BÀI 3 BIẾN NGẪU NHIÊN RỜI RẠC

Hướng dẫn học

Bài này tiếp tục nội dung của hai bài học trước nhưng kết hợp với các đại lượng đo lường. Các biến cố ngẫu nhiên trong bài trước được xét trở thành các biến ngẫu nhiên, từ đó tính toán các xác suất. Với các đại lượng đo lường được mô tả trong bài học này, không chỉ quan tâm đến xác suất xảy ra các giá trị mà còn quan tâm đến những đại lượng đặc trưng và cách tính các đại lượng đó. Để có thể nắm được bài học này, cần nhớ các khái niệm về xác suất trong các bài trước, cũng như cách tính các xác suất đó. Việc tính toán cần hết sức cẩn thận tránh nhầm lẫn.

Để học tốt bài này, sinh viên cần tham khảo các phương pháp học sau:

- Học đúng lịch trình của môn học theo tuần, làm các bài luyện tập đầy đủ và tham gia thảo luận trên diễn đàn.
- Đọc tài liệu: Giáo trình Lý thuyết xác suất và thống kê toán của NXB Đại học KTQD.
- Sinh viên làm việc theo nhóm và trao đổi với giảng viên trực tiếp tại lớp học hoặc qua email.
- Tham khảo các thông tin từ trang Web môn học.

Nội dung

- Khái niệm và phân loại biến ngẫu nhiên.
- Bảng phân phối xác suất của biến ngẫu nhiên rời rạc.
- Các tham số đặc trưng: kỳ vọng, phương sai, độ lệch chuẩn.
- Biến ngẫu nhiên phân phối Không – một và phân phối Nhị thức.
- Khái niệm và các tham số của biến ngẫu nhiên hai chiều rời rạc.

Mục tiêu

Sau khi học xong bài này sinh viên cần thực hiện được các việc sau:

- Hiểu khái niệm biến ngẫu nhiên và phân biệt được hai loại biến ngẫu nhiên.
- Lập được bảng phân phối xác suất của biến ngẫu nhiên rời rạc.
- Tính các tham số: kỳ vọng toán, phương sai, độ lệch chuẩn và áp dụng trong phân tích kinh tế.
- Biết sử dụng quy luật Không – Một và quy luật Nhị thức để tính xác suất và các đại lượng.
- Hiểu khái niệm biến ngẫu nhiên 2 chiều và tính được một số tham số đặc trưng của biến ngẫu nhiên 2 chiều rời rạc.

Tình huống dẫn nhập

Lựa chọn vị trí làm việc

Một người có thể lựa chọn giữa hai vị trí làm việc. Vị trí thứ nhất là tại một văn phòng và nhận một mức lương tháng cố định ở mức vừa phải. Vị trí thứ hai là tại một đơn vị kinh doanh và nhận lương theo kết quả làm việc: nếu kết quả rất tốt thì nhận lương cao, kết quả vừa phải thì lương bình thường, không hoàn thành thì nhận lương rất thấp, thậm chí không có lương. Với đơn vị kinh doanh việc kết quả công việc hay không là ngẫu nhiên, chỉ có thể nhận định được về khả năng đạt kết quả chứ không chắc chắn sẽ có kết quả tốt hay không hoàn thành.

?

1. Làm sao để phân tích tình huống và nhận định, đánh giá về mức lương?
2. Có cách nào so sánh lương trong hai vị trí trên?
3. Đánh giá về mặt trung bình thì có thể lựa chọn thế nào?
4. Đánh giá về sự rủi ro thì lựa chọn thế nào?

Mở rộng tình huống này cho các tình huống trong kinh doanh: mỗi phương án có các khả năng khác nhau: lỗ, hòa, có lãi. Mức lỗ có thể lỗ ít, lỗ nhiều; lãi có thể lãi ít, lãi nhiều… Trong trường hợp đó việc so sánh, đánh giá các phương án kinh doanh thực hiện thế nào dưới góc độ xem xét của môn Xác suất?

3.1. Khái niệm biến ngẫu nhiên

3.1.1. Khái niệm

Trong bài giảng trước ta đã làm quen với các biến cố ngẫu nhiên. Biến cố ngẫu nhiên thể hiện một hiện tượng nào đó mà ta quan tâm, có thể xảy ra hoặc không xảy ra khi thực hiện một phép thử. Với biến cố ngẫu nhiên ta phải dùng ngôn ngữ thông thường để mô tả, chẳng hạn các biến cố: "gieo hai đồng xu thì được hai mặt sấp", "được một mặt sấp", "được hai mặt ngửa". Trong hầu hết các trường hợp người ta nhận thấy có thể mô tả các biến cố ngẫu nhiên đó bằng cách dùng con số mô tả các đối tượng có thể lượng hóa trong biến cố đó.

Chẳng hạn với biến cố khi gieo hai đồng xu, có thể xét đến đại lượng là "số mặt sấp xuất hiện" và đặt tên nó là một đại lượng ký hiệu là X, khi đó biến cố "được hai mặt sấp" trở thành biến cố "X bằng 2", biến cố "được một mặt sấp" trở thành biến cố "X bằng 1", biến cố "được hai mặt ngửa" trở thành biến cố "X bằng 0". Đại lượng X đặt như vậy đã đơn giản hóa biến cố và có thể đánh giá được giá trị khi so sánh. Cách đặt như vậy cho ta khái niệm về biến ngẫu nhiên.

Định nghĩa 3.1 – Biến ngẫu nhiên: *Biến ngẫu nhiên là một biến số mà trong kết quả của phép thử nó sẽ nhận một và chỉ một trong các giá trị có thể có của nó tùy thuộc vào sự tác động của các nhân tố ngẫu nhiên.*

Định nghĩa trên cho thấy trước hết biến ngẫu nhiên phải nhận giá trị bằng số, hoặc là bản chất nó đã là số, hoặc là ta sẽ chuyển nó thành con số. Khi ta thực hiện phép thử, tương ứng với biến cố nào đó xảy ra, thì giá trị của biến ngẫu nhiên sẽ là một con số.

Thông thường, biến ngẫu nhiên được đặt bằng những chữ cái in hoa ở cuối bảng chữ cái như X, Y, Z. Trong một số trường hợp khác có thể đặt theo ý nghĩa của biến.

Ví dụ 3.1. Với các biến cố về số mặt sấp khi gieo hai đồng xu, đặt X là số chấm xuất hiện khi gieo hai đồng xu thì:

- X là biến số, có thể nhận các giá trị là 0; 1; 2.
- Sau khi gieo hai đồng xu thì X nhận đúng một trong ba giá trị trên.

Do đó X là biến ngẫu nhiên, có thể viết là $X = \{0; 1; 2\}$

Ví dụ 3.2. Đặt Y là số chấm xuất hiện khi gieo một con xúc sắc thì:

- Y là biến số, có thể nhận các giá trị là 1, 2, 3, 4, 5, 6.
- Sau khi gieo con xúc sắc thì Y nhận đúng 1 trong 6 giá trị trên.

Vậy Y là 1 biến ngẫu nhiên, có thể viết là $Y = \{1; 2; 3; 4; 5; 6\}$

Ví dụ 3.3. Khi làm việc tại doanh nghiệp kinh doanh, người lao động nếu hoàn thành hết công việc thì nhận lương tháng là 20 triệu đồng, nếu không thì chỉ được lương 5 triệu đồng. Đặt Z là tiền lương nhận được.

- Z là biến số, có thể nhận các giá trị 5, 20 (triệu đồng).
- Sau khi thực hiện phép thử (làm sau một tháng) thì Z nhận đúng một giá trị trong hai giá trị ở trên.

Vậy Z là biến ngẫu nhiên, có thể viết là $Z = \{5; 20\}$ (triệu đồng).

Ví dụ 3.4. Tại một bến xe buýt cứ 15 phút lại có một chuyến xe. Một hành khách tới bến vào một thời điểm ngẫu nhiên. Đặt *T* là thời gian hành khách đó phải chờ xe buýt, đơn vị là phút, thì:

- *T* là biến số, có thể nhận các giá trị bất kì thuộc nửa đoạn [0;15).
- Với mỗi hành khách trong một lần đợi xe buýt thì thời gian chờ xe T nhận đúng một giá trị trong khoảng trên.

Vậy T là biến ngẫu nhiên, có thể viết là $T \in [0;15)$.

3.1.2. Phân loại biến ngẫu nhiên

Qua các ví dụ trên có thể thấy trong ví dụ 3.1, ví dụ 3.2, ví dụ 3.3 các biến ngẫu nhiên có thể liệt kê được giá trị và số lượng giá trị chỉ là một số giá trị hữu hạn, cách rời nhau. Trong ví dụ 3.4 giá trị của biến ngẫu nhiên là lấp đầy một khoảng, không thể liệt kê hết được và các giá trị liền vào nhau. Từ đó có thể phân làm hai loại biến ngẫu nhiên là rời rạc và liên tục.

Biến ngẫu nhiên rời rạc: là biến ngẫu nhiên mà giá trị có thể có của nó lập thành một tập hợp hữu hạn hoặc đếm được các giá trị. Nói cách khác, ta có thể liệt kê tất cả các giá trị của biến ngẫu nhiên đó.

Biến ngẫu nhiên trong ví dụ 3.1 đến 3.3 thuộc loại rời rạc.

Biến ngẫu nhiên liên tục: là biến ngẫu nhiên mà tập các giá trị có thể có của nó nhận lấp đầy một khoảng trên trục số.

Biến ngẫu nhiên trong ví dụ 3.4 thuộc loại liên tục.

Trong bài giảng này ta chỉ xét biến ngẫu nhiên rời rạc, biến ngẫu nhiên liên tục sẽ được xét trong bài giảng số 4.

Với biến ngẫu nhiên rời rạc, ta có thể liệt kê được các giá trị có thể có của biến ngẫu nhiên. Nếu biến ngẫu nhiên rời rạc *X*, nếu có n giá trị có thể có, các giá trị có thể có được viết bằng chữ thường, chẳng hạn là $x_1, x_2, ..., x_n$, khi đó có thể viết $X = \{x_1, x_2, ..., x_n\}$.

Lưu ý: số giá trị có thể có của biến ngẫu nhiên rời rạc có thể lên đến vô hạn, chẳng hạn như *X* là số viên đạn phải dùng cho đến khi bắn trúng hồng tâm của một cái bia để rất xa, khi đó $X = \{1; 2; 3;...\}$ có thể kéo đến vô hạn. Tuy nhiên trường hợp này sẽ không xét đến trong bài giảng. Người học quan tâm có thể tham khảo trong giáo trình.

3.1.3. Biến ngẫu nhiên và biến cố

Như trên đã trình bày, giữa biến ngẫu nhiên và biến cố có mối liên hệ chặt chẽ. Việc biến ngẫu nhiên nhận một giá trị cụ thể hoặc một khoảng giá trị cụ thể tương ứng với một biến cố.

Với biến ngẫu nhiên rời rạc $X = \{x_1, x_2, ..., x_n\}$ thì việc "biến ngẫu nhiên X nhận giá trị bằng x" là một biến cố, ký hiệu là $(X = x)$. Tổng quát, $(X = x_i)$ với i từ 1 đến n là các biến cố.

Tương tự, việc "biến ngẫu nhiên X nhận giá trị nhỏ hơn con số x" là một biến cố, ký hiệu là $(X < x)$, việc X nhỏ hơn hoặc bằng con số x là biến cố ký hiệu là $(X \le x)$. Các quan hệ của X và con số đều tạo thành biến cố.

Ví dụ 3.5. Biến ngẫu nhiên X là số chấm xuất hiện khi gieo con xúc sắc, $X = \{1; 2; 3; 4; 5; 6\}$

- Biến cố ($X = 2$) là biến cố "được mặt có 2 chấm".
- Biến cố ($X = 2{,}5$) là biến cố không thể có.
- Biến cố ($X > 0$) là biến cố chắc chắn.
- Biến cố ($X < 2$) bằng với biến cố ($X = 1$) vì cùng là "được mặt có một chấm".
- Biến cố ($X \le 2$) bằng tổng hai biến cố ($X = 1$) + ($X = 2$).
- Biến cố ($X < 1{,}5$) bằng với biến cố ($X = 1$) và bằng với biến cố ($X \le 1{,}5$).

Qua ví dụ trên nhận thấy rằng với biến ngẫu nhiên rời rạc việc xét dấu bất đẳng thức < hay ≤ có thể khác nhau và cũng có thể giống nhau, tùy thuộc vào con số so sánh có nằm trong số các giá trị có thể có hay không. Trong ví dụ trên, biến cố ($X < 2$) khác với biến cố ($X \le 2$) vì con số 2 nằm trong số các giá trị có thể có và ($X = 2$) có thể xảy ra; trong khi đó biến cố ($X < 1{,}5$) bằng với biến cố ($X \le 1{,}5$) vì con số 1,5 không nằm trong số các giá trị có thể xảy ra.

3.2. Bảng phân phối xác suất

Với một phép thử ta có các biến cố, với mỗi biến cố có xác suất tương ứng. Bài giảng trước đã xét xác suất xảy ra các biến cố, với bài giảng này ta xét biến ngẫu nhiên rời rạc và xác suất tương ứng. Nếu liệt kê tất cả các giá trị có thể có của biến ngẫu nhiên rời rạc X dưới dạng $X = \{x_1, x_2, ..., x_n\}$ thì cũng có thể liệt kê các xác suất tương ứng với các giá trị có thể có đó. Cách liệt kê như vậy cho ta một cách đánh giá về xác suất được "phân phối" như thế nào cho các giá trị của biến ngẫu nhiên

3.2.1. Lập bảng phân phối xác suất

Với biến ngẫu nhiên rời rạc $X = \{x_1, x_2, ..., x_n\}$ thì $(X = x_i)$ là biến cố ngẫu nhiên, do đó có xác suất tương ứng. Ký hiệu: $p_i = P(X = x_i)$ thì với mỗi giá trị x_i có một xác suất p_i, có thể lập thành một bảng tương ứng, với x_i ở dòng trên và p_i ở dòng dưới, gọi là bảng phân phối xác suất.

Bảng phân phối xác suất của biến ngẫu nhiên X xác định như sau:

x	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

Việc liệt kê các giá trị và có được xác suất tương ứng phải dựa vào việc tính các xác suất của biến cố đã học trong bài trước. Cùng với bảng phân phối xác suất, có thể vẽ đồ thị các xác suất để nhận thấy rõ hơn xác suất được phân phối như thế nào.

3.2.2. Tính chất của bảng phân phối xác suất

Vì p_i là xác suất của một biến cố nên giá trị của p_i phải thuộc đoạn từ 0 đến 1. Các giá trị có thể có x_i là rời rạc tách rời nhau, nên $(X = x_1)$, $(X = x_2)$,…, $(X = x_n)$ là xung khắc nhau và tạo thành nhóm đầy đủ các biến cố, do đó tổng các xác suất tương ứng bằng 1. Từ đó ta có hai tính chất của bảng phân phối xác suất:

- Tính chất 1: $0 \le p_i \le 1$

- Tính chất 2: $\sum_{i=1}^{n} p_i = p_1 + p_2 + ... + p_n = 1$

Các bảng phân phối xác suất đều phải thỏa mãn hai tính chất này, nếu không thỏa mãn thì đó không phải là bảng phân phối xác suất.

Ví dụ 3.7. Đặt X là số chấm xuất hiện khi tung một con xúc sắc cân đối đồng chất trên mặt phẳng cứng 1 lần, ta có $X = \{1; 2; 3; 4; 5; 6\}$ và xác suất tương ứng đều bằng 1/6. Ta có bảng phân phối xác suất của X:

x	1	2	3	4	5	6
p	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Ví dụ 3.6. Hộp có 10 sản phẩm gồm 8 chính phẩm và 2 phế phẩm. Lấy đồng thời hai sản phẩm từ hộp. Đặt Y là số chính phẩm lấy được.

Do lấy hai sản phẩm từ hộp có 8 chính phẩm và 2 phế phẩm nên Y có thể nhận các giá trị là $\{0; 1; 2\}$. Do phép thử là lấy đồng thời 2 sản phẩm từ hộp có 10 sản phẩm nên ta dùng công thức xác suất cổ điển để tính các xác suất như sau:

$$P(Y=0) = \frac{C_2^2}{C_{10}^2} = \frac{1}{45} \approx 0,022$$

$$P(Y=1) = \frac{C_2^1 C_8^1}{C_{10}^2} = \frac{16}{45} \approx 0,356$$

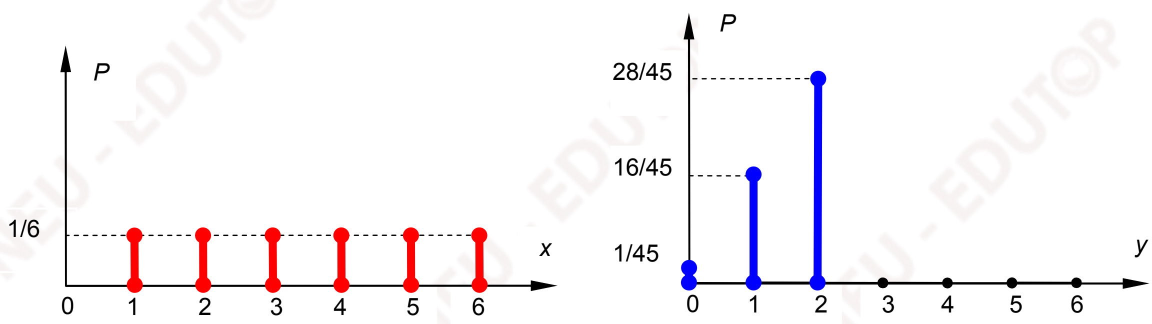
$$P(Y=2) = \frac{C_8^2}{C_{10}^2} = \frac{28}{45} \approx 0,622$$

Ta có $\frac{1}{45} + \frac{16}{45} + \frac{28}{45} = 1$. Từ đó suy ra bảng phân phối xác suất của biến ngẫu nhiên Y như sau:

Y	0	1	2
P	$\frac{1}{45}$	$\frac{16}{45}$	$\frac{28}{45}$

Minh họa phân phối xác suất của hai ví dụ 3.6 và 3.7 trên đồ thị, ta có hình 3.1.

Ví dụ 3.6. **Ví dụ 3.7**



Hình 3.1. ***Phân phối xác suất của biến ngẫu nhiên trong ví dụ 3.6 và 3.7***

Trong hình 3.1, đồ thị phía trên có 6 cột nằm tại vị trí các giá trị của biến ngẫu nhiên *X*, chiều cao các cột đúng bằng xác suất tương ứng với các giá trị đó; đồ thị phía dưới có 3 cột nằm tại ví trí các giá trị của biến ngẫu nhiên *Y*, chiều cao các cột cũng bằng xác suất tương ứng với các giá trị đó. Qua đồ thị có thể nhận thấy với biến *X*, xác suất được phân phối bằng nhau trên các con số từ 1 đến 6, hay có thể nói xác suất được rải đều. Với biến *Y*, xác suất chỉ có ba con số và phân phối không đều, xác suất tập trung nhiều nhất vào giá trị bằng 2, tiếp đó là giá trị bằng 1, và chỉ rất ít tại giá trị bằng 0. Việc nhìn vào bảng phân phối hoặc đồ thị sẽ cho biết xác suất phân bố thế nào vào các giá trị của biến ngẫu nhiên.

Trong một số trường hợp khác, giả định xác suất được tính từ các phương pháp khác (từ thống kê, đánh giá ý kiến chuyên gia), ta cũng có thể có được bảng phân phối xác suất.

3.3. Các tham số đặc trưng

Nghiên cứu các biến ngẫu nhiên rời rạc, nếu chỉ đánh giá thông qua bảng phân phối xác suất, ta chỉ có thể nhận biết các giá trị có thể có và xác suất tương ứng, mà không dễ dàng đánh giá so sánh các biến ngẫu nhiên với nhau. Do đó để đánh giá so sánh các biến ngẫu nhiên về một đặc tính, cần tính toán một số con số, gọi là các tham số đặc trưng.

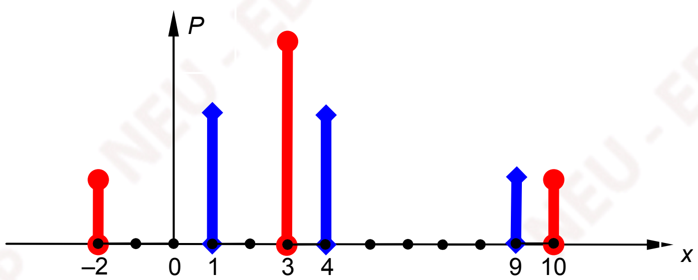
Ví dụ 3.8. Có hai dự án kinh doanh A và B, lợi nhuận của mỗi dự án là biến ngẫu nhiên rời rạc, đơn vị đều là tỷ đồng. Lợi nhuận dự án A ký hiệu là X_A, lợi nhuận dự án B ký hiệu là X_B và có bảng phân phối xác suất như sau:

X_A	– 2	3	10
P	0,2	0,6	0,2

X_B	1	4	9
P	0,4	0,4	0,2

Giá trị âm cho thấy lợi nhuận âm, dự án bị lỗ.

Đồ thị phân phối xác suất hai biến X_A, X_B trong hình 3.2.



Hình 3.2. ***So sánh hai biến ngẫu nhiên***

Trong hình 3.2, cột màu đỏ ứng với phân phối xác suất của lợi nhuận dự án A, cột màu xanh ứng với phân phối xác suất của lợi nhuận dự án B. Nếu nhìn trên đồ thị việc đánh giá so sánh không dễ dàng. Dự án A có lợi nhuận nhỏ nhất là âm, kém hơn dự án B, nhưng lợi nhuận lớn nhất lại là 10 trội hơn dự án B. Ta cần phải tính toán một số con số để so sánh được.

Để đánh giá về biến ngẫu nhiên, cần quan tâm hai đại lượng chính là đại lượng thể hiện xu thế trung tâm của biến ngẫu nhiên, và đại lượng thể hiện sự phân tán của biến ngẫu nhiên. Đại lượng thể hiện xu thế trung tâm được xét trong chương trình là kì vọng, đại lượng thể hiện sự phân tán là phương sai và độ lệch chuẩn. Tất cả đều được tính theo xác suất. Các tham số đặc trưng khác như trung vị, mốt, hệ số biến thiên, giá trị tới hạn… người học quan tâm có thể tự đọc giáo trình.

3.3.1. Kỳ vọng toán

Xét biến ngẫu nhiên lợi nhuận dự án kinh doanh A là X_A trong ví dụ 3.9 có ba giá trị có thể có là – 2; 3; 10. Tuy nhiên ba giá trị này lại không có khả năng xảy ra như nhau, do đó khi tính đại lượng mang tính chất "bình quân" thì không thể lấy trọng số bằng nhau. Trọng số mà ta lấy ở đây chính là khả năng xảy ra tương ứng với ba giá trị đó, hay chính là ba xác suất. Con số bình quân tính theo trọng số chính là xác suất được gọi là kỳ vọng toán, hay gọi tắt là kỳ vọng, và còn được gọi là trung bình.

Định nghĩa 3.2 – Kỳ vọng toán: *Biến ngẫu nhiên rời rạc X nhận một trong các giá trị có thể có* $x_1, x_2, \ldots, x_n$ *với xác suất tương ứng* $p_1, p_2, \ldots, p_n$ *thì kỳ vọng toán của X, ký hiệu là E(X) được tính theo công thức*:

$$E(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

Như vậy kỳ vọng toán là tổng của các tích giá trị có thể có của biến ngẫu nhiên và xác suất tương ứng với giá trị đó.

Lưu ý: Đơn vị của E(X) trùng với đơn vị của *X*.

Ví dụ 3.9 (Tiếp ví dụ 3.8). Lợi nhuận (tỷ đồng) hai dự án A và B, ký hiệu X_A và X_B có bảng phân phối xác suất như sau, hãy tính và so sánh hai kỳ vọng toán của lợi nhuận hai dự án.

X_A	– 2	3	10
P	0,2	0,6	0,2

X_B	1	4	9
P	0,4	0,4	0,2

Giải:

Theo công thức tính kỳ vọng toán ta có:

$$E(X_A) = -2 \times 0{,}2 + 3 \times 0{,}6 + 10 \times 0{,}2 = 3{,}4 \text{ (tỷ đồng)}$$

$$E(X_B) = 1 \times 0{,}4 + 4 \times 0{,}4 + 9 \times 0{,}2 = 3{,}3 \text{ (tỷ đồng)}$$

Như vậy về kỳ vọng, lợi nhuận dự án A lớn hơn lợi nhuận dự án B.

Ý nghĩa và ứng dụng của kì vọng toán

Kỳ vọng là giá trị trung bình theo xác suất của các giá trị mà biến ngẫu nhiên *X* nhận. Nó phản ánh giá trị trung tâm của các giá trị của biến ngẫu nhiên.

Trong kinh tế, kỳ vọng toán đặc trưng cho năng suất trung bình của một phương án sản suất, lợi nhuận trung bình của một danh mục đầu tư, trọng lượng trung bình của một loại sản phẩm, tuổi thọ trung bình của một chi tiết máy… Do đó trong kinh tế, kì vọng toán là một tiêu chuẩn để ra quyết định khi có nhiều phương án lựa chọn khác nhau.

Tính chất của kỳ vọng toán

Từ định nghĩa của kỳ vọng, ta có thể chứng minh được các tính chất sau, với *c* là hằng số, *X* và *Y* là các biến ngẫu nhiên.

Tính chất 1. Kỳ vọng của hằng số bằng chính nó:

$$E(c) = c$$

Tính chất 2. Tính kỳ vọng của tích hằng số và biến ngẫu nhiên thì có thể đưa hằng số ra ngoài:

$$E(cX) = cE(X)$$

Tính chất 3. Kỳ vọng của tổng các biến ngẫu nhiên bằng tổng các kỳ vọng của mỗi biến ngẫu nhiên thành phần:

$$E(X + Y) = E(X) + E(Y)$$

Tính chất 4. Kỳ vọng của tích hai biến ngẫu nhiên độc lập bằng tích các kỳ vọng của chúng:

$$E(XY) = E(X) \times E(Y).$$

Hai biến ngẫu nhiên X, Y độc lập được hiểu là biến cố $(X = x)$ và $(Y = y)$ là độc lập, với x là giá trị có thể có bất kỳ của X, và y là giá trị có thể có bất kỳ của Y.

Từ tính chất 1 và 3 dễ dàng suy ra: $E(c + X) = c + E(X)$

Từ tính chất 2 và 3 dễ dàng suy ra: $E(aX + bY) = aE(X) + bE(Y)$.

Ví dụ liên quan đến các tính chất này sẽ được xét cùng với các tính chất trong phần sau.

3.3.2. Phương sai và Độ lệch chuẩn

Trong ví dụ 3.9, ta thấy hai biến ngẫu nhiên X_A và X_B có kỳ vọng tương ứng là 3,4 và 3,4 tỷ đồng, do đó nếu đánh giá trên kỳ vọng thì dự án A tốt hơn dự án B. Tuy nhiên không phải ai cũng đồng ý rằng dự án A tốt hơn, vì xét thấy dự án A có sự chênh lệch giữa các giá trị lợi nhuận là nhiều hơn dự án B, có trường hợp dự án A bị lỗ, trong khi dự án B không bao giờ lỗ. Vì vậy sử dụng kỳ vọng để đánh giá so sánh là chưa đủ, còn cần đánh giá thêm mức độ phân tán của các giá trị của biến ngẫu nhiên.

Chẳng hạn với dự án A, để đánh giá mức độ phân tán, ta xét sự sai lệch giữa giá trị có thể có của biến ngẫu nhiên và chính kỳ vọng của nó để xem nó thay đổi thế nào. Khi đó với giá trị có thể có là 10 thì sai lệch là 10 – 3,4 = 6,6 và điều này xảy ra với xác suất là 0,2; với giá trị có thể có là 3 thì sai lệch là 3 – 3,4 = – 0,4 và điều này xảy ra với xác suất 0,6; với giá trị có thể có là –2 thì sai lệch là –2 – 3,4 = –5,4 và điều này xảy ra với xác suất 0,2. Việc xét các sai lệch mang cả dấu âm và dương cùng lúc gây nhiều phức tạp, vì vậy người ta bình phương lên để triệt tiêu dấu. Sau khi bình phương lên, tính kỳ vọng của đại lượng bình phương sai lệch đó, được giá trị gọi là phương sai, với nghĩa là "bình phương sai lệch".

Định nghĩa 3.3 – Phương sai: *Phương sai của biến ngẫu nhiên X, ký hiệu là V(X), là kỳ vọng toán của bình phương sai lệch của biến ngẫu nhiên so với kỳ vọng toán của nó*:

$$V(X) = E[X - E(X)]^2$$

Áp dụng tính chất của kỳ vọng, chứng minh được:

$$V(X) = E(X^2) - [E(X)]^2$$

Trong đó $E(X^2)$ là kỳ vọng của biến X bình phương, với X rời rạc thì:

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

Thông thường tính phương sai ta dùng theo cách *tính $E(X^2)$ trước rồi lấy $E(X^2)$ trừ đi bình phương của $E(X)$.*

Ý nghĩa của phương sai:

Phương sai đặc trưng cho độ phân tán các giá trị của biến ngẫu nhiên xung quanh $E(X)$. Nếu $V(X)$ lớn chứng tỏ sự biến động của X lớn, nếu $V(X)$ nhỏ thì các giá trị của X biến động ít, tương đối ổn định.

Trong kinh tế, phương sai đặc trưng cho độ rủi ro của các quyết định. Tùy từng bài toán, có thể cũng dùng nhiều danh từ khác để chỉ độ phân tán các giá trị của đại lượng ngẫu nhiên tương ứng như: độ dao động, độ biến động, độ bấp bênh, độ phân tán, độ ổn định, độ đồng đều, độ chính xác… Do đó trong kinh tế phương sai là một tiêu chuẩn để ra quyết định trong trường hợp có nhiều phương án lựa chọn khác nhau.

Lưu ý:

- Phương sai càng lớn thì ta nói biến càng biến động, càng dao động, càng phân tán.
- Phương sai càng nhỏ thì ta nói biến càng ổn định, càng tập trung, càng đồng đều.
- Đơn vị của phương sai là bình phương đơn vị của biến ngẫu nhiên. Nếu X có đơn vị là USD thì $V(X)$ đơn vị là USD^2; nếu X đơn vị là mét thì $V(X)$ có đơn vị là mét^2.

Vì phương sai liên quan đến phép tính bình phương, đơn vị của phương sai biến trở thành bình phương đơn vị của biến, nên không thể so sánh phương sai với kỳ vọng hay với giá trị của biến. Để dùng cho các phân tích tiếp theo, người ta tính một đại lượng là căn bậc hai của phương sai, gọi là độ lệch chuẩn.

Định nghĩa 3.4 – Độ lệch chuẩn: *Độ lệch chuẩn của biến ngẫu nhiên X, ký hiệu là σ_X hoặc $\sigma(X)$, là căn bậc hai của phương sai của X*:

$$E(X^2) = \sum_{i=1}^{n} x_i^2 p_i \ ; \ \sigma_x = \sqrt{V(X)}$$

Như vậy để tính được độ lệch chuẩn buộc phải tính phương sai trước.

Ý nghĩa của độ lệch chuẩn tương tự ý nghĩa của phương sai, chỉ có điều khác biệt lớn nhất là độ lệch chuẩn có đơn vị là đơn vị của X, và như vậy có thể so sánh độ lệch chuẩn với giá trị có thể có của X, so sánh với kỳ vọng của X.

Ví dụ 3.10 (tiếp ví dụ 3.8). Lợi nhuận (tỷ đồng) hai dự án A và B, ký hiệu X_A và X_B có bảng phân phối xác suất như sau, hãy tính và so sánh hai phương sai, độ lệch chuẩn của lợi nhuận hai dự án.

X_A	– 2	3	10
P	0,2	0,6	0,2

X_B	1	4	9
P	0,4	0,4	0,2

Giải:

Theo kết quả đã tính từ ví dụ 3.9 ta có: $E(X_A) = 3{,}4$ và $E(X_B) = 3{,}4$ (tỷ đồng)

Để tính phương sai của X_A là $V(X_A)$, ta cần tính $E(X_A^2)$

$$E(X_A^2) = (-2)^2 \times 0{,}2 + 3^2 \times 0{,}6 + 10^2 \times 0{,}2 = 26{,}2$$

Suy ra: $V(X_A) = E(X_A^2) - [E(X_A)]^2 = 26{,}2 - 3{,}4^2 = 14{,}64$

Và: $\sigma_{X_A} = \sqrt{V(X_A)} = \sqrt{14{,}64} \approx 3{,}83$

Vậy lợi nhuận dự án A có phương sai là 14,64 (tỷ đồng)^2 và độ lệch chuẩn là 3,83 (tỷ đồng).

Tương tự với dự án B ta có:

$$E(X_B^2) = 1^2 \times 0,4 + 4^2 \times 0,4 + 9^2 \times 0,2 = 16,5$$

$$V(X_B) = 16,5 - 3,3^2 = 5,61$$

$$\sigma_{X_B} = \sqrt{5,61} \approx 2,37$$

Lợi nhuận dự án B có phương sai là 5,61 (tỷ đồng)^2 và độ lệch chuẩn là 2,37 (tỷ đồng).

Nếu đánh giá trên phương sai thì ta thấy $V(X_A)$ lớn hơn $V(X_B)$, có thể nói lợi nhuận của dự án A là phân tán hơn, dao động nhiều hơn, biến động nhiều hơn; lợi nhuận dự án B là ổn định hơn, đồng đều hơn. Trong kinh tế phương sai lớn hơn còn được gọi là rủi ro nhiều hơn, nên dự án A có rủi ro nhiều hơn.

Kết hợp kỳ vọng và phương sai, có thể thấy lợi nhuận dự án A có kỳ vọng lớn hơn nhưng phương sai cũng lớn hơn, độ bất ổn lớn hơn; lợi nhuận dự án B tuy nhỏ hơn nhưng phương sai cũng nhỏ hơn, ổn định hơn. Việc lựa chọn dự án nào là quyết định của nhà đầu tư, môn học này chỉ cung cấp thông tin để đánh giá so sánh.

Tính chất của phương sai

Các tính chất của phương sai giúp cho việc tính toán một số đại lượng ngẫu nhiên dễ dàng hơn, khi đại lượng đó là tổng hợp từ các biến ngẫu nhiên. Với c là hằng số, X và Y là các biến ngẫu nhiên.

Tính chất 1. Phương sai hằng số bằng không:

$$V(c) = 0$$

Tính chất 2. Phương sai của tích biến ngẫu nhiên với hằng số bằng bình phương hằng số nhân với phương sai của biến ngẫu nhiên:

$$V(cX) = c^2 V(X)$$

Tính chất 3. Phương sai của tổng, hiệu các biến ngẫu nhiên độc lập đều bằng tổng các phương sai của hai biến ngẫu nhiên đó:

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

Từ tính chất 1 và 3 dễ dàng suy ra: $V(c + X) = V(X)$, hay khi cộng thêm một hằng số vào biến ngẫu nhiên thì phương sai không đổi, biến X đơn thuần chỉ chuyển vị trí, còn sự dao động của nó vẫn như cũ.

Ta có thể so sánh tính chất của kỳ vọng và phương sai qua bảng sau, với c là hằng số:

Kỳ vọng	Phương sai
$E(c) = c$	$V(c) = 0$
$E(c + X) = c + E(X)$	$V(c + X) = V(X)$
$E(cX) = cE(X)$	$V(cX) = c^2V(X)$
$E(X \pm Y) = E(X) \pm E(Y)$	$V(X \pm Y) = V(X) + V(Y)$

Ví dụ 3.11. (ví dụ tình huống dẫn nhập)

Một người đi làm có hai lựa chọn:

Lựa chọn thứ nhất: làm văn phòng với lương tháng cố định là 6 triệu đồng.

Lựa chọn thứ hai: làm kinh doanh và lương tính theo số hợp đồng kí được, mỗi hợp đồng được nhận 5 triệu đồng. Biết rằng số hợp đồng kí được trong 1 tháng có bảng phân phối xác suất như sau:

Số hợp đồng	0	1	2	3
Xác suất	0,1	0,3	0,4	0,2

Hãy đánh giá hai lựa chọn qua việc so sánh kì vọng, phương sai, độ lệch chuẩn của lương tháng.

Giải:

Ta thấy với lựa chọn thứ nhất thì lương tháng là một hằng số: c = 6, do đó kỳ vọng bằng chính số đó và phương sai là bằng 0:

$$E(c) = c = 6 \text{ (triệu đồng)}$$

$$V(c) = 0; \sigma_c = 0$$

Vậy lựa chọn thứ nhất không có rủi ro.

Với lựa chọn thứ hai, lương tháng bằng một hằng số nhân với số hợp đồng kí được, mà số hợp đồng kí được là biến ngẫu nhiên. Đặt X là số hợp đồng kí được, ta có Lương = 5X.

Theo tính chất của kỳ vọng và phương sai, ta có:

$$E(\text{Lương}) = E(5X) = 5E(X)$$

$$V(\text{Lương}) = V(5X) = 5^2V(X) = 25V(X)$$

Để tính được kỳ vọng, phương sai của lương, ta tính kỳ vọng và phương sai của X, với bảng phân phối xác suất của X chính là bảng phân phối xác suất của số hợp đồng.

$$E(X) = 0 \times 0,1 + 1 \times 0,3 + 2 \times 0,4 + 3 \times 0,2 = 1,7 \text{ (hợp đồng)}$$

$$E(X^2) = 0^2 \times 0,1 + 1^2 \times 0,3 + 2^2 \times 0,4 + 3^2 \times 0,2 = 3,7$$

$$V(X) = 3,7 - 1,7^2 = 0,81 \text{ (hợp đồng)}^2$$

$$\sigma_X = \sqrt{0,81} = 0,9 \text{ (hợp đồng)}$$

Từ đó suy ra:

$$E(\text{Lương}) = 5 \times 1,7 = 8,5 \text{ (triệu đồng)}$$

$$V(\text{Lương}) = 25 \times 0,81 = 20,25 \text{ (triệu đồng)}^2$$

$$\sigma_{\text{Lương}} = 4,5 \text{ (triệu đồng)}$$

So sánh kỳ vọng, phương sai, độ lệch chuẩn của lương tháng ta thấy lựa chọn làm trong kinh doanh có kỳ vọng cao hơn, tuy nhiên có phương sai cũng cao hơn, hay biến động nhiều hơn; làm văn phòng kỳ vọng thấp hơn nhưng không có rủi ro. Việc lựa chọn phương án nào do cá nhân người đi làm quyết định tùy thuộc vào tâm lý, sở thích, cá tính của người đó.

Lưu ý:

Ngoài việc áp dụng tính chất trong việc tính kỳ vọng và phương sai của lương tháng của trường hợp lựa chọn thứ hai, cũng có thể đặt lương tháng là Y (đơn vị: triệu đồng) và có bảng phân phối xác suất như sau:

Y	$0 \times 5 = 0$	$1 \times 5 = 5$	$2 \times 5 = 10$	$3 \times 5 = 15$
P	0,1	0,3	0,4	0,2

$$E(Y) = 0 \times 0{,}1 + 5 \times 0{,}3 + 10 \times 0{,}4 + 15 \times 0{,}2 = 8{,}5 \text{ (triệu đồng)}$$

$$E(Y^2) = 0^2 \times 0{,}1 + 5^2 \times 0{,}3 + 10^2 \times 0{,}4 + 15^2 \times 0{,}2 = 28{,}8$$

$$V(Y) = 28{,}8 - 8{,}5^2 = 20{,}25 \text{ (triệu đồng)}^2$$

$$\sigma_Y = \sqrt{20{,}25} = 4{,}5 \text{ (triệu đồng)}$$

Bạn đọc có thể tùy chọn cách làm nào thuận tiện cho mình.

Ví dụ 3.12. Một doanh nghiệp sản xuất đi chào hàng ở hai nơi, độc lập nhau. Xác suất mỗi nơi đặt hàng cho doanh nghiệp này lần lượt là 0,6 và 0,7.

(a) Lập bảng phân phối xác suất và tính kỳ vọng, phương sai, độ lệch chuẩn số đơn hàng mà doanh nghiệp này nhận được.

(b) Nếu chi phí cố định là 10 triệu đồng, chi phí sản xuất cho mỗi đơn hàng là 60 triệu đồng, hãy tính kỳ vọng, phương sai, độ lệch chuẩn của tổng chi phí của doanh nghiệp.

(c) Nếu doanh thu của mỗi đơn hàng là 80 triệu đồng, tìm kỳ vọng, phương sai, độ lệch chuẩn của lợi nhuận doanh nghiệp.

Giải:

(a) Đặt X là số đơn hàng mà doanh nghiệp nhận được, dễ thấy X có thể bằng 0 (không nơi nào đặt hàng), bằng 1 (chỉ có 1 nơi đặt hàng) hoặc bằng 2 (cả hai nơi đặt hàng). Như vậy $X = \{0; 1; 2\}$.

Xác suất $X = 0$ là xác suất để cả hai nơi đều từ chối, do đó:

$$P(X = 0) = (1 - 0{,}6) \times (1 - 0{,}7) = 0{,}3 \times 0{,}4 = 0{,}12$$

Xác suất để $X = 1$ là xác suất của hai trường hợp: nơi thứ nhất đặt hàng đồng thời nơi thứ hai từ chối, hoặc nơi thứ nhất từ chối đồng thời nơi thứ hai đặt hàng, do đó:

$$P(X = 1) = 0{,}6 \times (1 - 0{,}7) + (1 - 0{,}6) \times 0{,}7 = 0{,}18 + 0{,}28 = 0{,}46$$

Xác suất để $X = 2$ là xác suất để cả hai nơi đều đặt hàng, do đó:

$$P(X = 2) = 0{,}6 \times 0{,}7 = 0{,}42.$$

Kiểm tra lại tính chất, ta thấy:

$$P(X = 0) + P(X = 1) + P(X = 2) - 0{,}12 + 0{,}26 + 0{,}42 = 1$$

Như vậy kết quả thỏa mãn điều kiện của bảng phân phối xác suất. Bảng phân phối xác suất của số đơn hàng doanh nghiệp nhận được là:

***X* (Số đơn hàng)**	0	1	2
P	0,12	0,46	0,42

Từ đó tính kỳ vọng, phương sai, độ lệch chuẩn

$$E(X) = 0 \times 0{,}12 + 1 \times 0{,}46 + 2 \times 0{,}42 = 1{,}3 \text{ (đơn hàng)}$$

$$E(X^2) = 0^2 \times 0{,}12 + 1^2 \times 0{,}46 + 2^2 \times 0{,}42 = 2{,}14$$

$$V(X) = 2{,}14 - 1{,}3^2 = 0{,}45 \text{ (đơn hàng)}^2$$

$$\sigma_X = \sqrt{0{,}45} \approx 0{,}67 \text{ (đơn hàng)}$$

Nhận xét thấy rằng kỳ vọng của số đơn hàng chính là tổng hai xác suất: $E(X) = 1,3 = 0,6 + 0,7$. Tính chất này có thể được thấy rõ hơn qua các phần sau.

(b) Ta thấy tổng chi phí bằng chi phí cố định cộng với tích của chi phí cho mỗi đơn hàng nhân với số đơn hàng doanh nghiệp có được.

Đặt Y là tổng chi phí của doanh nghiệp, đơn vị là triệu đồng, ta thấy: $Y = 10 + 60X$

Theo tính chất của kỳ vọng và phương sai, ta có:

$E(Y) = E(10 + 60X) = 10 + E(60X) = 10 + 60E(X) = 10 + 60 \times 1,3 = 88$ (triệu đồng)

$V(Y) = V(10 + 60X) = V(60X) = 60^2V(X) = 3600 \times 0,45 = 1620$ (triệu đồng)^2

$\sigma_X = \sqrt{1620} \approx 40,23$ (triệu đồng).

Ta cũng có thể giải câu (b) bằng cách lập bảng phân phối xác suất của Y, với Y nhận các giá trị 10 (không có đơn hàng nào), 70 (có một đơn hàng), 130 (có hai đơn hàng) với xác suất tương ứng là 0,12; 0,46; 0,42. Việc làm cụ thể bạn đọc có thể tự thực hiện, kết quả tính toán sẽ giống như trên đã tính.

(c) Với câu này, lợi nhuận bằng tổng doanh thu trừ tổng chi phí. Ta phải lập bảng phân phối xác suất của lợi nhuận.

Đặt Z là lợi nhuận, đơn vị là triệu đồng, có các trường hợp sau:

- Không có đơn hàng nào: lợi nhuận bằng âm 10 do doanh thu bằng 0 và chi phí bằng 10. Xác suất xảy ra là 0,12
- Có một đơn hàng: lợi nhuận bằng 10 do doanh thu bằng 80 và chi phí bằng $10 + 60 = 70$. Xác suất xảy ra là 0,46
- Có hai đơn hàng: lợi nhuận bằng 30 do doanh thu bằng $2 \times 80 = 160$ và chi phí bằng $10 + 2 \times 60 = 130$. Xác suất xảy ra là 0,42

Bảng phân phối xác suất của lợi nhuận như sau:

***Z* (Lợi nhuận)**	– 10	10	30
P	0,12	0,46	0,42

Từ đó tính kỳ vọng, phương sai, độ lệch chuẩn

$E(Z) = -10 \times 0,12 + 10 \times 0,46 + 30 \times 0,42 = 16$ (triệu đồng)

$E(Z^2) = (-10)^2 \times 0,12 + 10^2 \times 0,46 + 30^2 \times 0,42 = 436$

$V(Z) = 436 - 16^2 = 180$ (triệu đồng)^2

$\sigma_Z = \sqrt{180} \approx 13,42$ (triệu đồng)

3.4. Biến ngẫu nhiên phân phối Không – Một

Trong các loại biến ngẫu nhiên, biến đơn giản nhất là loại chỉ nhận hai giá trị 0 và 1, gọi là biến Không – Một.

3.4.1. Định nghĩa

Định nghĩa 3.5: Biến ngẫu nhiên Không – Một: *Biến ngẫu nhiên X chỉ nhận một trong hai giá trị có thể có là 0 và 1 với xác suất nhận giá trị bằng 1 là bằng p, gọi là biến ngẫu nhiên phân phối Không – Một với tham số p, ký hiệu là X ~ A(p).*

Như vậy $X = \{0; 1\}$ và $P(X = 1) = p$, suy ra $P(X = 0) = 1 - p$.

Bảng phân phối xác suất của X có dạng:

X	0	1
P	1 – p	p

Biến ngẫu nhiên phân phối Không – Một gọi tắt là biến phân phối Không – Một. Tất nhiên $0 \le p \le 1$.

Người ta có thể định nghĩa biến phân phối Không – Một qua công thức như sau:

Định nghĩa 2: Biến ngẫu nhiên X được gọi là phân phối Không – Một nếu có công thức tính xác suất như sau:

$$P(X = x) = p^x(1-p)^{1-x}, x = 0;1$$

Dễ chứng minh hai định nghĩa là tương đương, vì theo định nghĩa thứ hai, X chỉ có thể là 0 và 1, và:

$$P(X = 0) = p^0(1-p)^{1-0} = 1-p$$
$$P(X = 1) = p^1(1-p)^{1-1} = p$$

3.4.2. Tham số đặc trưng

Theo bảng phân phối xác suất của X, ta có

Kỳ vọng toán: $E(X) = 0 \times (1-p) + 1 \times p = p$

$E(X^2) = 0^2 \times (1-p) + 1^2 \times p = p$

Phương sai: $V(X) = p - p^2 = p(1-p)$

Độ lệch chuẩn: $\sigma_X = \sqrt{p(1-p)}$

Ví dụ 3.13. Với một loại sản phẩm, doanh nghiệp S chiếm 70% thị phần. Kiểm tra ngẫu nhiên **một** sản phẩm trên thị trường, đặt X là số sản phẩm của doanh nghiệp S mà ta có. Khi đó X là biến ngẫu nhiên nhận một trong hai giá trị là 0 và 1; X nhận giá trị 0 khi sản phẩm đó không phải của doanh nghiệp S, và X nhận giá trị 1 khi sản phẩm đó là của doanh nghiệp S. Xác suất tương ứng như sau:

$$P(X = 1) = 0{,}7 \text{ và } P(X = 0) = 0{,}3$$

Có thể nói X là biến ngẫu nhiên phân phối Không – Một với tham số p = 0,7; ký hiệu $X \sim A(0{,}7)$.

Từ đó $E(X) = p = 0{,}7$ và $V(X) = p(1-p) = 0{,}3 \times 0{,}7 = 0{,}21$.

Kỳ vọng bằng 0,7 nghĩa là khi kiểm tra ngẫu nhiên một sản phẩm, ta sẽ kỳ vọng là có được 0,7 sản phẩm của doanh nghiệp S. Điều này nghe có vẻ lạ lùng, tuy nhiên trong kinh tế xã hội, giá trị kỳ vọng hoàn toàn có thể là số lẻ, thập phân.

Ví dụ 3.14. Một người đi chào hàng ở ba nơi độc lập nhau, xác suất mỗi nơi đặt hàng lần lượt là 0,6; 0,7; 0,8. Tìm kỳ vọng và phương sai tổng số đơn hàng người đó có được.

Giải:

Đặt Y là tổng số đơn hàng người đó có được, như vậy $Y = \{0; 1; 2; 3\}$

Cách 1.

Bài này có thể giải theo phương pháp thông thường, lập bảng phân phối xác suất của Y:

$P(Y = 0) = (1 - 0,6) \times (1 - 0,7) \times (1 - 0,8) = 0,024$

$P(Y = 1) = (0,6) \times (1 - 0,7) \times (1 - 0,8) + (1 - 0,6) \times (0,7) \times (1 - 0,8) + (1 - 0,6) \times (1 - 0,7) \times (0,8) = 0,188$

$P(Y = 2) = (0,6) \times (0,7) \times (1 - 0,8) + (1 - 0,6) \times (0,7) \times (0,8) + (0,6) \times (1 - 0,7) \times (0,8) = 0,452$

$P(Y = 3) = (0,6) \times (0,7) \times (0,8) = 0,336$

Bảng phân phối xác suất của Y:

Y	0	1	2	3
P	0,024	0,188	0,452	0,336

Từ đó tính được: $E(Y) = 2,1$ (đơn hàng); $V(Y) = 0,61$ (đơn hàng)^2; $\sigma_Y = 0,78$ (đơn hàng).

Cách 2.

Áp dụng biến phân phối Không – Một. Xét riêng từng nơi mà người đó đến chào hàng.

Đặt X_i là số đơn hàng người đó có được tại nơi thứ i, $i = 1, 2, 3$.

Khi đó tổng số đơn hàng là: $Y = X_1 + X_2 + X_3$

Suy ra: $E(Y)= E(X_1) + E(X_2) + E(X_3)$ và $V(Y)= V(X_1) + V(X_2) + V(X_3)$

Dễ thấy X_i chỉ có thể nhận giá trị là 0 (không chào được, không có đơn hàng) và 1 (chào được, có đơn hàng) và các X_i độc lập nhau.

X_1 nhận giá trị bằng 1 với xác suất 0,6 do đó $X_1 \sim$ A(0,6), suy ra $E(X_1) = 0,6$ và $V(X_1) = 0,6 \times 0,4 = 0,24$

Tương tự ta có:

$X_2 \sim$ A(0,7) $\Rightarrow E(X_2) = 0,7$ và $V(X_2) = 0,7 \times 0,3 = 0,21$

$X_3 \sim$ A(0,8) $\Rightarrow E(X_3) = 0,8$ và $V(X_3) = 0,8 \times 0,2 = 0,16$

Vậy $E(Y) = 0,6 + 0,7 + 0,8 = 2,1$

$V(Y) = 0,24 + 0,21 + 0,16 = 0,61$

Cách tính thứ hai trong ví dụ này không ngắn hơn so với cách thứ nhất. Tuy nhiên nếu người đi chào hàng không phải ở 3 nơi mà ở 4 nơi, 5 nơi thì cách thứ nhất rất dài và khó, trong khi cách thứ hai không khó hơn, chỉ tiếp tục thêm vào các nơi mà ta có xác suất.

Do đó biến phân phối Không – Một có ứng dụng khi ta xét riêng từng trường hợp, trong mỗi trường hợp một biến cố có thể xảy ra hoặc không xảy ra. Biến Không – Một thường đặc trưng cho các trường hợp nghiên cứu có dạng yếu tố định tính chỉ có hai trường hợp luân phiên: Có và Không có một tính chất nào đó, chẳng hạn như giới tính Nam – Nữ, Hoàn thành – Không hoàn thành.

3.5. Biến ngẫu nhiên phân phối Nhị thức

Xét ví dụ 3.14 về người đi bán hàng ở ba nơi, theo cách giải thứ hai ta đã thấy khi xét kỳ vọng, phương tổng số đơn hàng, có thể xét riêng từng trường hợp, mỗi trường hợp là biến ngẫu nhiên phân phối Không – Một. Trong ví dụ đó, xác suất bán được hàng ở mỗi nơi là khác nhau, lần lượt bằng 0,6; 0,7 và 0,8 thì kỳ vọng của tổng số đơn hàng là tổng của ba xác suất đó, và phương sai là tổng của ba phương sai thành phần. Nếu như xác suất bán được hàng ở mỗi nơi đều bằng nhau và bằng 0,6 thì dễ thấy kỳ vọng của tổng số đơn hàng là 3 × 0,6 và phương sai của tổng số đơn hàng là 3 × 0,6 × 0,4. Tổng quát hơn, nếu có *n* nơi và xác suất mỗi nơi đặt hàng đều bằng *p* thì ta dễ dàng suy ra kỳ vọng và phương sai của tổng số đơn hàng.

Nếu người đó đi bán hàng ở n nơi độc lập nhau, xác suất mỗi nơi đặt hàng đều bằng p, thì ta có trường hợp lược đồ Bernoulli đã xét trong bài số 2. Từ đó ta cũng có khái niệm về *biến ngẫu nhiên phân phối Nhị thức*.

3.5.1. Định nghĩa

Định nghĩa 3.6 – Biến phân phối Nhị thức: *Biến ngẫu nhiên rời rạc X gọi là có phân phối theo quy luật Nhị thức với tham số p, ký hiệu là X ~ B(n, p), nếu X nhận một trong các giá trị: 0, 1, 2,…, n với xác suất tương ứng cho bởi công thức Bernoulli:*

$$P(X=x)=C_n^x p^x(1-p)^{n-x}, x=0,1,2,...,n$$

Nếu bài toán thỏa mãn lược đồ Bernoulli, nghĩa là chỉ ra được:

- Có n phép thử độc lập.
- Trong mỗi phép thử, xác suất xuất hiện biến cố A không đổi là $P(\text{A}) = p$.
- X là số lần xuất hiện biến cố A trong n phép thử đó thì X phân phối theo quy luật Nhị thức.

3.5.2. Tham số đặc trưng

Từ ví dụ trên có thể thấy nếu X là biến ngẫu nhiên phân phối Nhị thức B(n; p) thì X chính là tổng của n trường hợp riêng độc lập nhau, mỗi trường hợp riêng đều phân phối theo quy luật Không – Một A(p) với cùng tham số p, do đó kỳ vọng của X sẽ bằng tổng các kỳ vọng thành phần và phương sai của X sẽ bằng tổng các phương sai thành phần, mà kỳ vọng và phương sai thành phần đều bằng nhau.

$$X \sim \text{B(n; p) thì } X=\sum_{i=1}^{n} X_i$$

Trong đó X_i độc lập, $X_i \sim$ A(p), $i = 1, 2,\ldots, n$.

- Kỳ vọng: $E(X)=E\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} E(X_i)=np$
- Phương sai: $V(X)=V\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} V(X_i)=np(1-p)$

Ví dụ 3.15. Một người đi chào hàng ở 5 nơi độc lập nhau, xác suất để mỗi nơi đặt hàng đều bằng nhau và bằng 0,6. Phân phối xác suất của tổng số đơn hàng, kỳ vọng và phương sai của số đơn hàng như thế nào?

Giải:

Đặt X là tổng số đơn hàng, $X = \{0; 1; 2; 3; 4; 5\}$

Vì có 5 nơi độc lập, xác suất đặt hàng đều bằng nhau, nên X được hình thành từ lược đồ Bernoulli, X có phân phối Nhị thức với hai tham số: $n = 5$ và $p = 0,6$.

Vậy $X \sim B(n = 5; p = 0,6)$.

Tham số đặc trưng là:

$E(X) = np = 5 \times 0,6 = 3$ (đơn hàng)

$V(X) = np(1 - p) = 5 \times 0,6 \times (1 - 0,6) = 1,2$ (đơn hàng)^2

Chỉ cần viết $X \sim B(n = 5; p = 0,6)$ là đủ thông tin, tuy nhiên ta có thể tính chi tiết phân phối xác suất để thấy rõ hơn, giá trị xác suất này có thể lấy từ bảng phụ lục.

$P(X = 0 \mid n = 5, p = 0,6) = 0,0102$

$P(X = 1 \mid n = 5, p = 0,6) = 0,0768$

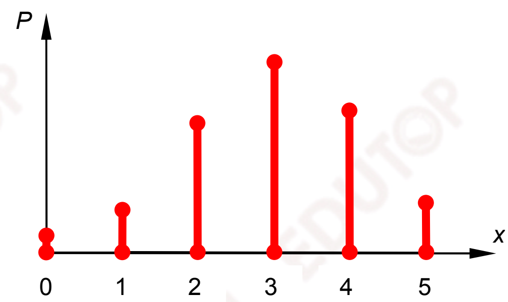
$P(X = 2 \mid n = 5, p = 0,6) = 0,2304$

$P(X = 3 \mid n = 5, p = 0,6) = 0,3456$

$P(X = 4 \mid n = 5, p = 0,6) = 0,2592$

$P(X = 5 \mid n = 5, p = 0,6) = 0,0778$

Cách tính này tính theo phương pháp xác suất biến cố đối lập.



Hình 3.3. ***Phân phối xác suất – Ví dụ 3.16***

Qua giá trị xác suất và đồ thị có thể thấy sự phân bố của xác suất vào các giá trị của biến ngẫu nhiên như thế nào.

3.6. Biến ngẫu nhiên hai chiều rời rạc

Trong phần trước ta đề cập các biến ngẫu nhiên được xét một cách riêng biệt, với mỗi biến ngẫu nhiên có các giá trị rời rạc, có phân phối xác suất, các tham số đặc trưng. Trong thực tế các biến ngẫu nhiên không hoàn toàn riêng biệt mà có thể kết hợp với nhau, và ta phải xét cùng lúc nhiều biến ngẫu nhiên. Chẳng hạn khi đánh giá về kết quả học một môn học của sinh viên KTQD, không chỉ xét trên một biến riêng biệt như điểm thi hết học phần, mà cần xét trên nhiều điểm thành phần: điểm chuyên cần, điểm kiểm tra, điểm hết học phần. Khi đánh giá so sánh về hai doanh nghiệp, không chỉ đánh giá trên một biến duy nhất, mà cần xét đồng thời nhiều biến: lợi nhuận, quy mô, thị phần. Khi so sánh đánh giá về nhân khẩu một hộ gia đình, có thể xét đồng thời số người đi làm có thu nhập và số người ăn theo, số người lớn và số trẻ em. Rất nhiều hiện tượng kinh tế xã hội phải được nhìn nhận với nhiều khía cạnh cùng lúc, do đó cần đến tập hợp nhiều biến ngẫu nhiên cùng lúc.

Trong phạm vi môn học này, ta sẽ xét trường hợp chỉ có hai biến ngẫu nhiên được xét cùng một lúc, gọi là biến ngẫu nhiên hai chiều, và các biến đều rời rạc nên gọi là biến ngẫu nhiên hai chiều rời rạc.

3.6.1. Khái niệm

Khi xét hai biến ngẫu nhiên X và Y cùng một lúc trên một đối tượng, ta có hệ hai biến ngẫu nhiên, và gọi hệ đó là biến ngẫu nhiên hai chiều, ký hiệu là (X,Y).

Với biến ngẫu nhiên hai chiều (*X*, *Y*), hai biến *X*, *Y* gọi là hai thành phần.

Ví dụ 3.16. Xét nhân khẩu trong một hộ gia đình, không chỉ quan tâm đến tổng số người, mà còn cần xem xét số người lớn (từ 18 tuổi trở lên) đồng thời với số trẻ em (dưới 18 tuổi), khi đó ký hiệu số người lớn là *X*, số trẻ em là *Y*, ta có một hệ hai biến ngẫu nhiên (*X*, *Y*), là một biến ngẫu nhiên hai chiều.

Để đơn giản, giả sử với thành phần *X*, có $X = \{1; 2; 3\}$, nghĩa là số người lớn tối thiểu phải bằng 1, có ít nhất một người lớn mới là một gia đình, và số người lớn tối đa bằng 3. Giả sử thành phần *Y* có $Y = \{0; 1; 2\}$ nghĩa là có thể không có trẻ em, và tối đa có 2 trẻ em. Khi đó biến ngẫu nhiên hai chiều (*X*, *Y*) là các cặp giá trị, từ (1;0), (1;1),…, đến (3;2), tổng cộng có 9 trường hợp.

Lưu ý:

- Biến hai chiều phải có thứ tự. Trong ví dụ trên, nếu viết (1;2) sẽ khác với (2;1) vì (1;2) là hộ có 1 người lớn và 2 trẻ em còn (2;1) là hộ có 2 người lớn và 1 trẻ em.
- Phân biệt biến hai chiều với tổng hai biến thành phần. Với hai trường hợp (1;2) và (2;1) thì tổng số người đều bằng 3, nhưng hai trường hợp đó là khác nhau về cấu trúc thành phần hộ gia đình.

Tổng quát:

Nếu thành phần X có n giá trị có thể có: $X = \{x_1, x_2, ..., x_n\}$

Và thành phần Y có m giá trị có thể có:

Thì biến ngẫu nhiên hai chiều sẽ là n×m cặp giá trị có thể có:

$$(X, Y) = \left\{(x_i, y_j), i = 1, 2, ..., n; j = 1, 2, ..., m\right\}$$

3.6.2. Bảng phân phối xác suất hai chiều

Nếu biến ngẫu nhiên rời rạc thông thường có bảng phân phối xác suất là bảng dạng hàng thì với biến ngẫu nhiên hai chiều có phân phối xác suất dưới dạng bảng hai chiều, một thành phần là dòng và một thành phần là cột. Trong mỗi ô của bảng là xác suất tương ứng, và theo điều kiện của bảng phân phối xác suất thì tổng các xác suất trong lòng bảng phải bằng 1.

Ví dụ 3.17. Xét bảng phân phối xác suất của số người lớn (*X*) và số trẻ em (*Y*) trong một hộ gia đình, để đơn giản mỗi biến chỉ có ba giá trị có thể có:

X \ Y	0	1	2
1	0,05	0,1	0,05
2	0,1	0,2	0,15
3	0,05	0,2	0,1

Với bảng này có thể giải thích như sau:

Với dòng $X = 1$, cột $Y = 0$ có ô với giá trị 0,01 nghĩa là xác suất để một hộ gia đình có 1 người lớn và không có trẻ em là 0,05, hay $P(X = 1; Y = 0) = 0{,}05$.

Tương tự, ô ứng với dòng $X = 2$ và cột $Y = 1$ bằng 0,2 cho biết $P(X = 2; Y = 1) = 0{,}2$.

Nếu tính tổng số 9 con số xác suất trong lòng bảng, tổng xác suất sẽ bằng 1. Nếu không bằng 1 thì đó không phải là bảng phân phối xác suất.

Tổng quát:

Nếu X có n giá trị: $X = \{x_1, x_2, ..., x_n\}$, Y có m giá trị: $Y = \{y_1, y_2, ..., y_m\}$ thì bảng phân phối xác suất hai chiều sẽ có n dòng và m cột, trong mỗi ô sẽ là:

$p_{ij} = P(X = x_i; Y = y_j)$ với i = 1,2,…, n; j = 1,2,…, m. Điều kiện của bảng là:

- $0 \le p_{ij} \le 1$
- $\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = 1$

3.6.3. Bảng phân phối xác suất biên

Từ bảng phân phối xác suất hai chiều, có thể tính được bảng phân phối xác suất của từng thành phần bằng cách tính tổng xác suất theo dòng hoặc theo cột. Các bảng phân phối xác suất của từng thành phần vì được cộng ra ngoài bảng chính nên được gọi là bảng phân phối xác suất biên. Với các bảng phân phối xác suất biên, ta có thể tính được các tham số đặc trưng của từng thành phần như với các biến ngẫu nhiên thông thường.

Ví dụ 3.17 (tiếp). Từ bảng phân phối xác suất hai chiều của X và Y có thể cộng ngang theo dòng và cộng dọc theo cột, được bảng sau:

Y / X	0	1	2	Tổng P_X
1	0,05	0,1	0,05	0,2
2	0,1	0,2	0,15	0,45
3	0,05	0,2	0,1	0,35
Tổng: P_Y	0,2	0,5	0,3	Tổng = 1

Ta thấy tổng của dòng thứ nhất được 0,2 chính là tổng các xác suất ứng với $X = 1$, bất kể Y thế nào. Như vậy $P(X = 1) = 0{,}2$. Tương tự có $P(X = 2) = 0{,}45$ và $P(X = 3) = 0{,}35$.

Tổng cột thứ nhất được 0,2 chính là tổng các xác suất ứng với $Y = 0$ bất kể X thế nào, hay $P(Y = 0) = 0{,}2$; tương tự $P(Y = 1) = 0{,}5$ và $P(Y = 2) = 0{,}3$.

Từ đó ta có hai bảng phân phối xác suất biên của X và của Y.

Bảng phân phối xác suất biên của X:

X	1	2	3
P_X	0,2	0,45	0,35

Do đó tính được:

$$E(X) = 1 \times 0{,}3 + 2 \times 0{,}45 + 3 \times 0{,}35 = 2{,}15$$

$$V(X) = 1^2 \times 0{,}3 + 2^2 \times 0{,}45 + 3^2 \times 0{,}35 - 2{,}15^2 = 0{,}5275$$

$$\sigma_Y = \sqrt{0{,}5275} \approx 0{,}73$$

Như vậy số người lớn có kỳ vọng là 2,15 (người), phương sai là 0,5275 (người)^2, độ lệch chuẩn là 0,73 (người).

Bảng phân phối xác suất biên của Y:

Y	0	1	2
P_Y	0,2	0,5	0,3

Do đó tính được:

$$E(Y) = 0 \times 0{,}2 + 1 \times 0{,}5 + 2 \times 0{,}3 = 1{,}1$$
$$V(Y) = 0^2 \times 0{,}2 + 1^2 \times 0{,}5 + 2^2 \times 0{,}3 - 1{,}1^2 = 0{,}49$$
$$\sigma_Y = \sqrt{0{,}49} = 0{,}7$$

Như vậy số trẻ em có kỳ vọng là 1,1 (người), phương sai là 0,49 (người)^2, độ lệch chuẩn là 0,7 (người).

3.6.4. Hệ số tương quan

Khi xét hai biến ngẫu nhiên cùng một lúc thành biến ngẫu nhiên hai chiều, một ý nghĩa quan trọng có thể phân tích là sự quan hệ qua lại giữa hai thành phần. Để đánh giá hai thành phần của biến ngẫu nhiên hai chiều liên quan với nhau có chặt chẽ hay không, người ta tính một tham số gọi là hệ số tương quan. Tuy nhiên để tính được hệ số tương quan, cần tính một tham số là hiệp phương sai của hai biến ngẫu nhiên thành phần.

Nếu như phương sai là đo độ phân tán, dao động của từng biến ngẫu nhiên thành phần, thì hiệp phương sai dùng để đo độ phân tán, dao động khi kết hợp hai biến ngẫu nhiên với nhau.

Định nghĩa 3.7 – Hiệp phương sai: *Hiệp phương sai của hai biến ngẫu nhiên X và Y, ký hiệu là Cov(X,Y), được tính theo công thức sau:*

$$Cov(X,Y) = E\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right]$$

- Chứng minh được: $Cov(X,Y) = E(XY) - E(X)E(Y)$
- Với biến rời rạc thì: $E(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p_{ij}$

Chi tiết ý nghĩa của hiệp phương sai bạn đọc có thể xem trong giáo trình. Tại đây ta tập trung việc tính hệ số tương quan.

Định nghĩa 3.8 – Hệ số tương quan: *Hệ số tương quan của hai biến ngẫu nhiên X và Y, ký hiệu là ρ_{XY} được tính theo công thức:*

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

(Chữ ρ đọc là "rô" trong tiếng Việt, "rho" trong tiếng Anh).

Chứng minh được hệ số tương quan luôn nằm trong đoạn 0 đến 1: $0 \le \rho_{XY} \le 1$.

Hệ số ρ_{XY} được gọi là hệ số đo mức độ tương quan giữa hai biến ngẫu nhiên. Tương quan ở đây được hiểu là có liên quan qua lại với nhau theo xu thế thẳng. Vì vậy có các trường hợp sau:

- $\rho_{XY} = 1$: X và Y có tương quan cùng chiều chặt chẽ nhất, dạng đường thẳng dốc lên.
- $0 < \rho_{XY} < 1$: X và Y có tương cùng chiều, càng gần 1 càng chặt chẽ, càng gần 0 càng yếu.
- $\rho_{XY} = 0$: X và Y không có tương quan.
- $-1 < \rho_{XY} < 0$: X và Y có tương quan ngược chiều, càng gần – 1 càng chặt chẽ, càng gần 0 càng yếu.
- $\rho_{XY} = -1$: X và Y có tương quan ngược chiều chặt chẽ nhất, dạng đường thẳng dốc xuống.

Ví dụ 3.17 (tiếp). Tính hệ số tương quan của số người lớn và số trẻ em.

Ta sẽ tính lần lượt từng đại lượng.

$$E(XY)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i y_j p_{ij}$$

Giá trị $x_i y_j p_{ij}$ tức là con số ở đầu hàng nhân với con số đầu cột nhân với xác suất ở cột tương ứng với hàng và cột đó.

Do đó với ví dụ này:

E(XY) = 1 × 0 × 0,05 + 1 × 1 × 0,1 + 1 × 2 × 0,05 + 2 × 0 × 0,1 + 2 × 1 × 0,2 + 2 × 2 × 0,15 + 3 × 0 × 0,05 + 3 × 1 × 0,2 + 3 × 2 × 0,1 = 2,4

Suy ra hiệp phương sai: Cov(X,Y) = E(XY) – E(X)E(Y) = 2,4 – 2,15 × 1,1 = 0,035

Hệ số tương quan: $\rho_{XY}=\dfrac{Cov(X,Y)}{\sigma_X\sigma_Y}=\dfrac{0,035}{0,73\times 0,7}\approx 0,07$

Như vậy số người lớn và số trẻ em có tương quan cùng chiều, nhưng rất yếu.

Ví dụ 3.18. Từ kết quả phân tích các số liệu thống kê trong tháng về doanh số bán hàng (D) và chi phí cho quảng cáo (Q) (đơn vị triệu đồng) của một công ty, thu được bảng phân phối xác suất đồng thời như sau:

D \ Q	100	200	300
1	0,15	0,1	0,04
1,5	0,05	0,2	0,15
2	0,01	0,05	0,25

(a) Tính giá trị trung bình và phương sai của doanh số bán hàng.

(b) Tính giá trị trung bình và phương sai của chi phí cho quảng cáo.

Giải: Từ bảng phân phối xác suất đồng thời ta có bảng phân phối xác suất của doanh số bán hàng và chi phí cho quảng cáo như sau:

D	100	200	300
P	0,21	0,35	0,44

Q	1	1,5	2
P	0,29	0,4	0,31

Từ đó ta có:

(a) Giá trị trung bình và phương sai của doanh số bán hàng là:

E(D) = 100 × 0,21 + 200 × 0,35 + 300 × 0,44 = 223

E(D^2) = 100^2 × 0,21 + 200^2 × 0,35 + 300^2 × 0,44 = 55700

V(D) = E(D^2) – [E(D)]^2 = 5971

(b) Giá trị trung bình và phương sai của chi phí cho quảng cáo là:

E(Q) = 1 × 0,29 + 1,5 × 0,4 + 2 × 0,31 = 1,51

E(Q^2) = 1^2 × 0,29 + 1,5^2 × 0,4 + 2^2 × 0,31 = 2,43

V(Q) = E(Q^2) – [E(Q)]^2 = 0,1499

Tóm lược cuối bài

- Biến ngẫu nhiên phân phối Không – một là biến ngẫu nhiên chỉ nhận một trong hai giá trị có thể có là 0 và 1 với xác suất nhận giá trị bằng 1 là bằng *p*, gọi là biến ngẫu nhiên phân phối Không – một với tham số *p*, ký hiệu là $X \sim A(p)$.
- Biến ngẫu nhiên rời rạc X gọi là có phân phối theo quy luật nhị thức với tham số *p*, ký hiệu là $X \sim B(n, p)$, nếu *X* nhận một trong các giá trị: 0, 1, 2,…, *n* với xác suất tương ứng cho bởi công thức Bernoulli:
$$P(X = x) = C_n^x p^x (1-p)^{n-x}, x = 0, 1, 2, ..., n$$
- Khi xét hai biến ngẫu nhiên *X* và *Y* cùng một lúc trên một đối tượng, ta có hệ hai biến ngẫu nhiên, và gọi hệ đó là biến ngẫu nhiên hai chiều, ký hiệu là (*X*,*Y*). Với *X*, *Y* là các biến ngẫu nhiên rời rạc thì (*X*,*Y*) là biến ngẫu nhiên hai chiều rời rạc. Xác suất của biến ngẫu nhiên hai chiều rời rạc phân phối dưới dạng bảng phân phối xác suất đồng thời. Từ bảng này có thể tính được kì vọng, phương sai của các thành phần, hiệp phương sai và hệ số tương quan của hai thành phần.
- Hệ số tương quan đo mức độ tương quan giữa hai biến ngẫu nhiên. Tương quan ở đây được hiểu là có liên quan qua lại với nhau theo xu thế thẳng.

Câu hỏi ôn tập

1. Biến ngẫu nhiên là gì? Có mấy loại biến ngẫu nhiên?

2. Bảng phân phối xác suất là gì? Tính chất của nó thế nào?

3. Kỳ vọng của biến ngẫu nhiên rời rạc tính như thế nào? Nó cho biết điều gì?

4. Phương sai của biến ngẫu nhiên rời rạc tính như thế nào? Ý nghĩa của phương sai?

5. Độ lệch chuẩn tính như thế nào? Tại sao phải tính độ lệch chuẩn?

6. Thế nào là biến ngẫu nhiên phân phối Không – Một, các tham số của nó tính thế nào?

7. Thế nào là biến ngẫu nhiên phân phối Nhị thức, các tham số tính thế nào, xác suất tính như thế nào?

8. Thế nào là biến ngẫu nhiên hai chiều?

9. Tính chất của bảng phân phối xác suất hai chiều và bảng phân phối biên thế nào?

10. Hệ số tương quan tính như thế nào và cho biết điều gì?