_3 độc lập toàn phần với nhau. Do đó

$$P(X_1\overline{X_2}X_3) = P(X_1)P(\overline{X_2})P(X_3) = \frac{2}{20}.\frac{18}{20}.\frac{2}{20} = 0,9\%$$

$$P(\overline{X_1}X_2X_3) = P(\overline{X_1})P(X_2)P(X_3) = \frac{18}{20}.\frac{2}{20}.\frac{2}{20} = 0,9\%$$

$P(D) = 1,8\%$

1.5.3. Công thức xác suất đầy đủ- Công thức Bayer

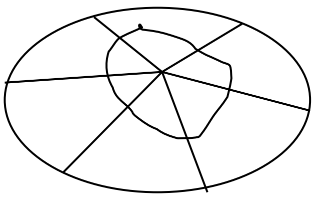
Công thức xác suất đầy đủ.

Cho $A_1, A_2, ..., A_n$ là một nhóm đầy đủ các biến cố và B là một biến cố bất kỳ. Khi đó

$$\begin{aligned} P(B) &= \sum_{i=1}^{n} P(A_i)P(B|A_i) \\ &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + ... + P(A_n)P(B|A_n) \end{aligned}$$

Giải thích

$$A_1 + A_2 + \ldots + A_n = \Omega$$
$$B = B\Omega = BA_1 + BA_2 + \ldots + BA_n$$
$$P(B) = P(BA_1) + P(BA_2) + ... + P(BA_n) = P(A_1)P(B\,|\,A_1) + ...$$



Công thức Bayer

Cho $A_1, A_2, ..., A_n$ là một nhóm đầy đủ các biến cố và B là một biến cố bất kỳ. Khi đó

$$P(A_i|B)=\frac{P(A_iB)}{P(B)}=\frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n}P(A_i)P(B|A_i)}\quad (i=1,2,....,n)$$

+ $P(A_1),P(A_2),.....,P(A_n)$: xác suất tiên nghiệm

+ $P(A_1|B),P(A_2|B),.....,P(A_n|B)$: xác suất hậu nghiệm

Ví dụ 1. Có hai hộp bi hình dáng giống nhau. Hộp 1 có 8 bi đỏ, 3 bi vàng. Hộp 2 có 10 bi đỏ, 4 bi vàng. Lấy ngẫu nhiên 1 hộp, từ đó lấy ngẫu nhiên ra 1 bi. Tính xác suất lấy được bi đỏ.

HD:

Gọi H_i là biến cố lấy được hộp i; i=1,2. Ta có H_1, H_2 là một nhóm đầy đủ và $P(H_i)=1/2$

Gọi A là biến cố lấy được bi đỏ. Khi đó

$P(A)=P(H_1).P(A|H_1) + P(H_2).P(A|H_2)$

Ví dụ 2

Một nhà máy sản xuất bóng đèn có ba phân xưởng. Phân xưởng I sản xuất 25%, phân xưởng II sản xuất 35%, phân xưởng III sản xuất 40% số bóng đèn. Tỉ lệ phế phẩm của từng phân xưởng lần lượt là 3%, 2%, 1%. Lấy ngẫu nhiên một bóng đèn do nhà máy sản xuất.

a. Tính xác suất để lấy được phế phẩm.

b. Biết rằng sản phẩm này là chính phẩm. Tính xác suất để nó do phân xưởng III sản xuất.

c. Nếu lấy được phế phẩm. Khả năng sản phẩm do phân xưởng nào sản xuất là nhiều nhất.

HD:

Đặt A_i là biến cố sản phẩm lấy được do phân xưởng i sản xuất. T là biến cố lấy được sản phế phẩm. A_1, A_2, A_3 là hệ đầy đủ các biến cố.

$$a) P(T) = P(A_1)P(T \mid A_1) + P(A_2)P(T \mid A_2) + P(A_3)P(T \mid A_3)$$
$$= 0,25.0,03 + 0,35.0,02 + 0,4.0,01$$
$$P(\overline{T}) = 1 - P(T)$$
$$b) P(A_3 \mid \overline{T}) = \frac{P(A_3\overline{T})}{P(\overline{T})} = \frac{P(A_3)P(\overline{T} \mid A_3)}{P(\overline{T})} = \frac{0,4.0,99}{....}$$
$$c) P(A_1 \mid T) = \frac{P(A_1)P(T \mid A_1)}{P(T)} = ...$$
$$P(A_2 \mid T) = \frac{P(A_2)P(T \mid A_2)}{P(T)} = ...$$
$$P(A_3 \mid T) = \frac{P(A_3)P(T \mid A_3)}{P(T)} = ...$$

Ví dụ 3: Hộp 1 có 10 quả cầu đỏ, 5 quả cầu vàng. Hộp 2 có 7 quả cầu đỏ, 3 quả cầu vàng. Từ mỗi hộp lấy ngẫu nhiên một quả cầu, sau đó lấy ngẫu nhiên 1 quả từ hai quả cầu này. Tìm xác suất quả cầu lấy sau cùng là quả cầu vàng.

HD:

Gọi Ci là biến cố trong hai quả cầu lấy ra từ hai hộp có i quả vàng (i=0,1,2). Ta có C0, C1, C2 là một nhóm đầy đủ.

Gọi A là biến cố quả cầu lấy từ hộp 1 màu vàng

B biến cố quả cầu lấy từ hộp 2 màu vàng.

A và B độc lập

Ta có

$$P(C_0) = P(\overline{A}\overline{B}) = P(\overline{A})P(\overline{B})$$
$$P(C_1) = P(\overline{A}B + A\overline{B})$$
$$P(C_2) = P(AB)$$

1.5.4. Dãy phép thử Bernoulli

Dãy *n* phép thử Bernoulli là dãy *n* phép thử thỏa mãn 3 điều kiện sau:

- Các phép thử của dãy độc lập với nhau (nghĩa là kết quả của mỗi phép thử không phụ thuộc vào kết quả của các phép thử khác)
- Mỗi phép thử có hai kết quả: A và $\overline{A}$
- Xác suất xảy ra biến cố *A* trong mỗi phép thử đều bằng p

Ví dụ 1

Gieo một con xúc xắc cân đối đồng chất 10 lần và theo dõi mặt 6 chấm có xuất hiện hay không.

Ví dụ 2:

Một xạ thủ bắn 5 phát súng vào mục tiêu, xác suất trúng đích của mỗi phát là 0,2.

Ví dụ 3. Một đề thi trắc nghiệm có 3 câu hỏi . Xác suất trả lời đúng mỗi câu của một sinh viên là 0,4. Tính xác suất sinh viên đó trả lời đúng 2 câu.

Công thức Bernoulli

Xác suất để trong dãy n phép thử Bernoulli biến cố A xuất hiện đúng k lần, kí hiệu $P_n(k)$, được tính bởi công thức

$$P_n(k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0,1,2...,n$$

Ví dụ 4: Trong phân xưởng có 5 máy hoạt động , xác suất để trong ca mỗi máy bị hỏng đều bằng 0,1. Tìm xác suất để trong ca có đúng 2 máy hỏng.

BÀI TẬP CHƯƠNG 1

1. Kiểm tra 4 sản phẩm . Gọi A_k là biến cố sản phẩm thứ k tốt. Hãy trình bày các biến cố sau qua A_k .
   a. A: tất cả đều xấu
   b. B: có ít nhất 1 sản phẩm xấu
   c. C: có ít nhất 1 sản phẩm tốt
   d. D: có đúng một sản phẩm xấu
   e. E: có 2 sản phẩm tốt
2. Bắn vào bia 5 phát. Gọi A_i là biến cố bắn trúng ít nhất i phát. B_j là biến cố bắn trúng đúng j phát.
   a. Diễn tả các biến cố $\overline{A_1}, \overline{B_1}, \overline{A_2}, \overline{B_2}$?
   b. Hai biến cố $\overline{A_1}, \overline{B_1}$ có xung khắc nhau không?
   c. Diễn tả các biến cố $A_1 + \overline{B_2}$; $B_1 + \overline{A_2}$; $A_1\overline{B_2}$; $\overline{A_2}B_1$?
3. Quan sát 4 sinh viên làm bài thi. Gọi B_j là biến cố sinh viên thứ j làm bài đạt yêu cầu. Hãy biểu diễn các biến cố sau đây qua B_j.
   a. Có đúng một sinh viên đạt yêu cầu.
   b. Có đúng 3 sinh viên đạt yêu cầu.
   c. Có ít nhất một sinh viên đạt yêu cầu.
   d. Không có sinh viên nào đạt yêu cầu.
4. Một lô hàng gồm 100 sản phẩm trong đó có 10 phế phẩm. Lấy ngẫu nhiên ra 20 sản phẩm. Tìm xác suất để cho trong 20 sản phẩm lấy ra:
   a. Có 5 phế phẩm
   b. Không có phế phẩm
5. Gieo đồng thời 2 con xúc sắc. Tìm xác suất để:
   a. Tổng số chấm xuất hiện trên 2 con là 7.
   b. Tổng số chấm xuất hiện trên 2 con là 8.
   c. Số chấm xuất hiện trên 2 con hơn kém nhau
6. Có hai hộp bi. Hộp 1 có 6 bi đỏ và 5 bi vàng. Hộp 2 có 7 bi đỏ và 6 bi vàng. Lấy ngẫu nhiên mỗi hộp 1 bi. Tìm xác suất
   a. Lấy được 2 bi cùng màu
   b. Lấy được 2 bi khác màu

c. Lấy được 2 bi đỏ

7. Xếp ngẫu nhiên 5 người lên 7 toa tàu. Tính xác suất để

   a. 5 người lên cùng toa đầu

   b. 5 người lên cùng một toa

   c. 5 người lên 5 toa khác nhau.

8. Một lô hàng gồm 20 sản phẩm trong đó có 5 phế phẩm. Người ta kiểm tra bằng phương pháp sau: kiểm tra lần lượt 4 sản phẩm (không hoàn lại). Nếu có ít nhất 1 trong 4 sản phẩm đó là phế phẩm thì loại lô hàng đó. Tìm xác suất để lô hàng đó được nhận?

9. Một xí nghiệp có hai ô tô hoạt động độc lập. Xác suất trong một ngày làm việc các ô tô này hỏng tương ứng là 0,08; 0,1. Tính xác suất trong 1 ngày làm việc xí nghiệp có

   a. 2 ô tô hỏng

   b. có 1 ô tô hỏng

   c. có ô tô hỏng

10. Một công ty cần tuyển 4 nhân viên. Biết rằng có 20 người dự tuyển trong đó có 12 nam và 8 nữ.

    a. Tính xác suất trong 4 người được tuyển có ít nhất 1 nam

    b. An là một trong 12 nam dự tuyển. Biết rằng trong 4 người được tuyển có ít nhất 1 nam. Tính xác suất để An được tuyển

11. Có 3 bình đựng bi trong đó :

    bình loại 1 đựng 6 bi trắng 3 bi đỏ.

    bình loại 2 đựng 5 bi trắng 4 bi đỏ.

    bình loại 3 đựng 2 bi trắng 7 bi đỏ.

    Lấy ngẫu nhiên 1 bình và từ bình đó lấy ra 1 bi.

    a. Tính xác suất để lấy được bi trắng?
    b. Biết rằng bi lấy ra là bi trắng. Tính xác suất để bình lấy ra là loại 3?

12. Có hai lô sản phẩm, mỗi lô có 10 sản phẩm. Lô thứ nhất có 3 sản phẩm loại 1, lố thứ hai có 4 sản phẩm loại 1. Tư lô thứ nhất lấy 1 sản phẩm bỏ sang lô thứ 2. Rồi từ lô thứ hai lấy 1 sản phẩm

    a. Tính xác suất để sản phẩm lấy ở lô thứ hai là loại 1.

b. Giả sử sản phẩm lấy ở lô thứ hai là loại 1. Nhiều khả năng sản phẩm bỏ từ lô thứ nhất sang lô thứ hai là loại gì

13. Trong một thúng cam có 42% cam Trung Quốc, 24% cam Thái Lan, 26% cam Campuchia và 8% cam Việt Nam. Trong số đó có một số cam hư gồm: 20% số cam của Trung Quốc, 10% số cam của Thái Lan, 12% số cam của Campuchia và 2% số cam của Việt Nam.
    a. Tính xác suất để một người mua phải 1 trái cam TQ hư?
    b. Tính xác suất để một người mua phải 1 trái cam hư?
    c. Biết một người đã mua phải 1 trái cam hư. Tính xác suất để trái cam ấy của CPC?
    d. Biết một người đã mua phải 1 trái cam hư. Tính xác suất để trái cam ấy không của Việt Nam?
14. Trong một cơ quan điều tra người ta dùng máy dò tìm tội phạm, kinh nghiệm cho biết cứ 10 người bị tình nghi thì có 7 người là tội phạm.
    Máy báo đúng người có tội với xác suất 0,85.
    Máy báo sai người vô tội với xác suất 0,1.
    Một người được máy phân tích. Hãy tìm xác suất.
    a. Người này là tội phạm?
    b. Máy báo người này là tội phạm?
    c. Người này thực sự có tội biết rằng máy đã báo có tội?
    d. Máy báo đúng?
15. Kiện hàng 1 có 5 sản phẩm loại A, 1 sản phẩm loại B. Kiện hàng 2 có 2 sản phẩm loại A, 4 sản phẩm loại B. Từ mỗi kiện chọn ra ngẫu nhiên 1 sản phẩm đem giao cho khách hàng. Sau đó các sản phẩm còn lại được dồn chung vào kiện hàng 3 đang trống.
    a. Nếu ta lấy ngẫu nhiên 1 sản phẩm từ kiện hàng 3 thì xác suất để chọn được sản phẩm loại B là bao nhiêu?
    b. Nếu ta lấy ngẫu nhiên 2 sản phẩm từ kiện hàng 3, hãy tính xác suất để có ít nhất 1 sản phẩm loại B trong 2 sản phẩm được chọn?
16. Xác suất tiêu thụ điện không quá mức quy định của một nhà máy trong một ngày là 0,8. Tính xác suất trong 1 tuần (6 ngày)nhà máy
    a. Có 4 ngày tiêu thụ điện không quá mức quy định

b. Có ngày tiêu thụ điện quá mức quy định

17. Một sinh viên thi 5 môn với xác suất đạu từng môn là 0,7. Tính xác suất sinh viên đó

    a. Đậu 3 môn

    b. Đậu ít nhất 1 môn

    c. Đậu từ 1 đến 3 môn

18. Có 20 hộp sản phẩm. Mỗi hộp có 8 sản phẩm loại A, 2 sản phẩm loại B. Lấy ngẫu nhiên mỗi hộp 1 sản phẩm để kiểm tra. Tính xác suất trong số sản phẩm lấy ra có 2 sản phẩm loại B.

19. Có 2 xe chở hàng độc lập về 1 xí nghiệp, xác suất để 2 xe chở hàng về xí nghiệp lần lượt là 0,7; 0,6. Tìm xác suất sao cho:

    a. Chỉ có 1 xe chở hàng về tới xí nghiệp?

    b. Xí nghiệp nhận được hàng?

20. Có 12 hộp sữa trong đó có 3 hộp hư được chia làm 3 gói mỗi gói 4 hộp. Tìm xác suất để trong mỗi gói đều có một hộp hư?

21. Một công nhân kĩ thuật đứng 4 máy hoạt động độc lập nhau. Xác suất để trong khoảng thời gian T các máy không cần người công nhân đến coi lần lượt là 0,8; 0,9; 0,85; 0,95. Tìm xác suất sao cho trong khoảng thời gian T đó:

    a. Không máy nào cần công nhân đến coi.

    b. Ít nhất 1 máy cần công nhân đến coi.

22. Gieo đồng thời 2 con xúc sắc. Tìm xác suất để:

    d. Tổng số chấm xuất hiện trên 2 con là 7.

    e. Tổng số chấm xuất hiện trên 2 con là 8.

    f. Số chấm xuất hiện trên 2 con hơn kém nhau 2.

23. (Bài toán về 4 người nói dối). Người thứ nhất nhận 1 thông tin dưới 2 tín hiệu “Có” hoặc “Không” và anh ta thông báo cho người thứ 2 biết, sau đó người thứ 2 thông báo cho người thứ 3 và người thứ 3 thông báo cho người thứ 4. Bằng cách tương tự người thứ 4 thông báo kết quả nhận được cho người khác. Biết rằng mỗi người trong 4 người đó chỉ nói đúng 1 lần trong 3 lần nói. Tìm xác suất để người 1 nói đúng với điều kiện người 4 nói đúng?

24. (Bài toán về cách lựa chọn tối ưu). Một cửa hàng bán áo len gồm 1000 chiếc. Một người đi mua áo len và cần chọn áo tốt nhất có thể lựa chọn được. Người

đó lần lượt xem áo, nếu ưng ý thì mua và việc lựa chọn kết thúc, nếu không ưng ý thì trả lại áo ấy để xem áo khác. Và cái đã trả lại thì không bao giờ xem xét lại. Giả sử theo qui tắc trên người đó chọn áo ở lần thứ 100 (tức là áo thứ 100 tốt hơn 99 áo đã xem). Tìm xác suất để áo đó là tốt nhất trong toàn bộ 1000 áo của cửa hàng?

25. Một đợt thi tuyển viên chức có 3 vòng thi. Vòng 1 lấy 80% thí sinh dự thi; vòng 2 lấy 70% thí sinh đã qua vòng 1 và vòng 3 lấy 90% thí sinh đã qua vòng 2. Giả sử khả năng trúng tuyển của các thí sinh là như nhau. Tìm xác suất để 1 thí sinh bất kì trúng tuyển?
26. Một hộp gồm 9 quả bóng. Mỗi lần chơi người ta lấy ra 3 quả, chơi xong lại bỏ vào hộp rồi lấy 3 quả ra chơi. Tìm xác suất để sau 3 lần lấy bóng ra chơi các bóng đều được sử dụng?
27. Một hộp gồm 9 quả bóng. Mỗi lần chơi người ta lấy ra 3 quả, chơi xong bỏ lại 2 quả vào hộp rồi lấy 3 quả ra chơi. Tìm xác suất để sau 3 lần lấy bóng ra chơi các bóng đều được sử dụng?
28. Một hộp gồm 24 sản phẩm gồm 2 sản phẩm loại 2. Lấy ngẫu nhiên ra từng sản phẩm kiểm tra (lấy không hoàn lại) đến khi nào được sản phẩm loại 2 thì dừng lại. Tìm xác suất để quá trình kiểm tra kết thúc sau không quá 3 lần lấy?
29. Một hộp gồm 4 bi đỏ và 5 bi xanh. Hai người lần lượt lấy ra từng viên theo phương thức không hoàn lại, ai lấy được bi xanh trước thì thắng cuộc. Tìm xác suất để người thứ 2 thắng cuộc?
30. Hai xí nghiệp hoạt động độc lập nhau, khả năng chỉ có 1 xí nghiệp hoàn thành kế hoạch là 0,46. Tìm xác suất hoàn thành kế hoạch của xí nghiệp thứ nhất biết xác suất hoàn thành kế hoạch của xí nghiệp thứ 2 là 0,6.
31. Một người mua xổ số cào , người đó mua liên tiếp từng vé xổ số đến khi nào được vé trúng thưởng thì dừng. Tìm xác suất sao cho người đó mua đến vé thứ 4 thì dừng, biết rằng xác suất trúng thưởng của mỗi lần mua là như nhau và bằng 0,01.
32. Học kì này sinh viên được thi môn LTXSTKT 3 lần. Xác suất để 1 sinh viên thi đỗ lần 1 là 0,5; lần 2 là 0,7 và lần 3 là 0,9. Tìm xác suất để sinh viên thi đỗ?
33. Trên bàn có 5 đồng xu gồm 3 sấp. Gieo tiếp 2 đồng xu lên bàn sau đó lấy ra ngẫu nhiên 4 đồng xu.

a. Tính xác suất trong 4 đồng xu lấy ra có 3 đồng xấp?

b. Giả sử lấy được 2 đồng xu ngửa, tính xác suất 2 đồng xu tung trước đó là 2 đồng xu ngửa?

34. Cho biết tỉ lệ người có nhóm máu O, A, B và AB tương ứng là 33,7%; 37,5%; 20,9% và 7,9%. Chọn nhẫu nhiên 1 người cần tiếp máu và 1 người cho máu. Tính xác suất để sự truyền máu thực hiện được?

35. CM: A, B độc lập khi và chỉ khi $A, \overline{B}$ độc lập

36. CM: A, B xung khắc kéo theo AC, BC xung khắc

37. CM: $P(A|B) = 1 - P(\overline{A}|B)$

38. CM: $P(A+B|C) = P(A|C) + P(B|C) - P(AB|C)$

39. Cho A, B, C sao cho $P(A|C) \ge P(B|C)$; $P(A|\overline{C}) \ge P(B|\overline{C})$.

Chứng minh: $P(A) \ge P(B)$

40. Cho 3 biến cố A, B, C sao cho A,B độc lập. $P(ABC) = 0,04$; $P(C|AB) = 0,25$; $P(B) = 4P(A)$. Hãy tính $P(A+B)$?

41. Cho $P(A) = 0,5$; $P(B) = 0,6$; $P(AB) = 0,4$. Tính $P(A|B)$; $P(B|\overline{A})$?

42. Cho A và B là hai biến cố sao cho $P(A) = \frac{1}{2}$; $P(B) = \frac{1}{3}$; $P(AB) = \frac{1}{6}$

a. $P(A+B)$ $P(\overline{A}+\overline{B})$ $P(\overline{A+B})$

b. $P(\overline{AB})$ $P(\overline{A}B)$ $P(A\overline{B})$

c. $P(\overline{A}+B)$ $P(A/B)$ $P(\overline{A}/B)$

d. $P(AB/B)$ $P(A\overline{B}/B)$ $P(A\overline{B}/\overline{B})$

43. Trong một hộp có 15 bóng đèn trong đó có 3 bóng hỏng. Lấy ngẫu nhiên có thứ tự không hoàn lại 3 bóng để dùng. Tìm xác suất để:

a. Cả 3 bóng đều hỏng?

b. Cả 3 bóng đều không hỏng?

c. Có ít nhất 1 bóng không hỏng?

d. Chỉ có bóng thứ 2 hỏng?

44. Xác suất vi trùng kháng mỗi loại thuốc A, B, C lần lượt là 5%, 10%, 20%. Nếu dùng cả 3 loại để diệt vi trùng . Tính xác suất để vi trùng bị tiêu diệt? (Giả sử 3 loại thuốc độc lập nhau).

45. Có 8 bình đựng bi trong đó có :

    + 2 bình loại 1: mỗi bình đựng 6 bi trắng 3 bi đỏ.

    + 3 bình loại 2: mỗi bình đựng 5 bi trắng 4 bi đỏ.

    + 3 bình loại 3: mỗi bình đựng 2 bi trắng 7 bi đỏ.

    Lấy ngẫu nhiên 1 bình và từ bình đó lấy ra 1 bi.

    c. Tính xác suất để lấy được bi trắng?

    d. Biết rằng bi lấy ra là bi trắng. Tính xác suất để bình lấy ra là loại 3?

46.

47. Một chàng trai viết thư cho 3 cô bạn gái, vì đãng trí nên bỏ thư vào phong bì một cách ngẫu nhiên. Tìm xác suất để cả 3 cô không ai nhận được đúng thư của mình?

CHƯƠNG 2. BIẾN NGẪU NHIÊN

2.1. Khái niệm biến ngẫu nhiên, phân loại biến ngẫu nhiên

Ví dụ

Gieo một con xúc sắc xuống mặt bàn. Gọi X là số nốt xuất hiện. X là một đại lượng (tức là có thể cân đo , đong , đếm được) biến đổi , ta không biết trước được một cách chắc chắn giá trị của X vì nó phụ thuộc vào kết quả của phép thử. Có thể số nốt là 1, là 2… chỉ biết rằng X có thể nhận các giá trị từ 1 đến 6 với xác suất tương ứng đều bằng 1/6 (chỉ Có thể dự đoán các giá trị với một xác suất nhất định).

2.1.1. Khái niệm

Biến ngẫu nhiên (hay đại lượng ngẫu nhiên) là một đại lượng mà trong kết quả của phép thử nó sẽ nhận một và chỉ một trong các giá trị có thể có của nó với một xác suất tương ứng nào đấy.

Biến ngẫu nhiên nhận giá trị phụ thuộc vào kết quả của phép thử ngẫu nhiên. Giá trị của nó là ngẫu nhiên không dự đoán trước được.

Ta kí hiệu biến ngẫu nhiên bằng các chữ in hoa X, Y, Z,… và các chữ thường x,y,z… để kí hiệu giá trị của chúng.

Ví dụ 1:

- Bắn một phát súng vào bia. Gọi X là khoảng cách từ điểm chạm của viên đạn đến tâm bia. X là BNN

- Trong bình có 7 quả cầu trắng và hai quả cầu vàng. Lấy ngẫu nhiên 3 quả. Gọi X là là số quả đen. X là BNN

Ví dụ 2. Các đại lượng sau là các biến ngẫu nhiên

- Số nốt xuất hiện khi gieo một con xúc xắc.
- Tuổi thọ của một thiết bị đang hoạt động.
- Số khách hàng đến siêu thị trong một ngày.
- Số cuộc gọi đến một tổng đài.
- Số lỗi trên một trang giấy A4.
- Sai số khi đo lường một đại lượng vật lí.
- Số cơn bão đổ bộ vào Việt Nam trong vòng một năm.

2.1.2. Phân loại biến ngẫu nhiên

Người ta chia các biến ngẫu nhiên thành 2 loại:

- Biến ngẫu nhiên rời rạc nếu nó chỉ nhận một số hữu hạn hoặc vô hạn đếm được các giá trị. Nghĩa là ta có thể liệt kê các giá trị thành một dãy $x_1, x_2, ...$
- Biến ngẫu nhiên liên tục nếu tập giá trị của nó có thể lấp đầy một khoảng trên trục số.

Ví dụ

- Gọi X là số nốt xuất hiện khi gieo một con xúc sắc thì X là biến ngẫu nhiên rời rạc nhận các giá trị 1,2,3,4,5,6.
- Gọi Y là tuổi thọ của một loại thiết bị đang hoạt động thì Y là biến ngẫu nhiên liên tục nhận giá trị trong một khoảng.
- Gọi Z là số khách hàng vào siêu thị trong một ngày thì Z là biến ngẫu nhiên rời rạc nhận các giá trị 0,1,2,3…
- Số cuộc gọi đến một tổng đài là biến ngẫu nhiên rời rạc nhận các giá trị 0,1,2,3,4…
- Sai số khi đo lường một đại lượng vật lí Y nào đó là một biến ngẫu nhiên liên tục nhận giá trị trong một khoảng.

2.2. Quy luật phân phối xác suất của biến ngẫu nhiên

Quy luật phân phối xác suất của biến ngẫu nhiên là một cách biểu diễn quan hệ giữa các giá trị của biến ngẫu nhiên với các xác suất tương ứng mà nó nhận các giá trị đó.

Để thiết lập qui luật phân phối xác suất của BNN ta có thể dùng: bảng phân phối xác suất, hàm phân phối xác suất và hàm mật độ xác suất.

- ***Biến ngẫu nhiên rời rạc:*** ta dùng bảng phân phối xác suất & hàm phân phối xác suất.
- ***Biến ngẫu nhiên liên tục:*** ta dùng hàm phân phối & hàm mật độ xác suất.

2.2.1. Bảng phân phối xác suất

Cho biến ngẫu nhiên rời rạc X nhận các giá trị $x_1, x_2, ..., x_n$ với các xác suất tương ứng $P\left(X = x_i\right) = p_i$ $(i = 1, 2, \ldots, n)$.

$(X = x_1), (X = x_2), ..., (X = x_n)$ là hệ đầy đủ các biến cố

Bảng phân phối xác suất của biến ngẫu nhiên rời rạc X có dạng

X	x_1	x_2	….	x_n
P	p_1	p_2	….	p_n

Tính chất

$$\begin{cases} 0 \le p_i \le 1 \\ \sum_{i=1}^{n} p_i = 1 \end{cases}$$

$$P(a < X < b) = \sum_{a<x_i<b} p_i$$

Ví dụ 1. Một hộp có 10 sản phẩm trong đó có 6 sản phẩm đạt loại A. Lấy ngẫu nhiên 2 sản phẩm. Lập bảng phân phối xác suất của số sản phẩm loại A lấy ra?

Giải

Gọi X là số sản phẩm loại A lấy ra. X nhận các giá trị 0,1,2. Ta có:

$$p_1 = P(X=0) = \frac{C_4^2}{C_{10}^2} = \frac{2}{15} \qquad p_2 = P(X=1) = \frac{C_6^1 C_4^1}{C_{10}^2} = \frac{8}{15} \qquad p_3 = P(X=2) = \frac{C_6^2}{C_{10}^2} = \frac{5}{15}$$

Vậy bảng phân phối xác suất của X là:

X	0	1	2
P	$\frac{2}{15}$	$\frac{8}{15}$	$\frac{5}{15}$

Ví dụ 2. Có 2 kiện hàng. Kiện 1 có 3 sản phẩm tốt, 2 sản phẩm xấu. Kiện 2 có 2 sản phẩm tốt, 3 sản phẩm xấu. Lấy ngẫu nhiên từ kiện 1 ra 2 sản phẩm và từ kiện 2 ra 1 sản phẩm. Lập luật phân phối xác suất của số sản phẩm tốt trong 3 sản phẩm lấy ra?

Giải:

Gọi X là số sản phẩm tốt lấy ra. X nhận các giá trị 0, 1, 2, 3.

$$P(X=0) = \frac{C_2^2}{C_5^2}.\frac{C_3^1}{C_5^1} = \frac{3}{50} \qquad P(X=1) = \frac{C_3^1C_2^1}{C_5^2}.\frac{C_3^1}{C_5^1} + \frac{C_2^2}{C_5^2}.\frac{C_2^1}{C_5^1} = \frac{20}{50}$$

$$P(X=2) = \frac{C_3^1C_2^1}{C_5^2}.\frac{C_2^1}{C_5^1} + \frac{C_3^2}{C_5^2}.\frac{C_3^1}{C_5^1} = \frac{21}{50} \qquad P(X=3) = \frac{C_3^2}{C_5^2}.\frac{C_2^1}{C_5^1} = \frac{6}{50}$$

Bảng phân phối xác suất:

X	0	1	2	3
P	$\frac{3}{50}$	$\frac{20}{50}$	$\frac{21}{50}$	$\frac{6}{50}$

Nếu hỏi thêm tính $P(-1 \le X \le 1)$? Ta có

$$P(-1 \le X \le 1) = P\big((X=0)+(X=1)\big) = P(X=0) + P(X=1) = \frac{23}{50}$$

Ví dụ 3. Có 3 hộp trong đó có 2 hộp loại 1 và 1 hộp loại 2. Hộp loại 1 có 3 bi trắng, 2 bi vàng. Hộp loại 2 có 3 bi trắng, 3 bi vàng. Chọn ngẫu nhiên 1 hộp rồi từ hộp đó chọn ngẫu nhiên ra 2 bi. Gọi X là số bi trắng lấy được. Lập bảng phân phối xác suất của X?

2.2.2. Hàm phân phối xác suất

Định nghĩa: Cho X là một biến ngẫu nhiên. Hàm phân phối xác suất của biến ngẫu nhiên X, kí hiệu $F(x)$ là hàm số xác định như sau:

$$F(x) = P(X < x), \qquad x \in \mathrm{R}$$

Ví dụ . Cho biến ngẫu nhiên rời rạc X có bảng phân phối xác suất (Ví dụ 1 ở trên)

X	0	1	2
P	$\frac{2}{15}$	$\frac{8}{15}$	$\frac{5}{15}$

Tìm hàm phân phối xác suất của BNN X?

Giải:

Ta có:

- $x \le 0 \qquad \Rightarrow F(x) = P(X < x) = 0$
- $0 < x \le 1 \qquad \Rightarrow F(x) = P(X < x) = P(X = 0) = \frac{2}{15}$
- $1 < x \le 2 \qquad \Rightarrow F(x) = P(X < x) = P(X = 0) + P(x = 1) = \frac{2}{15} + \frac{8}{15} = \frac{2}{3}$
- $2 < x \qquad \Rightarrow F(x) = P(X < x) = P(X = 0) + P(X = 1) + P(X = 2) = 1$

Vậy hàm phân phối xác suất

$$F(x) = \begin{cases} 0 & khi \ \ x \le 0 \\ \frac{2}{15} & khi \ \ 0 < x \le 1 \\ \frac{2}{3} & khi \ \ 1 < x \le 2 \\ 1 & khi \ \ 2 < x \end{cases}$$

$$F(x) = P(X \le x), \qquad x \in \mathrm{R}$$

- $x < 0 \qquad \Rightarrow F(x) = P(X \le x) = 0$
- $0 \le x < 1 \qquad \Rightarrow F(x) = P(X \le x) = P(X = 0) = \frac{2}{15}$
- $1 \le x < 2 \qquad \Rightarrow F(x) = P(X \le x) = P(X = 0) + P(x = 1) = \frac{2}{15} + \frac{8}{15} = \frac{2}{3}$

- $2 \le x \qquad \Rightarrow F(x) = P(X \le x) = P(X=0) + P(X=1) + P(X=2) = 1$

$$F(x) = P(X \le x) = \begin{cases} 0 & khi\ \ x < 0 \\ \dfrac{2}{15} & khi\ \ 0 \le x < 1 \\ \dfrac{2}{3} & khi\ \ 1 \le x < 2 \\ 1 & khi\ \ 2 \le x \end{cases} \qquad F(x) = P(X < x) = \begin{cases} 0 & khi\ \ x \le 0 \\ \dfrac{2}{15} & khi\ \ 0 < x \le 1 \\ \dfrac{2}{3} & khi\ \ 1 < x \le 2 \\ 1 & khi\ \ 2 < x \end{cases}$$

Tính chất

1) $0 \le F(x) \le 1, \forall x \in R$

2) $F(x)$ là hàm không giảm $x_1 < x_2 \Rightarrow F(x_1) \le F(x_2)$

3) $\lim\limits_{x \to -\infty} F(x) = 0 \qquad \lim\limits_{x \to +\infty} F(x) = 1$

4) $P(a \le X < b) = F(b) - F(a)$ với mọi *a*, *b* thuộc R và *a* < *b*

$P(a < X \le b) = F(b) - F(a)$

" Các tính chất thứ *2*, *3* còn được gọi là các tính chất đặc trưng của hàm phân phối xác suất. Một hàm $F(x)$ xác định trên R có các tính chất *2, 3* đều là hàm phân phối xác suất của một biến ngẫu nhiên nào đó.

Ví dụ 1: Chứng tỏ rằng hàm sau đây: $F(x) = \dfrac{1}{2} + \dfrac{1}{\pi}\arctan x$ là hàm phân phối xác suất của một biến ngẫu nhiên.

Giải:

Do $F'(x) = \dfrac{1}{\pi}.\dfrac{1}{1+x^2} > 0$ nên $F(x)$ là hàm tăng.

Ta lại có: $\lim\limits_{x \to -\infty} F(x) = \dfrac{1}{2} + \dfrac{1}{\pi}\left(-\dfrac{\pi}{2}\right) = 0 \qquad \lim\limits_{x \to +\infty} F(x) = \dfrac{1}{2} + \dfrac{1}{\pi}\left(\dfrac{\pi}{2}\right) = 1$

Do đó $F(x)$ là hàm phân phối xác suất của một biến ngẫu nhiên.

2.2.3. Hàm mật độ xác suất

Định nghĩa. Cho *X* là một biến ngẫu nhiên liên tục có hàm phân phối $F(x)$ là hàm có đạo hàm. Khi đó hàm mật độ xác suất của biến ngẫu nhiên *X*, kí hiệu $f(x)$, là hàm xác định bởi

$$f(x) = F'(x) \quad (x \in \mathbb{R})$$

Tính chất:

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

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

iii) $P(a < X < b) = \int_{a}^{b} f(x)dx$

iv) $F(x) = \int_{-\infty}^{x} f(t)dt$

Các tính chất *i), ii)* là tính chất đặc trưng của hàm mật độ xác suất. Một hàm *f(x)* xác định trên R thỏa mãn *i), ii)* là hàm mật độ xác suất của một biến ngẫu nhiên liên tục *X* nào đó.

Chú ý.

Nếu *X* là biến ngẫu nhiên liên tục thì $P(X = a) = 0$. Do đó nếu *X* là biến ngẫu nhiên liên tục thì $P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b)$

Nếu *X* là biến ngẫu nhiên rời rạc có bảng phân phối xác suất

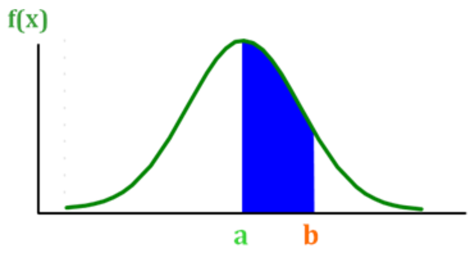
X	x_1	x_2	....	x_n
P	p_1	p_2	....	p_n

Với $x_1 < x_2 < \ldots < x_n$ thì hàm phân phối của *X* là

$$F(x) = \begin{cases} 0 & x < x_1 \\ p_1 & x_1 \leq x < x_2 \\ p_1 + p_2 & x_2 \leq x < x_3 \\ \ldots & \\ p_1 + p_2 + \ldots + p_{n-1} & x_{n-1} \leq x < x_n \\ 1 & x \geq x_n \end{cases}$$

Nếu *X* là biến ngẫu nhiên liên tục thì $F(x) = \int_{-\infty}^{x} f(t)dt$

Về mặt hình học, xác suất để biến ngẫu nhiên liên tục *X* nhận giá trị nằm trong khoảng (a,b) chính là diện tích của miền giới hạn bởi: $\begin{cases} Ox \\ y = f(x) \\ x = a;\ x = b \end{cases}$



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

Diện tích hình học của phần giới hạn bởi $\begin{cases} Ox \\ y=f(x) \\ x=-\infty;\ x=+\infty \end{cases}$ bằng 1.

Ví dụ 1. Cho biến ngẫu nhiên liên tục *X* có hàm mật độ xác suất

$$f(x)=\begin{cases} 0 & ,x<1 \\ \dfrac{k}{x^2} & ,x\geq 1 \end{cases}$$

a) Tìm *k*?

b) Tìm hàm phân phối xác suất *F*(*x*)?

c) Tính $P(2<X<3)$?

d) Tính xác suất để trong 4 phép thử độc lập biến ngẫu nhiên *X* đều không lấy giá trị trong (2;3).

Giải:

a)Vì *f*(*x*) là hàm mật độ xác suất, do đó:

$$\begin{cases} f(x)\geq 0 \\ \int_{-\infty}^{+\infty} f(x)dx=1 \end{cases}$$

$$1=\int_{-\infty}^{+\infty} f(x)dx=\int_{-\infty}^{1} f(x)dx+\int_{1}^{+\infty} f(x)dx=\int_{1}^{+\infty}\frac{k}{x^2}dx=\lim_{A\to+\infty}\left(\int_1^A\frac{k}{x^2}dx\right)=k.$$

$\lim_{A\to+\infty}\left(\int_1^A\frac{k}{x^2}dx\right)=\lim_{A\to+\infty}\left(-\frac{k}{x}\Big|_1^A\right)=\lim_{A\to+\infty}\left(k-\frac{k}{A}\right)=k$. Vậy $k=1$

vậy $f(x)=\begin{cases} 0 & ,x<1 \\ \dfrac{1}{x^2} & ,x\geq 1 \end{cases}$

b) Ta có :

TH1. $x<1$

$$F(x)=\int_{-\infty}^{x} f(t)\,dt=0$$

TH2. $x\geq 1$

$$F(x)=\int_{-\infty}^{x} f(t)\,dt=\int_{-\infty}^{1} f(t)\,dt+\int_{1}^{x} f(t)\,dt=\int_{1}^{x}\frac{1}{t^2}dt=-\frac{1}{t}\Big|_{1}^{x}=\frac{x-1}{x}$$

$$F(x)=\int_{-\infty}^{x} f(t)\,dt=\begin{cases}0 & ,x<1\\ \dfrac{x-1}{x} & ,x\geq 1\end{cases}$$

c) Ta có: $P(2<X<3)=F(3)-F(2)=\dfrac{3-1}{3}-\dfrac{2-1}{2}=\dfrac{1}{6}$

hoặc $P(2<X<3)=\int_{2}^{3} f(x)dx=\int_{2}^{3}\dfrac{1}{x^2}dx=\dfrac{1}{6}$

d) Xác suất để X không lấy giá trị trong khoảng (2;3) trong một phép thử là $1-\dfrac{1}{6}=\dfrac{5}{6}$.

Như vậy để trong bốn phép thử độc lập biến ngẫu nhiên X đều không nhận giá trị trong khoảng (2;3) là $\left(\dfrac{5}{6}\right)^4\approx 0,48$

Ví dụ 2. Cho hàm

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

a) Chứng tỏ $f(x)$ là hàm mật độ xác suất của một biến ngẫu nhiên *X*.

b) Tìm hàm phân phối xác suất $F(x)$ của *X*?

c) Tính xác suất $P\left(0<X<\dfrac{1}{2}\right)$

Giải:

a) Hiển nhiên $f(x)\geq 0$. Ta có: $\int_{-\infty}^{+\infty} f(x)dx=\int_{0}^{1} 2xdx=1$. Vậy $f(x)$ là hàm mật độ xác suất của một biến ngẫu nhiên X.

b) Ta có: $\mathrm{F}(x)=\int_{-\infty}^{x} f(t)\,dt=\begin{cases}0 & ,khi\ x\leq 0\\ x^2 & ,khi\ 0\leq x\leq 1\\ 1 & ,khi\ x>1\end{cases}$

c) $P\left(0 < X < \frac{1}{2}\right) = \int_0^{\frac{1}{2}} 2x dx = x^2 \Big|_0^{1/2} = \frac{1}{4}$

Ví dụ 3. Giả sử hàm phân phối xác suất của một biến ngẫu nhiên liên tục *X* có dạng:

$$F(x) = \begin{cases} 0 & , x \le 0 \\ kx^2 & , 0 < x \le 1 \\ 1 & , x > 1 \end{cases}$$

(*k* là hằng số)

a) Xác định hằng số *k*?

b) Tìm hàm mật độ xác suất của *X*?

Giải:

a) Do $F(x)$ liên tục, nên tại $x = 1$ ta phải có $1 = F(1) = \lim\limits_{x \to 1^-} F(x) = \lim\limits_{x \to 1^-} kx^2 = k$

b) Theo tính chất của hàm mật độ xác suất ta có:

$$f(x) = \begin{cases} 0 & , x \le 0 \\ 2x & , 0 < x \le 1 \\ 0 & , x > 1 \end{cases}$$

Sự độc lập của các biến ngẫu nhiên

- Hai biến ngẫu nhiên gọi là độc lập với nhau nếu quy luật phân phối xác suất của biến ngẫu nhiên này không phụ thuộc vào biến ngẫu nhiên kia nhận giá trị bằng bao nhiêu.

"Hai BNN gọi là độc lập với nhau nếu mọi biến cố liên quan đến X độc lập với biến cố bất kỳ liên quan đến Y. Ví dụ chọn ngẫu nhiên 1 người trong TP. Gọi X là chiều cao. Y là ngày sinh (không tính năm tháng), Z là cân nặng. Y và Z độc lập. X và Z không độc lập vì không thể có người cao 1,8 mà cân nặng 10 kg".

- Các biến ngẫu nhiên gọi là độc lập lẫn nhau nếu quy luật phân phối xác suất của một số bất kỳ các biến ngẫu nhiên nào đó không phụ thuộc vào việc các biến ngẫu nhiên còn lại nhận giá trị bằng bao nhiêu.

2.2.4. Phân phối xác suất của hàm của biến ngẫu nhiên

Cho hàm $\varphi(x)$ và biến ngẫu nhiên *X*. Khi đó phân phối xác suất của $Y = \varphi(X)$ được xác định như sau:

Trường hợp *X* là biến ngẫu nhiên rời rạc

Từ tập giá trị của X tìm tập giá trị của Y. Xác suất tương ứng để Y nhận giá trị y là $P(Y=y)=\sum_{\varphi(x_i)=y} p_i$. Từ đó ta có bảng phân phối xác suất của Y.

Ví dụ

Cho X có bảng phân phối xác suất

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

Tìm bảng phân phối xác suất của $Y=\varphi(X)=X^2 \quad \left(\varphi(x)=x^2\right)$

Giải

Ta có

X	-1	0	1	2
Y	1	0	1	4

Y nhận các giá trị 0,1,4.

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

$P(Y=1)=P\left((X=-1)+(X=1)\right)=P(X=-1)+P(X=1)=0,5$

$P(Y=4)=P(X=2)=0,2$

Vậy

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

Trường hợp X là biến ngẫu nhiên liên tục

-Tìm hàm phân phối của Y

$F_Y(x)=P(Y\le x)=P(\varphi(X)\le x)=\int_A f_X(u)du$ với $A=\{u:\varphi(u)\le x\}$

- Lấy đạo hàm của $F_Y(x)$ ta có $f_Y(x)$.

Ví dụ

Cho X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f_X(x)$.Tìm hàm mật độ xác suất của $Y=\varphi(X)=2X+3$.

Giải

$F(x)=P(X\le x)\Rightarrow F(\varphi(x))=P(X\le\varphi(x))$

$F_Y(x)=P(Y\le x)=P(2X+3\le x)=P(X\le\frac{x-3}{2})=F_X(\frac{x-3}{2})$

Từ đó $f_Y(x)=F_Y'(x)=\left(F_X(\frac{x-3}{2})\right)'=\frac{1}{2}f_X(\frac{x-3}{2})$

2.3. Các tham số của biến ngẫu nhiên

2.3.1. Kì vọng, phương sai, độ lệch chuẩn

Kì vọng

Kì vọng của biến ngẫu nhiên *X*, kí hiệu *E*(*X*) được xác định như sau:

Nếu *X* là biến ngẫu nhiên rời rạc có bảng phân phối xác suất

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

thì $E(X)=\sum_{i=1}^{n} x_i p_i$

Nếu *X* là biến ngẫu nhiên liên tục có hàm mật độ xác suất *f(x)*

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

Tính chất:

- $E(C)=C$,C là hằng số
- $E(CX)=C.E(X)$,C là hằng số
- $E(X\pm Y)=E(X)\pm E(Y)$
- *Nếu X, Y là hai biến ngẫu nhiên độc lập thì* $E(X.Y)=E(X).E(Y)$

Cho hàm số $y=\varphi(x)$ và xét biến ngẫu nhiên $Y=\varphi(X)$. Khi đó

$E(Y)=\sum_{i=1}^{n}\varphi(x_i)\,p_i$ nếu *X* là biến ngẫu nhiên rời rạc có $P(X=x_i)=p_i$

$E(Y)=\int_{-\infty}^{+\infty}\varphi(x)\,f(x)dx$ nếu *X* là biến ngẫu nhiên liên tục có hàm mật độ xác suất *f*(*x*)

Ví dụ 1. Chọn ngẫu nhiên 3 bi từ một túi có 4 bi trắng 6 bi đen. Gọi *X* là số bi trắng có trong 3 bi. Tìm kì vọng của *X*?

Giải

Gọi *X* là số bi trắng có trong 3 bi. Bảng phân phối xác suất của *X* là:

X	**0**	**1**	**2**	**3**
P	$\frac{5}{30}$	$\frac{15}{30}$	$\frac{9}{30}$	$\frac{1}{30}$

Kì vọng của biến ngẫu nhiên *X* là :

$$E(X)=\frac{5}{30}\times 0+\frac{15}{30}\times 1+\frac{9}{30}\times 2+\frac{1}{30}\times 3=\frac{6}{5}$$

Ví dụ 2. Cho *X* là đại lượng ngẫu nhiên có hàm mật độ

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

Tìm kì vọng của *X*?

Giải:

Ta có $E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{0}^{1}2x^2dx=\frac{2}{3}$

Ý nghĩa của kì vọng

"Tiến hành *n* phép thử. Giả sử *X* nhận các giá trị có thể $x_1,x_2,...,x_n$ với tần số tương ứng $k_1,k_2,...,k_n\left(k_1+k_2+.....+k_n=n\right)$. Ta có giá trị trung bình mà biến *X* nhận được qua *n* phép thử là:

$\overline{x}=\frac{k_1x_1+k_2x_2+.....+k_nx_n}{n}=\frac{k_1}{n}x_1+\frac{k_2}{n}x_2+......+\frac{k_n}{n}x_n\approx f_1x_1+f_2x_2+.......+f_nx_n$ Với $f_i=\frac{k_i}{n}$ là tần suất để biến *X* nhận giá trị x_i $P(X=x_i)=p_i$

Ta thấy rằng khi *n* đủ lớn thì theo định nghĩa thống kê về xác suất ta có thể xem tần suất như xác suất của biến cố. Vậy khi *n* đủ lớn ta có:

$\overline{x}\approx\sum_{i=1}^{n}p_ix_i=E(X)$"

Kì vọng của BNN là giá trị trung bình mà biến ngẫu nhiên nhận.

"Nó phản ánh giá trị trung tâm của phân phối xác suất. Khái niệm kỳ vọng được áp dụng rộng rãi trong nhiều lĩnh vực kinh doanh và quản lí. Chẳng hạn nếu cần chọn phương án cho lợi nhuận cao ta chọn Phuong án cho lợi nhuận kỳ vọng cao "

Ví dụ 4. Tuổi thọ của một loại côn trùng nào đó là biến ngẫu nhiên *X* (đơn vị: tháng) với hàm mật độ như sau:

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

Tìm hàm phân phối và tuổi thọ trung bình của loại côn trùng trên?

HD:

Do ta có: $\int_0^4 kx^2(4-x)dx = \frac{64k}{3} = 1 \Rightarrow k = \frac{3}{64}$. Hàm phân phối xác suất:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0 & ,x \le 0 \\ \frac{3x^3}{64}\left(\frac{4}{3} - \frac{x}{4}\right) & ,0 < x \le 4 \\ 1 & ,x > 4 \end{cases}$$

Tuổi thọ trung bình $E(X) = \frac{3}{64}\int_0^4 x^3(4-x)\,dx = \frac{12}{5}$ (tháng).

Ví dụ 6: Biến ngẫu nhiên liên tục *X* có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} \sin x & x \in \left(0, \frac{\pi}{2}\right) \\ 0 & x \notin \left(0, \frac{\pi}{2}\right) \end{cases}$$

Tìm kỳ vọng toán của $Y = \varphi(X) = X^2$.

Ta có $E(Y) = E(\varphi(X)) = E(X^2) = \int_{-\infty}^{+\infty} \varphi(x) f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_0^{\pi/2} x^2 \sin x dx = \pi - 2$

Phương sai

Định nghĩa: Phương sai của biến ngẫu nhiên *X* kí hiệu *V*(*X*) được xác định như sau:

$$\begin{aligned} V(X) &= E\left((X - E(X))^2\right) \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

Trong đó

$E(X^2) = \sum_i x_i^2 . p_i$ nếu *X* là biến ngẫu nhiên rời rạc có $P(X = x_i) = p_i$

$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx$ nếu *X* là biến ngẫu nhiên liên tục có hàm mật độ *f*(*x*)

Nếu *X* rời rạc nhận các giá trị với các xác suất tương ứng $p_i = P(X = x_i)$ thì **:**

$$V(X) = E(X^2) - [E(X)]^2 = \sum_i x_i^2 . p_i - [E(X)]^2$$

Nếu *X* liên tục có hàm mật độ *f*(*x*) thì:

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

Tính chất: Cho 2 biến ngẫu nhiên *X*, *Y* và hằng số *C*

$i)\ V(C)=0$

$ii)\ V(CX)=C^2V(X)$

iii) Nếu *X*, *Y* là hai biến ngẫu nhiên độc lập thì

$$V(X\pm Y)=V(X)+V(Y)$$

$iv)\ V(C+X)=V(X)$

Ví dụ 1. Tính phương sai của biến ngẫu nhiên trong Ví dụ 1 phần

$$E(X^2)=\frac{5}{30}\times 0^2+\frac{15}{30}\times 1^2+\frac{9}{30}\times 2^2+\frac{1}{30}\times 3^2=2$$

$$V(X)=E(X^2)-(E(X))^2=2^2-\left(\frac{6}{5}\right)^2=\frac{64}{25}$$

Ví dụ 2. Cho *X* có hàm mật độ xác suất

$$f(x)=\begin{cases}3x^2 & ,khi\ x\in[0,1]\\ 0 & ,khi\ x\notin[0,1]\end{cases}$$

Tìm kì vọng và phương sai của biến ngẫu nhiên đã cho.

Giải:

Ý nghĩa của phương sai

Phương sai phản ánh mức độ phân tán của các giá trị của biến ngẫu nhiên xung quanh giá trị trung bình của nó. Nếu *V*(*X*) nhỏ thì ta nói các giá trị của *X* tập trung quanh *E*(*X*); nếu *V*(*X*) lớn thì ta nói các giá trị của *X* phân tán ra xa *E*(*X*). "Trong công nghiệp, phương sai biểu thị độ chính xác của sản xuất, trong chăn nuôi nó biểu thị độ đồng đều của các con gia súc, trong trồng trọt nó biểu thị mức độ ổn định của năng suất, Trong quản lí và kinh doanh phương sai đặc trưng cho mức độ rủi ro của các quyết định".

Độ lệch chuẩn

Định nghĩa. Độ lệch chuẩn của biến ngẫu nhiên *X* kí hiệu là $\sigma(X)$, được định nghĩa như sau: $\sigma(X)=\sqrt{V(X)}$

2.3.2. Mode, trung vị, hệ số biến thiên

Mode

Với biến ngẫu nhiên rời rạc : Mode của X, kí hiệu ModX, là giá trị của X ứng với xác suất lớn nhất.

Ta có $\text{ModX} = x_{i_0} \Leftrightarrow p_{i_0} = max\{p_1, p_2, ...\}$

Với biến ngẫu nhiên liên tục có hàm mật độ *f(x)*thì ModX là giá trị mà tại đó f(x) đạt cực đại

Ví dụ 1. Cho biến ngẫu nhiên X có bảng phân phối xác suất như sau:

X	1	2	3	4	5	6	7	8	9	10
P	0,01	0,2	0,15	0,07	0,35	0,02	0,05	0,1	0,02	0,03

Ta thấy p_5 lớn nhất nên ModX=5

Ví dụ 2. Tung một đồng xu S, N 3 lần. Gọi X là số lần được mặt sấp.

X	0	1	2	3
P	1/8	3/8	3/8	1/8

Theo bảng phân phối ta có: ModX=1 hoặc ModX=2

Trung vị (Median)

Trung vị của biến ngẫu nhiên X, kí hiệu MedX là một số thỏa

$$\begin{cases} P\left(X < MedX\right) \le \dfrac{1}{2} \\ P\left(X \le MedX\right) \ge \dfrac{1}{2} \end{cases} \text{hay} \begin{cases} P\left(X < MedX\right) \le \dfrac{1}{2} \\ P\left(X > MedX\right) \le \dfrac{1}{2} \end{cases}$$

Nếu X là biến ngẫu nhiên liên tục thì $MedX = x_0 \;\, \textit{nếu} \;\, F(x_0) = \int_{-\infty}^{x_0} f(x)dx = \dfrac{1}{2}$

Ví dụ 1. Giả sử biến ngẫu nhiên rời rạc X có bảng phân phối xác suất như sau:

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

Tìm ModX và MedX?

Dễ thấy ModX=4

Hàm phân phối xác suất của X

$$F\left(x\right) = \begin{cases} 0 & khi\, x \le 1 \\ 0,1 & khi\, 1 < x \le 2 \\ 0,3 & khi\, 2 < x \le 3 \\ 0,6 & khi\, 3 < x \le 4 \\ 1 & khi\, 4 < x \end{cases}$$

Từ đó ta có MedX =3.

Ví dụ 2. Giả sử biến ngẫu nhiên rời rạc X có bảng phân phối xác suất như sau:

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

Dễ thấy ModX=2

Hàm phân phối xác suất của X

$$F(x)=\begin{cases}0 & khi\, x\le 1\\ 0,1 & khi\, 1<x\le 2\\ 0,5 & khi\, 2<x\le 3\\ 0,8 & khi\, 3<x\le 4\\ 1 & khi\, 4<x\end{cases}$$

Trung vị của biến ngẫu nhiên X là $\mathrm{MedX}=\mathrm{m}, \forall 2\le m\le 3$

Ví dụ 3. Giả sử biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f(x)=\begin{cases}4x^3 & khi\, x\in(0,1)\\ 0 & khi\, x\notin(0,1)\end{cases}$$

Hàm phân phối xác suất

$$F(x)=\begin{cases}0 & khi\, x\le 0\\ x^4 & khi\, 0<x\le 1\\ 1 & khi\, 1<x\end{cases}$$

Đặt $MedX=m$.Ta có : $F(m)=\frac{1}{2}$

Do đó ta có: $MedX=\frac{1}{\sqrt[4]{2}}$

Moment, hệ số biến thiên, hệ số bất đối xứng, hệ số nhọn

- Moment gốc cấp k $\quad m_k=E\left(X^k\right), k=1,2,...$
- Moment quy tâm cấp k $\quad \mu_k=E\left(\left[X-E(X)\right]^k\right), k=1,2,...$
- Hệ số biến thiên $\quad CV=\left|\frac{\sigma(X)}{E(X)}\right|.100\%$

  Dùng để so sánh mức độ phân tán của các biến ngẫu nhiên có kì vọng và phương sai khác nhau.
- Hệ số bất đối xứng $\quad \alpha_3=\frac{\mu_3}{\sigma^3},\ \sigma=\sqrt{D(X)}$
- Hệ số nhọn $\quad \alpha_4=\frac{\mu_4}{\sigma^4}$

Nhận xét:

+ $m_1 = E(X)$, $\mu_1 = 0$, $\mu_k = D(X)$

+ α_3 đo mức độ bất đối xứng của luật phân phối.

Nếu $\alpha_3 < 0$ thì phân phối xác suất và đồ thị của hàm mật độ sẽ lệch về bên trái.

$\alpha_3 = 0$ thì phân phối xác suất và đồ thị của hàm mật độ đối xứng.

$\alpha_3 > 0$ thì phân phối xác suất và đồ thị của hàm mật độ sẽ lệch về bên phải.

+ Hệ số nhọn α_4 dùng để xét độ tập trung của phân phối của biến ngẫu nhiên

$\alpha_4 < 3$: ít tập trung

$\alpha_4 = 3$: tập trung ở mức độ thường

$\alpha_4 > 3$: rất tập trung

Khi phân phối của biến ngẫu nhiên X đối xứng hoặc gần đối xứng thì ta dùng kì vọng để định vị là tốt nhất, song nếu phân phối của X quá lệch thì nên dùng Median và Mod để định vị.

Ví dụ . Tung một đồng xu 2 lần xấp ngửa độc lập. Gọi X là số lần được mặt sấp. Tính hệ số bất đối xứng, hệ số nhọn

Giải:

X	0	1	2
P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Ta tính được $E(X) = 1$ $Var(X) = \frac{1}{2}$

$$\mu_3 = E(X-1)^3 = (0-1)^3 0,25 + (1-1)^3 0,5 + (2-1)^3 0,25 = 0$$

$$\sigma = \sqrt{V(X)} = \sqrt{\frac{1}{2}}$$

Ta có: $\alpha_3 = 0$ nên phân phối đối xứng qua giá trị $E(X) = 1$

$$\mu_4 = E(X-1)^4 = (0-1)^4 0,25 + (1-1)^4 0,5 + (2-1)^4 0,25 = \frac{1}{2}$$

$$\alpha_4 = \frac{t_4}{\sigma^4} = \frac{1/2}{1/4} = 2 < 3$$

CHƯƠNG 3. MỘT SỐ QUY LUẬT PHÂN PHỐI XÁC SUẤT THÔNG DỤNG

3.1. Phân phối xác suất của biến ngẫu nhiên rời rạc

3.1.1. Phân phối nhị thức $B(n, p)$

Định nghĩa

Biến ngẫu nhiên *X* được gọi là có phân phối nhị thức tham số *n, p*. Kí hiệu $X \sim B(n, p)$ nếu *X* nhận một trong các giá trị 0, 1, 2,…, *n* với xác suất tương ứng như sau:

$$P(X = k) = C_n^k p^k (1-p)^{n-k} \quad ,k = 0,1,...,n$$

trong đó *p* là một hằng số, $0 < p < 1$, $n \in N, n > 0$

Xét dãy *n* phép thử Bernouli với xác suất xảy ra biến cố *A* trong mỗi phép thử là *p*. Gọi *X* là số lần xuất hiện biến cố *A* trong *n* phép thử Bernoulli trên. Khi đó $X \sim B(n, p)$

Tính chất: Cho $X \sim B(n, p)$.

- $E(X) = np$
- $V(X) = np(1-p)$
- $np - q \le ModX \le np - q + 1$ với $q = 1 - p$

Ví dụ 1. Tung một con xúc sắc cân đối, đồng chất 10 lần. Gọi *X* là số lần xuất hiện mặt số 5. Hãy lập bảng phân phối xác suất của biến ngẫu nhiên *X*?

HD

Ta có xác suất xuất hiện mặt số 5 khi gieo con xúc sắc là $\frac{1}{6}$

X là số lần xuất hiện mặt số 5. Ta có $X \sim B\left(10, \frac{1}{6}\right)$

Ví dụ 2. Một bài thi trắc nghiệm gồm 50 câu hỏi, mỗi câu hỏi có 4 phương án trả lời, trong đó chỉ có một phương án đúng. Một học sinh làm bài bằng cách chọn ngẫu nhiên một phương án trả lời cho mỗi câu hỏi.

a) Tính xác suất để học sinh này làm được đúng 25 câu.

b) Trung bình học sinh này trả lời được bao nhiêu câu?

c) Tính phương sai của số câu học sinh này trả lời được.

Giải:

Xác suất làm đúng một câu là: $\frac{1}{4}$

Gọi X là số câu học sinh này làm đúng.

Theo đề bài ta có $X \sim B\left(50;\frac{1}{4}\right)$

a)Ta có $P\left(X=25\right)=P_{50}\left(25;\frac{1}{4}\right)=C_{50}^{25}\left(\frac{1}{4}\right)^{25}\left(\frac{3}{4}\right)^{25}$

b) Số câu học sinh này trả lời được trung bình chính là kì vọng của biến ngẫu nhiên X. Do đó số câu học sinh này trả lời được trung bình là:

$$E\left(X\right)=np=50.\frac{1}{4}=12,5$$

c) Phương sai số câu học sinh trả lời được là:

$$V\left(X\right)=npq=50.\frac{1}{4}.\frac{3}{4}=\frac{150}{16}=9,375$$

Ví dụ 3. Xác suất để một bệnh nhân sống sót sau khi mắc một loại bệnh hiếm thấy về máu là 0,4. Nếu biết rằng đã có 12 người mắc loại bệnh này, tìm xác suất để

a) có ít nhất 10 người sống sót;

b) có từ 3 đến 8 người sống sót;

c) có đúng 5 người sống sót.

Đáp số

a) $P\ \ X\geq 10\ \ =P_{12}\ \ 10;0,4\ \ +P_{12}\ \ 11;0,4\ \ +P_{12}\ \ 12;0,4\ \ =...$

b) $P\ \ 3\leq X\leq 8\ \ =P_{12}\ \ 3;0,4\ \ +P_{12}\ \ 4;0,4\ \ +P_{12}\ \ 5;0,4\ \ +P_{12}\ \ 6;0,4\ \ +P_{12}\ \ 7;0,4\ \ +P_{12}\ \ 8;0,4\ \ =...$

c) $P\ \ X=5\ \ =P_{12}\ \ 5;0,4\ \ =...$

Ví dụ 4. Một bưu cục nhận 10 loại báo khác nhau, xác suất bán hết báo hàng ngày cho mỗi loại là 0,8. Vậy nếu trong một năm với 300 ngày mở cửa thì có khoảng bao nhiêu ngày bưu điện không bán hết báo?

Giải:

Gọi X là số loại báo bán hết trong ngày thì $X\sim B\ \ 10;0,8$. Vậy xác suất để một ngày bán không hết báo là:

$$P=P\ \ X<10\ \ =1-P\ \ X=10\ \ =1-C_{10}^{10}\ \ 0,8\ ^{10}=0,8926$$

Tương tự trong một năm với 300 ngày bán báo tương ứng với 300 phép thử Bernoulli mà kết quả của mỗi lần thử là ngày bán không hết báo. Gọi Y là số ngày trong năm không bán hết báo thì $Y \sim B(300;0,8926)$. Vậy số ngày trong năm mà bưu cục không bán hết báo là kì vọng của biến ngẫu nhiên Y.

Ta có: $E(Y) = np = 300 \times 0,8926 = 267,78$ ngày.

3.1.2. Phân phối không - một A(p)

Biễn ngẫu nhiên X gọi là có phân phối không - một tham số p (hay phân phối Bernoulli), kí hiệu $X \sim A(p)$ nếu X chỉ nhận 2 giá trị 0 và 1 theo bảng phân phối xác suất như sau:

X	0	1
P	q	p

trong đó p là hằng số $0 < p < 1$, q=1-p

Phân phối A(p) là trường hợp đặc biệt của phân phối B(n,p) khi n=1

Giả sử ta tiến hành một phép thử trong đó xác suất xuất hiện biến cố A bằng p. Gọi X là số lần biến cố A xuất hiện trong phép thử đó . Ta có $X \sim A(p)$

Tính chất

$$E(X) = p \qquad V(X) = pq \qquad \sigma(X) = \sqrt{pq}$$

3.1.3. Phân phối Poisson $P(\lambda)$

Định nghĩa

Biến ngẫu nhiên *X* gọi là có phân phối Poisson với tham số λ, kí hiệu $X \sim P(\lambda)$ nếu *X* nhận một trong các giá trị 0,1,2,…với xác suất tương ứng được tính bằng công thức

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

Tính chất

- $E(X) = V(X) = \lambda$
- $\lambda - 1 \le ModX \le \lambda$
- Nếu X_1 , X_2 là hai biến ngẫu nhiên độc lập có phân phối Poisson với tham số lần lượt là λ_1, λ_2 thì $X_1 + X_2$ cũng có phân phối Poisson với tham số $(\lambda_1 + \lambda_2)$

$$X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$$

Phân phối Poisson được mô tả lần đầu tiên bởi Simeon Denis Poisson năm 1887. Phân phối Poisson có nhiều ứng dụng liên quan đến *số quan sát một hiện tượng thực tiễn* nào đó trong một đơn vị thời gian hoặc khoảng không gian nhất định. Chẳng hạn số các cuộc điện thoại gọi đến một trạm nào đó trong một phút, số khách hàng đến nhà băng trong mỗi chu kỳ 01 giờ, số máy điện thoại hỏng trong một ngày ở Tp. HCM, số sinh viên vào forum khoa KT-ĐHQG Tp.HCM trong mỗi chu kỳ 30 phút ... Nói chung *dòng vào của một hệ phục vụ* (một quán cà phê, hiệu cắt tóc, tiệm sửa xe, trạm điện thoại thẻ, siêu thị, ...) đều là các biến ngẫu nhiên có phân phối Poison.

Trong Lý thuyết xác suất thống kê Toán người ta chứng minh được rằng:

+ Giả sử biến cố A xuất hiện ở các thời điểm ngẫu nhiên.

+ Số lần xuất hiện A trong các khoảng thời gian rời nhau là độc lập.

+ Số lần xuất hiện trung bình trong một đơn vị thời gian là như nhau và bằng λ.

Khi đó biến ngẫu nhiên X chỉ số lần xuất hiện A trong khoảng thời gian t sẽ là biến ngẫu nhiên có phân phối Poisson với tham số λt.

Ví dụ 1. Biết trung bình trong một ngày có 600 người đến siêu thị.

a. Tính xác suất trong một ngày có 700 người đến siêu thị.

b. Xác định số người chắc chắn nhất có thể đến siêu thị trong 1 ngày?

Giải:

Gọi X là số người đến siêu thị trong 1 ngày. Theo đề ta có: $X \sim P(600)$

a. $P(X = 700) = \dfrac{600^{700}}{700!} e^{-600}$

b. $600 - 1 \le ModX \le 600$. Vậy số người tin chắc nhất là 599 hoặc 600 người.

Ví dụ 2. Ở một tổng đài điện thoại các cuộc điện thoại gọi đến là ngẫu nhiên và độc lập nhau. Tốc độ trung bình là 2 cuộc trên 1 phút. Tìm xác suất để

a. Có đúng 5 cuộc gọi trong vòng 4 phút? (biến cố A)

b. Không có cuộc nào trong khoảng thời gian 30 giây? (biến cố B)

c. Có ít nhất 1 cuộc trong khoảng thời gian 15 giây? (biến cố C)

Giải

Nếu gọi $X(t)$ là số cuộc gọi đến tổng đài trong thời gian t phút

a. $X\ 4 \sim P\ 8 \Rightarrow P\ A = P(X = 5) = \dfrac{e^{-8}8^5}{5!}$

b. $X\left(\frac{1}{2}\right) \sim P(1) \Rightarrow P(B) = P(X=0) = \frac{e^{-1}1^0}{0!} = \frac{1}{e}$

c. $X\left(\frac{1}{4}\right) \sim P\left(\frac{1}{2}\right) \Rightarrow P(C) = P(X \geq 1) = 1 - P(X<1) = 1 - P(X=0) = 1 - \frac{e^{-\frac{1}{2}}\left(\frac{1}{2}\right)^0}{0!} = 1 - \frac{1}{\sqrt{e}}$

3.1.4. Phân phối siêu bội H(N,M,m)

Xét một tập hợp gồm N phần tử, trong đó có M phần tử có tính chất A nào đó. Lấy ngẫu nhiên từ tập hợp ra n phần tử. Gọi X là số phần tử có tính chất A có trong n phần tử lấy ra. Khi đó X có phân phối xác suất được xác định như sau:

$$P(X=k) = \frac{C_M^k . C_{N-M}^{n-k}}{C_N^n}, \quad k = 0,1,2,\ldots,\min(n,M) \quad (*)$$

Biến ngẫu nhiên X có phân phối xác suất được xác định bởi công thức (*) gọi là có phối siêu bội với các tham số là N, M và n. Kí hiệu $X \sim H(N,M,n)$

Ví dụ 1

Một lô hàng có 10 sản phẩm, trong đó có 6 sản phẩm tốt. Lấy ngẫu nhiên từ lô hàng ra 4 sản phẩm. Tìm xác suất để có 3 sản phẩm tốt trong 4 sản phẩm được lấy ra.

Giải

Gọi X là số sản phẩm tốt có trong 4 sản phẩm lấy ra. X phân phối theo qui luật siêu bội với tham số N =10, M = 6, n = 4.

Xác suất có 3 sản phẩm tốt trong 4 sản phẩm lấy ra là

$$P(X=3) = \frac{C_6^3 . C_4^1}{C_{10}^4} = 0,3809$$

Tính chất

Nếu $X \sim H(N,M,n)$ thì

$$E(X) = np; \quad V(X) = npq.\frac{N-n}{N-1}$$

trong đó $p = \frac{M}{N}; q = 1-p$

Số $\frac{N-n}{N-1}$ được gọi là hệ số *hiệu chỉnh*

Chú ý: Khi n rất bé so với N thì phân phối siêu bội thực tế không khác biệt so với phân phối nhị thức vì:

$$\lim_{\frac{n}{N}\to 0}\frac{C_M^x.C_{N-M}^{n-x}}{C_N^n} = C_n^x\left(\frac{M}{N}\right)^x\left(1-\frac{M}{N}\right)^{n-x}$$

$X \sim H(N,M,n)$ Khi N khá lớn, n rất nhỏ so với N thì

$$P(X=k) \approx C_n^k p^k q^{n-k} \ (p=\frac{M}{N}, q=1-p)$$

Công thức xấp xỉ tốt khi $n < 0{,}05N$

Khi n khá bé so với N thì sự khác biệt giữa lấy có hoàn lại và không hoàn lại gần như không khác nhau

Ta có thể xem $X \in B(n,p)$ với p là tỷ lệ phần tử có tính chất A của tập hợp

3.2. Phân phối xác suất của biến ngẫu nhiên liên tục

3.2.1. Phân phối chuẩn $N\left(\mu,\sigma^2\right)$

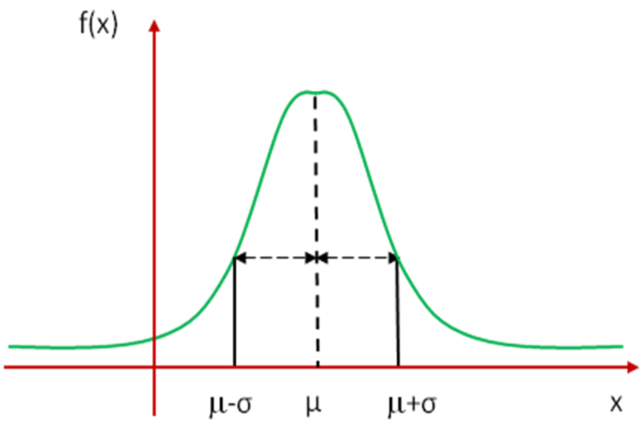
Định nghĩa

Biến ngẫu nhiên X được gọi là có phân phối chuẩn với tham số μ và $\sigma^2(\sigma>0)$, kí hiệu: $X \sim N\left(\mu,\sigma^2\right)$ nếu hàm mật độ có dạng:

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

"Phân phối chuẩn được Gauss tìm ra năm 1809 nên nó còn được gọi là phân phối Gauss".

Đồ thị hàm mật độ của phân phối chuẩn



Đồ thị $f(x)$ có dạng hình chuông, nhận đường thẳng $x=\mu$ làm trục đối xứng, trục hoành làm tiệm cận ngang, các điểm uốn $\left(\mu\pm\sigma;\frac{1}{\sigma\sqrt{2\pi e}}\right)$, điểm cực đại của đồ thị $\left(\mu;\frac{1}{\sigma\sqrt{2\pi}}\right)$

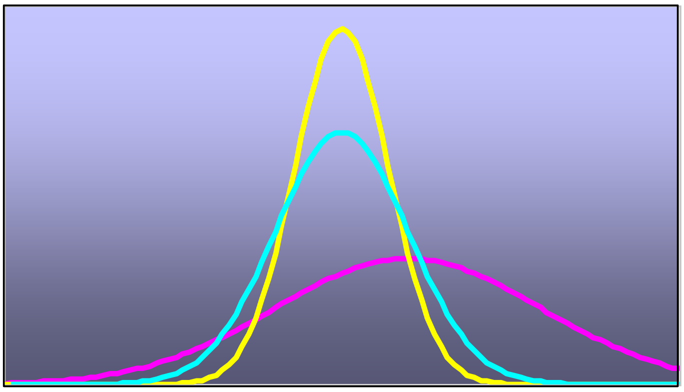
Trung bình = Trung vị = Mode

Vị trí của phân phối được xác định bởi kỳ vọng, μ

Với các giá trị μ,σ khác nhau ta có các dạng phân phối khác nhau:

Khi σ không đổi, μ tăng (giảm) thì đồ thị tịnh tiến sang phải (trái);

Khi μ không đổi, σ tăng thì đồ thị thấp xuống, phình ra; σ giảm thì đồ thị cao lên và nhọn ra



Hàm phân phối của một biến ngẫu nhiên có phân phối chuẩn

$$F(x) = \int_{-\infty}^{x} f(t)dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \left(e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} \right) dt$$

Tính chất

- $E(X) = \mu$
- $V(X) = \sigma^2$
- $ModX = MedX = \mu$
- Nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân phối chuẩn $X_1 \sim N(\mu_1, \sigma_1^2)$; $X_2 \sim N(\mu_2, \sigma_2^2)$ thì

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Tổng quát: Tổng của *n* biến ngẫu nhiên độc lập có phân phối chuẩn vẫn là một biến ngẫu nhiên có phân phối chuẩn, nghĩa là nếu $X_i \sim N(\mu_i, \sigma_i^2)$, các X_i độc lập $i = 1, 2, \ldots, n$ thì

$$T = X_1 + X_2 + \ldots + X_n \sim N(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$$

Nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân phối chuẩn $X_1 \sim N(\mu_1, \sigma_1^2)$; $X_2 \sim N(\mu_2, \sigma_2^2)$ thì một tổ hợp tuyến tính bất kì của hai biến ngẫu nhiên này cũng có phân phối chuẩn. $aX_1 + bX_2 \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$ $(ab \neq 0)$

Phân phối chuẩn tắc

Biến ngẫu nhiên *X* được gọi là có phân phối chuẩn tắc nếu $X \sim N(0,1)$. Khi đó hàm mật độ của X :

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

gọi là hàm mật độ Gauss.

Hàm phân phối $F(x)=\int_{-\infty}^{x}\varphi(t)dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$, gọi là hàm phân phối Gauss

Hàm mật độ Gauss là hàm chẵn, $\int_{-\infty}^{0}\varphi(x)dx=\int_{0}^{+\infty}\varphi(x)dx=0,5$

Tích phân Laplace: $\phi(x)=\int_{0}^{x}\varphi(t)dt=\frac{1}{\sqrt{2\pi}}\int_{0}^{x}e^{-\frac{t^2}{2}}dt$

Tính chất

“ $\phi(-x)=-\phi(x)$

$x>5:\phi(x)\approx 0,5$

$F(x)=0,5+\phi(x)$

+ $P(a<X<b)=F(b)-F(a)=\phi(b)-\phi(a)$, $X\sim N(0,1)$

+ $P(X>a)=0,5-\phi(a)=1-F(a)$

+ $P(X<b)=0,5+\phi(b)=F(b)$

Chú ý

Nếu $X\sim N(\mu,\sigma^2)$ thì biến ngẫu nhiên $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$

Nếu $X\sim N(\mu,\sigma^2)$ thì

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

$$P(X>a)=0,5-\phi\left(\frac{a-\mu}{\sigma}\right)=1-F\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X<b)=0,5+\phi\left(\frac{b-\mu}{\sigma}\right)=F\left(\frac{b-\mu}{\sigma}\right)$$

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

$$Z=\frac{X-\mu}{\sigma}\Rightarrow Z\sim N(0,1)$$

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

Qui tắc 2σ và qui tắc 3σ:

Từ (*) , nếu $X \sim N(\mu;\sigma^2)$ thì

$$P\left(|X-\mu|<\varepsilon\right)=2\phi\left(\frac{\varepsilon}{\sigma}\right) \quad (**)$$

$$P\left(|X-\mu|<\varepsilon\right)=P\left(-\varepsilon<X-\mu<\varepsilon\right)=P\left(\mu-\varepsilon<X<\mu+\varepsilon\right)$$
$$=\phi\left(\frac{\mu+\varepsilon-\mu}{\sigma}\right)-\phi\left(\frac{\mu-\varepsilon-\mu}{\sigma}\right)=\phi\left(\frac{\varepsilon}{\sigma}\right)-\phi\left(-\frac{\varepsilon}{\sigma}\right)=\phi\left(\frac{\varepsilon}{\sigma}\right)+\phi\left(\frac{\varepsilon}{\sigma}\right)=2\phi\left(\frac{\varepsilon}{\sigma}\right)$$

Với $\varepsilon=2\sigma$ thì $P\left(|X-\mu|<2\sigma\right)=2\phi(2)=0,9544$ hay $P\left(\mu-2\sigma<X<\mu+2\sigma\right)=0,9544$

Tương tự với $\varepsilon=3\sigma$ ta được $P\left(\mu-3\sigma<X<\mu+3\sigma\right)=0,9973$

Nếu X có phân bố chuẩn $X \sim N(\mu;\sigma^2)$ thì có đến 95,44% giá trị của X nằm trong khoảng (μ-2σ; μ+2σ) và hầu như toàn bộ giá trị của X nằm trong khoảng (μ-3σ; μ+3σ). Hai quy tắc này gọi là quy tắc 2σ và 3σ .

« Trong thực tế quy tắc tắc 2σ và qui tắc 3σ được áp dụng như sau: Nếu quy luật phân phối xác suất của biến ngẫu nhiên được nghiên cứu chưa biết, xong nó thỏa mãn quy tắc 2σ hoặc qui tắc 3σ thì có thể xem như biến ngẫu nhiên đó có phân phối chuẩn »

Giá trị tới hạn chuẩn

Giá trị tới hạn chuẩn mức α , kí hiệu z_α là giá trị xác định bởi

$$P\left(Z>z_\alpha\right)=\alpha \text{ với } Z \sim N(0,1).$$

Ta có $z_\alpha=-z_{1-\alpha}$

$P\left(Z>z_\alpha\right)=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{z_\alpha}^{+\infty}e^{-\frac{u^2}{2}}du=\alpha$. Do đó $\phi(z_\alpha)=0,5-\alpha$ $\left(F(z_\alpha)=1-\alpha\right)$ nên cho trước α người ta tính được z_α và ngược lại.

Bảng tính các giá trị tới hạn z_α được tính trong bảng phụ lục

Ví dụ 1. Giả sử X ~ $N\left(\mu;\sigma^2\right)$ với μ = 2100 và σ = 200. Hãy tìm :

$a)\mathrm{P}\left(\mathrm{X}>2400\right)$

$b)\mathrm{P}\left(1700<\mathrm{X}<2200\right)$

$c)$Tìm a sao cho $\mathrm{P}\left(\mathrm{X}>\mathrm{a}\right)=0,03$

Giải :

a) $P(X > 2400) = 0,5 - \phi\left(\dfrac{2400-2100}{200}\right) = 0,5 - \phi(1,5) = 0,5 - 0,43319 = 0,06681$

$P(X > 2400) = 1 - F\left(\dfrac{2400-2100}{200}\right) = 1 - F(1,5) = 1 - 0,93319 = 0,06681$

b) $P(1700 < X < 2200) = \phi\left(\dfrac{2200-2100}{200}\right) - \phi\left(\dfrac{1700-2100}{200}\right) = \phi(0,5) - \phi(-2)$

$= 0,19146 + 0,47725 = 0,66871$

$P(1700 < X < 2200) = F\left(\dfrac{2200-2100}{200}\right) - F\left(\dfrac{1700-2100}{200}\right) = F(0,5) - F(-2)$

$= 0,69146 - 0,02275 = 0,66871$

c)Ta có

$P(X > a) = 0,5 - \phi\left(\dfrac{a-2100}{200}\right) = 0,03 \Rightarrow \phi\left(\dfrac{a-2100}{200}\right) = 0,47 = \phi(1,88) \Rightarrow \dfrac{a-2100}{200} \approx 1,88$

$P(X > a) = 1 - F\left(\dfrac{a-2100}{200}\right) = 0,03 \Rightarrow F\left(\dfrac{a-2100}{200}\right) = 0,97 = F(1,88) \Rightarrow \dfrac{a-2100}{200} \approx 1,88$

Do $\phi(x)$ đồng biến nên $\Rightarrow \dfrac{a-2100}{200} \approx 1,88 \Rightarrow a = 2476$

Ví dụ 2. Chiều dài của một loại chi tiết máy có phân phối chuẩn với $\mu = 30cm$, độ lệch chuẩn $\sigma = 2cm$.

a. Một chi tiết máy sản xuất ra được xem là đạt yêu cầu khi có chiều dài nằm trong khoảng từ 28 cm đến 31 cm. Chọn ngẫu nhiên 1 chi tiết máy, tính xác suất chọn được chi tiết máy đạt yêu cầu.
b. Một chi tiết máy được xem là quá dài khi chiều dài của nó lớn hơn 34,5 cm. Chọn ngẫu nhiên 1 chi tiết máy, tính xác suất chi tiết máy này quá dài?
c. Một chi tiết máy được xem là quá ngắn nếu chiều dài của nó nhỏ hơn 20cm. Chọn ngẫu nhiên 1 chi tiết máy, tính xác suất chi tiết máy này quá ngắn?

Giải:

Gọi X là chiều dài chi tiết máy sản xuất ra (đơn vị:cm)

Theo đề bài ta có: $X \sim N(30, 2^2)$

a. $P(28 < X < 31) = \phi\left(\dfrac{31-30}{2}\right) - \phi\left(\dfrac{28-30}{2}\right) = \phi(0,5) - \phi(-1)$

$= \phi(0,5) + \phi(1) = 0,19146 + 0,34134 = 0,5328$

b. $P(34,5 < X) = 0,5 - \phi\left(\frac{34,5-30}{2}\right) = 0,5 - \phi(2,25) = 0,5 - 0,48778 = 1 - 0,9878 = 0,01222$

c. $P(X < 20) = 0,5 + \phi\left(\frac{20-30}{2}\right) = 0,5 + \phi(-5) = 0,5 - \phi(5) \approx 0$

Ví dụ 3. Các quả bóng do một nhà máy sản xuất ra được coi là đạt tiêu chuẩn nếu đường kính của nó sai lệch so với đường kính thiết kế không quá 0,2 cm. Biết rằng độ sai lệch này là biến ngẫu nhiên phân phối theo qui luật chuẩn với $\mu = 0$ cmvà $\sigma = 0,1$ cm.Tìm tỉ lệ bóng đạt tiêu chuẩn do nhà máy sản xuất ra.

Giải:

Gọi X là độ sai lệch đường kính của bóng do máy sản xuất ra và bóng mẫu.

Theo đề bài ta có: $X \sim N\left(0;(0,1cm)^2\right)$

Tỉ lệ bóng đạt tiêu chuẩn do nhà máy sản xuất ra cũng chính là xác suất chọn được bóng đạt tiêu chuẩn khi chọn ngẫu nhiên 1 bóng do nhà máy sản xuất ra.

Ta có: $P(|X| < 0,2) = P\left(-2 < \frac{X}{0,1} < 2\right)$

Vậy tỉ lệ bóng đạt tiêu chuẩn do nhà máy sản xuất ra là 95,46%

3.2.2. Phân phối $\chi^2(n)$

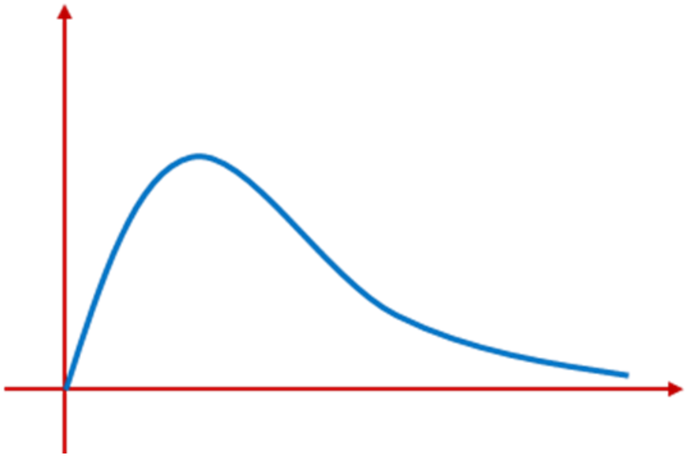
Định nghĩa: Biến ngẫu nhiên X gọi là có phân phối $\chi^2(n)$ (khi bình phương với n bậc tự do), kí hiệu $X \sim \chi^2(n)$ nếu hàm mật độ xác suất của nó có dạng:

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

Trong đó $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ (hàm Gamma).

« Phân phối Khi bình phương do Karl Pearson đưa ra vào năm 1990. »

Đồ thị hàm mật độ của của quy luật "Khi bình phương"



Tính chất: Cho biến ngẫu nhiên $X \sim \chi^2(n)$

- $E(X) = n$
- $V(X) = 2n$
- Nếu $X_1 \sim \chi^2(n), X_2 \sim \chi^2(m)$, X_1, X_2 độc lập thì $X_1 + X_2 \sim \chi^2 \; n+m$
- Nếu $X_i \, (i = 1, 2, ..., n)$ là các biến ngẫu nhiên độc lập có phối chuẩn tắc $N(0,1)$ thì biến ngẫu nhiên $X = \sum_{i=1}^{n} X_i^2$ có phân phối khi bình phương với n bậc tự do.
- Khi n đủ lớn thì $\frac{X-n}{\sqrt{2n}}$ có phân phối xấp xỉ phân phối N(0,1) với $X \sim \chi^2(n)$ (công thức dùng tốt khi n>30)

Giá trị tới hạn χ^2

Giá trị tới hạn mức α của phân phối khi bình phương với n bậc tự do, kí hiệu $\chi^2_\alpha(n)$ là giá trị xác định bởi

$$P\left(X > \chi^2_\alpha(n)\right) = \alpha \text{ trong đó } X \sim \chi^2(n)$$

Bảng các giá trị tới hạn $\chi^2_\alpha(n)$ được tính sẵn ở bảng phụ lục

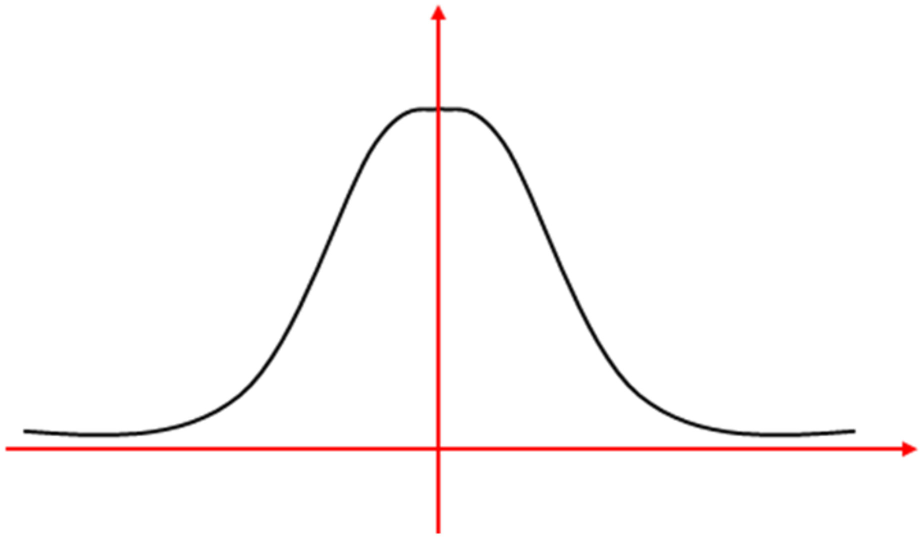
3.2.3. Phân phối student $T(n)$

Biến ngẫu nhiên X gọi là có phân phối Student với n bậc tự do, kí hiệu $X \sim T(n)$ nếu hàm mật độ của X có dạng:

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

Trong đó $\Gamma(x)$ là hàm Gamma.

Đồ thị hàm mật độ của phân phối Student như sau



Tính chất

Cho $X \sim T(n)$.Khi đó

$E(X) = 0$ (n>1)

$V(X) = \dfrac{n}{n-2}$ (n>2)

Nếu $X \sim N(0,1)$, $Y \sim \chi^2(n)$, X và Y độc lập thì biến ngẫu nhiên $T = \dfrac{X}{\sqrt{Y/n}} \sim T(n)$

Hàm mật độ là hàm số chẵn nên đồ thị đối xứng qua trục tung. Khi số bậc tự do n tăng lên phân phối Student hội tụ rất nhanh về phân phối chuẩn tắc N(0,1). Do đó khi n đủ lớn (n≥30) ta có thể dùng phân phối chuẩn tắc thay cho phân phối Student.

Giá trị tới hạn Student

Giá trị tới hạn mức α của phân phối Student n bậc tự do, kí hiệu $t_\alpha(n)$ là giá trị xác định bởi :

$P(X > t_\alpha(n)) = \alpha$ với $X \sim T(n)$

Ta có $t_\alpha(n) = -t_{1-\alpha}(n)$

Bảng tính các giá trị tới hạn $t_\alpha(n)$ được tính trong bảng phụ lục.

3.2.4. Phân phối Fisher-Snedecor $F(n_1, n_2)$

Biến ngẫu nhiên X gọi là có phân phối Fisher – Snedecor với n_1 và n_2 bậc tự do, kí hiệu $X \sim F(n_1, n_2)$, nếu hàm mật độ xác suất có dạng:

$$f(x)=\begin{cases}0 & ,x\le 0\\ C\dfrac{x^{\frac{n_1-2}{2}}}{\left(n_2+n_1x\right)^{\frac{n_1+n_2}{2}}} & ,x>0\end{cases}$$

Trong đó $C=\dfrac{\Gamma\left(\dfrac{n_1+n_2}{2}\right)n_1^{\frac{n_1}{2}}n_2^{\frac{n_2}{2}}}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)}$

Nếu $U\sim\chi^2(n_1)$, $V\sim\chi^2(n_2)$ và U, V độc lập thì biến ngẫu nhiên: $F=\dfrac{U/n_1}{V/n_2}\sim F(n_1,n_2)$

Giá trị tới hạn Fisher – Snedecor

Giá trị tới hạn mức α của phân phối Fisher – Snedecor với n_1 và n_2 bậc tự do kí hiệu $f_\alpha(n_1,n_2)$ là giá trị xác định bởi:

$$\mathrm{P}\left(\mathrm{X}>f_\alpha(\mathrm{n}_1,\mathrm{n}_2)\right)=\alpha \text{ với } X\sim F(n_1,n_2)$$

Ta có $f_\alpha(\mathrm{n}_1,\mathrm{n}_2)=\dfrac{1}{f_{1-\alpha}(\mathrm{n}_1,\mathrm{n}_2)}$

Các giá trị tới hạn $f_\alpha(\mathrm{n}_1,\mathrm{n}_2)$ được cho trong bảng phụ lục

3.2.5. Phân phối mũ $\mathrm{E}(\lambda)$

Định nghĩa

Biến ngẫu nhiên X gọi là có phân phối mũ với tham số λ $(\lambda>0)$, ký hiệu $X\sim E(\lambda)$ nếu hàm mật độ xác suất của nó có dạng

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

Nếu X có phân phối theo quy luật mũ thì hàm phân bố xác suất của X là

$$F(x)=\begin{cases}1-e^{-\lambda x} & ,x\ge 0\\ 0 & ,x<0\end{cases}$$

Tính chất

Nếu $X\sim E(\lambda)$ thì

$$E(X)=\frac{1}{\lambda}\ ,\quad V(\mathrm{X})=\frac{1}{\lambda^2}$$

- Phân phối mũ có mặt trong nhiều ứng dụng thực tiễn. Chẳng hạn thời gian phục vụ của hệ phục vụ đám đông là các biến ngẫu nhiên có phân phối mũ (thời gian nói chuyện của các cuộc đàm thoại, thời gian chờ của khách hàng để được phục vụ…). Tuổi thọ của các thiết bị, tuổi thọ của nhiều loài sinh vật… cũng có phân phối mũ.

- Quy luật mũ có nhiều ứng dụng trong nhiều lĩnh vực khác nhau. Người ta chứng minh được rằng thời gian giữa hai lần xuất hiện yêu cầu của một dòng yêu cầu tối giản trong các hệ thống phục vụ công cộng phân phối theo quy luật mũ. Trong các hệ thống kỹ thuật, thời gian làm việc liên tục của máy móc thiết bị giữa hai lần sửa chữa cũng thường phân phối theo quy luật mũ.

Ví dụ

Giả sử tuổi thọ (tính bằng năm) của một mạch điện tử trong máy tính là một đại lượng ngẫu nhiên có phân phối mũ với kỳ vọng là 6,25. Thời gian bảo hành của mạch điện tử này là 5 năm. Hỏi có bao nhiêu phần trăm mạch điện tử bán ra phải thay thế trong thời gian bảo hành

Giải

Gọi X là tuổi thọ của mạch. Khi đó X có phân phối mũ. Ta có

$$\lambda = \frac{1}{E(X)} = \frac{1}{6,25}$$

$$P(X \le 5) = F(5) = 1 - e^{-\lambda.5} = 1 - e^{-0,8} = 0,5506$$

Vậy có khoảng 55% số mạch điện tử bán ra phải thay thế trong thời gian bảo hành

3.2.6. Qui luật phân phối đều $U(a,b)$

Định nghĩa

Biến ngẫu nhiên X được gọi là có phân phối đều trên đoạn $[a;b]$, kí hiệu $X \sim U(a;b)$ nếu hàm mật độ xác suất của X có dạng:

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

Hàm phân bố : $F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{khi } x < a \\ \dfrac{x-a}{b-a} & \text{khi } a \le x < b \\ 1 & \text{khi } b \le x \end{cases}$

Vậy X có khả năng nhận giá trị trong đoạn $[a;b]$ là đều nhau và không nhận các giá trị nằm ngoài $[a;b]$.

Tính chất

Cho biến ngẫu nhiên $X \sim U(a;b)$. Khi đó

$$E(X)=\frac{a+b}{2} \qquad V(X)=\frac{(b-a)^2}{12}$$

$$Med(X)=E(X)=\frac{a+b}{2} \qquad Mod(X)=m, voi\ \ m\in[a;b]$$

Qui luật phân phối đều có nhiều ứng dụng trong toán thống kê như mô phỏng thống kê, đặc biệt là phương pháp phi tham số. Trong một số lí thuyết kết luận thống kê người ta thường xuất phát từ qui tắc sau đây: Nếu ta không biết gì về giá trị của tham số cần ước lượng thì mỗi giá trị có thể có của tham số đó là đồng khả năng. Điều đó dẫn đến việc xem tham số cần ước lượng như là một biến ngẫu nhiên có phân phối đều.

Ví dụ.

Khi thâm nhập vào thị trường mới, doanh nghiệp không thể khẳng định được một cách chắc chắn doanh số hàng tuần có thể đạt được sẽ là bao nhiêu mà chỉ dự kiến được rằng doanh số tối thiểu sẽ là 20 triệu đồng/tuần và tối đa là 40 triệu đồng/tuần. Tìm xác suất để doanh nghiệp đạt được doanh số tối thiểu là 35 triệu đồng/tuần.

Giải

Gọi X là doanh số hàng tuần mà doanh nghiệp có thể đạt được ở thị trường đó. Do không có thông tin gì hơn nên có thể xem X là biến ngẫu nhiên liên tục phân phối đều trên khoảng (20;40).

Vậy X có hàm mật độ xác suất như sau

$$f(x)=\begin{cases}\dfrac{1}{40-20}=0,05 & khi\ x\in(20,40)\\ 0 & khi\ x\notin(20,40)\end{cases}$$

Từ đó xác suất để doanh nghiệp đạt được doanh số tối thiểu là 35 triệu đồng/ tuần được tính như sau:

$$P(X\geq 35)=1-P(X<35)=\int_{-\infty}^{+\infty} f(x)dx-\int_{-\infty}^{35} f(x)dx$$

$$=\int_{35}^{+\infty} f(x)dx=\int_{35}^{40} 0,05dx=0,05x\Big|_{35}^{40}=0,25$$

3.2.7. Các công thức xấp xỉ

Xấp xỉ phân phối siêu bội bằng phân phối nhị thức

Xấp xỉ phân phối nhị thức bằng phân phối Poisson

Định lí: Cho $X_n \sim B(n,p)$. Khi $n \to +\infty$, $p \to 0$, $np = \lambda(const)$ thì

$$\lim_{n\to\infty} P(X_n = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Ý nghĩa

Trong thực hành, $X \sim B(n,p)$, nếu với n khá lớn, và p khá bé (sao cho $npq \approx np$) thì

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

Trong thưc tế dùng công thức xấp xỉ trên khi $n > 50$ và $p < 0,1$

Ví dụ

Một máy dệt có 4000 ống sợi. Xác suất để mỗi ống sợi ấy bị đứt trong một phút là 0,0005. Tính xác suất trong 1 phút

a) Có 3 ống sợi bị đứt

b) Có ít nhất hai ống sợi bị đứt

HD

Gọi X là số ống sợi bị đứt trong một phút thì $X \sim B(4000;0,0005)$. Có thể coi $X \sim P(2)$

a)

$$P(X=3) = C_{4000}^3 (0,0005)^3 (0,9995)^{3997}$$
$$\approx e^{-2}.\frac{2^3}{3!} = 0,1804$$

b)

$$P(X \ge 2) = 1 - (P(X=0) + P(X=1))$$
$$\approx 1 - \left(e^{-2}.\frac{2^0}{0!} + e^{-2}.\frac{2^1}{1!}\right) = 1 - (0,1353 + 0,2707) = 0,594$$

Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

Định lí giới hạn địa phương Moivre-Laplace

Gỉa sử $P_n(k)$ là xác suất hiện k lần biến cố A trong dãy n phép thử Bernoulli với P(A)=p, p không quá gần 0 và 1, $x_k = \dfrac{k-np}{\sqrt{npq}}$ bị chặn đều. Khi đó

$$\lim_{n\to\infty} \frac{P_n(k)\sqrt{npq}}{\varphi(x_k)} = 1$$

Trong đó φ là hàm mật độ Gauss. $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, q = 1 - p$

Định lí giới hạn Moivre-Laplace tích phân

Giả sử $X \sim B(n, p)$, p không quá gần 0 và 1. Khi đó

$$\lim_{n\to\infty} \frac{P(k_1 \le X \le k_2)}{\phi(\alpha_2) - \phi(\alpha_1)} = 1$$

trong đó $\alpha_i = \frac{k_i - np}{\sqrt{npq}}, i = 1,2$ và $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{\frac{-t^2}{2}} dt$ (tích phân Laplace)

Ý nghĩa

Trong thực hành, với $X \sim B(n, p)$, khi n đủ lớn và p không quá lớn và không quá bé ta có:

$P(X = k) \approx \frac{1}{\sqrt{npq}} \varphi\left(\frac{k - np}{\sqrt{npq}}\right)$ với $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$P(k_1 \le X \le k_2) \approx \phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \phi\left(\frac{k_1 - np}{\sqrt{npq}}\right)$ (*) với $\phi(x)$ là tích phân Laplace

Việc xấp xỉ tốt khi $np \ge 5$ và $nq \ge 5$ hoặc $npq \ge 20$ hoặc $n > 5$ và

$$\left|\sqrt{\frac{p}{1-p}} - \sqrt{\frac{1-p}{p}}\right| \frac{1}{\sqrt{n}} < 0,3$$

Chú ý: Khi sử dụng công thức (**) để giảm sai số ta cần có sự hiệu chỉnh

$$P(k_1 \le X \le k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k} \approx \left[\phi\left(\frac{k_2 - np - 0,5}{\sqrt{npq}}\right) - \phi\left(\frac{k_1 - np - 0,5}{\sqrt{npq}}\right)\right]$$

Ví dụ.

Một xạ thủ bắn 64 phát sung vào bia. Xác suất bắn trúng mỗi phát là 0,8. Tính xác suất

a) Có từ 45 đến 52 phát trúng bia

b) Có không dưới 51 phát trúng bia

Giải

Gọi X là số đạn trúng bia. Ta có $X \sim B(64; 0,8)$.

a) $P(45 \le X \le 52) \approx \phi\left(\frac{52 - 64.0,8}{\sqrt{64.0,8.0,2}}\right) - \phi\left(\frac{52 - 64.0,8}{\sqrt{64.0,8.0,2}}\right)$

b) $P(51 \le X \le 64) \approx \phi\left(\frac{64 - 64.0,8}{\sqrt{64.0,8.0,2}}\right) - \phi\left(\frac{51 - 64.0,8}{\sqrt{64.0,8.0,2}}\right)$

Ví dụ

Một khách sạn nhận đặt chỗ của 325 khách hàng cho 300 phòng vào ngày 1 tháng 7 vì theo kinh nghiệm của những năm trước cho thấy tỷ lệ khách đặt chỗ nhưng không đến là 10%. Tính xác suất

a) Có 300 khách đến vào ngày 1 tháng 7 để nhận phòng

b) Tất cả các khách đến vào ngày 1 tháng 7 đều nhận được phòng

Giải

Gọi X là số khách đặt phòng và đến vào ngày 1 tháng 7. $X \sim B(325;0,9)$ ta có $np = 292,5; \sqrt{npq} = 5,4081$

a) $P(X = 300) \approx \frac{1}{\sqrt{npq}} \varphi\left(\frac{300 - np}{\sqrt{npq}}\right) \approx \frac{1}{5,4081} \varphi(1,39) = \frac{0,1518}{5,4081} = 0,0281$

b)Xác suất tất cả các khách đến ngày 1 tháng 7 đều có phòng

$$P(X \le 300) = P(0 \le X < 301) \approx \phi\left(\frac{301 - 292,5 - 0,5}{5,4081}\right) - \phi\left(\frac{-292,5 - 0,5}{5,4081}\right) = 0,5 + \phi(1,48)$$

$= 0,93056 \approx 93,06\%$

3.3. Luật số lớn và định lý giới hạn trung tâm

Luật số lớn Trê bư sép

Nếu dãy các biến ngẫu nhiên (X_n) độc lập có kỳ vọng hữu hạn và phương sai bị chặn đều ($(V(X_i) \le C \forall i)$, khi đó với mọi $\varepsilon > 0$

$$\lim_{n \to \infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \varepsilon\right) = 1$$

Hệ quả1: Nếu (X_n) độc lập, cùng phân phối, có kỳ vọng μ và phương sai σ^2 thì

$$\lim_{n \to \infty} P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| < \varepsilon\right) = 1$$

Luất số lớn Bernoulli

Gọi $f_n(A)$ là tần xuất xuất hiện biến cố A trong dãy n phép thử độc lập và P(A)=p trong mỗi phép thử thì với mọi $\varepsilon > 0$

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

Ý nghĩa

- Từ luật số lớn Trê bư sép ta kết luận: Từng biến nhẫu nhiên X_i có thể nhận giá trị khác nhiều so với EX_i nhưng khi n đủ lớn trung bình cộng của chúng lại rất gần với trung bình cộng của các kỳ vọng với xác suất gần 1. Điều đó cho phép ta dự đoán về trugn bình cộng của các biến ngẫu nhiên.

- Từ hệ quả 1 rút ra kết luận: Có thể ước lượng kì vọng của biến ngẫu nhiên bằng trung bình cộng các kết quả đo đạc độc lập của biến ngẫu nhiên đó với xác suất gần 1. Đó cũng là cơ sở của phương pháp mẫu được sử dụng trong thống kê.
- Luật số lớn Bernoulli cho chúng ta cơ sở định nghĩa xác suất theo thống kê.

Định lý giới hạn trung tâm

Giả sử các biến ngẫu nhiên $X_1, X_2, \ldots, X_n, \ldots$ độc lập cùng phân phối với kỳ vọng $E(X_i) = \mu$, phương sai $V(X_i) = \sigma^2$ (hữu hạn khác 0) .

Đặt $T_n = \dfrac{\dfrac{X_1 + X_2 + \ldots + X_n}{n} - \mu}{\sigma}\sqrt{n}$.

Khi đó với mọi *x* thuộc *R* ta có

$$\lim_{n\to\infty} P\left(T_n < x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$$

Nói cách khác khi *n* đủ lớn phân phối xác suất của T_n xấp xỉ phân phối chuẩn tắc.

Ý nghĩa của định lý

Nếu $X_1, X_2, \ldots, X_n, \ldots$ độc lập cùng phân phối trong đó $E(X_i) = \mu$, $V(X_i) = \sigma^2$, với *n* đủ lớn thì

+ $X = \sum_{i=1}^{n} X_i$ có phân phối xấp xỉ phân phối chuẩn $N(n\mu, n\sigma^2)$ (*) (“dù các X_i không có phân phối chuẩn)

+ $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ có phân phối xấp xỉ phân phối chuẩn $N(\mu, \dfrac{\sigma^2}{n})$

Từ (*) ta rút ra kết luận: tổng của một số rất lớn BNN độc lập cùng phân phối mà sự đóng góp của mỗi thành phần trong tổng đó rất nhỏ thì tổng sẽ có phân phối xác suất xấp xỉ phân phối chuẩn.

CHƯƠNG 4. BIẾN NGẪU NHIÊN HAI CHIỀU

4.1. Khái niệm

Từ trước đến giờ ta chỉ xét các biến ngẫu nhiên 1 chiều nhưng trong thực tế để mô tả kết quả của một thí nghiệm cần đồng thời hai hay nhiều hơn hai biến ngẫu nhiên. Ví dụ nghiên cứu đặc tính chiều cao và cân nặng của phụ nữ Việt Nam, nghiên cứu nhiệt độ và áp suất trong một thí nghiệm vật lí và nghiên cứu nhiệt độ ở các tháng trong năm của một vùng nào đó. Trong những trường hợp này đồng thời hai hay nhiều hơn hai biến ngẫu nhiên được xét đến.

Biến ngẫu nhiên hai chiều (véc tơ ngẫu nhiên hai chiều) là cặp (X,Y) trong đó X, Y là các biến ngẫu nhiên.

Nếu X, Y đều rời rạc (hay liên tục) ta gọi (X,Y) là rời rạc (liên tục). Ở đây ta không xét trường hợp một thành phần rời rạc, một thành phần là liên tục.

Một bộ có n biến ngẫu nhiên $X_1, X_2, ..., X_n$ gọi là một vectơ n chiều.

Ví dụ 1. Một nhà máy sản xuất một loại sản phẩm. Nếu kích thước của sản phẩm được đo bằng chiều dài X và chiều rộng Y thì ta có biến ngẫu nhiên 2 chiều. Còn nếu xét thêm cả chiều cao Z thì ta có biến ngẫu nhiên 3 chiều. Nếu ta chỉ quan tâm đến trọng lượng và thể tích của một sản phẩm thì ta cũng có biến ngẫu nhiên 2 chiều.

Ví dụ 2. Tung một con xúc sắc 2 lần độc lập. Gọi X= số nút xuất hiện ở lần tung 1, Y= số nút xuất hiện ở lần tung 2. Khi đó ta có biến ngẫu nhiên 2 chiều (X,Y)

4.2. Quy luật phân phối xác suất của biến ngẫu nhiên hai chiều

4.2.1. Bảng phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc

Bảng phân phối xác suất đồng thời

Để mô tả vecto ngẫu nhiên (X,Y) nhận giá trị nào đó với xác suất tương ứng là bao nhiêu người ta sử dụng bảng phân phối xác suất sau

X \ Y	y_1	y_2	……	y_j	……	y_m	$\sum_j$
x_1	$P(x_1, y_1)$	$P(x_1, y_2)$	……	$P(x_1, y_j)$	……	$P(x_1, y_m)$	$P_X(x_1)$

x_2	$P(x_2, y_1)$	$P(x_2, y_2)$	……	$P(x_2, y_j)$	……	$P(x_2, y_m)$	$P_X(x_2)$
……	……	……	……	……	……	……	……
x_i	$P(x_i, y_1)$	$P(x_i, y_2)$	……	$P(x_i, y_j)$	……	$P(x_i, y_m)$	$P_X(x_i)$
……	……	……	……	……	……	……	……
x_n	$P(x_n, y_1)$	$p(x_n, y_2)$	……	$p(x_n, y_j)$	……	$p(x_n, y_m)$	$P_X(x_n)$
$\sum_i$	$P_Y(y_1)$	$P_Y(y_2)$	……	$P_Y(y_j)$	……	$P_Y(y_m)$	1

Bảng phân phối xác suất đồng thời của (X,Y)

Trong đó:

- $x_i \quad (i=\overline{1,n})$ là các giá trị có thể có của X; $y_j\,(j=\overline{1,m})$ là các giá trị có thể có của Y.
- $P(x_i, y_j) = P(X = x_i, Y = y_j)$ gọi là xác suất đồng thời của biến ngẫu nhiên hai chiều (X,Y) nhận giá trị *(x_i,y_j).*
- Các xác suất này thoả mãn :

$$\begin{cases} p(x_i, y_j) \geq 0, \quad \forall i = \overline{1,n}, j = \overline{1,m} \\ \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) = 1 \end{cases}$$

Ví dụ 1. Biến ngẫu nhiên rời rạc X nhận các giá trị 1,2 và 3. Biến ngẫu nhiên Y nhận các giá trị 1,2,3,4. Phân phối đồng thời của véc tơ ngẫu nhiên (X,Y) cho bởi bảng sau.

X \ Y	1	2	3	4

6	0,1	0,0	0,1	0,0
7	0,3	0,0	0,1	0,2
8	0,0	0,2	0,0	0,0

Ta sẽ xác định giá trị $P\ X \geq 7, Y \geq 2$ và $P\ X = 6$

Bằng cách lấy tổng các hàm giá trị xác suất $P(x,y)$ tương ứng với $X \geq 7$ và $Y \geq 2$ chúng ta tính được:

$$P(X \geq 7, Y \geq 2) = p(7,2) + p(7,3) + p(7,4) + p(8,2) + p(8,3) + p(8,4) = 0,5$$

Cộng các giá trị xác suất của dòng đầu tiên ta có :

$$P(X = 6) = p(6,1) + p(6,2) + p(6,3) + p(6,4) = 0,2$$

Phân phối biên

Nếu biết được phân phối đồng thời của vectơ ngẫu nhiên (X,Y) ta cũng sẽ xác định được phân phối biên của mỗi thành phần

Phân phối biên của biến ngẫu nhiên X (cộng theo hàng)

$$P_X(x_i) = \sum_{j=1}^{m} p(x_i, y_j)$$

Bảng phân phối biên của biến ngẫu nhiên X

X	x_1	x_2	……..	x_n
P	$p_X(x_1)$	$p_X(x_2)$	……..	$p_X(x_n)$

- Phân phối biên của biến ngẫu nhiên Y (cộng theo cột)

$$p_Y(y_j) = \sum_{i=1}^{n} p(x_i, y_j)$$

Bảng phân phối biên của biến ngẫu nhiên Y

X	y_1	y_2	……..	y_m
P	$p_Y(y_1)$	$p_Y(y_2)$	……..	$p_Y(y_m)$

Ví dụ 2. Tung 3 đồng xu cân đối A,B,C.Gọi X là số mặt ngửa xuất hiện của 2 đồng xu A,B và Y là số mặt ngửa của cả 3 đồng tiền A, B, C. Hãy lập bảng phân phối xác suất đồng thời của vectơ ngẫu nhiên (X,Y) và các phân phối biên của X,Y?

Giải :

Chúng ta có bảng 8 kết quả đồng khả năng của việc gieo 3 đồng tiền cân đối và tính giá trị X, Y tương ứng. Trong đó N, S là kí hiệu mặt ngửa hoặc mặt sấp.

A	B	C	X	Y
S	S	S	0	0
S	S	N	0	1
S	N	S	1	1
S	N	N	1	2
N	S	S	1	1
N	S	N	1	2
N	N	S	2	2
N	N	N	2	3

Sử dụng công thức xác suất cổ điển ta có phân bố xác suất đồng thời của (X,Y) như sau:

X \ Y	0	1	2	3	Tổng
0	1/8	1/8	0	0	2/8
1	0	2/8	2/8	0	4/8
2	0	0	1/8	1/8	2/8
Tổng	1/8	3/8	3/8	1/8	1

- Cộng các cột ta được phân bố xác suất của biến ngẫu nhiên X

X	0	1	2
P	2/8	4/8	2/8

- Cộng các hàng ta được phân bố xác suất của biến ngẫu nhiên Y

Y	0	1	2	3
P	1/8	3/8	3/8	1/8

4.2.2. Hàm phân phối xác suất của biến ngẫu nhiên hai chiều

Hàm phân phối xác suất của biến ngẫu nhiên hai chiều (X,Y), kí hiệu là F(X,Y) là hàm được xác định như sau:

$F(x,y)=P(X<x,Y<y)$

Tính chất

1) $0 \le F(x,y) \le 1, \forall x, y \in R$

2) F(x,y) không giảm theo từng biến số, nghĩa là

$F(x_2, y) \ge F(x_1, y)$ nếu $x_2>x_1$

$F(x, y_2) \ge F(x, y_1)$ nếu $y_2>y_1$

3) Ta có các biểu thức giới hạn sau

$$F(-\infty,-\infty)=0; F(+\infty,+\infty)=1;$$
$$F(-\infty, y)=0; F(x,-\infty)=0;$$
$$F(x,+\infty)=F_X(x); F(+\infty; y)=F_Y(y)$$

4) $P\left(x_1 < X < x_2; y_1 < Y < y_2\right) = F(x_2, y_2) + F(x_1, y_1) - F(x_2, y_1) - F(x_1, y_2)$

4.2.3. Hàm mật độ xác suất của biến ngẫu nhiên hai chiều liên tục

Giả sử biến ngẫu nhiên liên tục hai chiều (X,Y) có hàm phân phối xác suất F(x,y) liên tục và có đạo hàm hỗn hợp cấp 2 liên tục (có thể trừ một số hữu hạn đường cong)

Hàm mật độ xác suất của biến ngẫu nhiên hai chiều liên tục (X,Y), kí hiệu f(x,y) là hàm

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

Ta có

$P\left((X,Y) \in D\right) = \iint\limits_D f(x,y)dxdy$ với $D \subset R^2$

$$F(x,y)=\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(x,y)dxdy$$

Tính chất

1) $f(x,y) \ge 0; \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)dxdy = 1;$

2) Các hàm mật độ của X, của Y

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy;\ f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx$$

3) X,Y độc lập $\Leftrightarrow f(x,y)=f_X(x).f_Y(y)$

4.3. Quy luật phân phối xác suất có điều kiện của các biến ngẫu nhiên thành phần

Bảng phân phối xác suất có điều kiện

Xét biến ngẫu nhiên rời rạc (X,Y) trong đó các giá trị có thể có của thành phần X là $x_1,x_2,\dots,x_n$, còn các giá trị có thể có của thành phần Y là $y_1,y_2,\dots,y_n$. Gọi $P(x_i/y_j)\ \ (i=\overline{1,n};\ j=\overline{1,m})$ là xác suất có điều kiện để thành phần X nhận giá trị bằng x_i với điều kiện thành phần Y nhận giá trị bằng y_j.

Bảng phân phối xác suất có điều kiện của thành phần X với điều kiện $Y=y_j$ có dạng

X/y_j	x_1	x_2	$\dots$	x_i	$\dots$	x_n
P	$P(x_1/y_j)$	$P(x_2/y_j)$	$\dots$	$P(x_i/y_j)$	$\dots$	$P(x_n/y_j)$

Trong đó các xác suất có điều kiện được tính bằng công thức

$$P(x_i/y_j)=\frac{P(x_i,y_j)}{P(y_j)}\quad i=\overline{1,n};\ j=\overline{1,m}$$

Ta chú ý rằng

$$\sum_{i=`}^{n}P(x_i/y_j)=\sum_{i=1}^{n}\frac{P(x_i,y_j)}{P(y_j)}=\frac{P(y_j)}{P(y_j)}=1\quad j=\overline{1,m}$$

Tức là các xác suất có điều kiện $P(x_i/y_j)$ cũng phải thỏa mãn các yêu cầu của một quy luật phân phối xác suất

Bảng phân phối xác suất có điều kiện của thành phần Y với điều kiện $X=x_i$ có dạng

Y/x_i	y_1	y_2	$\dots$	y_j	$\dots$	y_m
P	$P(y_1/x_i)$	$P(y_2/x_x)$	$\dots$	$P(y_j/x_i)$	$\dots$	$P(y_m/x_i)$

Trong đó

$$P(y_i/x_i)=\frac{P(x_i,y_j)}{P(y_i)}\quad i=\overline{1,n};\ j=\overline{1,m}$$

$$\sum_{i=`}^{n} P(y_j / x_i) = \sum_{i=1}^{n} \frac{P(x_i, y_j)}{P(x_i)} = \frac{P(x_i)}{P(x_i)} = 1 \quad i = \overline{1,n}$$

Ví dụ 1. Phân phối xác suất có điều kiện của biến ngẫu nhiên Y với điều kiện X=1 ở ví dụ 2 tung 3 đồng xu ở trên

Y|X=1	0	1	2	3
P	0	2/8	2/8	0

Ví dụ 2

Phân phối xác suất của lương tháng Y (triệu đồng) và giới tính X của coong nhân một số công ty như sau

X \ Y	0,5	1	1,5
Nữ: 0 Nam:1	0,1 0,06	0,3 0,18	0,2 0,16

Tìm phân phối xác suất của lương tháng của nữ công nhân

Giải

Ta có $P(x_1) = 0,1 + 0,3 + 0,2 = 0,6$. Từ đó

$$P(y_1 / x_1) = \frac{P(x_1, y_1)}{P(x_1)} = \frac{0,10}{0,60} = \frac{1}{6}$$

$$P(y_2 / x_1) = \frac{P(x_1, y_2)}{P(x_1)} = \frac{0,30}{0,60} = \frac{1}{2}$$

$$P(y_3 / x_1) = \frac{P(x_1, y_3)}{P(x_1)} = \frac{0,20}{0,60} = \frac{1}{3}$$

Vậy bảng phân phối xác suất của lương tháng của nữ công nhân là

$Y / X = x_1$	0,5	1	1,5
P	1/6	1/2	1/3

Hàm mật độ xác suất có điều kiện

Giả sử (X,Y) là biến ngẫu nhiên hai chiều liên tục.

Hàm mật độ xác suất có điều kiện của thành phần X với Y=y, kí hiệu f(x/y) là biểu thức

$$f(x/y)=\frac{f(x,y)}{f_Y(y)}=\frac{f(x,y)}{\int\limits_{-\infty}^{\infty} f(x,y)dx}$$

Hàm mật độ xác suất có điều kiện của thành phần Y với X=x, kí hiệu f(y/x) là biểu thức

$$f(y/x)=\frac{f(x,y)}{f_X(x)}=\frac{f(x,y)}{\int\limits_{-\infty}^{\infty} f(x,y)dy}$$

Hàm mật độ xác suất có điều kiện cũng thỏa mãn các điều kiện

$$f(x/y)\geq 0; \int\limits_{-\infty}^{+\infty} f(x/y)dx=1;$$

$$f(y/x)\geq 0; \int\limits_{-\infty}^{+\infty} f(y/x)dy=1$$

Công thức tổng hợp hệ hai biến ngẫu nhiên theo phân phối xác suất của các thành phần như sau

Nếu (X,Y) là biến ngẫu nhiên hai chiều rời rạc thì

$$P(x_i,y_j)=P(x_i)P(y_j/x_i)=P(y_j)P(x_i/y_j) \quad i=\overline{1,n},\ j=\overline{1,m}$$

Nếu (X,Y) là biến ngẫu nhiên hai chiều liên tục thì

$$f(x,y)=f_X(x)f(y/x)=f_Y(y)f(x/y)$$

Nếu hai thành phần X và Y độc lập với nhau thì phân phối xác suất có điều kiện cũng bằng phân phối xác suất không điều kiện . Khi đó

$P(x_i,y_j)=P(x_i)P(y_j) \quad (*) i=\overline{1,n},\ j=\overline{1,m}$ nếu (X,Y) là biến ngẫu nhiên rời rạc

$f(x,y)=f_X(x)f_Y(y)(**)$ nếu (X,Y) là biến ngẫu nhiên liên tục

(*) và (**) là điều kiện cần và đủ để X và Y độc lập.

4.4. Các tham số đặc trưng của biến ngẫu nhiên 2 chiều

4.4.1. Kỳ vọng, phương sai của biến ngẫu nhiên thành phần.

Từ bảng phân bố xác suất thành phần ta có công thức tính kì vọng và phương sai sau:

Nếu X,Y là biến ngẫu nhiên rời rạc

$$E(X)=\sum_{i=1}^{m}x_i p(x_i)=\sum_{i=1}^{m}\sum_{j=1}^{n}x_i p(x_i,y_j)$$

$$E(Y)=\sum_{j=1}^{n}y_j p(y_j)=\sum_{j=1}^{n}\sum_{i=1}^{m}y_j p(x_i,y_j)$$

$$V(X)=E(X^2)-[E(X)]^2=\sum_{i=1}^{m}\sum_{j=1}^{n}x_i^2 p(x_i,y_j)-[E(X)]^2$$

$$V(Y)=E(Y^2)-[E(Y)]^2=\sum_{j=1}^{n}\sum_{i=1}^{m}y_j^2 p(x_i,y_j)-[E(Y)]^2$$

Nếu X,Y là biến ngẫu nhiên liên tục

$$E(X)=\int_{-\infty}^{+\infty}xf_1(x)dx=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy$$

$$E(Y)=\int_{-\infty}^{+\infty}yf_2(y)dy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)dxdy$$

$$V(X)=\int_{-\infty}^{+\infty}(x-E(X))^2 f_1(x)dx=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x^2 f(x,y)dxdy-(E(X))^2$$

$$V(Y)=\int_{-\infty}^{+\infty}(y-E(X))^2 f_2(y)dy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}y^2 f(x,y)dxdy-(E(y))^2$$

4.4.2. Hiệp phương sai, hệ số tương quan

Hiệp phương sai

Hiệp phương sai (hay còn gọi là Covariance) của 2 biến ngẫu nhiên X,Y kí hiệu Cov(X,Y) là kì vọng toán của tích các sai lệch của các biến ngẫu nhiên đó với kì vọng của chúng.

$$\mathrm{Cov}(\mathrm{X},\mathrm{Y})=\mathrm{E}\{[X-E(X)][Y-E(Y)]\}$$

Nếu X,Y rời rạc có bảng phân bố đồng thời như ở phần 4.2.1 thì :

$$\mathrm{Cov}(\mathrm{X},\mathrm{Y})=\sum_{i=1}^{m}\sum_{j=1}^{n}x_i y_j p(x_i,y_j)-E(X)E(Y)$$

Nếu X, Y liên tục thì $Cov(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)dxdy-E(X)E(Y)$

Tính chất:

(i) Cov(X,Y)=E(XY)-E(X)E(Y)

(ii) Cov(X,Y)=Cov(Y,X)

(iii) Cov(X,X)=V(X)

(iv) Cov(aX+b,cY+d)= abCov(X,Y) với mọi hằng số a,b,c,d.

(v) Nếu X,Y độc lập thì Cov(X,Y)=0. Ngược lại chưa chắc đúng.

Hệ số tương quan

Hệ số tương quan của hai biến ngẫu nhiên X,Y được kí hiệu và định nghĩa bởi công thức: $$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Tính chất:

1) $-1 \le \rho_{XY} \le 1$ với mọi X,Y.

2) Nếu X,Y độc lập thì $\rho_{XY} = 0$.Điều ngược lại chưa chắc đúng.

3) Với mọi hằng số a,b,c,d. Ta có: $\rho_{aX+b,cY+d} = \begin{cases} \rho_{X,Y} & khi\, ab > 0 \\ -\rho_{X,Y} & khi\, ab < 0 \end{cases}$

4) Với Y=aX+b,a≠0 ta có: $\rho_{X,Y} = \begin{cases} 1 & khi\, a > 0 \\ -1 & khi\, a < 0 \end{cases}$

Ý nghĩa của hệ số tương quan

- Hệ số tương quan đo lường mức độ phụ thuộc tuyến tính giữa X và Y. Khi $|\rho_{X,Y}|$ càng gần 1 thì tính chất phụ thuộc tuyến tính càng chặt, khi $|\rho_{X,Y}|$ càng gần 0 thì tính phụ thuộc càng lỏng lẻo.
- Khi $|\rho_{X,Y}| = 0$ ta nói X, Y không tương quan.

4.4.3. Kì vọng toán có điều kiện, hàm hồi quy

Kỳ vọng toán có điều kiện

X,Y rời rạc

Từ luật phân bố có điều kiện, kì vọng toán của biến ngẫu nhiênY với điều kiện { X=x_i}được kí hiệu và tính theo công thức sau:

$$E(Y/X = x_i) = \sum_{j=1}^{m} y_j p(y_j / x_i)$$

X, Y liên tục

$$E(Y/X = x_i) = \int_{-\infty}^{+\infty} y(f(y/x_i)dy$$

Trong đó $f(y/x_i)$ là hàm mật độ xác suất có điền kiện của Y với $X = x_i$

Tương tự ta có định nghĩa kì vọng toán có điều kiện của X với điều kiện{ Y=y_j} khi X, Y rời rạc và liên tục lần lượt bằng

$$E\left(X/Y=y_j\right)=\sum_{i=1}^{n}x_i p\left(x_i/y_j\right)$$

$$E\left(X/Y=y_j\right)=\int_{-\infty}^{+\infty}x(f(x/y_j)dx$$

Hàm hồi quy

Hàm hồi qui của Y đối với X là.kì vọng toán có điều kiện của Y đối với X :

$f\left(x\right)=E\left(Y/X=x\right)$

Hàm hồi qui của X đối với Y là.kì vọng toán có điều kiện của X đối với Y :

$g\left(y\right)=E\left(X/Y=y\right)$

Các hàm hồi quy cho biết giá trị trung bình của biến ngẫu nhiên này phụ thuộc vào biến kia như thế nào

Miền xác định của hàm hồi qui của X (Y) là miền giá trị của X (Y).

Khi X,Y liên tục thì miền xác định của hàm hồi qui là một khoảng nào đó.

Ví dụ . Thống kê dân cư ở một thành phố nọ ở độ tuổi trưởng thành về thu nhập hàng tháng X và độ tuổi Ythu được kết quả trong bảng sau:

Y / X	30	45	70
1	0,01	0,02	0,05
2	0,03	0,06	0,10
3	0,18	0,21	0,15
4	0,07	0,08	0,21

Trong đó X có đơn vị là triệu đồng/tháng.

Y=30, 45, 70 là chỉ độ tuổi người dân trong khoảng 25-35; 35-55; 55-85.

Giải : Thu nhập bình quân theo lứa tuổi là kì vọng có điều kiện của X theo Y.

Với Y=30 ta có bảng phân phối xác suất điều kiện tương ứng:

X/Y=30	1	2	3	4
P	$\frac{0,01}{0,29}$	$\frac{0,03}{0,29}$	$\frac{0,18}{0,29}$	$\frac{0,07}{0,29}$

Từ đó :

$$E\left(X/Y=30\right)=1.\frac{0.01}{0,29}+2.\frac{0,03}{0,29}+3.\frac{0,18}{0,29}+4.\frac{0,07}{0,29}=2,069$$

Tương tự : $E\left(X/Y=45\right)=1,496$ $E\left(X/Y=70\right)=1,529$

Vậy :

- Thu nhập bình quân ở độ tuổi 30 là 2,069 triệu đồng / tháng.
- Thu nhập bình quân ở độ tuổi 45 là 1,496 triệu đồng / tháng.
- Thu nhập bình quân ở độ tuổi 70 là 1,529 triệu đồng / tháng.

Khoa học thống kê là khoa học về thu thập, phân tích, diễn giải và trình bày các dữ liệu để từ đó tìm ra bản chất và tính quy luật của các hiện tượng kinh tế, xã hội - tự nhiên.
Khoa học thống kê dựa vào lý thuyết thống kê, một loại toán học ứng dụng. Trong lý thuyết thống kê, tính chất ngẫu nhiên và sự không chắc chắn có thể làm mô hình dựa vào lý thuyết xác suất. Vì mục đích của khoa học thống kê là để tạo ra thông tin "đúng nhất" theo dữ liệu có sẵn, có tác giả nhìn khoa thống kê như một loại lý thuyết quyết định.
Thông tin thống kê bao gồm số liệu thống kê và bản phân tích các số liệu đó, là sản phẩm thu được của hoạt động thống kê do một tổ chức hoặc cá nhân nhất định tiến hành trong một không gian và thời gian cụ thể.

Vai trò của thống kê trong kinh tế, chính trị, xã hội học

Thống kê là một trong những công cụ quản lý vĩ mô quan trọng, cung cấp các thông tin thống kê trung thực, khách quan, chính xác, đầy đủ, kịp thời trong việc đánh giá, dự báotình hình, hoạch định chiến lược, chính sách, xây dựng kế hoạch phát triển kinh tế - xã hội và đáp ứng nhu cầu thông tin thống kê của các tổ chức, cá nhân.
Chẳng hạn các bạn hay nghe trên TV trần lãi suất, hay chính sách thả nổi đồng tiền này tiền kia, giá vàng … mỗi một quyết định đều ảnh hưởng đến nền kinh tế vĩ mô.
Câu chuyện về Goeorghe Soros
Hoặc là cơ cấu dân số già, trẻ… tất cả đều ảnh hưởng đến nền kinh tế
Hoặc là gần đây các bạn cứ nghe tỷ lệ sinh con trai và con gái ở VN rất đáng ngại nên sẽ có những điều chỉnh Chẳng hạn sinh con gái được thưởng, không cho tiết lộ giới tính thai nhi khi siêu âm.
Tất cả những nghiên cứu đó dựa trên khoa học thống kê, khoa học thống kê Khoa học thống kê dựa vào lý thuyết thống kê, một loại toán học ứng dụng. Trong lý thuyết thống kê, tính chất ngẫu nhiên và sự không chắc chắn có thể làm mô hình dựa vào lý thuyết xác suất. Vì mục đích của khoa học thống kê là để tạo ra thông tin "đúng nhất" theo dữ liệu có sẵn, có tác giả nhìn khoa thống kê như một loại lý thuyết quyết định.

CHƯƠNG 5. MẪU NGẪU NHIÊN

5.1. Khái niệm mẫu ngẫu nhiên

Tổng thể và mẫu

Tổng thể: Tập hợp tất cả các phần tử đồng nhất theo một dấu hiệu nghiên cứu định tính hay định lượng nào đó gọi là tổng thể, hay đám đông. Số phần tử của tổng thể gọi là kích thước của tổng thể, kí hiệu N.

Mỗi phần tử thuộc tổng thể gọi là cá thể.

Tổng thể được đặc trưng bởi dấu hiệu nghiên cứu X. Dấu hiệu này thay đổi qua các phần tử của tổng thể, được xem như một biến ngẫu nhiên.

Ví dụ: Nghiên cứu về:

- Thu nhập của các hộ gia đình ở HCM→Tổng thể: các hộ gia đình ở HCM. Dấu hiệu nghiên cứu X: thu nhập.

- Điểm trung bình của sinh viên ở FTU →Tổng thể: các sinh viên ở FTU. Đặc Dấu hiệu nghiên cứu X: điểm số.

- Tỉ lệ hài lòng của khách hàng sử dụng dịch vụ internet banking →Tổng thể: các khách hàng dùng dịch vụ. Dấu hiệu nghiên cứu X: hài lòng hay không.

Mẫu: Tập hợp các phần tử lấy ra từ một tổng thể được gọi là mẫu. Số lượng phần tử của mẫu gọi là kích thước mẫu, kí hiệu n.

Ví dụ : Nghiên cứu về:

- Thu nhập của các hộ gia đình ở HCM. Khảo sát 1000 hộ gia đình ở Tp Hồ Chí Minh
- Điểm trung bình của sinh viên ở FTU . Điều tra trên 1000 SV

Có hai cách lấy mẫu: Lấy mẫu hoàn lại và lấy mẫu không hoàn lại.

Lấy mẫu không hoàn lại: là mẫu được chọn bằng cách phần tử đã lấy ra quan sát thì loại khỏi tổng thể rồi mới lấy phần tử tiếp theo.Mẫu không hoàn lại còn gọi là mẫu không lặp.

Lấy mẫu hoàn lại: là mẫu được chọn bằng cách phần tử đã lấy ra quan sát được bỏ trở lại tổng thể rồi mới lấy ra phần tử tiếp theo.Mẫu hoàn lại còn gọi là mẫu lặp. Trong mẫu lặp, một phần tử của tổng thể có thể được chọn nhiều lần.

Trường hợp kích thước tổng thể lớn thì mẫu không hoàn lại và mẫu hoàn lại có thể coi như nhau. Trong lý thuyết tổng quát, chúng ta luôn giả thiết mẫu là mẫu hoàn lại.

Ví dụ

Khi lấy chai bia ra điều tra thì người ta phải khui nó ra để phân tích thành phần hóa học, khi lấy nó ra để phân tích ta không thể trả nó về tổng thể

-Khi nghiên cứu cá trong hồ, bắt 20 con cá lên xem

+ ta có thể bắt cùng lúc 20 con lên xem-> lấy mẫu không hoàn lại

+Bắt lần lượt từng con, xem xét xong thả lại hồ rồi mới con tiếp theo-> mẫu có hoàn lại

Mẫu ngẫu nhiên

Tiến hành *n* quan sát độc lập về biến ngẫu nhiên *X* nào đó . Gọi X_i là việc quan sát lần thứ *i* về biến ngẫu nhiên *X*. Khi đó $(X_1, X_2,..., X_n)$ được gọi là mẫu ngẫu nhiên, *n* gọi là kích thước mẫu hay cỡ mẫu.

Như vậy mẫu ngẫu nhiên cỡ n thực chất là *n* biến ngẫu nhiên độc lập, cùng phân phối như biến ngẫu nhiên *X*. *X* gọi là biến ngẫu nhiên gốc (tức là $X_1, X_2,..., X_n$ là các BNN độc lập, cùng phân phối với BNN *X*)

Gọi x_i là kết quả quan sát được ở lần thứ i. Khi đó $(x_1, x_2, \ldots, x_n)$ là n giá trị cụ thể mà ta quan sát được, gọi là mẫu cụ thể.

Cách trình bày mẫu cụ thể

Bảng phân phối tần số thực nghiệm- Bảng phân phối tần suất thực nghiệm

Giả sử mẫu cụ thể $(x_1, x_2, \ldots, x_n)$ trong đó có thể có nhiều giá trị giống nhau. Giả sử giá trị x_1 xuất hiện n_1 lần, x_2 xuất hiện n_2 lần,… , x_k xuất hiện n_k lần , $x_1<\ldots<x_k$ và $n_1+\ldots.+n_k = n$.

Khi đó n_i gọi là tần số xuất hiện x_i

$f_i = \dfrac{n_i}{n}$ gọi là tần suất xuất hiện x_i

Bảng phân phối tần số thực nghiệm

x_i	x_1 x_2 ………… x_k
n_i	n_1 n_2 .... n_k

Bảng phân phối tần suất thực nghiệm:

x_i	x_1 x_2 ………… x_k
f_i	f_1 f_2 .... f_k

Ta có $f_1 + f_2 \cdots + f_k = 1$

Ví dụ 1. Nghiên cứu điểm thi môn XSTK của SV sinh viên Ngoại Thương, người ta chọn ngẫu nhiên 50 sinh viên, thu được kết quả như sau : 2 SV đạt 4 điểm; 15 SV đạt 7điểm, 20 sinh viên đạt 8 điểm, 12 sinh viên đạt 9 điểm, 1 sinh viên đạt 10 điểm.
Bảng phân phối tấn số thực nghiệm điểm thi của SV

x_i	*4*	*7*	*8*	*9*	*10*
n_i	*2*	*15*	*20*	*12*	*1*

Bảng phân phối tấn suất thực nghiệm điểm thi của SV

x_i	*4*	*7*	*8*	*9*	*10*
f_i	*2/50*	*15/50*	*20/50*	*12/50*	*1/50*

Bảng phân bố ghép lớp

Trong những trường hợp mẫu điều tra có kích thước lớn, các giá trị cụ thể của mẫu khá gần nhau, người ta chia các giá trị cụ thể mẫu thành các lớp và lập bảng phân phối thực nghiệm ghép lớp.

Độ rộng các khoảng không nhất thiết bằng nhau.

Ví dụ 2. Một mẫu về chiều cao của 400 cây con trong một vườn ươm được trình bày trong bảng như sau:

Khoảng	**Tần số n_i**	**Tần suất f_i**
4,5-9,5	18	0,045
9,5-11,5	58	0,145
11,5-13,5	62	0,155
13,5-16,5	72	0,180
16,5-19,5	57	0,1425
19,5-22,5	42	0,1050
22,5-26,5	36	0,090
26,5-36,5	55	0,1375

❖ **Qui ước :**

Hai lớp liền nhau $x_{i-1} - x_i$; $x_i - x_{i+1}$ thì x_i thuộc lớp $x_{i-1} - x_i$.

Đầu mút bên phải của mỗi khoảng thuộc vào khoảng đó mà không thuộc vào khoảng tiếp theo khi tính tần số của lớp.

5.2. Các tham số đặc trưng của tổng thể

Trung bình tổng thể

Giả sử trong tổng thể kích thước N, dấu hiệu định lượng X nhận các giá trị $x_1, x_2, \ldots, x_N$. Trung bình tổng thể, kí hiệu là μ, xác định bởi

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Nếu trong tổng thể dấu hiệu X chỉ nhận các giá trị $x_1, x_2, \ldots, x_k$ với các tần số tương ứng là $N_1, N_2, \ldots, N_k$ với $N_1 + N_2 + \ldots + N_k = N$ thì trung bình tổng thể được xác định bằng biểu thức

$$\mu = \frac{1}{N}\sum_{i=1}^{k} x_i N_i$$

Phương sai tổng thể

Phương sai tổng thể, kí hiệu là σ^2, xác định:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2$$

Nếu trong tổng thể dấu hiệu X chỉ nhận các giá trị $x_1, x_2, \ldots, x_k$ với các tần số tương ứng là $N_1, N_2, \ldots, N_k$ với $N_1 + N_2 + \ldots + N_k = N$ thì

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{k} N_i (x_i - \mu)^2 = \frac{1}{N}\sum_{i=1}^{k} N_i x_i^2 - \mu^2$$

Độ lệch chuẩn của tổng thể

Lấy căn bậc hai của phương sai ta thu được độ lệch chuẩn của tổng thể, kí hiệu σ

Tỷ lệ tổng thể

Ta xét tổng thể về mặt định tính: tổng thể có kích thước N trong đó có M phần tử mang tính chất A. Khi đó $p = \frac{M}{N}$ gọi là tỉ lệ phần tử mang tính chất A trong tổng thể.

Các tham số đặc trưng khác của X như Mốt, trung vị, Hệ số biến thiên… cũng đều là các tham số đặc trưng của tổng thể. Cách tính các tham số này giống như đã trình bày ở chương 2.

Ví dụ 1. Khảo sát thu nhập trong 1 tháng của các nhân viên làm việc ở 1 công ty. Thu được kết quả như sau Tổng thể là tập hợp các nhân viên làm việc ở công ty này

Thu nhập (triệu đồng/tháng)	5	10	15	20	25
Số người	15	25	30	20	10

Hãy tính các đặc trưng của tổng thể

Giải

Ta có bảng sau

x_i	n_i	$x_i n_i$	$x_i^2 n_i$
5 10 15 20 25	15 25 30 20 10	75 250 450 400 250	375 2500 6750 8000 6250
$\sum$	100	1425	23875

Từ bảng ta có

$$\mu = \frac{1425}{100} = 14,25$$

$$\sigma^2 = \frac{23875}{100} - (14,25)^2 = 35,6875$$

Ví dụ 2: Những người có thu nhập từ 15 triệu trở lên được xem là những người "có thu nhập cao". Tìm tỷ lệ người có thu nhập cao

5.3. Thống kê

5.3.1. Khái niệm

Một hàm của mẫu ngẫu nhiên $G = G\ X_1, X_2, ..., X_n$ được gọi là một thống kê.

Như vậy thống kê G cũng là một biến ngẫu nhiên tuân theo một qui luật phân phối xác suất nhất định và có các số đặc trưng như kì vọng, phương sai…Mặt khác khi mẫu ngẫu nhiên nhận một giá trị cụ thể $x_1, x_2, ..., x_n$ thì G cũng nhận một giá trị cụ thể gọi là giá trị quan sát được của thống kê $g = G\ x_1, x_2, ..., x_n$

5.3.2. Một số thống kê đặc trưng thường gặp

Giả sử từ biến ngẫu nhiên gốc *X* trong tổng thể lập mẫu ngẫu nhiên kích thước n $X_1, X_2, ..., X_n$

Trung bình mẫu

Trung bình của mẫu ngẫu nhiên $X_1, X_2, \ldots, X_n$ kí hiệu $\bar{X}$ được xác định như sau:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Khi mẫu ngẫu nhiên nhận một giá trị cụ thể $x_1, x_2, \ldots, x_n$ thì trung bình mẫu cũng nhận giá trị cụ thể bằng:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Tính chất Nếu $E(X) = \mu$, $V(X) = \sigma^2$, thì $E\left(\bar{X}\right) = \mu, \quad V\left(\bar{X}\right) = \frac{\sigma^2}{n}$

Giải thích

$$E\left(\bar{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\left(\sum_{i=1}^{n} E(X_i)\right) = \frac{1}{n}.n.\mu = \mu$$

Vì $X_1, X_2, \ldots, X_n$ là các BNN độc lập cùng phương sai σ^2 nên

$$V\left(\bar{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n} V(X_i)\right) = \frac{\sigma^2}{n}$$

Phương sai mẫu

- Phương sai mẫu S^2

$$\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

- Phương sai mẫu có hiệu chỉnh S^2

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

- Phương sai mẫu S^{*2}. Khi dấu hiệu nghiên cứu X của tổng thể có kì vọng xác định $E(X) = \mu$

$$S^{*2} = \frac{1}{n}\sum_{i=1}^{n} X_i - \mu^2$$

Tính chất.: Nếu $E(X) = \mu$, $V(X) = \sigma^2$, thì

$$E\left(\hat{S}^2\right) = \frac{n-1}{n}\sigma^2$$

$$E\left(S^2\right) = \sigma^2$$

$$E(S^{*2}) = \sigma^2$$

Độ lệch chuẩn mẫu

Độ lệch chuẩn mẫu $S = \sqrt{S^2}$

Độ lệch chuẩn mẫu hiệu chỉnh $S = \sqrt{S^2}$

Độ lệch chuẩn mẫu $S^* = \sqrt{S^{*2}}$

Khi mẫu ngẫu nhiên nhận một giá trị cụ thể $x_1, x_2, \ldots, x_n$ ta cũng có các giá trị cụ thể

$\hat{s}^2, s^2, \hat{s}, s$

Ví dụ

Để nghiên cứu nhu cầu mua gạo ở một thành phố, người ta tiến hành điều tra một số gia đình và ghi kết quả ở bảng sau đây

Nhu cầu (kg/tháng)	30	32	35	36	38	42	50
Số gia đình	12	7	8	20	32	15	26

Tính các đặc trưng của mẫu

Giải

Từ mẫu ta có bảng

x_i	n_i	$n_i x_i$	$n_i x_i^2$
30	12	360	10 800
32	7	224	7 168
35	8	280	9 800
36	20	720	25 920
38	32	1 216	46 208
42	15	630	26 460
50	26	1 300	65 000
Tổng	**120**	**4 730**	**191 356**

Ta có

$$\bar{x} = \frac{4730}{120} = 39,4168$$
$$\hat{s}^2 = \frac{191356}{120} - 39,4168^2 = 40,9492$$
$$s^2 = \frac{n}{n-1}\hat{s}^2 = \frac{120}{119}.40,9492 = 41,2933$$
$$\hat{s} = 6,3992$$
$$s = 6,4260$$

Tỷ lệ mẫu

Xét tổng thể gồm N phần tử, trong đó mỗi cá thể có thể mang hoặc không mang dấu hiệu A nào đó. Gọi p là tỷ lệ các cá thể mang dấu hiệu A**.** Nếu cá thể có dấu hiệu A ta cho nhận giá trị là 1, trường hợp ngược lại ta cho nhận giá trị 0. Lúc đó dấu hiệu nghiên cứu có thể xem là biến ngẫu nhiên X có qui luật phân phối A(p) có kì vọng và phương sai

$$E\ X\ = p \qquad\qquad V\ X\ = pq \text{ với } q = 1 - p$$

Từ biến ngẫu nhiên gốc X trong tổng thể lập mẫu ngẫu nhiên kích thước n

$X_1, X_2, \ldots, X_n$. $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân phối A(p).

Tỷ lệ mẫu $F = \frac{1}{n}\sum_{i=1}^{n} X_i$

Khi mẫu ngẫu nhiên nhận một giá trị cụ thể $x_1, x_2, \ldots, x_n$ ta có giá trị cụ thể f

Tính chất Nếu $X \sim A(p)$ thì

$$E(F) = p \qquad\qquad V(F) = \frac{pq}{n}$$

Ví dụ

Trước kỳ bầu cử người ta phỏng vấn 1575 cử tri thấy có 1212 người trả lời là ủng hộ ứng cử viên A. Tìm tỷ lệ mẫu ủng hộ ứng cử viên A

Chú ý

- Với bảng phân phối ghép lớp ta thay lớp $x_{i-1} - x_i$ bằng một đại diện $x_i' = \frac{x_{i-1} + x_i}{2}$ để tính các thống kê mẫu

Ví dụ

Lượng xăng hao phí của một ô tô đi từ A đến B sau 30 lần chạy, kết quả cho trong bảng

Lượng xăng hao phí (lít)	9,6-9,8	9,8-10	10-10,2	10,2-10,4	10,4-10,6
Số lần tương ứng	3	5	10	8	4

Lập bảng

Lớp	x_i	n_i	$x_i n_i$	$n_i x_i^2$
9,6-9,8	9,7	3	29,1	282,27
9,8-10	9,9	5	49,5	490,05
10-10,2	10,1	10	101	1020,1
10,2-10,4	10,3	8	82,4	848,72
10,4-10,6	10,5	4	42	441
Tổng		**30**	**304**	**3082,14**

$\overline{x} = 10,1333$

$s^2 = 0,0561$

- Khi số liệu phức tạp ta có thể đổi biến để giảm độ phức tạp của tính toán:

Với $h \neq 0$, đặt $u_i = \dfrac{x_i - a}{h}$. Vì $x_i = hu_i + a$, nên theo tính chất của kì vọng và phương sai ta có:

$$\overline{x} = h\overline{u} + a$$

$$\hat{s}_x^2 = h^2\left(\overline{u^2} - \left(\overline{u}\right)^2\right)$$

$$s^2 = \frac{n}{n-1}h^2\left(\overline{u^2} - \left(\overline{u}\right)^2\right)$$

Do đó nếu chọn được a và h sao cho việc tính $\overline{u}$ và $\hat{s}_u^2$ một cách đơn giản thì ta sẽ có một cách đơn giản để tính $\overline{x}$ và $\hat{s}_x^2$.

Trong trường hợp các giá trị x_i cách đều nhau thì ta chọn h bằng khoảng cách giữa hai số liệu của 1lớp và chọn $a = x_{i0}$ với $n_{i0} = max\left(n_1, n_2, ..., n_k\right)$

Ví dụ

Trở lại ví dụ trên, chọn $h = 0,2, a = 10,1$. Khi đó ta có

Lớp	x_i	$u_i = \dfrac{x_i - 10,1}{0,2}$	n_i	$n_i u_i$	$n_i u_i^2$
9,6-9,8	9,7	-2	3	-6	12

9,8-10	9,9	-1	5	-5	5
10-10,2	10,1	0	10	0	0
10,2-10,4	10,3	1	8	8	8
10,4-10,6	10,5	2	4	8	16
Tổng			**30**	**5**	**41**

Ta có

$\bar{u} = \frac{5}{30}$. Do đó

$\bar{x} = 0,2.\frac{5}{30} + 10,1 \approx 10,1333$

$s^2 = 0,0561$

5.3.3. Qui luật phân phối xác suất của một số thống kê đặc trưng mẫu

Trường hợp *X* có phân phối chuẩn $X \sim N(\mu, \sigma^2)$. Khi đó

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$G = \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N(0,1)$$

$$G = \frac{nS^{*2}}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$$

$$G = \frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$$

$G = \frac{\overline{X} - \mu}{S}\sqrt{n} \sim T(n-1)$ (khi n<30) , σ^2 chưa biết

Trường hợp *X* không có phân phối chuẩn

Khi n đủ lớn $(n \geq 30)$ ta có các phân phối xấp xỉ chuẩn sau

- Khi phương sai tổng thể σ^2 đã biết

$G = \frac{\overline{X} - \mu}{\sigma}\sqrt{n}$ có phân phối xấp xỉ phân phối chuẩn tắc N(0,1)

$\overline{X}$ có phân phối xấp xỉ phân phối chuẩn $N(\mu, \frac{\sigma^2}{n})$

- Khi phương sai tổng thể σ^2 chưa biết

$G = \frac{\overline{X} - \mu}{S}\sqrt{n}$ có phân phối xấp xỉ phân phối chuẩn tắc N(0,1)

$\overline{X}$ có phân phối xấp xỉ phân phối chuẩn $N(\mu,\frac{S^2}{n})$

Khi X có phân phối không-một $X \sim A(p)$

Tỉ lệ mẫu $F = \frac{1}{n}\sum_{i=1}^{n} X_i$

Với $np \geq 5; n(1-p) \geq 5$ thì $G = \frac{F-p}{\sqrt{\frac{p(1-p)}{n}}}$ có phân phối xấp xỉ phân phối chuẩn tắc N(0,1); F có phân phối xấp xỉ phân phối chuẩn $N(p,\frac{p(1-p)}{n})$

Trường hợp có hai biến ngẫu nhiên gốc phân bố theo qui luật chuẩn

Giả sử ta xét cùng một lúc hai tổng thể. Ở tổng thể thứ nhất dấu hiệu nghiên cứu được xem như biến ngẫu nhiên X_1 có phân phối chuẩn $N\left(\mu_1;\sigma_1^2\right)$. Ở tổng thể thứ hai dấu hiệu nghiên cứu được xem như biến ngẫu nhiên X_2 có phân phối chuẩn $N\left(\mu_2;\sigma_2^2\right)$

Từ hai tổng thể nói trên rút ra hai mẫu độc lập kích thước tương ứng là n_1, n_2.

$$W_1 = \left(X_{11}, X_{12}, \ldots, X_{1n_1}\right) \qquad W_2 = \left(X_{21}, X_{22}, \ldots, X_{2n_2}\right)$$

Thống kê hiệu của hai trung bình mẫu $\overline{X}_1 - \overline{X}_2$

Đây là một tổ hợp tuyến tính của các phân bố chuẩn do đó cũng có qui luật phân bố chuẩn. Mặt khác kì vọng và phương sai:

$$E\left(\overline{X}_1 - \overline{X}_2\right) = E\left(\overline{X}_1\right) - E\left(\overline{X}_2\right) = \mu_1 - \mu_2$$

$$V\left(\overline{X}_1 - \overline{X}_2\right) = V\left(\overline{X}_1\right) + V\left(\overline{X}_2\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Do đó $Z = \dfrac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \simeq N\left(0,1\right)$

$$\overline{X}_1 - \overline{X}_2 \simeq N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Thống kê của hai phương sai thành phần

Ta có

$$\chi_1^2 = \frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1-1);\ \chi_2^2 = \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2-1);$$ và độc lập nhau nên

$$\chi^2 = \frac{(n_1-1)S_1^2}{\sigma_1^2} + \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_1+n_2-2);$$

$$G = \frac{\dfrac{\chi_1^2}{n_1-1}}{\dfrac{\chi_2^2}{n_2-1}} = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2} \sim F\left[(n_1-1);(n_2-1)\right]$$

Thống kê hiệu hai tỉ lệ mẫu $F_1 - F_2$

Giả sử ta xét cùng một lúc hai tổng thể trong đó dấu hiệu nghiên cứu trong hai tổng thể có thể được xem như các biến ngẫu nhiên phân phối không –một với các tham số tương ứng là p_1, p_2.

Từ hai tổng thể nói trên rút ra hai mẫu độc lập kích thước tương ứng là n_1, n_2.

$$W_1 = (X_{11}, X_{12}, ..., X_{1n_1}) \qquad W_2 = (X_{21}, X_{22}, ..., X_{2n_2})$$

Nếu $n_1 > 30; n_2 > 30$ thì $F_1 - F_2$ sẽ có phân bố xấp xỉ chuẩn theo định lý giới hạn trung tâm với các tham số đặc trưng là:

$$E(F_1 - F_2) = p_1 - p_2 \qquad V(F_1 - F_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

Do đó thống kê sau xấp xỉ phân bố chuẩn tắc

$$U = \frac{(F_1 - F_2) - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}} \sim N(0;1)$$

CHƯƠNG 6. BÀI TOÁN ƯỚC LƯỢNG

6.1. Ước lượng điểm

Giả sử X là biến ngẫu nhiên hình thành từ tổng thể, có dạng phân phối xác suất $P(x,\theta)$ đã biết, nhưng θ chưa biết và ta cần tìm. Việc tìm giá trị thực của tham số θ rất khó khăn nên chỉ có thể ước lượng θ căn cứ theo kết quả của mẫu. Muốn vậy người ta lấy mẫu $(X_1, X_2,..., X_n)$ và lập một đại lượng thống kê $\hat{\theta}(X_1, X_2,..., X_n)$ để dùng thay θ.

Đại lượng thống kê $\hat{\theta}(X_1, X_2,..., X_n)$ dùng để ước lượng θ gọi là hàm ước lượng của θ hay ước lượng của θ.

$\hat{\theta}(X_1, X_2,..., X_n)$ là một biến ngẫu nhiên. $\hat{\theta}(X_1, X_2,..., X_n)$ nhận một giá trị cụ thể (một điểm) khi mẫu $(X_1, X_2,..., X_n)$ nhận một giá trị cụ thể $(x_1, x_2,..., x_n)$ nên $\hat{\theta}(x_1, x_2,..., x_n)$ được gọi là ước lượng điểm của θ.

$\hat{\theta} = \hat{\theta}(X_1, X_2,..., X_n)$ chỉ phụ thuộc vào $(X_1, X_2,..., X_n)$ mà không phụ thuộc vào θ.

Cùng với một mẫu ngẫu nhiên có nhiều cách xây dựng thống kê $\hat{\theta}$ khác nhau để ước lượng θ .Vì vậy ta cần lựa chọn thống kê tốt nhất để ước lượng cho tham số θ dựa vào các tiêu chuẩn sau:

Ước lượng không chệch

Thống kê $\hat{\theta}$ gọi là ước lượng không chệch của θ nếu $E(\hat{\theta}) = \theta$. Nếu $E(\hat{\theta}) \neq \theta$ thì $\hat{\theta}$ gọi là một ước lượng chệch của θ.

Với mọi mẫu ta có:

+ Tỉ lệ mẫu F là ước lượng không chệch của tỉ lệ tổng thể p. (Vì $E(F) = p$)

+ Trung bình mẫu $\overline{X}$ là ước lượng không chệch của kì vọng μ của biến ngẫu nhiên gốc.

+ Phương sai mẫu hiệu chỉnh S^2 là ước lượng không chệch của phương sai tổng thể σ^2.

+ Phương sai mẫu $\hat{S}{}^2$ là ước lượng chệch của σ^2 do $E\left(\hat{S}{}^2\right) \neq \sigma^2$

Ước lượng hiệu quả

Thống kê $\hat{\theta}$ được gọi là ước lượng hiệu quả của θ nếu nó là ước lượng không chệch và có phương sai bé nhất trong các ước lượng không chệch của θ.

Nếu $X \sim N(\mu, \sigma^2)$ thì $\overline{X}$ là ước lượng hiệu quả của μ

Nếu $X \sim A(p)$ thì tỷ lệ mẫu F là ước lượng hiệu quả của p

Ước lượng vững

Thống kê $\hat{\theta} = \hat{\theta}\left(X_1, X_2, ..., X_n\right)$ được gọi là ước lượng vững của θ nếu với mọi $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}\left(X_1, X_2, ..., X_n\right) - \theta\right| < \varepsilon\right) = 1$$

$F, \overline{X}, S^2$ tương ứng là ước lượng vững của p, μ, σ^2. S^{2} cũng là ước lượng vững của σ^2

Phương pháp ước lượng điểm

Phương pháp hợp lí cực đại

Giả sử đã biết quy luật phân phối xác suất tổng quát của biến ngẫu nhiên gốc *X* dưới dạng hàm mật độ $f\left(x, \theta\right)$ (hoặc có thể là biểu thức xác suất nếu *X* là biến ngẫu nhiên rời rạc). Cần ước lượng tham số θ nào đó của *X*.

Lập mẫu ngẫu nhiên kích thước n: $W = \left(X_1, X_2, ..., X_n\right)$

Xây dựng hàm đối số θ tại một giá trị cụ thể của mẫu :

$L\left(\theta\right) = L\left(x_1, x_2, ..., x_n, \theta\right) = f\left(x_1, \theta\right) f\left(x_2, \theta\right) ... f\left(x_n, \theta\right)$

Hàm *L* gọi là hàm hợp lí của tham số θ.

Số $\hat{\theta} = \hat{\theta}\left(X_1, X_2, ..., X_n\right)$ được gọi là ước lượng hợp lí cực đại của θ nếu hàm hợp lý đạt giá trị cực đại tại $\hat{\theta} = \hat{\theta}\left(x_1, x_2, ..., x_n\right)$

Vì hàm $L\left(x_1, x_2, ..., x_n, \theta\right)$ và hàm $\ln L\left(x_1, x_2, ..., x_n, \theta\right)$ đạt cực đại tại cùng giá trị θ, do đó có thể tìm giá trị của θ để lnL đạt giá trị cực đại với các bước như sau:

- Tìm đạo hàm bậc nhất của lnL theo θ
- Giải phương trình $\dfrac{d \ln L}{d\theta} = 0$. Giả sử nó có nghiệm $\theta = \theta = f(x_1, x_2, \ldots, x_n)$
- Tìm đạo hàm bậc hai $\dfrac{d^2 \ln L}{d\theta^2} = 0$
- Nếu tại điểm $\theta = \theta$ đạo hàm bậc hai âm thì tại điểm này hàm lnL đạt cực đại, và do đó $\theta = f(x_1, x_2, \ldots, x_n)$ là ước lượng hợp lý tối đa cầm tìm của θ

Ví dụ

Tìm ước lượng hợp lý của tham số p trong quy luật phân phối A(p)

Giải

Biểu thức xác suất tổng quát của quy luật không – một là

$p_x = p^x(1-p)^{1-x}$

Lập mẫu ngẫu nhiên kích thước n: W=$(X_1, X_2, \dots, X_n)$ và tại giá trị w=$(x_1, x_2, \dots, x_n)$ của nó lập hàm hợp lý với mỗi giá trị x_i ($x_i = 0$ hoặc 1)

$$L=(x_1, x_2, \dots, x_n, p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

Từ đó

$$\ln L = \sum_{i=1}^{n}\left(x_i \ln p + (1-x_i)\ln(1-p)\right)$$

$$\frac{d\ln L}{dp} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}\sum_{i=1}^{n}(1-x_i)$$

$$\frac{d\ln L}{dp} = 0 \Rightarrow p = \overline{x} = f$$

Dễ thấy rằng $\frac{d^2\ln L}{dp^2} = -\frac{n}{p(1-p)} < 0$, do đó ước lượng hợp lý tố ta của p là f

Định lí :

- Cho tổng thể có phân phối Poisson và mẫu . Khi đó $\overline{X}$ là ước lượng hợp lí cực đại của tham số μ của tổng thể.

- Cho tổng thể có phân phối chuẩn và mẫu $X = \left(X_1, X_2, ..., X_n\right)$. Khi đó $\overline{X}$ là ước lượng hợp lí cực đại của kì vọng μ và S^2 là ước lượng hợp lí cực đại của phương sai σ^2 của tổng thể.

6.2. Ước lượng bằng khoảng tin cậy

Các phương pháp ước lượng điểm có nhược điểm là khi kích thước mẫu bé thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng. Mặt khác phương pháp trên không thể đánh giá được khả năng mắc sai lầm khi ước lượng là bao nhiêu. Do đó khi kích thước mẫu bé người ta thường dùng phương pháp ước lượng khoảng tin cậy.

Định nghĩa

Cho $\left(X_1, X_2, ..., X_n\right)$ là mẫu ngẫu nhiên từ biến ngẫu nhiên gốc X có phân phối phụ thuộc vào tham số θ và α là một số thực thuộc $(0,1)$.

Khoảng (θ_1,θ_2) với $\theta_1=\theta_1(X_1,X_2,...,X_n),\theta_2=\theta_2(X_1,X_2,...,X_n)$ là hai thống kê nào đó gọi là khoảng tin cậy (khoảng ước lượng) của θ với độ tin cậy $1-\alpha$ nếu $P\left(\theta_1(X_1,X_2,...,X_n)<\theta<\theta_2(X_1,X_2,...,X_n)\right)=1-\alpha$

- $1-\alpha$ gọi là độ tin cậy của ước lượng.
- $I=\theta_2-\theta_1$ được gọi là độ dài khoảng tin cậy.
- $I=\theta_2-\theta_1=2\varepsilon$ (độ dài khoảng tin cậy) thì ε được gọi là độ chính xác của ước lượng.

Trong thực tế thường yêu cầu $1-\alpha$ khá lớn. Thông thường ta chọn độ tin cậy (1- α) ở các mức: 0,95; 0,99; 0,999

Rõ ràng là với cùng một độ tin cậy thì khoảng tin cậy càng hẹp càng giúp chúng ta xác định chính xác được tham số cần tìm.

Các bước giải bài toán tìm khoảng tin cậy

+ Từ tổng thể lập mẫu ngẫu nhiên kích thước n $W=(X_1,X_2,...,X_n)$

+ Xây dựng thống kê $G=f(X_1,X_2,...,X_n,\theta)$ *sao cho phân phối xác suất của nó không phụ thuộc vào tham số và hoàn toàn xác định.*

+ Cho độ tin cậy $1-\alpha$*, tìm được các số* $\alpha_1,\alpha_2:\alpha_1+\alpha_2=\alpha$. *Tìm cặp số* $g_{\alpha_1},g_{\alpha_2}$ *thỏa*

$$P(G<g_{\alpha_1})=\alpha_1$$
$$P(G>g_{\alpha_2})=\alpha_2$$

Khi đó $P(g_{\alpha_1}<G<g_{\alpha_2})=1-(\alpha_1+\alpha_2)=1-\alpha$ (*)

+ Biến đổi tương đương (*) đưa về dạng $P(\theta_1<\theta<\theta_2)=1-\alpha$

+ Tiến hành một phép thử với mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ *ta thu được mẫu cụ thể* $w=(x_1,x_2,...,x_n)$*, ta tính được các giá trị cụ thể của* θ_1,θ_2

Ước lượng trung bình của tổng thể

Bài toán: Giả sử trong tổng thể biến ngẫu nhiên X có $E(X)=\mu$ chưa biết, $V(X)=\sigma^2$. Hãy ước lượng μ với độ tin cậy $1-\alpha$.

Trường hợp 1: $n\geq 30,\sigma^2$ đã biết. Khi đó $G=\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\sim N(0,1)$ nếu X có phân phối chuẩn hoặc $G=\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}$ có phân phối xấp xỉ $N(0,1)$ nếu X không có giả thiết này.

Chọn thống kê $G=\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}$ với $\overline{X}$ là trung bình mẫu. Ta có $G=\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma} \sim N(0,1)$

Với độ tin cậy $1-\alpha$ cho trước, chọn được cặp giá trị α_1, α_2 không âm sao cho $\alpha_1+\alpha_2=\alpha$ và tìm các giá trị tới hạn mức $1-\alpha_1, \alpha_2$ của phân phối chuẩn tắc:

$P(G>z_{1-\alpha_1})=1-\alpha_1; P(G>z_{\alpha_2})=\alpha_2;$ Khi đó $P(z_{1-\alpha_1}<G<z_{\alpha_2})=1-\alpha$

Do G có phân phối chuẩn tắc nên $z_{\alpha_1}=-z_{1-\alpha_1}$ suy ra $P(-z_{\alpha_1}<G<z_{\alpha_2})=1-\alpha$ hay

$$P\left(\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha_1}<\mu<\overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha_2}\right)=1-\alpha$$

Với độ tin cậy $1-\alpha$ tham số μ của biến ngẫu nhiên gốc X sẽ nằm trong khoảng

$$\left(\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha_1}; \overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha_2}\right) \quad (*)$$

"Như vậy với độ tin cậy $1-\alpha$ có vô số cặp số α_1, α_2 thỏa mãn $\alpha_1+\alpha_2=\alpha$, do đó có vô số khoảng tin cậy, sau đây ta xét một số trường hợp đặc biệt".

Ta có

Khoảng tin cậy đối xứng: Lấy $\alpha_1=\alpha_2=\alpha/2$ ta có khoảng tin cậy của μ với độ tin cậy $1-\alpha$ là $\left(\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}; \overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$ với $F(z_{\alpha/2})=1-\alpha/2$, $F(x)=\int_{-\infty}^{x}\varphi(t)dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$

Khoảng tin cậy bên phải (ước lượng giá trị tối thiểu cho μ**)**: Lấy $\alpha_1=\alpha; \alpha_2=0$ thì $z_{\alpha_2}=+\infty$. Ta có khoảng tin cậy của μ là $\left(\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha}; +\infty\right)$

Khoảng tin cậy bên trái (ước lượng giá trị tối đa cho μ**)**: Lấy $\alpha_1=0; \alpha_2=\alpha$ thì $z_{\alpha_1}=+\infty$. Ta có khoảng tin cậy của μ là $\left(-\infty; \overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha}\right)$

Khi có mẫu cụ thể ta tìm được $\overline{x}$, ta có các khoảng tin cậy cụ thể.

Trường hợp 2: $n<30, X \sim N(\mu,\sigma^2); \sigma^2$ đã biết. Khi đó $G=\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma} \sim N(0,1)$

Làm tương tự như trường hợp 1. Ta có

Khoảng tin cậy đối xứng: $\left(\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}; \overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$

Khoảng tin cậy bên phải: $\left(\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha}; +\infty\right)$

Khoảng tin cậy bên trái: $\left(-\infty; \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha}\right)$

Với cùng độ tin cậy $1-\alpha$, khoảng tin cậy nào ngắn hơn sẽ tốt hơn. Trong trường hợp này độ dài khoảng tin cậy sẽ là ngắn nhất khi khoảng tin cậy đối xứng. Lúc đó độ dài khoảng tin cậy sẽ bằng hai lần độ chính xác và được xác định bằng biểu thức

$$I = 2\varepsilon = \frac{2\sigma}{\sqrt{n}} z_{\alpha/2} \quad \left(\varepsilon = \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right) \quad (*)$$

Từ (*) ta thu được công thức xác định kích thước mẫu tối thiểu *n* sao cho với độ tin cậy $1-\alpha$ cho trước, độ dài khoảng tin cậy không vượt quá giá trị I_0 cho trước. Công thức có dạng

$$n \geq \frac{4\sigma^2}{I_0^2} z_{\alpha/2}^2 \geq \frac{\sigma^2}{\varepsilon_0^2} z_{\alpha/2}^2$$

Trường hợp 3: $n \geq 30, \sigma^2$ **chưa biết.** Ta có thể thay σ bởi *S* Khi đó $G = \frac{(\overline{X}-\mu)\sqrt{n}}{S}$ có phân phối xấp xỉ $N(0,1)$. Tương tự trường hợp 1. Ta có

Khoảng tin cậy đối xứng: $\left(\overline{X} - \frac{S}{\sqrt{n}} z_{\alpha/2}; \overline{X} + \frac{S}{\sqrt{n}} z_{\alpha/2}\right)$

Khoảng tin cậy bên phải: $\left(\overline{X} - \frac{S}{\sqrt{n}} z_{\alpha}; +\infty\right)$

Khoảng tin cậy bên trái: $\left(-\infty; \overline{X} + \frac{S}{\sqrt{n}} z_{\alpha}\right)$

Trường hợp 4: $n < 30, X \sim N(\mu, \sigma^2); \sigma^2$ **chưa biết**. Chọn thống kê $G = \frac{(\overline{X}-\mu)\sqrt{n}}{S}$. Ta có $G = \frac{(\overline{X}-\mu)\sqrt{n}}{S} \sim T(n-1)$

Với độ tin cậy $1-\alpha$ cho trước chọn được cặp giá trị α_1, α_2 sao cho $\alpha_1 + \alpha_2 = \alpha$ và tìm các giá trị tới hạn mức $1-\alpha_1, \alpha_2$ của phân phối phân phối Student (*n*-1) bậc tự do tương ứng là $t_{1-\alpha_1}^{(n-1)}, t_{\alpha_2}^{(n-1)}$ thỏa mãn

$P(G > t_{1-\alpha_1}^{(n-1)}) = 1-\alpha_1; P(G > t_{\alpha_2}^{(n-1)}) = \alpha_2;$ Từ đó $P(t_{1-\alpha_1}^{(n-1)} < G < t_{\alpha_2}^{(n-1)})) = 1-\alpha$

Do $t_{\alpha_1}^{(n-1)} = -t_{1-\alpha_1}^{(n-1)}$ suy ra $P(-t_{\alpha_1}^{(n-1)} < G < t_{\alpha_2}^{(n-1)}) = 1-\alpha$ hay

$$P\left(\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha_1}^{(n-1)}<\mu<\overline{X}+\frac{S}{\sqrt{n}}t_{\alpha_2}^{(n-1)}\right)=1-\alpha$$

Với độ tin cậy $1-\alpha$ tham số μ của biến ngẫu nhiên gốc X sẽ nằm trong khoảng

$$\left(\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha_1}^{(n-1)};\overline{X}+\frac{S}{\sqrt{n}}t_{\alpha 2}^{(n-1)}\right) \quad (**)$$

Khoảng tin cậy đối xứng: $\left(\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}^{(n-1)};\overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}^{(n-1)}\right)$, với $t_{\alpha/2}^{(n-1)}$ là giá trị tới hạn mức $\alpha/2$ của phân phối Student n-1 bậc tự do

Khoảng tin cậy bên phải: $\left(\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha}^{(n-1)};+\infty\right)$

Khoảng tin cậy bên trái: $\left(-\infty;\overline{X}+\frac{S}{\sqrt{n}}t_{\alpha}^{(n-1)}\right)$

với $t_{\alpha}^{(n-1)}$ là giá trị tới hạn mức α của phân phối student n-1 bậc tự do.

Ví dụ

Trọng lượng một loại sản phẩm là biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn là 1gam. Cân thử 25 sản phẩm này ta thu được kết quả như sau

Trọng lượng (gam)	18	19	20	21
Số SP tương ứng	3	5	15	2

Với độ tin cậy 0,95 hãy tìm khoảng tin cậy cho trọng lượng trung bình của loại sản phẩm nói trên.

Giải

Gọi X là trọng lượng sản phẩm (đơn vị: gam). Theo giả thiết X có phân phối chuẩn với $\sigma=1$. Gọi trọng lượng trung bình của sản phẩm là μ.

Do $\alpha=0,05 \Rightarrow z_{\alpha/2}=z_{0,025}=1,96$ (do $F(z_{\alpha/2})=1-\alpha/2=1-0,05/2=0,975$)

Từ mẫu cụ thể ta có n = 25, $\bar{x}=\frac{3.18+5.19+15.20+2.21}{25}=19,64$

Vậy với độ tin cậy 0,95 khoảng tin cậy của μ là

$$(19,64-\frac{1}{\sqrt{25}}1,96;19,64+\frac{1}{\sqrt{25}}1,96) \text{ hay } (19,248;20,032)$$

Ví dụ 2 Ở một cửa hàng chế biến thủy sản, theo dõi nhu cầu của mặt hàng nước mắm trong một số ngày ta có kết quả

Số bán ra (lít)	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100	100-110
Số ngày	3	8	30	45	20	25	17	9	3

Với độ tin cậy 99% hãy ước lượng nước mắm trung bình bán một ngày

Giải

Gọi *X* là lượng nước mắm bán ra một ngày (đơn vị: lít)

Do số liệu cho dưới dạng khoảng nên ta thay mỗi khoảng bởi điểm giữa của nó. Ta có

x_i	n_i
25	3
35	8
45	30
55	45
65	20
75	25
85	17
95	9
105	3
	$n=160$

Ta có $n=160>30;\quad \overline{x}=62,3;\quad s^2=311,95;\quad s=17,66$

$1-\alpha=99\%\Rightarrow\alpha=0,01\Rightarrow F(z_{\alpha/2})=1-\alpha/2=0,995\Rightarrow z_{0,005}=2,58$

Với độ tin cậy 99% khoảng ước lượng cần tìm là

$$\left(62,3-2,58\frac{17,66}{\sqrt{160}};62,3+2,58\frac{17,66}{\sqrt{160}}\right)\quad \text{hay}\dots$$

Ước lượng tỉ lệ tổng thể

Bài toán: Giả sử tỷ lệ *p* của tổng thể chưa biết, với độ tin cậy $1-\alpha$ tìm khoảng tin cậy cho *p*.

Chọn thống kê $G=\dfrac{(F-p)\sqrt{n}}{\sqrt{F(1-F)}}\sim N(0,1)$ với *n* đủ lớn, *F* là tần suất mẫu

Với độ tin cậy $1-\alpha$ cho trước tìm được cặp giá trị α_1 và α_2 $(\alpha_1+\alpha_2=\alpha)$ và hai giá trị tới hạn chuẩn tương ứng là $z_{1-\alpha_1}, z_{\alpha_2}$ thỏa

$P(G<z_{1-\alpha_1})=\alpha_1,\ P(G>z_{\alpha_2})=\alpha_2$

Từ đó $P(z_{1-\alpha_1} < G < z_{\alpha_2}) = 1-(\alpha_1+\alpha_2) = 1-\alpha$

$$P\left(-z_{\alpha_1} < \frac{(F-p)\sqrt{n}}{\sqrt{p(1-p)}} < z_{\alpha_2}\right) = 1-\alpha$$

Từ đó ta thu được khoảng tin cậy của *p*: $\left(F - \frac{\sqrt{F(1-F)}}{\sqrt{n}} z_{\alpha_2}; F + \frac{\sqrt{F(1-F)}}{\sqrt{n}} z_{\alpha_1}\right)$

Khoảng tin cậy đối xứng khi $\alpha_1 = \alpha_2 = \alpha/2$: $\left(F - \frac{\sqrt{F(1-F)}}{\sqrt{n}} z_{\alpha/2}; F + \frac{\sqrt{F(1-F)}}{\sqrt{n}} z_{\alpha/2}\right)$

Khoảng tin cậy bên phải khi $\alpha_1 = 0; \alpha_2 = \alpha$: $\left(F - \frac{\sqrt{F(1-F)}}{\sqrt{n}} z_{\alpha}; +\infty\right)$

Khoảng tin cậy bên trái khi $\alpha_1 = \alpha; \alpha_2 = 0$: $\left(-\infty; F + \frac{\sqrt{F(1-F)}}{\sqrt{n}} z_{\alpha}\right)$

Với cùng độ tin cậy $1-\alpha$, khoảng tin cậy nào ngắn hơn sẽ tốt hơn. Trong trường hợp này độ dài khoảng tin cậy sẽ là ngắn nhất khi khoảng tin cậy đối xứng. Lúc đó độ dài khoảng tin cậy sẽ bằng hai lần độ chính xác và được xác định bằng biểu thức

$$I = 2\varepsilon = \frac{2\sqrt{F(1-F)}}{\sqrt{n}} z_{\alpha/2} \quad \left(\varepsilon = \frac{\sqrt{F(1-F)}}{\sqrt{n}} z_{\alpha/2}\right) \quad (*)$$

Từ (*) ta thu được công thức xác định kích thước mẫu tối thiểu *n* sao cho với độ tin cậy $1-\alpha$ cho trước, độ dài khoảng tin cậy không vượt quá giá trị I_0 cho trước. Công thức có dạng

$$n \geq \frac{F(1-F)}{\varepsilon_0^2} z_{\alpha/2}^2$$

Ví dụ 1

Kiểm ta ngẫu nhiên 400 sản phẩm do một máy sản suất thấy có 20 phế phẩm. Với độ tin cậy 0,95 hãy ước lượng tỷ lệ phế phẩm tối đa của nhà máy đó.

Giải

Gọi *p* là tỷ lệ phế phẩm của nhà máy đó

Khoảng tin cậy của *p* có dạng $\left(-\infty; F + \frac{\sqrt{F(1-F)}}{\sqrt{n}} z_{\alpha}\right)$

Qua mẫu cụ thể ta có $f = \frac{20}{400} = 0,05$

Với $1-\alpha = 0,95 \Rightarrow \alpha = 0,05 \Rightarrow z_\alpha = z_{0,05} = 1,645$.

Với độ tin cậy 0,95 khoảng tin cậy của p là $\left(-\infty; 0,05 + \frac{\sqrt{0,05(1-0,05)}}{\sqrt{400}}.1,645\right)$ hay

$(-\infty; 0,0679)$ tức là $p < 0,0679$ hay tỷ lệ phế phẩm tối đa của máy đó là 6,79%

Ví dụ 2

Mẫu điều tra về trọng lượng của 1 loại trái cây (đơn vị :gam) cho kết quả trong bảng

Trọng lượng	0-5	5-10	10-15	15-20	20-25	25-30	30-35
Số trái	5	7	13	22	30	18	10

Những trái cây có trọng lượng trên 25g là loại A. Hãy ước lượng tỷ lệ trái cây loại A với độ tin cậy 98%.

Giải

$1-\alpha = 0,98 \Rightarrow \alpha = 0,02 \Rightarrow z_{\alpha/2} = z_{0,01} = 2,33$

Từ bảng ta có n = 105. Số trai cây có trọng lượng trên 25g là 28

Tỷ lệ mẫu trái cây loại A $f = \frac{28}{105} = 0,2667$

Khoảng tin cậy cho tỷ lệ trái cây loại A với độ tin cậy 98% là

$\left(0,2667 - 2,33.\frac{\sqrt{0,2667.(1-0,2667)}}{\sqrt{105}}; 0,2667 + 2,33.\frac{\sqrt{0,2667.(1-0,2667)}}{\sqrt{105}}\right)$ hay

$(0,1661; 0,3673)$

Ước lượng phương sai tổng thể

Bài toán: Giả sử biến ngẫu nhiên gốc $X \sim N(\mu,\sigma^2), \sigma^2$ chưa biết. Với độ tin cậy $1-\alpha$, tìm khoảng ước lượng cuả σ^2

Trường hợp μ đã biết

Khi đó $G = \frac{nS^{*2}}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2(n)$

Ta có

$$P(G > \chi^2_{\alpha/2}(n)) = \frac{\alpha}{2}$$

$$P(G > \chi^2_{1-\alpha/2}(n)) = 1 - \frac{\alpha}{2}$$

với $\chi^2_{\alpha/2}(n)$ là giá trị tới hạn mức $\alpha/2$ của phân phối "Khi bình phương" *n* bậc tự do. Do đó

$$P(G > \chi^2_{\alpha/2}(n)) = \frac{\alpha}{2}$$
$$P\left(\chi^2_{1-\alpha/2}(n) < G < \chi^2_{\alpha/2}(n)\right) = 1-\alpha$$

Khoảng ước lượng của σ^2 : $\left(\frac{\sum_{i=1}^{n}\left(X_i-\mu\right)^2}{\chi^2_{\alpha/2}(n)};\frac{\sum_{i=1}^{n}\left(X_i-\mu\right)^2}{\chi^2_{1-\alpha/2}(n)}\right)$ với $\chi^2_{\alpha/2}(n)$ là giá trị tới hạn mức $\alpha/2$ của phân phối khi bình phương với *n* bậc tự do

Trường hợp μ chưa biết

Khi đó $G = \frac{n-1}{\sigma^2}S^2 \sim \chi^2\left(n-1\right)$

Khoảng tin cậy của σ^2: $\left(\frac{\left(n-1\right)S^2}{\chi^2_{\alpha/2}(n-1)};\frac{\left(n-1\right)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$ với $\chi^2_{\alpha/2}(n-1)$ là giá trị tới hạn mức $\alpha/2$ của phân phối khi bình phương với *n*-1 bậc tự do

Ví dụ 1

Kích thước của một chi tiết máy là một BNN có phân phối chuẩn. Trong một mẫu gồm 30 chi tiết được kiểm tra ta tính được $\overline{x} = 0,47$ và $s = 0,032.$ Với độ tin cậy 95%, hãy ước lượng khoảng cho phương sai và độ lệch chuẩn của kích thước chi tiết máy.

Giải

Ta có

$$n = 30; s = 0,032$$
$$1-\alpha = 0,95 \Rightarrow \alpha = 0,05$$
$$\Rightarrow \chi^2_{\alpha/2}(n-1) = \chi^2_{0,025}(29) = 45,772$$
$$\chi^2_{1-\alpha/2}(n-1) = \chi^2_{0,975}(29) = 16,047$$

Với độ tin cậy 95%, khoảng tin cậy cho phương sai là

$$\left(\frac{29.0,032^2}{45,772};\frac{29.0,032^2}{16,047}\right) \text{ hay } \left(0,000649;0,001851\right)$$

Từ đó ,với độ tin cậy 95%, khoảng tin cậy cho độ lệch chuẩn là

$$\left(\sqrt{0,000649};\sqrt{0,001851}\right) \text{ hay} \left(0,025;0,043\right)$$

Ví dụ

Trọng lượng của một loại mì gói là BNN có phân phối chuẩn. Kiểm tra ngẫu nhiên 15 gói mì ta được số liệu

Trọng lượng (g)	84	84,5	85	85,5
Số gói	2	3	8	2

Hãy tìm khoảng ước lượng cho phương sai với độ tin cậy 90%

HD

Từ mẫu ta có

$n = 15; s^2 = 0,02$

$1-\alpha = 0,9 \Rightarrow \alpha = 0,1$

$\Rightarrow \chi^2_{\alpha/2}(n-1) = \chi^2_{0,05}(14) = 23,685$

$\chi^2_{1-\alpha/2}(n-1) = \chi^2_{0,95}(14) = 6,571$

Mức hao phí nguyên liệu cho một đơn vị sản phẩm là biến ngẫu nhiên có phân phối chuẩn. Để ước lượng mức độ phân tán của mức hao phí này người ta cân thử 25 sản phẩm và thu được kết quả như sau:

Hao phí nguyên liệu (gam)	19,5	20	20,5
Số sản phẩm tương ứng	5	18	2

Với độ tin cậy 90% hãy tìm khoảng ước lượng cho σ^2

Ước lượng hiệu hai kì vọng toán của hai biến ngẫu nhiên phân phối chuẩn (Chưa xem từ đây trở xuống)

Giả sử có hai tổng thể nghiên cứu, trong đó các biến ngẫu nhiên X_1 và X_2 cùng phân phối chuẩn với các tham số đặc trưng tương ứng là μ_1, μ_2 và σ_1^2, σ_2^2 với μ_1, μ_2 chưa biết. Để ước lượng sự chênh lệch $\mu_1 - \mu_2$ giữa hai kỳ vọng toán từ hai tổng thể lập hai mẫu ngẫu nhiên độc lập kích thước tương ứng là n_1 và n_2

$W_1 = (X_{11}, X_{12}, \ldots, X_{1n_1})$

$W_2 = (X_{21}, X_{22}, \ldots, X_{2n_2})$

Từ đó ta tìm được các thống kê đặc trưng mẫu tương ứng là $\bar{X}_1, \bar{X}_2$ và S_1^2, S_2^2. Để chọn thống kê G thích hợp xét các trường hợp sau

Nếu đã biết phương sai σ_1^2, σ_2^2 của các tổng thể. Khi đó

$$G = U = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

U phân phối chuẩn tắc N(0,1). Với độ tin cậy $1-\alpha$ cho trước tìm được cặp giá trị α_1 và $\alpha_2(\alpha_1+\alpha_2=\alpha)$ và hai giá trị tới hạn chuẩn tương ứng là $u_{1-\alpha_1}, u_{1-\alpha_2}$ thỏa

$P(U < u_{1-\alpha_1}) = \alpha_1$

$P(U > u_{\alpha_2}) = \alpha_2$

Ta thu được khoảng tin cậy mức $1-\alpha$ của $\mu_1-\mu_2$ như sau:

$$P\left((\bar{X}_1-\bar{X}_2)-\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}u_{\alpha_2} < \mu_1-\mu_2 < (\bar{X}_1-\bar{X}_2)+\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}u_{\alpha_1} \right) = 1-\alpha$$

Nếu chưa biết các phương sai σ_1^2, σ_2^2 của các tổng thể song giả thiết rằng $\sigma_1^2=\sigma_2^2$

Chọn thống kê

$$G = T = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_p.\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \text{ trong đó } S_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

T phân phối T(n_1+n_2-2). Từ đó ta thu được khoảng tin cậy $1-\alpha$ của $\mu_1-\mu_2$ như sau:

$$P\left((\bar{X}_1-\bar{X}_2)-S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}t_{\alpha_2}^{(n_1+n_2-2)} < \mu_1-\mu_2 < (\bar{X}_1-\bar{X}_2)+S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}t_{\alpha_1}^{(n_1+n_2-2)} \right) = 1-\alpha$$

Nếu chưa biết các phương sai σ_1^2, σ_2^2 của các tổng thể và không có căn cứ để cho rằng chúng bằng nhau

Chọng thống kê

$$G = T = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

Thống kê T phân phối Student với số bậc tự do là $k = \dfrac{(n_1-1)(n_2-1)}{(n_2-1)C^2+(1-C)^2(n_1-1)}$

Với $C = \dfrac{\frac{S_1^2}{n_1}}{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$. Từ đó ta thu được khoảng tin cậy $1-\alpha$ của $\mu_1-\mu_2$ như sau:

$$P\left((\bar{X}_1-\bar{X}_2)-\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}t_{\alpha_2}^{(k)} < \mu_1-\mu_2 < (\bar{X}_1-\bar{X}_2)+\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}t_{\alpha_1}^{(k)} \right) = 1-\alpha$$

Ví dụ

Từ một chuồng nuôi lợn lấy ngẫu nhiên bốn con đem cân và thu được trọng lượng tương ứng của chúng là 64, 66, 89 và 77kg. Từ một chuồng khác lấy ra ba con đem cân thu được trọng lượng là 56, 71 và 53kg. Với độ tin cậy 95% hãy ước lượng sự khác biệt về trọng lượng trung bình của hai chuồng lơn đó. Giả thiết trọng lượng của lợn có phân phối chuẩn

Giải

Gọi X_1, X_2 tương ứng là trọng lượng của lợn ở hai chuồng nói trên. Theo giả thiết X_1, X_2 phân phối chuẩn. Gọi trọng lượng trung bình lần lượt là μ_1, μ_2. Đây là bài toán ước lượng hiệu số $\mu_1 - \mu_2$ khi chưa biết phương sai của tổng thể.

Nếu có thể cho rằng phương sai của chúng bang nhau (chẳng hạn cả hai chuồng cùng nuôi một giống lợn và được chăm sóc như nhau), từ đó suy ra khoảng tin cậy đối xứng như sau:

$$P\left((\bar{X}_1 - \bar{X}_2) - S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}t_{\alpha_2}^{(n_1+n_2-2)}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}t_{\alpha_1}^{(n_1+n_2-2)}} \right) = 1 - \alpha$$

Từ hai mẫu cụ thể ta có

$$n_1 = 4; \bar{x}_1 = 74; s_1^2 = 132,67$$
$$n_2 = 3; \bar{x}_2 = 60; s_2^2 = 93$$

Từ đó $S_p^2 = \dfrac{3.132,67 + 2.93}{4+3-2} = 116,8$

Ta có $t_{\alpha/2}^{(n_1+n_2-2)} = t_{0,025}^{(5)} = 2,57$

Do đó $14 - \sqrt{116,8}\sqrt{\frac{1}{4} + \frac{1}{3}}.2,57 < \mu_1 - \mu_2 < 14 + \sqrt{116,8}\sqrt{\frac{1}{4} + \frac{1}{3}}.2,57$ hay

$$-7,21 < \mu_1 - \mu_2 < 35,21$$

Ước lượng hiệu hai tham số p của hai biến ngẫu nhiên phân phối không- một

Giả sử có hai tổng thể nghiên cứu, trong đó các biến ngẫu nhiên X_1 và X_2 cùng phân phối không-một với tham số tương ứng là p_1, p_2. Nếu từ hai tổng thể trên rút ra hai mẫu ngẫu nhiên độc lập kích thước n_1 và n_2 và tìm được các tần suất tương ứng là f_1, f_2 thì có thể chọn lập thống kê

$G = U = \dfrac{(f_1 - f_2) - (p_1 - p_2)}{S_f}$ với

$$S_f = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \approx \sqrt{\frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2}}$$

Ta đã biết nếu n_1>30 và n_2>30 thì biến ngẫu nhiên U sẽ phân phối xấp xỉ N(0,1). Do đó ta thu được khoảng tin cậy mức $1-\alpha$ của hiệu $p_1 - p_2$ như sau

$$P\Big((f_1 - f_2) - S_f u_{\alpha_2} < p_1 - p_2 < (f_1 - f_2) + S_f u_{\alpha_1}\Big) = 1-\alpha$$

Từ công thức này có thể xây dựng các khoảng tin cậy đối xứng, khoảng tin cậy bên phải, khoảng tin cậy bên trái cho hiệu $p_1 - p_2$

Công thức tìm kích thước mẫu tối thiểu cần điều tra với cả hai mẫu sao cho với độ tin cậy $1-\alpha$ cho trước, độ dài khoảng tinc ậy không vượt quá $I_0 = 2\varepsilon_0$ cho trước có dạng

$$n \geq \frac{\big(f_1(1-f_1) + f_2(1-f_2)\big)}{\varepsilon_0^2} u_{\alpha/2}^2$$

Với f_1 và f_2 là tần suất của các mẫu sơ bộ kích thước m.

Ví dụ

Doanh nghiệp dự định đưa sản phẩm của mình vào hai thị trường khác nhau. Bán thử sản phẩm cho 100 khách hàng tiềm năng của thị trường thứ nhất thì có 50 người mua. Còn với thị trường thứ hai, khi bán thử sản phẩm cho 50 khách hàng tiềm năng thì có 20 người mua. Với độ tin cậy 95% hãy ước lượng mức độ chênh lệch về thị phần mà doanh nghiệp có thể đạt được tại hai thị trường đó

Gọi p_1, p_2 tương ứng là thị phần mà doanh nghiệp có thể đạt được ở hai thị trường. Vậy mức độ chênh lệch về thị phần là $p_1 - p_2$. Khoảng tin cậy đối xứng của $p_1 - p_2$

$$(f_1 - f_2) - S_f u_{\alpha/2} < p_1 - p_2 < (f_1 - f_2) + S_f u_{\alpha/2}$$

Với hai mẫu cụ thể

$n_1 = 100; n_2 = 50$

$f_1 = 0,5; f_2 = 0,4$

$$S_f = \sqrt{\frac{0,5,0,5}{100} + \frac{0,4.0,6}{50}} = 0,0084$$

Với $1-\alpha = 0,95 \Rightarrow \alpha = 0,05 \Rightarrow u_{\alpha/2} = u_{0,025} = 1,96$

Vậy $(0,5-0,4) - 0,0854.1,96 < p_1 - p_2 < (0,5-0,4) + 0,0854.1,96$

hay $0,0674 < p_1 - p_2 < 0,2674$

Ước lượng tỷ số của hai phương sai của hai biến ngẫu nhiên phân phối chuẩn

Giả sử có hai tổng thể nghiên cứu, trong đó các biến ngẫu nhiên X_1 và X_2 cùng phân phối chuẩn với phương sai σ_1^2, σ_2^2 chưa biết. Từ hai tổng thể trên lập hai mẫu ngẫu nhiên độc lập kích thước n_1 và n_2, từ đó tìm được các phương sai mẫu S_1^2, S_2^2 và tạo thống kê

$$G = F = \frac{S_1^2}{S_2^2} . \frac{\sigma_2^2}{\sigma_1^2} (S_1^2 > S_2^2)$$

Thống kê F phân phối $F(n_1-1,n_2-1)$. Với độ tin cậy $1-\alpha$ cho trước tìm được cặp giá trị α_1 và $\alpha_2 (\alpha_1 + \alpha_2 = \alpha)$ và hai giá trị tới hạn ch là $f_{1-\alpha_1}^{(n_1-1,n_2-1)}, f_{1-\alpha_2}^{(n_1-1,n_2-1)}$ thỏa

$$P(F < f_{1-\alpha_1}^{(n_1-1,n_2-1)}) = \alpha_1$$

$$P(F > f_{\alpha_2}^{(n_1-1,n_2-1)}) = \alpha_2$$

Từ đó

$$P\left(\frac{S_1^2}{S_2^2} f_{1-\alpha_1}^{(n_1-1,n_2-1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha_2}^{(n_1-1,n_2-1)} \right) = 1-\alpha. \text{ Từ đó}$$

$$P\left(\frac{S_1^2}{S_2^2} f_{1-\alpha_1}^{(n_1-1,n_2-1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha_2}^{(n_1-1,n_2-1)} \right) = 1-\alpha (*).$$

Từ (*) có thể xây dựng đucợ các khoảng tin cậy tương ứng với mỗi cặp giá trị α_1 và α_2. Việc tìm khoảng tin cậy của của $\frac{\sigma_1}{\sigma_2}$ được thực hiện bằng cách lấy căn bậc hai của cả ba vế của (*)

Ví dụ

Giá cổ phiếu của hai công ty A và B là các biến ngẫu nhiên có phân phối chuẩn. Thep dõi giá cổ phiếu của hai công ty đó trong 10 ngày tìm được phương sai mẫu tương ứng là 0,51 và 0,2. Với độ tin cậy 0,9 hãy ước lượng tỷ số của hai phương sai của giá cổ phiếu của hai công ty đó.

Giải

Gọi X_1, X_2 là giá cổ phiếu của hai công ty A và B. Theo giả thiết X_1, X_2 có phân phối chuẩn. Gọi các phương sau lần lượt là σ_1^2, σ_2^2. Khoảng tin cậy cần tìm là

$$\frac{S_1^2}{S_2^2} f_{1-\alpha_1}^{(n_1-1,n_2-1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha_2}^{(n_1-1,n_2-1)}$$

Từ mẫu cụ thể ta có

$n_1 = 10; s_1^2 = 0,51$

$n_2 = 10; s_2^2 = 0,2$

Với $1-\alpha = 0,9$ thì $f_{0,05}^{(9,9)} = 3,18,\ f_{0,95}^{(9,9)} = 3,18 = \dfrac{1}{f_{0,05}^{(9,9)} = 3,18} = \dfrac{1}{3,18} = 0,31$

Từ đó

$$\frac{0,51}{0,2}.0,31 < \frac{\sigma_1^2}{\sigma_2^2} < \frac{0,51}{0,2}.3,18 \text{ hay } 0,79 < \frac{\sigma_1^2}{\sigma_2^2} < 8,11$$

BÀI TẬP CHƯƠNG 6

Bài 1. Tuổi thọ của một loại bóng đèn do một dây chuyền công nghệ sản xuất ra có độ lệch chuẩn là 95 giờ. Điều tra 50 bóng đèn loại này tính được tuổi thọ trung bình là 350 giờ.

a) Với độ tin cậy 95% hãy ước lượng tuổi thọ trung bình của loại bóng đèn

b) Nếu muốn ước lượng tuổi thọ trung bình của loại bóng đèn đạt độ chính xác là 25 giờ và độ tin cậy 98% thì cần điều tra thêm bao nhiêu bóng nữa?

Bài 2. Một công ty sản xuất bột giặt muốn thăm dò mức độ tiêu thụ sản phẩm này trong một khu vực. Công ty tiến hành điều tra 500 hộ gia đình và có kết quả sau

Nhu cầu (kg/tháng)	0,5-1	1-1,5	1,5-2	2-2,5	2,5-3	3-3,5	3,5-4
Số hộ gia đình	21	147	192	78	34	16	12

Giả sử khu vực đó có 5000 hộ gia đình

a) Hãy ước lượng nhu cầu bột giặt gia đình của toàn khu vực trong một năm với độ tin cậy 95%

b) Những hộ có nhu cầu trên 2kg trong một tháng được gọi là những hộ có nhu cầu cao. Ước lượng tỷ lệ những hộ có nhu cầu cao với độ tin cậy 95%

c) Để ước lượng nhu cầu bột giặt trung bình của một hộ trong một tháng với độ chính xác 0,05kg và độ tin cậy 95% thì cần phải tra thêm bao nhiêu hộ gia đình nữa.

Bài 3. Để đánh giá trữ lượng cá trong hồ người ta bắt 2000 con cá đánh dấu rồi thả xuống hồ. Sau đó bắt lại 400 con thì thấy có 80 con đánh dấu.

a) Với độ tin cậy 95% hãy ước lượng trữ lượng cá trong hồ.

b)Nếu muốn sai số của ước lượng giảm đi một nửa thì cần phải đánh bắt thêm bao nhiêu con cá

Bài 4. X (đơn vị tính bằng %) là chỉ tiêu của một loại sản phẩm. Điều tra ở một số sản phẩm ta có kết quả

x_i	5-7	7-9	9-11	11-13	13-15	15-17	17-19	19-21
n_i	2	8	14	19	22	20	10	5

a) Để ước lượng trung bình chỉ tiêu X với độ tin cậy 92% và độ chính xác là 0,3% thì cần phải điều tra thêm bao nhiêu sản phẩm

b. Sản phẩm có chỉ tiêu X càng lớn càng được ưa chuộng. Người ta xem các sản phẩm có chỉ tiêu X dưới một mức quy định là loại II. Từ số liệu điều tra trên , bằng phương pháp ước lượng khoảng tỷ lệ (loại II), người ta tính được khoảng tin cậy là (4%,16%). Tìm độ chính xác và độ tin cậy của ước lượng này

c. Để ước lượng trung bình chỉ tiêu X với độ tin cậy 92% và độ chính xác 0,3% đồng thời ước lượng tỷ lệ sản phẩm loại II với độ chính xác 2,8% và độ tin cậy 90% thì cần điều tra thêm bao nhiêu sản phẩm.

Bài 5. Một khách hàng nhận được một lô hàng bút bi. Để ước lượng tỷ lệ bút bi hỏng, khách hàng lấy ngẫu nhiên 300 bút từ lô hàng để kiểm tra và thấy có 30 bút hỏng.

a) Ước lượng tỷ lệ bút bi hỏng với độ tin cậy 95%

b) Nếu muốn ước lượng tỷ lệ bút bi hỏng đạt độ tin cậy độ tin cậy 96% và độ chính xác 3% thì cần phải điều tra thêm bao nhiêu bút nữa

c) Nếu sử dụng mẫu điều tra tra, để ước lượng tỷ lệ bút bi hỏng đạt độ chính xác là 2,5% thì độ tin cậy là bao nhiêu.

Bài 6. Lượng nguyên liệu dùng để chế tạo sản phẩm là BNN có phân phối chuẩn với độ lệch chuẩn là σ. Quan sát một mẫu gồm 16 sản phẩm ta tính được $s = 24(g)$ Hãy ước lượng σ với độ tin cậy 95%

Bài 7. Trước bầu cử tổng thống người ta phỏng vấn ngẫu nhiên 1800 cử tri thấy có 1180 người ủng hộ ứng cử viên A. Với độ tin cậy 95% hỏi ứng viên đó thu được tối thiểu bao nhiêu phần trăm số phiếu bầu.

Bài 8. Đường kính trục của một loại sản phẩm là BNN tuân theo quy luật chuẩn. Đo đạc 6 trục ngẫu nhiên ta thu được kết quả như sau 7,1; 6,5; 9,7; 10,6; 7,5; 9,1. Với độ tin cậy 95%, tìm ước lượng cho phương sai.

1. Để xác định độ chính xác của một chiếc cân tạ không có sai số hệ thống, người ta tiến hành cân 5 lần độc lập (cùng một vật).

   Kết quả như sau: 94,1 94,8 96,0 95,4 95,2 (kg).

   Xác định ước lượng không chệch của phương sai số đo khi chưa biết trọng lượng của vật cân?

2. Điều tra năng suất lúa trên 100 hecta diện tích trồng lúa của một cùng người ta thu được bảng số liệu sau:

Năng suất (tạ/ha)	41	44	45	46	48	52	54
Diện tích có năng suất tương ứng	10	20	30	15	10	10	5

a. Tìm ước lượng không chệch của năng suất lua trung bình của vùng đó?

b. Tìm khoảng ước lượng (khoảng tin cậy) của năng suất lúa trung bình ở vùng đó với độ tin cậy 95%?

3. Điểm trung bình môn Toán của 100 sinh viên dự thi môn XSTK là 6 với độ lệch chuẩn mẫu đã điều chỉnh là 1,5.

   a. Ước lượng điểm trung bình môn XSTK của toàn thể sinh viên với độ tin cậy 95%?

   b. Với sai số 0,5 điểm. Hãy xác định độ tin cậy?

4. Tuổi thọ của một loại bóng đèn được biết theo qui luật chuẩn với độ lệch chuẩn 100h.

   a. Chọn ngẫu nhiên 100 bóng để thử nghiệm thì thấy mỗi bóng tuổi thọ trung bình là 1000h. Hãy ước lượng tuổi thọ trung bình của bóng đèn với độ tin cậy 95%?

   b. Với độ chính xác là 15h. Hãy xác định độ tin cậy?

   c. Với độ chính xác là 25h và định độ tin cậy là 95% thì cần thử nghiệm bao nhiêu bóng?

5. Một lô hàng có 5000 sản phẩm. Chọn ngẫu nhiên 400 sản phẩm từ lô hàng thì thấy có 360 sản phẩm loại A.

   a. Hãy ước lượng số sản phẩm loại A trong lô hàng với độ tin cậy 96%?

   b. Nếu muốn ước lượng số sản phẩm loại A của lô hàng đạt độ chính xác 150 sản phẩm và độ tin cậy 99% thì phải kiểm tra bao nhiêu sản phẩm?

6. Trọng lượng các bao xi măng ở một của hàng bán vật liệu xây dựng tuân theo qui luật chuẩn. Kiểm tra 20 bao thấy trọng lượng trung bình của mỗi bao là 48kg và phương sai mẫu điều chỉnh là $s^2 = (0,5kg)^2$

   a. Với độ tin cậy 95% hãy ước lượng trọng lượng trung bình của một bao xi măng thuộc của hàng đó?

   b. Với độ chính xác 0,26 kg, xác định độ tin cậy?

   c. Với độ chính xác 0,16 kg , độ tin cậy 95%. Hãy xác định cỡ mẫu n?

7. Để ước lượng tỷ lệ sản phẩm xấu của một kho đồ hộp, người ta kiểm tra ngẫu nhiên 100 hộp thấy có 11 hộp xấu.
   a. Ước lượng tỉ lệ sản phẩm xấu của kho đồ hộp với độ tin cậy 94%?
   b. Với sai số cho phép 3%, hãy xác định độ tin cậy?
8. Đo đường kính của 100 chi tiết do một máy sản xuất kết quả cho ở bảng sau:

Đường kính (mm)	Số chi tiết
19,80 – 19,85	3
19,85 – 19,90	5
19,90 – 19,95	16
19,95 – 20,00	28
20,00 – 20,05	23
20,05 – 20,10	14
20,10 – 20,15	7
20,15 – 20,20	4

Qui định những chi tiết có đường kính 19,9 mm đến 20,10 mm là những chi tiết đạt chuẩn.
   a. Ước lượng đường kính trung bình của những chi tiết đạt chuẩn với độ tin cậy 95%?
   b. Ước lượng tỷ lệ chi tiết đạt chuẩn với độ tin cậy 95,5%?
   c. Khi ước lượng đường kính của chi tiết đạt tieu chuẩn muốn độ chính xác đạt 0,02 mm và khi ước lượng tỉ lệ chi tiết đạt tiêu chuẩn muốn độ chính xác là 5%, với cùng độ tin cậy 99% thì cần đo thêm bao nhiêu chi tiết nữa?
9. Để ước lượng số cá trong hồ người ta đánh bắt 2000 con, đánh dấu rồi thả xuống hồ. Sau đó người ta đánh lên 400 con thì thấy có 40 con bị đánh dấu. Với độ tin cậy 95%, số cá trong hồ khoảng bao nhiêu con?

   ĐS: $(8368;12420)$
10. Chọn ngẫu nhiên 36 công nhân của một xí nghiệp thì thấy lương trung bình là 1,35 triệu/tháng. Giả sử đã biết lương công nhân xí nghiệp ấy tuân theo qui luật phân phối chuẩn với σ=0,2 triệu đồng. Với độ tin cậy 95%, háy ước lượng mức lương trung bình của công nhân xí nghiệp ấy?

11. Cần phải lập một mẫu ngẫu nhiên kích thước là bao nhiêu để tỉ lệ phế phẩm của mẫu là 0,2; độ chính xác là 0,025 và độ tin cậy của ước lượng là 95%? ĐS: $n \geq 985$

12. Đo chiều dài 25 chi tiết do một máy sản xuất ta có phương sai $\sigma^2 = 100cm^2$ và $\overline{x} = 100cm$. Giả sử chiều dài tuân theo qui luật phân phối chuẩn. Hãy ước lượng chiều dài trung bình của loại chi tiết đó với độ tin cậy 99%? ĐS: $(94,9;105,1)$

13. Cân thử 25 bao gạo người ta tính được trọng lượng trung bình của một bao gạo là $\overline{x} = 40kg$; độ lệch chuẩn hiệu chỉnh mẫu là $s = 5kg$. Với độ tin cậy 95% hãy tìm khoảng ước lượng cho trọng lượng trung bình của bao gạo, biết rằng trọng lượng bao gạo tuân theo qui luật phân phối chuẩn?

ĐS: $(37,936;42,064)$

14. Để định mức thời gian gia công một chi tiết máy người ta theo dõi ngẫu nhiên quá trình gia công 25 chi tiết máy và thu được số liệu sau:

Thời gian gia công (phút)	Số chi tiết tương ứng
15-17	1
17-19	3
19-21	4
21-23	12
23-25	3
25-27	2

Hãy ước lượng thời gian trung bình để gia công một chi tiết máy với độ tin cậy 95% biết thời gian gia công là đại lượng ngẫu nhiên tuân theo qui luật chuẩn.

ĐS: $(20,53;22,51)$

15. Điều tra năng suất lúa trên diện tích 100ha trồng lúa của một vùng người ta thu được bảng số liệu sau:

Năng suất (tạ/ha)	41	44	45	46	48	52	54
Diện tích có năng suất tương ứng	10	20	30	15	10	10	5

a) Tìm ước lượng không chệch của năng suất lúa vùng đó?

b) Tìm khoảng ước lượng của năng suất lúa trung bình với độ tin cậy 95%?

ĐS: a. 46 tạ/ha b. (45,356; 46,644)

CHƯƠNG 7. KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

7.1. Khái niệm

Giả thuyết thống kê: là giả thuyết về dạng phân phối xác suất của biến ngẫu nhiên gốc của tổng thể, các tham số đặc trưng của biến ngẫu nhiên này, tính độc lập của các biến ngẫu nhiên. Giả thuyết thống kê là những điều ta muốn bảo vệ hoặc ta nghi ngờ muốn bác bỏ, được phát biểu dưới dạng H_0. Đối lập với giả thuyết này (gọi là đối thuyết) kí hiệu H_1 , theo nghĩa rằng nếu bác bỏ H_0 thì chấp nhận H_1 và ngược lại.

Kiểm định giả thuyết thống kê: là phương pháp dùng công cụ của thống kê, từ các thông tin thu được trên mẫu điều tra cho kết luận về việc chấp nhận hay bác bỏ một giả thuyết thống kê.

Để kiểm định giả thuyết thống kê H_0, người ta tìm một miền W sao cho khi mẫu $(X_1, X_2, \ldots, X_n) \in \mathrm{W}$ thì ta bác bỏ giả thuyết H_0, còn khi $(X_1, X_2, \ldots, X_n) \notin \mathrm{W}$ thì ta chấp nhận giả thuyết H_0.

Miền W nói trên gọi là miền bác bỏ

Các loại sai lầm

Khi bác bỏ hay chấp nhận giả thuyết H_0 , ta có thể mắc phải một trong hai loại sai lầm sau:

- Sai lầm loại 1: H_0 đúng nhưng ta bác bỏ nó .
- Sai lầm loại 2: H_0 sai nhưng ta chấp nhận nó

Ta mong muốn chọn miền W sao cho khả năng mắc các sai lầm này là nhỏ nhất. Tuy nhiên, nếu ta muốn giảm khả năng phạm sai lầm loại 1 thì sẽ làm tăng khả năng phạm sai lầm loại 2 và ngược lại. Trong thực tế người ta tiến hành như sau: Người ta ấn định cho xác suất mắc lầm loại 1 không vượt quá một số α cho trước, trong các miền bác bỏ, chọn miền W sao cho khả năng phạm sai lầm loại 2 là nhỏ nhất.

α gọi là mức ý nghĩa của kiểm định.

3 dạng bài toán kiểm định chủ yếu:

- Kiểm định hai phía $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$
- Kiểm định một phía (phía trái) $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$

- Kiểm định một phia (phía phải) $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$

7.2. Các bài toán kiểm định: tham số, phi tham số

7.2.1. Kiểm định tham số

Kiểm định giả thuyết về trung bình tổng thể

Bài toán:

Giả sử biến ngẫu nhiên X trong tổng thể có $E(X) = \mu$ chưa biết, $V(X) = \sigma$. Với mức ý nghĩa α, kiểm định giả thuyết $H_0: \mu = \mu_0$ (μ_0 đã biết)

Từ tổng thể lập mẫu ngẫu nhiên kích thước n $(X_1, X_2, \ldots, X_n)$. Ta xét các trường hợp xảy ra như sau

Trường hợp 1: Mẫu lớn $(n \geq 30)$, đã biết σ^2. Giả sử H_0 đúng, khi đó

$G = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1)$ nếu X có phân phối chuẩn hoặc $G = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}$ có phân phối xấp xỉ N(0,1) nếu X không có giả thiết này.

Ta có các miền bác bỏ tương ứng với các dạng bài toán như sau

$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$; $W = \left\{ (X_1, X_2, \ldots, X_n) : \left| G = \dfrac{\overline{X} - \mu_0}{\sigma}\sqrt{n} \right| > z_{\frac{\alpha}{2}} \right\}$ với $z_{\frac{\alpha}{2}}$ là giá trị tới hạn chuẩn mức $\dfrac{\alpha}{2}$

$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$; $W = \left\{ (X_1, X_2, \ldots, X_n) : G = \dfrac{\overline{X} - \mu_0}{\sigma}\sqrt{n} > z_{\alpha} \right\}$

$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$; $W = \left\{ (X_1, X_2, \ldots, X_n) : G = \dfrac{\overline{X} - \mu_0}{\sigma}\sqrt{n} < -z_{\alpha} \right\}$

Trường hợp 2: Mẫu lớn $(n \geq 30)$, chưa biết σ^2. Khi đó $\dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$ có phân phối xấp xỉ N(0,1)

Các miền bác bỏ tương ứng với các dạng bài toán như sau

$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$; $W = \left\{ (X_1, X_2, \ldots, X_n) : \left| G = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n} \right| > z_{\frac{\alpha}{2}} \right\}$

$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$ $W = \left\{ (X_1, X_2, \ldots, X_n) : G = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n} > z_{\alpha} \right\}$

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \quad \mathrm{W} = \left\{ (X_1, X_2, \ldots, X_n) : G = \frac{\bar{X} - \mu_0}{S}\sqrt{n} < -z_\alpha \right\}$$

Trường hợp 3: Mẫu nhỏ $(n < 30)$, Tổng thể có phân phối chuẩn $N(\mu;\sigma^2)$, σ^2 đã biết.

Khi đó $\frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \sim N(0;1)$

Ta có các miền bác bỏ tương ứng với các dạng bài toán như sau

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} ; \mathrm{W} = \left\{ (X_1, X_2, \ldots, X_n) : \left| G = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \right| > z_{\frac{\alpha}{2}} \right\}$$

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} ; \mathrm{W} = \left\{ (X_1, X_2, \ldots, X_n) : G = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} > z_\alpha \right\}$$

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} ; \mathrm{W} = \left\{ (X_1, X_2, \ldots, X_n) : G = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} < -z_\alpha \right\}$$

Trường hợp 4: Mẫu nhỏ $(n < 30)$, Tổng thể có phân phối chuẩn $N(\mu;\sigma^2)$, σ^2 chưa biết. Khi đó $\frac{\bar{X} - \mu_0}{S}\sqrt{n} \sim T(n-1)$

Ta có các miền bác bỏ tương ứng với các dạng bài toán như sau

$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} ; \mathrm{W} = \left\{ (X_1, X_2, \ldots, X_n) : \left| G = \frac{\bar{X} - \mu_0}{S}\sqrt{n} \right| > t_{\frac{\alpha}{2}}(n-1) \right\}$ với $t_{\frac{\alpha}{2}}(n-1)$ là giá trị tới hạn mức $\frac{\alpha}{2}$ của phân phối Student n-1 bậc tự do

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} ; \mathrm{W} = \left\{ (X_1, X_2, \ldots, X_n) : G = \frac{\bar{X} - \mu_0}{S}\sqrt{n} > t_\alpha(n-1) \right\}$$

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} ; \mathrm{W} = \left\{ (X_1, X_2, \ldots, X_n) : G = \frac{\bar{X} - \mu_0}{S}\sqrt{n} < -t_\alpha(n-1) \right\}$$

Ví dụ 1: Trong năm trước trọng lượng trung bình trước khi xuất chuồng của bò ở một trang trại chăn nuôi là 380kg. Năm nay người ta áp dụng thử một chế độ chăn nuôi mới với hy vọng là bò sẽ tăng trọng nhanh hơn. Sau thời gian áp dụng thử, người ta lấy ngẫu nhiên 50 con bò trước khi xuất chuồng đem cân và tính được trọng lượng trung bình của chúng là 390kg. Vậy với mức ý nghĩa $\alpha = 0,01$ có thể cho rằng trọng lượng trung bình của bò trước khi xuất chuồng đã tăng lên hay không? Giả thiết trọng lượng của bò là biến ngẫu nhiên phân phối chuẩn với độ lệch chuẩn là 35,2kg.

Giải

"Lưu ý khi đặt bài toán kiểm định. Giả thuyết Ho là dấu =. Còn đối thuyết H1 thì dựa vào từ đề bài đã cho. Đọc kỹ đề bài để xem bài toán muốn quan tâm tới tình huống <,>, hoặc khác. Và có lúc ta cũng cần xem ý nghĩa thực tế để đặt H1 »

Gọi X là trọng lượng của bò trước khi xuất chuồng (đơn vị:kg), μ. là trọng lượng trung bình trước khi xuất chuồng.Theo giả thiết X có phân phối chuẩn với $\sigma = 35,2.$.

Bài toán kiểm định

$$\begin{cases} H_0 : \mu = 380 \\ H_1 : \mu > 380 \end{cases}$$

Ta có $\overline{x} = 390$;

$z_\alpha = z_{0,01} = 2,33$

Từ mẫu kiểm định ta có $g = \dfrac{390-380}{35,2}\sqrt{50} = 2,01 < 2,33$, do đó với mức ý nghĩa $\alpha = 0,01$, qua mẫu cụ thể đã cho chưa có cơ sở để bác bỏ H_0

Ví dụ 2:

Một công ty sản xuất hạt giống tuyên bố rằng một loại giống mới của họ có năng suất trung bình 21,5 tạ/ha. Gieo thử giống này tại 16 vườn thí nghiệm thì thu được kết quả:

19,2	18,7	22,4	20,3
16,8	25,1	17,0	15,8
21,0	18,6	23,7	24,1
23,4	19,8	21,7	18,9

Dựa vào kết quả này hãy nhận xét xem quảng cáo của công ty có đúng không với mức ý nghĩa 5%. Biết rằng năng suất giống cây trồng là một biến ngẫu nhiên có phân phối chuẩn

Giải

Gọi X là năng suất giống mới. Ta có X có phân phối chuẩn. Gọi μ là năng suất trung bình của loại giống mới.

Ta cần kiểm định giả thuyết: $\begin{cases} H_0 : \mu = 21{,}5 \\ H_1 : \mu \neq 21{,}5 \end{cases}$

Với α=0,05 ta có: $t_{\alpha/2}(n-1) = t_{0,025}(15) = 2{,}131$.

Từ mẫu trên ta tính được

$\bar{x} = 20{,}4062; s = 2{,}7999 \Rightarrow g = \dfrac{(20{,}4062 - 21{,}5)}{2{,}7999}\sqrt{16} = -1{,}5626$. Ta có $|g| < 2{,}131$

Vậy chưa có cơ sở để bác bỏ H_0. Có nghĩa là với số liệu này thì có thể chấp nhận lời quảng cáo của công ty

Kiểm định tỉ lệ

Kiểm định giả thuyết về tỉ lệ tổng thể

Bài toán: Giả sử p là tỷ lệ của tổng thể X, chưa biết. Với mức ý nghĩa α, kiểm định giả thuyết $H_0 : p = p_0$ (p_0 đã biết)

Từ tổng thể lập mẫu ngẫu nhiên kích thước n $(X_1, X_2, \ldots, X_n)$.

Khi $np_0 \geq 5$ và $n(1-p_0) \geq 5$, nếu H_0 đúng thì $G = \dfrac{(F - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$ có phân phối xấp xỉ N(0,1).

Với mức ý nghĩa α cho trước ta có miền bác bỏ tương ứng với các dạng bài toán như sau:

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}; \ \mathrm{W} = \left\{ (X_1, X_2, \ldots, X_n) : \left| G = \frac{(F - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \right| > z_{\alpha/2} \right\}$$

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \end{cases}; \ \mathrm{W} = \left\{ (X_1, X_2, \ldots, X_n) : G = \frac{(F - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} > z_{\alpha} \right\}$$

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases}; \ \mathrm{W} = \left\{ (X_1, X_2, \ldots, X_n) : G = \frac{(F - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} < -z_{\alpha} \right\}$$

Ví dụ

Một công ty tuyên bố rằng 40% dân chúng ưa thích sản phẩm của công ty. Một cuộc điều tra 400 người tiêu dùng có 120 ưa thích sản phẩm của công ty. Với mức ý nghĩa $\alpha = 5\%$ hãy kiểm định xem tuyên bố trên có cao hơn thực tế không

Giải

Gọi p là tỷ lệ khách hàng ưa thích sản phẩm của công ty.

Bài toán kiểm định

$$\begin{cases} H_0 : p = 0,4 \\ H_1 : p < 0,4 \end{cases}$$

Ta có $np_0 = 400.0,4 = 160 > 5; n(1 - p_0) = 400.(1 - 0,4) = 240 > 5$

$\alpha = 5\% \Rightarrow z_\alpha = z_{0,05} = 1,64$.

Từ mẫu ta có $f = \dfrac{120}{400} = 0,3; g = \dfrac{0,3 - 0,4}{\sqrt{\dfrac{0,4.0,6}{400}}} = -4,082 < -1,64$

Do đó ta bác bỏ H_0 chấp nhận H_1 hay tuyên bố của công ty cao hơn thực tế.

Ví dụ 2

Tỷ lệ khách hàng tiêu dùng thường xuyên một loại sản phẩm ở 1 địa phương nào đó là 60%. Sau một chiến dịch tiếp thị, người ta thấy dường như tỷ lệ này đã tăng lên. Ban lãnh đạo công ty quyết định điều tra xem quả thực tỷ lệ này đã tăng lên hay chưa, nếu không sẽ tiến hành quảng cáo trên các phương tiện thông tin đại chúng. Phỏng vấn ngẫu nhiên 400 người thì thấy có 250 người sử dụng thường xuyên sản phẩm đó. Với mức ý nghĩa $\alpha = 5\%$, xem tỷ lệ cũ đã tăng lên chưa

Giải

Gọi tỷ lệ khách hàng thường xuyên tiêu dùng sản phẩm của công ty là p.

Bài toán kiểm định

$$\begin{cases} H_0 : p = 0,6 \\ H_1 : p > 0,6 \end{cases}$$

$np_0 = 400.0,6 = 240 > 5; n(1 - p_0) = 400.(1 - 0,6) = 160 > 5$

$\alpha = 5\% \Rightarrow z_\alpha = z_{0,05} = 1,64$;

Từ mẫu ta có $f = \dfrac{250}{400}; g = \dfrac{\dfrac{250}{400} - 0,6}{\sqrt{\dfrac{0,6.(1 - 0,6)}{400}}} = 1,02 < z_\alpha = 1,64$ nên ta chưa có cơ sở để bác bỏ H_0 hay tỷ lệ chưa tăng đáng kể (hãy tiến hành quảng cáo nếu có thể)

Kiểm định phương sai của tổng thể

Bài toán: Giả sử tổng thể X có phân phối chuẩn $N(\mu,\sigma^2)$, phương sai σ^2 chưa biết.. Với mức ý nghĩa α hãy kiểm định giả thuyết sau: $H_0:\sigma^2=\sigma^2{}_0$.($\sigma^2{}_0$ đã biết)

Từ tổng thể rút ra mẫu ngẫu nhiên kích thước n $(X_1,X_2,\ldots,X_n)$.

Trường hợp $E(X)=\mu$ chưa biết.

Nếu giả thuyết H_0 đúng thì thống kê $G=\dfrac{(n-1)S^2}{\sigma_0^2}$ phân phối theo quy luật Khi bình phương với (n-1) bậc tự do.

Với mức ý nghĩa α cho trước và tùy thuộc dạng của đối thuyết H_1, miền bác bỏ W tương ứng với các dạng bài toán như sau :

$$\begin{cases}H_0:\sigma^2=\sigma_0^2\\ H_1:\sigma^2\neq\sigma_0^2\end{cases};\quad \text{W}=\left\{(X_1,X_2,\ldots,X_n):G=\frac{(n-1)S^2}{\sigma_0^2}<\chi^2_{1-\frac{\alpha}{2}}(n-1)\ hoac\ G>\chi^2_{\frac{\alpha}{2}}(n-1)\right\}$$

$$\begin{cases}H_0:\sigma^2=\sigma_0^2\\ H_1:\sigma^2>\sigma_0^2\end{cases};\quad \text{W}=\left\{(X_1,X_2,\ldots,X_n):G=\frac{(n-1)S^2}{\sigma_0^2}>\chi^2_{\alpha}(n-1)\right\}$$

$$\begin{cases}H_0:\sigma^2=\sigma_0^2\\ H_1:\sigma^2<\sigma_0^2\end{cases};\quad \text{W}=\left\{(X_1,X_2,\ldots,X_n):G=\frac{(n-1)S^2}{\sigma_0^2}<\chi^2_{1-\alpha}(n-1)\right\}$$

Trường hợp $E(X)=\mu$ đã biết.

Nếu giả thuyết H_0 đúng thì thống kê $G=\dfrac{\sum(X_i-\mu)^2}{\sigma_0^2}$ phân phối theo quy luật Khi bình phương với n bậc tự do.

Với mức ý nghĩa α cho trước và tùy thuộc dạng của đối thuyết H_1 miền bác bỏ W tương ứng với các dạng bài toán như sau :

$$\begin{cases}H_0:\sigma^2=\sigma_0^2\\ H_1:\sigma^2\neq\sigma_0^2\end{cases};\quad \text{W}=\left\{(X_1,X_2,\ldots,X_n):G=\frac{\sum(X_i-\mu)^2}{\sigma_0^2}<\chi^2_{1-\frac{\alpha}{2}}(\text{n})\ hoac\ G>\chi^2_{\frac{\alpha}{2}}(n)\right\}$$

$$\begin{cases}H_0:\sigma^2=\sigma_0^2\\ H_1:\sigma^2>\sigma_0^2\end{cases};\quad \text{W}=\left\{(X_1,X_2,\ldots,X_n):G=\frac{\sum(X_i-\mu)^2}{\sigma_0^2}>\chi^2_{\alpha}(n)\right\}$$

$$\begin{cases}H_0:\sigma^2=\sigma_0^2\\ H_1:\sigma^2<\sigma_0^2\end{cases};\quad \text{W}=\left\{(X_1,X_2,\ldots,X_n):G=\frac{\sum(X_i-\mu)^2}{\sigma_0^2}<\chi^2_{1-\alpha}(n)\right\}$$

Ví dụ

Để kiểm tra độ chính xác của một máy, người ta đo ngẫu nhiên kích thước của 15 chi tiết do máy đó sản xuất và tính được $s^2 = 14,6$. Với mức ý nghĩa $\alpha = 0,01$ hãy kết luận máy có hoạt động bình thường không, biết rằng kích thước chi tiết là biến ngẫu nhiên phân phối chuẩn có dung sai theo thiết kế là $\sigma^2 = 12$ (dung sai kích thước: phạm vi cho phép của sai số về kích thước)

Giải

Gọi X là kích thước chi tiết. Theo giả thuyết X có phân phối chuẩn.

Bài toán kiểm kịnh

$$\begin{cases} H_0 : \sigma^2 = 12 \\ H_1 : \sigma^2 > 12 \end{cases};$$

Ta có n=15; $g = \dfrac{(15-1).14,6}{12} = 17,033 < 29,145$

Do $\alpha = 0,01 \Rightarrow \chi^2_\alpha(14) = \chi^2_{0,01}(14) = 29,142$

Vậy chưa có cơ sở để bác bỏ H_0 hay máy móc hoạt động bình thường

"Kiểm định giả thuyết về phương sai có vai trò quan trọng trong thực tiễn vì phương sai đặc trưng cho nhiều chỉ tiêu kĩ thuật như độ chính xác của các chi tiết máy, của dụng cụ đo đạc. Thông thường trong thực tê phương sai lớn hơn giá trị xác định là bất lợi nên thường kiểm định bài toán $\begin{cases} H_0 : \sigma^2 = \sigma_0{}^2 \\ H_1 : \sigma^2 > \sigma_0{}^2 \end{cases};$ nhưng cũng tùy bài toán cụ thể"

Kiểm định so sánh hai trung bình

Bài toán: Giả sử có hai tổng thể X_1, X_2 có $E(X_1) = \mu_1, E(X_2) = \mu_2$ chưa biết. Với mức ý nghĩa α, kiểm định giả thuyết $H_0 : \mu_1 = \mu_2$

Từ hai tổng thể trên rút ra hai mẫu độc lập

$$\mathrm{W}_1 = (X_{11}, X_{12}, \ldots, X_{1n_1})$$
$$\mathrm{W}_2 = (X_{21}, X_{22}, \ldots, X_{2n_2})$$

Trường hợp 1 $n_1, n_2 \geq 30, \sigma_1^2, \sigma_2^2$ đã biết. Khi đó tiêu chuẩn kiểm định được chọn là thống kê $G = \dfrac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \approx N(0,1)$.

Do đó miền bác bỏ cho các dạng bài toán tương ứng như sau

$a) H_o: \mu_1 = \mu_2; H_1: \mu_1 > \mu_2$

$$W = \left\{ G = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}; G > z_\alpha \right\}$$

$b) H_o: \mu_1 = \mu_2; H_1: \mu_1 < \mu_2$

$$W = \left\{ G = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}; G < -z_\alpha \right\}$$

$c) H_o: \mu_1 = \mu_2; H_1: \mu_1 \neq \mu_2$

$$W = \left\{ G = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}; |G| > z_{\alpha/2} \right\}$$

Trường hợp 2 $n_1, n_2 \geq 30, \sigma_1^2, \sigma_2^2$ chưa biết. Khi đó tiêu chuẩn kiểm định được chọn là thống kê $G = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx N(0,1)$.

Do đó miền bác bỏ cho các dạng bài toán tương ứng như sau

$a) H_o: \mu_1 = \mu_2; H_1: \mu_1 > \mu_2$

$$W = \left\{ G = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}; G > z_\alpha \right\}$$

$b) H_o: \mu_1 = \mu_2; H_1: \mu_1 < \mu_2$

$$W = \left\{ G = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}; G < -z_\alpha \right\}$$

$c) H_o: \mu_1 = \mu_2; H_1: \mu_1 \neq \mu_2$

$$W = \left\{ G = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}; |G| > z_{\alpha/2} \right\}$$

Trường hợp $n_1, n_2 < 30$**,** X_1, X_2 có phân phối chuẩn, σ_1^2, σ_2^2 đã biết. Khi đó tiêu chuẩn kiểm định được chọn là thống kê $G = \dfrac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \approx N(0,1)$.

Do đó miền bác bỏ cho các dạng bài toán tương ứng như sau

$a) H_o : \mu_1 = \mu_2; H_1 : \mu_1 > \mu_2$

$$\mathrm{W} = \left\{ G = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}; G > z_\alpha \right\}$$

$b) H_o : \mu_1 = \mu_2; H_1 : \mu_1 < \mu_2$

$$\mathrm{W} = \left\{ G = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}; G < -z_\alpha \right\}$$

$c) H_o : \mu_1 = \mu_2; H_1 : \mu_1 \neq \mu_2$

$$\mathrm{W} = \left\{ G = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}; |G| > z_{\alpha/2} \right\}$$

Trường hợp $n_1, n_2 < 30 <$ X_1, X_2 có phân phối chuẩn, σ_1^2, σ_2^2 chưa biết nhưng giả định $\sigma_1^2 = \sigma_2^2$. Khi đó tiêu chuẩn kiểm định được chọn là thống kê $G = \dfrac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{S_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \approx T(n_1 + n_2 - 2)$. Với

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Do đó miền bác bỏ cho các dạng bài toán tương ứng như sau

$a) H_o: \mu_1 = \mu_2; H_1: \mu_1 > \mu_2$

$$W = \left\{ G = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}; G > t_\alpha(n_1 + n_2 - 2) \right\}$$

$b) aH_o: \mu_1 = \mu_2; H_1: \mu_1 < \mu_2$

$$W = \left\{ G = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}; G < -t_\alpha(n_1 + n_2 - 2) \right\}$$

$c) H_o: \mu_1 = \mu_2; H_1: \mu_1 \neq \mu_2$

$$W = \left\{ G = \frac{\left(\overline{X_1} - \overline{X_2}\right)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}; |G| > t_{\alpha/2}(n_1 + n_2 - 2) \right\}$$

Qua các mẫu cụ thể ta tiến hành các giá trị quan sát của kiểm định rồi so sánh với W và kết luận

Ví dụ

Tại một xí nghiệp người ta xây dựng hai phương án gia công cùng một loại chi tiết . Để đánh giá xem chi phí trung bình về nguyên liệu theo hai phương án ấy có khác nhau hay không người ta tiến hành sản xuất thử và thu được kết quả như sau:

Phương án 1: 2,5 3,2 3,5 3,8 3,5

Phương án 2: 2,0 2,7 2,5 2,9 2,3 2,6

Với mức ý nghĩa $\alpha = 0,05$ hãy kết luận về vấn đề trên biết rằng chi phí nguyên liệu theo cả hai phương án gia công đều là các biến ngẫu nhiên phân phối chuẩn với $\sigma_1^2 = \sigma_2^2 = 0,16$

Giải

Gọi X_1, X_2 tương ứng là chi phí nguyên liệu theo hai phương án gia công trên . Theo giả thiết X_1, X_2 phân phối chuẩn.

Gọi chi phí nguyên liệu trung bình theo các phương án đó là μ_1, μ_2. Ta có bài toán kiểm định

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

Do $\alpha = 0,05 \Rightarrow z_{\alpha/2} = z_{0,025} = 1,96.$ Vậy miền bác bỏ là $(-\infty, -1,96)$ và $(1,96, +\infty)$

Từ mẫu cụ thể ta tính được

$$\overline{x_1} = \frac{2,5+3,2+3,5+3,8+3,5}{5} = 3,3$$

$$\overline{x_2} = \frac{2,0+2,7+2,5+2,9+2,3+2,6}{6} = 2,5$$

$$g = \frac{3,3-2,5}{\sqrt{\frac{0,16}{5}+\frac{0,16}{6}}} = 3,33 > 1,96$$

Do đó ta bác bỏ H_0, thừa nhận H_1 tức là chi phí nguyên liệu theo hai phuong án đó thực sự khác nhau.

Kiểm định so sánh hai tỉ lệ

Bài toán: Gỉa sử từ hai tổng thể X_1, X_2 có cùng đặc trưng định tính A. X_1 có tỷ lệ p_1 về đặc tính A, X_2 có tỷ lệ p_2 về đặc tính A. Hãy kiểm định giả thiết sau $H_0 : p_1 = p_2$ với mức ý nghĩa α.

Với Giả thiết thêm $n_1, n_2 \geq 30$.

Từ hai tổng thể rút ra hai mẫu ngẫu nhiên độc lập khích thước tương ứng là n_1 và n_2:

$$\mathrm{W}_1 = (X_{11}, X_{12}, \ldots, X_{1n_1})$$

$$\mathrm{W}_2 = (X_{21}, X_{22}, \ldots, X_{2n_2})$$

Chọn lập tiêu chuẩn thống kê $G = \dfrac{(F_1 - F_2) - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$

Khi $n_1, n_2 \geq 30$ thì $G = \dfrac{(F_1 - F_2) - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$ có phân phối xấp xỉ N(0,1). Nếu H_0 đúng $(p_1 = p_2 = p)$ thì G trở thành

$$G = \frac{(F_1 - F_2)}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Thông thường p chưa biết nên thay bằng ước lượng của nó là $\overline{F} = \dfrac{n_1F_1 + n_2F_2}{n_1 + n_2}$.

Do đó ta có $G = \dfrac{(F_1 - F_2)}{\sqrt{\overline{F}(1-\overline{F})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ phân phối xấp xỉ chuẩn N(0,1) nếu $n_1, n_2 \geq 30$. Do đó với mức ý nghĩa α cho trước và tùy thuộc dạng của đối thuyết H_1 miền bác bỏ W được xây dựng theo các trường hợp sau:

$a) H_0 : p_1 = p_2; H_1 : p_1 > p_2;$

$$\mathrm{W} = \left\{ G = \frac{(F_1 - F_2)}{\sqrt{\overline{F}(1-\overline{F})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}; G > z_\alpha \right\}$$

$b) H_0 : p_1 = p_2; H_1 : p_1 < p_2$

$$\mathrm{W} = \left\{ G = \frac{(F_1 - F_2)}{\sqrt{\overline{F}(1-\overline{F})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}; G < -z_\alpha \right\}$$

$c) H_0 : p_1 = p_2; H_1 : p_1 \neq p_2$

$$\mathrm{W} = \left\{ G = \frac{(F_1 - F_2)}{\sqrt{\overline{F}(1-\overline{F})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}; |G| > z_{\alpha/2} \right\}$$

Với hai mẫu cụ thể ta tính được các giá trị cụ thể của $f_1, f_2, \overline{f}$ và giá trị g , so sánh với W để kết luận

"Chú ý: Ta có

$$E(X_1) = p_1; V(X_1) = p_1(1 - p_1);$$
$$E(X_2) = p_2; V(X_2) = p_2(1 - p_2)$$

->Đây là bài toán so sánh hai trung bình với phương sai chưa biết"

Ví dụ

Kiểm tra ngẫu nhiên các sản phẩm cùng loại do hai nhà máy sản xuất được các số liệu sau:

Nhà máy	Số sản phẩm được kiểm tra	Số phế phẩm
A	$n_1 = 1000$	$x_1 = 20$

B $n_2 = 900$ $x_2 = 30$

Với mức ý nghĩa $\alpha = 5\%$ có thể coi tỷ lệ phế phẩm của hai nhà máy như nhau được không

Giải

Gọi p_1, p_2 tương ứng là tỷ lệ phế phẩm của hai nhà máy A và B.

Cặp giả thuyết

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$$

Tiêu chuẩn kiểm định

$$G = \frac{\left(F_1 - F_2\right)}{\sqrt{\overline{F}(1-\overline{F})\left(\frac{1}{1000} + \frac{1}{900}\right)}}$$

$\alpha = 5\% \Rightarrow z_{\alpha/2} = z_{0,025} = 1,96$ nên miền bác bỏ $(-\infty; -1,96)$ và $(1,96; +\infty)$

Với hai mẫu cụ thể ta có

$$f_1 = \frac{20}{1000} = 0,02; \ f_2 = \frac{30}{900} = 0,033; \ \overline{f} = \frac{20+30}{1000+900} = 0,0263$$

$$g = \frac{\left(f_1 - f_2\right)}{\sqrt{\overline{f}(1-\overline{f})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0,02 - 0,033}{\sqrt{0,0263(1-0,0263)\left(\frac{1}{1000} + \frac{1}{900}\right)}} = -1,81$$

Vậy chưa có cơ sở để bác bỏ H_0, tức là có thể coi tỷ lệ phế phẩm ở hai nhà máy là như nhau.

Kiểm định giả thuyết về sự bằng nhau của hai phương sai

Bài toán: Giả sử có hai tổng thể nghiên cứu trong đó biến ngẫu nhiên gốc X_1 trong tổng thể thứ nhất phân phối $N_1(\mu_1, \delta_1^2)$ và biến ngẫu nhiên gốc X_2 trong tổng thể thứ hai phân phối $N_1(\mu_2, \delta_2^2)$. σ_1^2 và σ_2^2 chưa biết. Hãy kiểm định giả thuyết

$$H_0 : \sigma_1^2 = \sigma_2^2$$

Bước 1: Từ hai tổng thể rút ra hai mẫu ngẫu nhiên độc lập khích thước tương ứng là n_1 và n_2:

$$W_1 = (X_{11}, X_{12}, \ldots, X_{1n_1})$$
$$W_2 = (X_{21}, X_{22}, \ldots, X_{2n_2})$$

Khi H_0 đúng , chọn tiêu chuẩn kiểm định thống kê

$G = \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \dfrac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$ nếu $S_1^2 > S_2^2$. Do đó với mức ý nghĩa α cho trước và tùy thuộc dạng của đối thuyết H_1 miền bác bỏ W được xây dựng theo các trường hợp sau:

$a) H_o : \sigma_1^2 = \sigma_2^2; H_1 : \sigma_1^2 > \sigma_2^2$

$$\mathrm{W} = \left\{ G = \frac{S_1^2}{S_2^2} : G > f_\alpha(n_1 - 1, n_2 - 1) \right\}$$

$b) H_o : \sigma_1^2 = \sigma_2^2; H_1 : \sigma_1^2 \neq \sigma_2^2$

$$\mathrm{W} = \left\{ G = \frac{S_1^2}{S_2^2}; G < f_{1-\alpha/2}(n_1 - 1, n_2 - 1) hoac\ G > f_{\alpha/2}(n_1 - 1, n_2 - 1) \right\}$$

Với hai mẫu cụ thể tìm được s_1^2, s_2^2 và các giá trị quan sát F_{qs} của tiêu chuẩn kiểm định, so sánh với W và kết luận

Chú ý: Nếu $S_1^2 < S_2^2$. ta lấy $G = \dfrac{S_2^2}{S_1^2} \sim F(n_2 - 1, n_1 - 1)$

Ví dụ

Có hai giống lúa có năng suất trung bình xấp xỉ như nhau song mức độ phân tán về năng suất có thể khác nhau. Để kiểm tra điều đó người ta gặt mẫu tại hai vùng trồng hai giống lúa đó và thu được kết quả sau

Giống lúa	Số điểm gặt	Phương sai
A	$n_1 = 41$	$s_1^2 = 11,41$
B	$n_2 = 30$	$s_2^2 = 6,52$

Với mức ý nghĩa $\alpha = 5\%$ hãy kết luận về vấn đề trên, biết năng suất lúa là biến ngẫu nhiên có phân phối chuẩn

Giải

Gọi X_1, X_2 là năng suất của hai giống lúa A và B. X_1, X_2 có phân phối chuẩn.

Bài toán kiểm định cặp giả thuyết

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

Tiêu chuẩn kiểm định $F = \dfrac{S_1^2}{S_2^2}$

Do

$$\alpha = 5\% \Rightarrow f_{\alpha/2}(n_1 - 1, n_2 - 1) = f_{0,025}(40,29) = 2,03;\ f_{1-\alpha/2}(40,29) = \frac{1}{f_{\alpha/2}(29,40)} = \frac{1}{1,94} = 0,52$$

Miền bác bỏ $(0;0,52)$ và $(2,03;+\infty)$

Với mẫu cụ thể ta có $g = \frac{11,41}{6,52} = 1,75 \in [0,52;2,03]$

Vậy chưa có cơ sở để bác bỏ H_0 hay độ phân tán của năng suất hai giống lúa trên là như nhau.

7.2.2. Bài toán kiểm định phi tham số

Kiểm định giả thuyết về qui luật phân bố xác suất của biến ngẫu nhiên

Giả sử chưa biết qui luật phân phối xác suất của biến ngẫu nhiên gốc X trong tổng thể song ta có cơ sở để giả thiết X phân phối theo qui luật A nào đó. Ta có thể kiểm định:

Giả thiết H_0: X có phân phối xác suất theo qui luật A.

Đối thiết H_1: X có phân phối xác suất không theo qui luật A.

Sau đây ta sử dụng tiêu chuẩn phù hợp Pearson để kiểm định giả thuyết trên.

Phân hoạch miền giá trị của X thành k khoảng rời nhau $C_1, C_2, ..., C_k$. Từ tổng thể rút ra mẫu ngẫu nhiên kích thước n và gọi $n_1, n_2, ..., n_k$ là tần số tương ứng của mẫu nằm trong k khoảng này (ta gọi là tần số thực nghiệm).

Nếu giả thiết H_0 đúng ta có thể tính được các xác suất $p_i = P\left(X \in C_i\right), \left(i = \overline{1,k}\right)$

Từ đó ta có tần số lí thuyết $n_i = np_i$

Tiêu chuẩn được chọn kiểm định là thống kê $T = \sum_{i=1}^{k} \frac{\left(n_i - n_i\right)^2}{n_i}$

Người ta chứng minh được rằng không phụ thuộc vào qui luật phân phối của biến ngẫu nhiên gốc X, khi $n \to +\infty$ thống kê T sẽ hội tụ về phân phối Khi bình phương với *(k-r-1)* bậc tự do, trong đó r là số tham số cần ước lượng của qui luật cần kiểm định. Chẳng hạn nếu qui luật cần kiểm định là phân phối chuẩn thì r=2; còn qui luật cần kiểm định là Poisson thì r=1. *Các tham số này được tính bằng ước lượng hợp lí cực đại.*

Qui tắc kiểm định

- Bước 1: Tính các tham số cần ước lượng bằng phương pháp hợp lí cực đại.

- Bước 2: Tính thống kê $T = \sum_{i=1}^{k} \frac{\left(n_i - n_i\right)^2}{n_i}$ và $\chi_\alpha^2\left(k-r-1\right)$ với α là mức ý nghĩa.
- Bước 3: So sánh T và $\chi_\alpha^2\left(k-r-1\right)$

  + Nếu T> $\chi_\alpha^2\left(k-r-1\right)$ thì bác bỏ H_0.

  + Nếu T≤ $\chi_\alpha^2\left(k-r-1\right)$ thì chưa có cơ sở bác bỏ H_0.

Chú ý rằng tiêu chuẩn phù hợp của Pearson chỉ áp dụng được khi kích thước mẫu đủ lớn $\left(n > 50\right)$ và các tần số thực nghiệm cũng đủ lớn $\left(n_i \geq 5\right)$. Khi các tần số quá nhỏ ta phải ghép các giá trị hay các khoảng giới hạn của mẫu để tăng giá trị tần số thực nghiệm lên.

Ví dụ: Số cuộc gọi đến một trạm điện thoại cho bởi bảng sau

Số cuộc gọi	0	1	2	3	4	5	≥6
Số phút tương ứng	18	22	25	19	11	3	2

Với mức ý nghĩa $\alpha = 0,01$ có thể coi X phân phối theo qui luật Poisson được không?

Giải:

Giả thiết H_0: X có phân phối xác suất theo qui luật Poisson.

Đối thiết H_1: X có phân phối xác suất không theo qui luật Poisson.

Từ mẫu ta tính được $\overline{x} = 2$ là ước lượng hợp lí cực đại của tham số λ trong phân phối Poisson. Để áp dụng được tiêu chuẩn kiểm định ta phải nhập mẫu lại như sau:

Số cuộc gọi	0	1	2	3	4	≥5
Số phút tương ứng	18	22	25	19	11	5

Để tính giá trị quan sát ta lập bảng sau:

x_i	n_i	$p_i = \frac{e^{-2}2^{x_i}}{x_i!}$	$n_i = np_i$	$n_i - n_i$	$\frac{\left(n_i - n_i\right)^2}{n_i}$
0	18	0,1353	13,53	4,47	1,477
1	22	0,2707	27,07	-5,07	0,95

2	25	0,2707	27,07	-2,07	0,158
3	19	0,1804	18,04	0,96	0,051
4	11	0,0902	9,02	1,98	0,435
≥5	5	0,0527	5,27	-0,27	0,014
$\sum$	100	1,0000	100		3,085

Do $\alpha = 0,01 \Rightarrow \chi^2_\alpha(k-r-1) = \chi^2_{0,01}(6-1-1) = 13,277$

Do $T < \chi^2_{0,01}(4)$ nên ta chưa có cơ sở để bác bỏ H$_0$; hay ta có thể coi X có phân phối Poisson.

Kiểm định giả thiết về tính độc lập của hai dấu hiệu định tính

Giả sử cần nghiên cứu đồng thời hai dấu hiệu định tính A và B. Ta chia dấu hiệu A thành h mức độ $A_1, A_2, ..., A_h$ và chia B thành k cấp độ $B_1, B_2, ..., B_k$.

Ta kiểm định giả thiết

Giả thiết H$_0$: A và B độc lập.

Đối thiết H$_1$: A và B phụ thuộc.

Để kiểm định giả thiết trên từ tổng thể ta lập mẫu kích thước n và trình bày dưới dạng sau:

	B_1	B_2	……..	B_k	$\sum$
A_1	n_{11}	n_{12}		n_{1k}	n_{1*}
A_2	n_{21}	n_{22}		n_{2k}	n_{2*}
….	….	….	…..	…..	
A_h	n_{h1}	n_{h2}	….	n_{hk}	n_{h*}
$\sum$	n_{*1}	n_{*2}		n_{*k}	n

Với n đủ lớn theo định nghĩa thống kê về xác suất ta có:

$$P(A_iB_j) \approx \frac{n_{ij}}{n} \qquad P(A_i) \approx \frac{n_{i*}}{n} \qquad P(B_j) \approx \frac{n_{*j}}{n}$$

Nếu giả thiết H$_0$ đúng thì ta có:

$$P(A_iB_j) = P(A_i).P(B_j),\ \forall i = \overline{1,h};\ \forall j = \overline{1,k}$$

Tức là $\frac{n_{ij}}{n} = \frac{n_{i*}}{n}.\frac{n_{*j}}{n}$

Xét thống kê $T = n\left(\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{n_{ij}^2}{n_{i*}.n_{*j}} - 1\right)$

Ta có thể chứng minh được rằng khi $n \to \infty$ phân phối xác suất của T dần tới phân phối Khi bình phương (h-1)(k-1) bậc tự do. Vì vậy khi n đủ lớn, mức ý nghĩa α ta có qui tắc kiểm định sau:

Bước 1: Tính thống kê $T = n\left(\sum_{i=1}^{h}\sum_{j=1}^{k}\frac{n_{ij}^2}{n_{i*}.n_{*j}} - 1\right)$ và số $\chi^2(h-1)(k-1)$

Bước 2: Nếu $T > \chi^2(h-1)(k-1)$ thì bác bỏ H_0.

Bước 3: Ngược lại ta chấp nhận H_0.

Ví dụ: Bảng sau cho ta tần số sinh viên ứng với 3 mức của A khả năng học toán: thấp, trung bình, cao) và 3 mức của B (sự quan tâm đến xác suất thống kê). Với mức ý nghĩa 0,01 hãy kiểm định giả thiết A và B độc lập.

	Thấp	Trung bình	Cao	Tổng
Thấp	43	24	15	82
Trung bình	31	46	37	114
Cao	12	20	49	81
Tổng	86	90	101	277

Với $\alpha = 0,01 \Rightarrow \chi_\alpha^2(h-1)(k-1) = \chi_{0,01}^2(4) = 13,28$

Từ mẫu cụ thể ta tính được giá trị kiểm định là T=33,533>13,28 nên H_0 bị bác bỏ.

BÀI TẬP CHƯƠNG 7

Bài 1. Ở một đại lý bán hàng, lúc đầu trung bình một ngày bán được 300kg hàng. Người ta tiến hành một biện pháp quảng cáo và với ảnh hưởng quảng cáo này, theo dõi 20 ngày về số hàng bán trong ngày và tính được $\overline{x} = 306,2kg;\ s^2 = 54,6$.

Cho nhận xét về hiệu quả quảng cáo với mức ý nghĩa $\alpha = 5\%$. (Giả thiết số hàng bán trong ngày có phân phối chuẩn)

Bài 2. Một nhà máy sản xuất xăm lốp ô tô tuyên bố rằng tuổi thọ trung bình một chiếc lốp ô tô của họ là 30 000 dặm. Cơ quan giám định chất lượng nghi ngờ lời tuyên bố này đã kiểm tra 100 chiếc lốp và tìm được trung bình mẫu là $\overline{x} = 29000$ và độ lệch chuẩn là 5000 dặm. Với mức ý nghĩa $\alpha = 0,05$ cơ quan giám định có bác bỏ được lời quảng cáo của nhà máy nói trên không

Bài 3 Một nhà máy sản xuất sản phẩm với tỷ lệ sản phẩm hỏng là 0,08. Người ta tiến hành một cải tiến kỹ thuật và cho máy sản xuất 320 sản phẩm thấy có 16 sản phẩm hỏng. Cho nhận xét về hiệu quả của cải tiến với mức ý nghĩa $\alpha = 5\%$

Bài 4. Một kho hạt giống có tỷ lệ này mầm xác định $p_o = 0,9$. Ngẫu nhiên có một thiết bị bị hỏng làm thay đổi điều kiện bên trong của kho. Hỏi tỷ lệ nảy mầm của kho còn giữ nguyên hay không, với mức ý nghĩa $\alpha = 5\%$, biết rằng để có thông tin về tỷ lệ này mới ,người ta làm thí nghiệm 200 hạt thấy có 140 hạt nảy mầm

Bài 5. Điều tra doanh số bán hàng X (đơn vị: triệu đồng/tháng) của các hộ kinh doanh một loại hàng nă, nay cho số liệu

Doanh số (triệu đồng/tháng)	11	11,5	12	12,5	13	13,5
Số hộ	10	15	20	30	15	10

a. Những hộ có doanh số trên 12,5 triệu đồng/tháng là những hộ có doanh số cao. Có tài liệu cho rằng tỷ lệ hộ có doanh số cao là 35%. Cho nhận xét về tỷ lệ trong tài liệu đó với mức ý nghĩa $\alpha = 5\%$

b. Năm trước doanh số bán hàng trung bình của các hộ này là 120 triệu/năm. Có thể cho rằng doanh số bán hàng của các hộ này tăng lên không với mức ý nghĩa 5%.

Bài 6. Trọng lượng sản phẩm của một công ty là biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn là 0,9 (kg). Lấy một mẫu gồm 10 sản phẩm và tính được $s = 1,2(kg)$. Người ta nghi ngờ máy đóng gói sản phẩm hoạt động không ổn định và cho rằng độ lệch chuẩn đã tăng lên. Hãy cho nhận xét với mức ý nghĩa 0,05

Bài 7. Giám đốc một xí nghiệp cho biết lương trung bình của 1 công nhân thuộc xí nghiệp là 2 triệu đồng một tháng.

Chọn ngẫu nhiên 40 công nhân thấy lương trung bình là 1,8 triệu một tháng và độ lệch chuẩn là 500 ngàn. Lời bá cáo của giám đốc có tin cậy được không với mức ý nghĩa là 5%?

1. Theo một nguồn tin thì tỉ lệ hộ dân thích xem ca nhạc trên tivi là 80%. Thăm dò 49 hộ dân thấy có 25 hộ thích ca nhạc.

   Với mức ý nghĩa 5%. Kiểm định xem nguồn tin ấy có đáng tin không?

2. Một cửa hàng tạp hoá nhận thấy thời gian vừa qua trung bình một khách hàng mua 25 ngàn thuốc lá trong ngày. Nay cửa hàng chọn ngẫu nhiên 15 khách hàng thấy trung bình một khách hàng mua 24 ngàn đồng và phương sai mẫu điều chỉnh là 2 ngàn đồng.

   Với mức ý nghĩa 5%, thử xem có phải sức mua của khách hàng hiện nay đã giảm sút?
3. Một máy sản xuất tự động lúc đầu tỉ lệ sản phẩm loại A là 20%. Sau khi áp dụng một phương pháp sản xuất mới, người ta lấy 40 mẫu, mỗi mẫu gồm 10 sản phẩm để kiểm tra. Kết quả kiểm tra cho ở bảng sau :

Số sp loại A trong mẫu	1	2	3	4	5	6	7	8	9	10
Số mẫu	2	0	4	6	8	10	4	5	1	0

Với mức ý nghĩa 5%. Hãy cho kết luận về phương pháp sản xuất này?

4. Nếu công ty hoạt động bình thường thì lương của một nhân viên là một đại lượng ngẫu nhiên phân phối theo qui luật chuẩn với độ lệch chuẩn $\sigma = 2,5$ (triệu đồng).

   Kiểm tra ngẫu nhiên lương của một vài nhân viên ta có kết quả (tính theo triệu đồng) sau:

60	60,2	70	60,8	50,6	50,8	50,9	60,1
50,3	70	60,5	60	50,8	60,3	60,1	60,2

   a. Với mức ý nghĩa 5% hãy cho biêt công ty có hoạt động bình thường không?
   b. Ước lượng mức lương trung bình của nhân viên công ty này với độ tin cậy 95%?

5. Trong thập niên 80 trọng lượng trung bình của thanh niên là 48kg. Nay để xác định lại trọng lượng ấy người ta chọn ngẫu nhiên 100 thanh niên đo trọng lượng được trung bình là 50kg và phương sai mẫu hiệu chỉnh là $s^2 = 10kg^2$
   a. Thử xem trọng lượng thanh niên hiện nay có thay đổi, với mức ý nghĩa 1%?
   b. Nếu trọng lượng thực tế của thanh niên là a=51kg thì xác suất mắc sai lầm loại 2 là bao nhiêu?
   c. Nếu muốn xác suất mắc sai lầm loại 1 là 1% và xác suất mắc sai lầm loại 2 không vượt quá 5% thì phải đo trọng lượng của bao nhiêu thanh niên nếu trọng lượng trung bình thực tế của thanh niên hiện nay không vượt quá 52kg?
   d. Nếu muốn xác suất mắc sai lầm loại 1 là 1% và xác suất mắc sai lầm loại 2 không vượt quá 5% thì phải đo trọng lượng của bao nhiêu thanh niên nếu trọng lượng trung bình thực tế của thanh niên hiện nay trong khoảng (44 ; 52) kg
6. Tỉ lệ phế phẩm của một nhà máy trước đây là 5%. Năm nay nhà máy áp dụng một biện pháp kĩ thuật mới. Để nghiên cứu tác dụng của biện pháp này người ta lấy một mẫu gồm 800 sản phẩm để kiểm tra và thấy có 24 sản phẩm. Với α = 0,01. Hãy cho kết luận về biện pháp kĩ thuật mới này? Nếu nhà máy báo cáo tỉ lệ phế phẩm sau khi áp dụng biện pháp kĩ thuật mới này là 2% thì có chấp nhận được không? (với α = 0,05).

7. Nếu máy móc hoạt động bình thường thì kích thước của một loại sản phẩm tính theo cm là một đại lượng ngẫu nhiên phân phối theo qui luật chuẩn với phương sai $\sigma^2 = 25cm^2$. Nghi ngờ máy hoạt động không bình thường , người ta đo thử 20 sản phẩm và tính được $s^2 = 27,5cm^2$. Với α = 0,02 hãy kết luận về điều nghi ngờ này.

TÀI LIỆU THAM KHẢO

1. Nguyễn Cao Văn – Trần Thái Ninh, *Giáo trình Lý thuyết xác suất & thống kê Toán*, NXB Đại học Kinh tế quốc dân, 2008

2. Lê Sỹ Đồng – *Xác suất thống kê và ứng dụng*, NXB Giáo dục Việt Nam, 2011

3. Đinh Văn Gắng – Lý *thuyết xác suất và thống kê*, NXB Giáo dục, 2003

4. Trần Mạnh Tuấn – *Xác suất & Thống kê Lý thuyết và thực hành tính toán*, NXB Đại học Quốc gia Hà Nội, 2004

5. Trần Gia Tùng – *Giáo trình Lý thuyết xác suất & Thống kê toán học Lý thuyết cơ bản và ứng dụng*, NXB Đại học quốc gia TP. Hồ Chí Minh, 2009