BÀI 4 BIẾN NGẪU NHIÊN LIÊN TỤC

Hướng dẫn học

Bài này tiếp tục bài giảng trước, đề cập sâu hơn đến khái niệm biến ngẫu nhiên liên tục, do đó cần nắm vững các khái niệm, ý nghĩa của bài học trước. Mối liên hệ giữa hai loại biến ngẫu nhiên có thể được nhận ra qua hình ảnh hình học của phân phối xác suất. Ý nghĩa các tham số kỳ vọng toán, phương sai vẫn như cũ, nhưng cách tính toán có thay đổi. Để làm được bài tập tính toán cần theo dõi kĩ các ví dụ, đặc biệt là ví dụ về biến ngẫu nhiên phân phối Chuẩn, là loại biến ngẫu nhiên quan trọng và thông dụng nhất. Cần phải nắm cụ thể từ việc áp dụng công thức, tra bảng số, và tốt hơn là hiểu được ý nghĩa hình học của cách tính toán. Ứng dụng bài toán biến ngẫu nhiên phân phối Chuẩn vào các tình huống thực tế là điều quan trọng của bài học này. Bên cạnh đó, với biến ngẫu nhiên phân phối Khi – bình phương và phân phối Student cần nắm được cách ký hiệu và tra bảng số để sử dụng trong các bài sau.

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

- Biến ngẫu nhiên liên tục và hàm mật độ xác suất.s
- Biến ngẫu nhiên phân phối Chuẩn: khái niệm, định nghĩa, ý nghĩa.
- Biến ngẫu nhiên phân phối Chuẩn hóa: khái niệm, tra bảng số để tìm xác suất.
- Tính xác suất của biến ngẫu nhiên phân phối Chuẩn qua biến Chuẩn hóa.
- Tra bảng giá trị tới hạn các biến phân phối Chuẩn, Khi – bình phương, Student.

Mục tiêu

- Hiểu khái niệm biến ngẫu nhiên liên tục, đánh giá đồ thị hàm mật độ xác suất;
- Tra bảng chính xác để tìm xác suất biến phân phối Chuẩn hóa;
- Áp dụng công thức tính xác suất của biến phân phối Chuẩn, áp dụng trong các bài toán thực tế;
- Tra bảng tìm chính xác các giá trị tới hạn.

Tình huống dẫn nhập

Bảo hành sản phẩm và lợi nhuận

Một doanh nghiệp sản xuất hàng điện tử bán ra thị trường. Các sản phẩm đều có thời hạn bảo hành xác định. Nếu sản phẩm hỏng sau thời hạn bảo hành thì doanh nghiệp không phải chịu trách nhiệm và do đó có lợi nhuận. Nếu sản phẩm hỏng trước thời hạn bảo hành thì doanh nghiệp phải sửa chữa, thay thế và có thể bị lỗ. Tuổi thọ của sản phẩm là ngẫu nhiên, và doanh nghiệp biết một số thông tin cơ bản về tuổi thọ của sản phẩm như kỳ vọng, phương sai.

?

1. Làm thế nào để doanh nghiệp tính được tỷ lệ sản phẩm bị hỏng trong thời gian còn bảo hành?
2. Làm thế nào để doanh nghiệp tính toán lợi nhuận phù hợp?
3. Nếu doanh nghiệp bỏ chi phí để tăng chất lượng sản phẩm thì lợi nhuận có bị giảm đi không?
4. Nếu doanh nghiệp bỏ chi phí để tăng chất lượng sản phẩm thì mức chi phí chấp nhận được là bao nhiêu để đảm bảo lợi nhuận không bị ảnh hưởng?

4.1. Biến ngẫu nhiên liên tục và hàm mật độ xác suất

Trong bài học trước, ta đã xét khái niệm về biến ngẫu nhiên, và nghiên cứu chi tiết về biến ngẫu nhiên rời rạc. Trong bài học này ta sẽ đề cập đến biến ngẫu nhiên liên tục và đi sâu nghiên cứu một loại biến ngẫu nhiên liên tục thông dụng, phổ biến nhất trong kinh tế xã hội.

4.1.1. Biến ngẫu nhiên liên tục

Trong kinh tế xã hội, các biến ngẫu nhiên ngoài dạng rời rạc như đã xét trong bài trước, còn có thể có dạng liên tục. Nếu với biến ngẫu nhiên rời rạc ta có thể liệt kê các giá trị có thể, thì biến ngẫu nhiên liên tục các giá trị có thể có của nó lấp đầy một khoảng và không thể liệt kê chi tiết ra được.

Định nghĩa 4.1 – Biến ngẫu nhiên liên tục: *Biến ngẫu nhiên X là liên tục nếu giá trị có thể có của nó lấp đầy một khoảng trên trục số.*

Vì không liệt kê chi tiết ra được nên với biến ngẫu nhiên liên tục X, nếu giá trị nhỏ nhất có thể có là $x_{\min}$, giá trị lớn nhất có thể có là $x_{\max}$, thì thường viết dưới dạng $X \in (x_{\min}; x_{\max})$.

Ví dụ 4.1.

- Nhiệt độ trong ngày là một biến ngẫu nhiên liên tục, không thể liệt kê hết tất cả các giá trị có thể, và thường nói giá trị nhỏ nhất và giá trị lớn nhất: chẳng hạn như nhiệt độ trong khoảng 20 độ C đến 30 độ C; khi đó X đơn vị là độ C, và $X \in (20; 30)$.
- Thời gian hoàn thành một bài thi là biến ngẫu nhiên liên tục, trong khoảng 1 đến 2 giờ. Khi đó X có đơn vị là giờ, và $X \in (1; 2)$.
- Khối lượng một gói gia vị (đơn vị là gam) là biến ngẫu nhiên liên tục: $X \in (100; 110)$.

Trong thực tế có nhiều biến ngẫu nhiên bản chất là rời rạc, tuy nhiên vì số lượng giá trị của nó là rất nhiều nên cũng có thể xét như là biến ngẫu nhiên liên tục.

Ví dụ 4.2.

- Thu nhập của người lao động Việt Nam có đơn vị là VND, bản chất là biến ngẫu nhiên rời rạc, không có ai có thu nhập với số đơn vị là lẻ 0,4 hay 0,5 đồng. Như vậy biến ngẫu nhiên X nhận các giá trị 1.000.000 VND, 1.000.120 VND nhưng không có 1.000.000,234 VND và là rời rạc. Tuy nhiên vì số lượng giá trị có thể có rất nhiều nên thu nhập vẫn thường được xét như biến ngẫu nhiên liên tục: $X \in [0; \infty)$.
- Tương tự với thu nhập, các biến như chi tiêu, tiết kiệm, giá cả, sản lượng, doanh thu, lợi nhuận… cũng có thể xét như biến ngẫu nhiên liên tục.

4.1.2. Hàm mật độ xác suất

Với biến ngẫu nhiên rời rạc, ta có bảng phân phối xác suất để phản ánh về xác suất phân phối vào từng giá trị như thế nào. Với biến ngẫu nhiên liên tục, ta cũng có một cách để thể hiện phân phối của xác suất tương ứng với các giá trị. Tuy nhiên, vì các giá trị có thể có của biến ngẫu nhiên liên tục là vô hạn, không thể liệt kê được nên không thể có một bảng giống như với biến ngẫu nhiên rời rạc.

Trước khi xét cách mô tả phân phối xác suất của biến ngẫu nhiên liên tục, ta xét đồ thị tương ứng với bảng phân phối xác suất của biến ngẫu nhiên rời rạc.

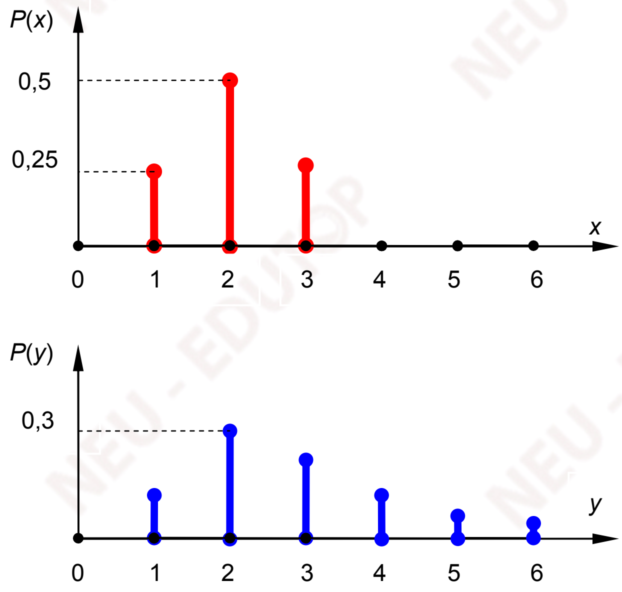
Ví dụ 4.3.

Xét hai biến ngẫu nhiên rời rạc X và Y có bảng phân phối xác suất như sau:

X	0	1	2
***P*(*x*)**	0,25	0,5	0,25

Y	1	2	3	4	5	6
***P*(*y*)**	0,15	0,3	0,25	0,15	0,1	0,05

Nếu vẽ trên đồ thị với chiều cao các cột đúng bằng giá trị xác suất thì ta có hình vẽ:



Hình 4.1. ***Đồ thị phân phối xác suất của hai biến ngẫu nhiên rời rạc***

Với hình 4.1 ta nhận thấy với biến ngẫu nhiên rời rạc, tổng chiều cao các cột bằng 1. Nếu biến ngẫu nhiên nhận nhiều giá trị khác nhau hơn thì các cột sẽ phân chia bớt cho các giá trị khác, phân phối xác suất sẽ có hình dạng khác nhưng tính chất tổng bằng 1 luôn được đảm bảo.

Với biến ngẫu nhiên rời rạc, xác suất để nó nhận giá trị nhỏ hơn một con số bằng tổng chiều cao các cột ở bên trái con số đó.

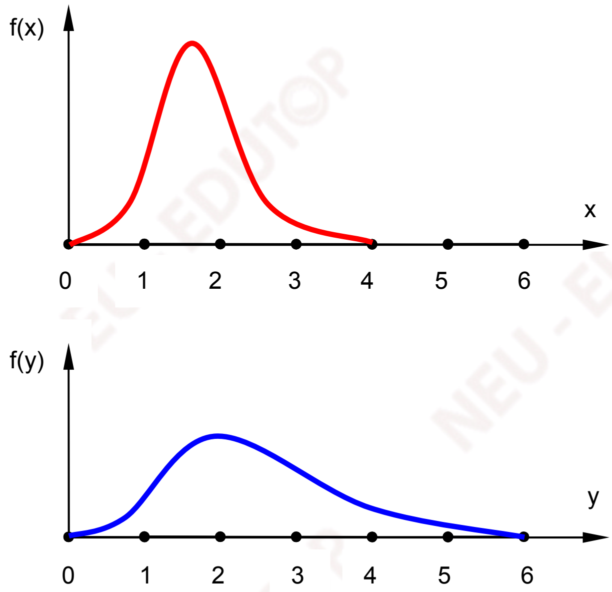
Với biến X ở trên xác suất để X nhỏ hơn 2,2 là:

$$P(X < 2{,}2) = P(X = 1) + P(X = 2) = 0{,}25 + 0{,}5 = 0{,}75$$

Với biến Y, xác suất để Y nhỏ hơn 2,2 là:

$$P(Y < 2{,}2) = P(Y = 1) + P(Y = 2) = 0{,}15 + 0{,}3 = 0{,}45$$

Đối với biến ngẫu nhiên liên tục, vì các giá trị không thể liệt kê nên không thể có đồ thị dưới dạng các cột cách rời nhau như trường hợp trên. Các giá trị của biến ngẫu nhiên liên tục là lấp đầy một khoảng, nên để thể hiện sự phân phối của xác suất, người ta dùng một đường liền, có dạng như hình 4.2.



Hình 4.2. ***Đường liền nét của biến ngẫu nhiên liên tục***

Với biến ngẫu nhiên rời rạc, các cột phải có chiều cao lớn hơn 0 thì với biến ngẫu nhiên liên tục, đường liền nét có dạng hàm số, trong khoảng giá trị có thể có nhận giá trị lớn hơn 0 và bằng 0 tại hai đầu ứng với giá trị nhỏ nhất và lớn nhất.

Với biến ngẫu nhiên rời rạc, tổng chiều cao các cột bằng 1 thì với biến ngẫu nhiên liên tục, tổng diện tích bên dưới đường liền nét cũng bằng 1. Do đó nếu đường liền nét cao như hình vẽ ứng với x thì sẽ hẹp hơn, và đường ứng với y thấp thì sẽ thoải và bẹt hơn. Đường liền nét đó được xác định bởi một hàm số gọi là hàm mật độ xác suất.

Để xây dựng hàm mật độ xác suất đúng theo chuẩn mực toán học cần phải qua nhiều bước. Tại đây ta không xét các bước đó, vì vậy có khái niệm dưới đây.

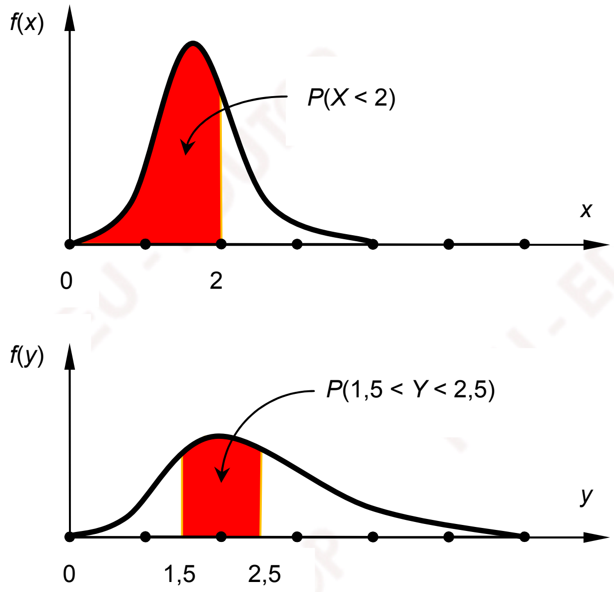
Khái niệm: *Hàm mật độ xác suất của biến ngẫu nhiên liên tục X, ký hiệu là $f(x)$, là hàm số không âm trong khoảng giá trị của X và diện tích tạo bởi hàm số đó và trục hoành bằng 1, thể hiện sự phân phối xác suất của X.*

4.1.3. Tính chất hàm mật độ xác suất

Để tính diện tích, phải tính bởi tích phân của hàm số, do đó nếu X liên tục nhận giá trị trong khoảng ($x_{\min}$; $x_{\max}$) thì hàm mật độ xác suất $f(x)$ là một hàm liên tục phải thỏa mãn:

- $f(x_{\min}) = f(x_{\max}) = 0$
- $f(x) > 0$ với $x_{\min} < x < x_{\max}$
- $\int_{x_{\min}}^{x_{max}} f(x)dx = 1$

Để tính xác suất biến ngẫu nhiên liên tục nhận giá trị nhỏ hơn một con số nào đó, không tính bằng tổng chiều cao các cột, mà tính bằng diện tích của hình giới hạn bởi hàm mật độ, minh họa trong hình 4.3.



Hình 4.3. ***Tính xác suất biến ngẫu nhiên liên tục***

Trong hình 4.3 với biến ngẫu nhiên liên tục X, xác suất để X nhận giá trị nhỏ hơn 2 là diện tích của phần hình vẽ bên trái điểm 2. Dễ thấy với diện tích thì $(X < 2)$ hay $(X \le 2)$ là không khác gì nhau, nên xác suất để $(X < 2)$ và xác suất để $(X \le 2)$ là bằng nhau:

Với X liên tục thì $P(X < 2) = P(X \le 2)$.

Điều này khác với biến ngẫu nhiên rời rạc. Với biến ngẫu nhiên liên tục, việc xác định dấu ≤ hay dấu < không làm thay đổi xác suất. Cũng từ đó có thể thấy xác suất để X liên tục nhận giá trị tại một điểm duy nhất là bằng 0.

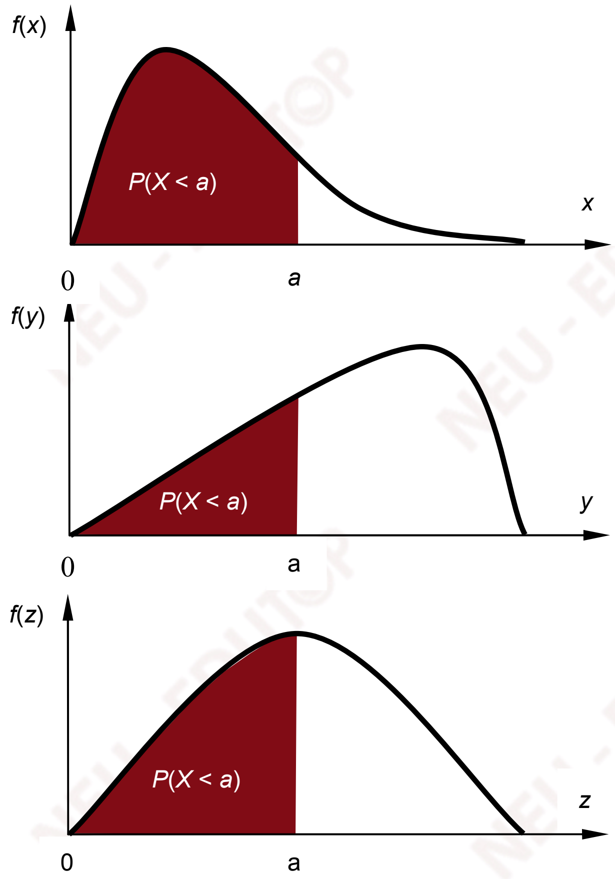
Với biến ngẫu nhiên liên tục Y, tương tự, xác suất để Y nhận giá trị trong khoảng (1,5; 2,5) là diện tính hình thang cong giới hạn bởi điểm 1,5 và điểm 2,5. Diện tích được tính bởi tích phân của hàm mật độ xác suất.

Từ nhận xét đó, có thể tổng quát thành tính chất sau:

Với biến ngẫu nhiên liên tục X thì:

- $P(a < X < b) = \int_a^b f(x)dx$
- $P(X = a) = 0$
- $P(a < X < b) = P(a \le X < b) = P(a < X \le b) = P(a \le X \le b)$

Dựa vào tính chất trên, để hiểu rõ hơn ý nghĩa và việc sử dụng hàm mật độ xác suất trong tính toán, có thể minh họa trên hình 4.4.



Hình 4.4. ***So sánh các dạng hàm mật độ xác suất***

Với hình 4.4, với biến X, hàm mật độ có đỉnh nằm ở phía bên trái, dồn về bên trái và thoải về bên phải, khi đó xác suất $P(X < a)$ sẽ lớn hơn so với $P(X > a)$. Ngược lại, với biến Y, hàm mật độ cao ở bên phải, thoải xuống ở bên trái, khi đó $P(Y < a)$ sẽ nhỏ hơn so với $P(Y > a)$. Trường hợp biến ngẫu nhiên Z hàm mật độ có tính đối xứng và đỉnh nằm ở giữa, tương ứng với điểm a, khi đó $P(Z < a)$ sẽ bằng với $P(Z > a)$.

Từ các tính chất trên có thể rút ra các nhận xét sau:

- Với biến ngẫu nhiên liên tục X, chỉ xét xác suất nhận giá trị trong một khoảng. Xác suất X nhận giá trị tại một điểm bằng 0.
- Khi xét xác suất X nhận giá trị trong một khoảng, không cần quan tâm đến cận.
- Hình ảnh hàm mật độ xác suất cho biết sự tập trung của xác suất, chỗ nào hàm mật độ càng cao thì xác suất tập trung ở khoảng vây quanh giá trị đó càng nhiều. Hàm mật độ xác suất bằng 0 là giá trị xảy ra với xác suất bằng 0.

Có vô số các hàm mật độ xác suất có thể xảy ra và có thể tính toán, tương ứng với rất nhiều dạng của biến ngẫu nhiên liên tục. Tuy nhiên trong nội dung chương trình ta sẽ chỉ xét một số biến ngẫu nhiên liên tục quan trọng có tính ứng dụng cao. Biến ngẫu nhiên liên tục quan trọng nhất sẽ được xét kĩ, cũng là biến ngẫu nhiên sẽ được dùng trong tất cả các bài sau, là biến phân phối Chuẩn.

4.2. Biến ngẫu nhiên phân phối Chuẩn

Qua phần trên ta thấy các biến ngẫu nhiên liên tục có thể phân phối khác nhau, thể hiện qua hàm mật độ có đỉnh nằm ở vị trí khác nhau. Qua quan sát và thu thập số liệu thực tế, người ta nhận thấy các biến ngẫu nhiên trong tự nhiên, kinh tế – xã hội thường có dạng phân phối gần giống như biến Z trong hình 4.4, nghĩa là đỉnh nằm ở giữa và có tính đối xứng ra hai bên. Đối với phân phối dạng này, xác suất để biến nhận giá trị nhỏ là bé, xác suất nhận giá trị lớn cũng bé, và xác suất tập trung nhiều nhất quanh giá trị ở giữa – giá trị trung bình.

Năm 1809, nhà toán học Friederich Gauss người Đức đã xác định phương trình của hàm mật độ một biến ngẫu nhiên liên tục có dạng phù hợp với các nghiên cứu thực nghiệm. Các chứng minh sau này cho thấy các dạng biến ngẫu nhiên khác đều sẽ hội tụ về dạng hàm mật độ này, do đó được gọi là phân phối Chuẩn.

4.2.1. Định nghĩa phân phối Chuẩn

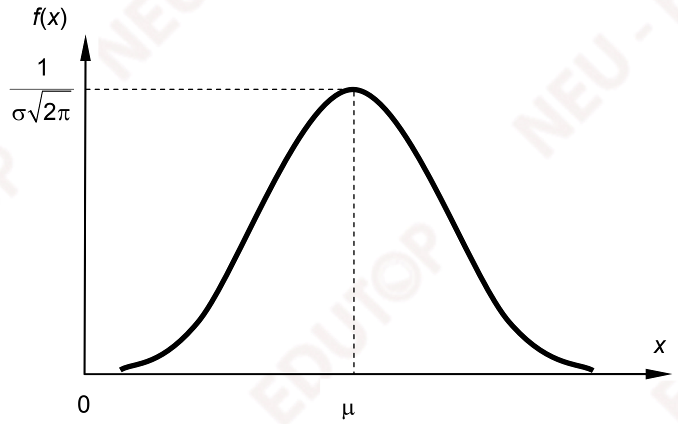
Định nghĩa 4.2 – Biến phân phối Chuẩn: *Biến ngẫu nhiên liên tục X phân phối Chuẩn với hai tham số μ và σ², ký hiệu là* $X \sim N(\mu;\sigma^2)$, *nếu hàm mật độ xác suất của nó có dạng:*

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

Trong công thức trên, x là giá trị của biến ngẫu nhiên; μ và σ^2 là các tham số; π và e là các hằng số của tự nhiên, $\pi \approx 3{,}14$; $e \approx 2{,}718$. Công thức khá phức tạp, tuy nhiên việc tính toán sẽ đơn giản vì các giá trị cần tìm sẽ được cho sẵn trong bảng số.

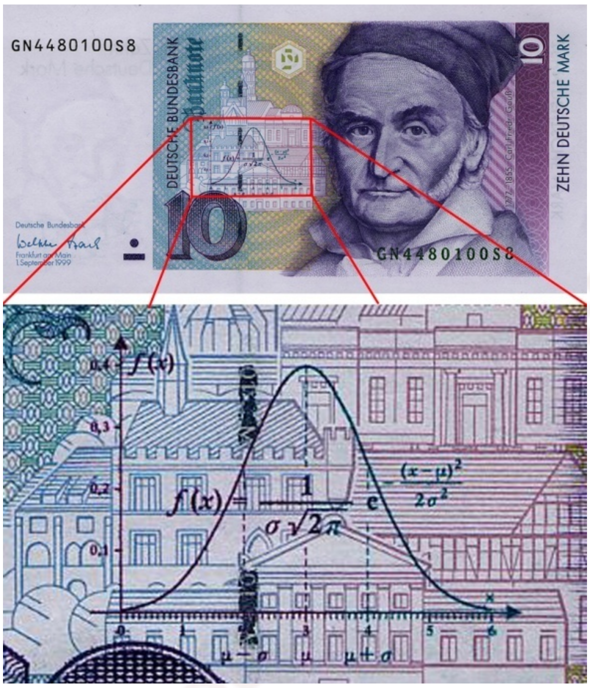
Biến ngẫu nhiên liên tục phân phối Chuẩn gọi tắt là biến phân phối Chuẩn.

Với công thức trên, đường cong đồ thị hàm mật độ có dạng quả chuông, đối xứng qua đường $x = \mu$ và nhận Ox làm tiệm cận ngang. Đỉnh của quả chuông phân phối Chuẩn có hoành độ là $x = \mu$, thay vào hàm mật độ thì phần bên phải bằng 1 nên giá trị hàm $f(x)$, cũng là tung độ của đỉnh, và bằng $1/\sigma\sqrt{2\pi}$. Hình dáng của hàm mật độ được minh họa trong hình 4.5.



Hình 4.5. ***Hàm mật độ biến ngẫu nhiên phân phối Chuẩn X ~ N(μ; σ²)***

Vì tầm quan trọng của phân phối Chuẩn, hình ảnh của quả chuông hàm mật độ chuẩn hóa và công thức đã được in lên đồng tiền 10 Deutsche Mark của Đức (lưu hành đến hết năm 2001, trước khi đồng Euro thay thế).



Hình 4.6. ***Đồng tiền 10 Deutsche Mark của Đức với chân dung Friederich Gauss và đồ thị, phương trình hàm mật độ phân phối Chuẩn***

4.2.2. Tính chất biến ngẫu nhiên phân phối Chuẩn

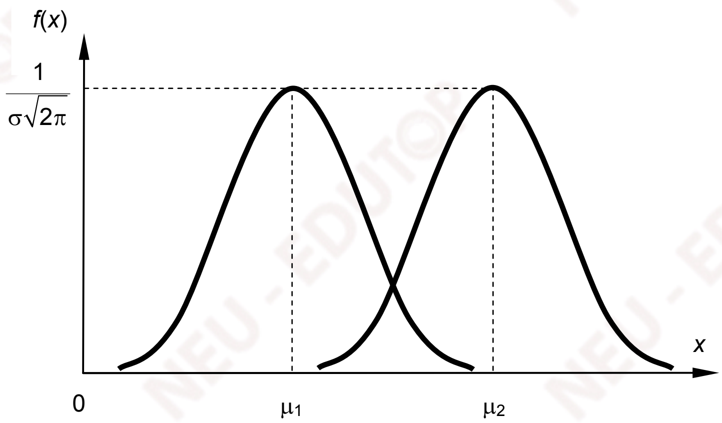
Biến ngẫu nhiên liên tục phân phối Chuẩn $X \sim N(\mu; \sigma^2)$ đối xứng qua $x = \mu$, do đó giá trị kỳ vọng chính bằng μ. Chứng minh được phương sai của biến ngẫu nhiên X cũng bằng chính σ^2 và do đó độ lệch chuẩn bằng σ.

$$X \sim N(\mu;\sigma^2) \text{ suy ra: } E(X) = \mu\,;\ V(X) = \sigma^2\,;\ \sigma(X) = \sigma\,.$$

Ví dụ 4.4.

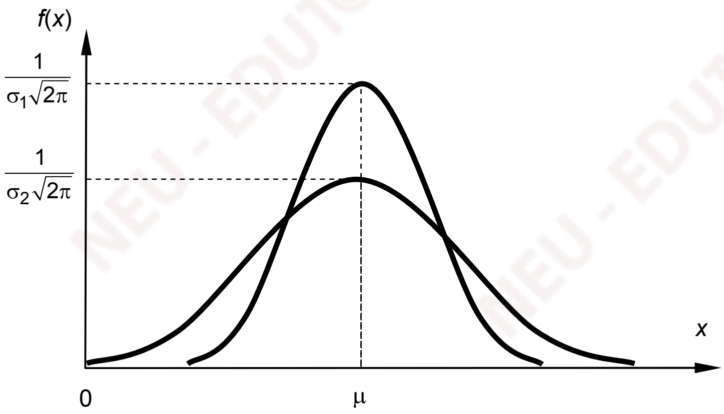
- Khi viết $X \sim N(200;\ 16)$ thì có nghĩa là X phân phối Chuẩn với hai tham số $\mu = 200$ và $\sigma^2 = 16$; hay X có kỳ vọng bằng 200 và phương sai bằng 16, độ lệch chuẩn bằng 4. Do đó cũng có thể viết là $X \sim N(200;\ 4^2)$.
- Khi nói: X có đơn vị là gam, phân phối Chuẩn với kỳ vọng, hay trung bình, bằng 100g; phương sai $25g^2$, thì có thể viết: $X \sim N(100;\ 25)$ hoặc $X \sim N(100;\ 5^2)$.

Để hiểu rõ hơn ý nghĩa của hình ảnh hàm mật độ phân phối Chuẩn, xét hình minh họa 4.7 và 4.8 về hai trường hợp khi kỳ vọng thay đổi và phương sai thay đổi.



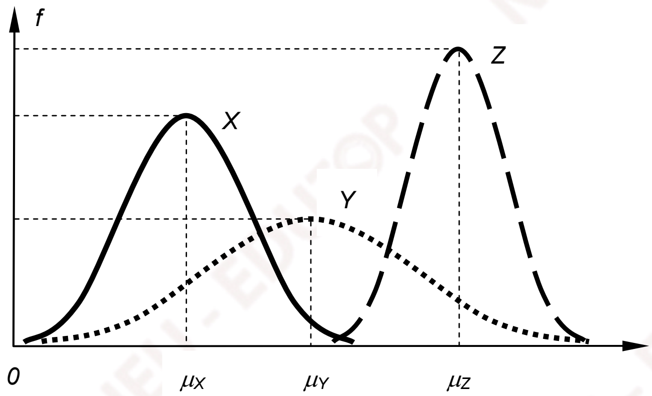
Hình 4.7. ***Hàm mật độ phân phối Chuẩn khi kỳ vọng thay đổi***

Trong hình 4.7, khi chỉ có kỳ vọng (trung bình) μ thay đổi, tăng từ μ_1 lên μ_2, phương sai σ^2 không đổi thì đồ thị sẽ dịch chuyển sang phải, hình dáng quả chuông và độ cao không thay đổi.



Hình 4.8. ***Hàm mật độ phân phối Chuẩn khi phương sai thay đổi***

Trong hình 4.8, khi phương sai σ^2 thay đổi, kỳ vọng μ không đổi thì độ cao của quả chuông thay đổi, vị trí của quả chuông không đổi. Khi phương sai tăng từ σ_1^2 lên σ_2^2 thì chiều cao của đỉnh quả chuông sẽ giảm xuống (do σ nằm ở mẫu số). Quả chuông thấp xuống, nhưng vì diện tích tạo bởi quả chuông luôn phải bằng 1 nên quả chuông sẽ rộng hơn, độ phân tán lớn hơn.



Hình 4.9. ***So sánh ba hàm mật độ phân phối Chuẩn của X, Y, Z***

Trong hình 4.9 ta có ba hàm mật độ của ba biến phân phối Chuẩn *X*, *Y*, *Z*. Qua đồ thị nhận thấy về kỳ vọng (trung bình) thì biến *X* là nhỏ nhất, biến *Y* có kỳ vọng lớn hơn *X* nhưng nhỏ hơn *Z* và biến *Z* có kỳ vọng là lớn nhất. Về phương sai thì biến *Z* là nhỏ nhất, tiếp đó là biến *X*, biến *Y* có phương sai lớn nhất.

Như vậy với biến ngẫu nhiên phân phối Chuẩn có thể rút ra nhận xét về hàm mật độ xác suất có dạng quả chuông như sau:

- Khi kỳ vọng (trung bình) μ thay đổi thì vị trí của quả chuông thay đổi; μ tăng thì quả chuông dịch sang phải; μ giảm thì quả chuông dịch sang trái.
- Khi phương sai σ^2 thay đổi thì độ cao và độ rộng của quả chuông thay đổi; σ^2 tăng thì quả chuông thấp xuống, rộng và bẹt hơn; σ^2 giảm thì quả chuông cao lên, hẹp và nhọn hơn.
- Quả chuông càng thấp thì độ phân tán càng lớn, biến dao động càng nhiều.

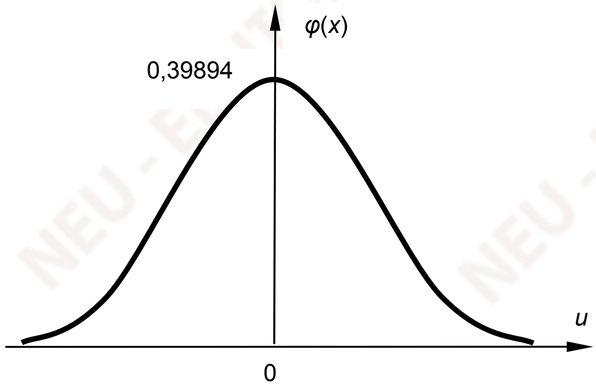
4.2.3. Biến ngẫu nhiên phân phối Chuẩn hóa

Các biến ngẫu nhiên phân phối Chuẩn có đồ thị quả chuông tại vị trí khác nhau, độ cao thấp khác nhau, do đó không thuận lợi trong tính toán các xác suất. Để việc tính toán được thuận lợi, ta xét một biến ngẫu nhiên phân phối Chuẩn đặc biệt là biến ngẫu nhiên phân phối Chuẩn hóa, được ký hiệu là U.

Định nghĩa 4.3 – Biến phân phối Chuẩn hóa: *Biến ngẫu nhiên U phân phối Chuẩn hóa nếu nó phân phối Chuẩn với kỳ vọng bằng 0 và phương sai bằng 1, ký hiệu là* $U \sim N(0; 1)$.

Như vậy $U \sim N(\mu_U; \sigma_U^2)$ với $\mu_U = 0$, $\sigma_U^2 = 1$ nên $\sigma_U = 1$.

Biến ngẫu nhiên phân phối Chuẩn hóa được gọi tắt là *biến phân phối Chuẩn hóa*. Hàm mật độ của biến Chuẩn hóa thay vì dùng hàm $f(x)$, sẽ dùng ký hiệu là hàm $\varphi(u)$. Đồ thị hàm $\varphi(u)$ được minh họa trong hình 4.10.



Hình 4.10. ***Hàm mật độ biến phân phối Chuẩn hóa N(0;1)***

Hàm mật độ Chuẩn hóa trong hình 4.10 cho thấy hàm $\varphi(u)$ đối xứng qua trục tung, đỉnh đạt tại điểm có tung độ là: $\dfrac{1}{\sigma_U\sqrt{2\pi}} = \dfrac{1}{\sqrt{2\pi}} \approx 0,39894$.

Xác suất của biến ngẫu nhiên Chuẩn hóa

Với biến ngẫu nhiên Chuẩn hóa $U \sim N(0;1)$, các giá trị xác suất để U nhỏ hơn một con số u đã được tính sẵn và cho trong bảng Phụ lục 2. Đối chiếu con số cần tra theo dòng và cột trong bảng để tìm giá trị xác suất.

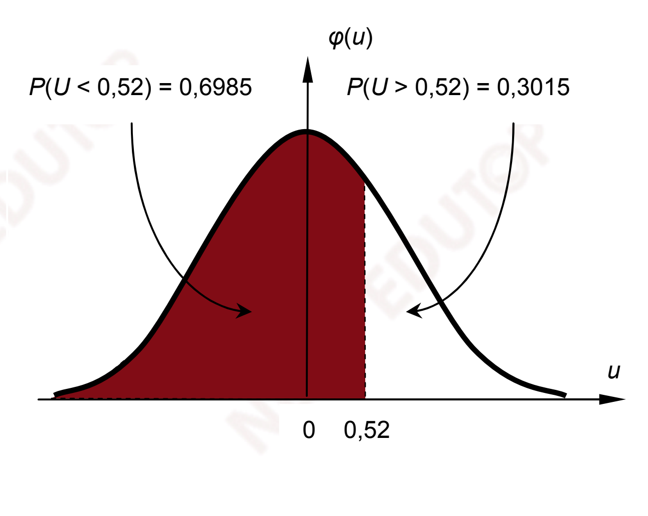
Ví dụ 4.5:

Để tìm xác suất để U nhỏ hơn 0,52 ta tìm dòng 0,5 và cột 0,02 vì 0,5 + 0,02 = 0,52. Vị trí của dòng và cột tương ứng cho con số 0,6985. Như vậy $P(U < 0,52) = 0,6985$.

Từ đó suy ra xác suất U lớn hơn 0,52 bằng 1 trừ xác suất U nhỏ hơn 0,52 do đó:

$$P(U > 0,52) = 1 - P(U < 0,52) = 1 - 0,6985 = 0,3015$$

u	0.00	0.01	0.02	0.03	…
0.0	0.5000	0.5040	0.5080	0.5120	…
0.1	0.5398	0.5438	0.5478	0.5517	…
0.2	0.5793	0.5832	0.5871	0.5910	…
0.3	0.6179	0.6217	0.6255	0.6293	…
0.4	0.6554	0.6591	0.6628	0.6664	…
0.5	0.6915	0.6950	0.6985	0.7019	…
0.6	0.7257	0.7291	0.7324	0.7357	…
0.7	0.7580	0.7611	0.7642	0.7673	…
…	…	…	…	…	…



Hình 4.11. *Xác suất của biến ngẫu nhiên Chuẩn hóa P(U < 0,52) = 0,6985*

Tương tự, có thể tra bảng để thấy xác suất để U nhỏ hơn 1,25 là:

$$P(U < 1{,}25) = 0{,}8944$$

Do tính chất đối xứng của hàm mật độ phân phối Chuẩn hóa quanh trục tung, nên có thể dễ dàng suy ra xác suất để U nhỏ hơn một số âm.

Ví dụ để tính $P(U < -0{,}52)$, do tính chất đối xứng nên:

$$P(U < -0{,}52) = P(U > 0{,}52) = 1 - P(U < 0{,}52) = 0{,}3015.$$

Tương tự: $P(U < -2) = P(U > 2) = 1 - P(U < 2) = 1 - 0{,}9772 = 0{,}0228$.

Tuy nhiên để giảm bớt việc tính toán, các giá trị ứng với số âm cũng được cho sẵn trong bảng phụ lục. Bạn đọc có thể tự kiểm tra lại các xác suất trong các ví dụ trên.

Khi tính xác suất chính xác đến 4 số thập phân, tra bảng với con số lớn hơn 3,99 thì xác suất xấp xỉ bằng 1 và với con số nhỏ hơn –3,99 thì xác suất xấp xỉ bằng 0.

4.2.4. Công thức tính xác suất

Có thể sử dụng các giá trị xác suất đã có với biến ngẫu nhiên Chuẩn hóa để tính xác suất của biến ngẫu nhiên phân phối Chuẩn bất kỳ. Giữa biến ngẫu nhiên phân phối Chuẩn bất kỳ và biến ngẫu nhiên phân phối Chuẩn hóa có mối liên hệ đặc biệt như sau:

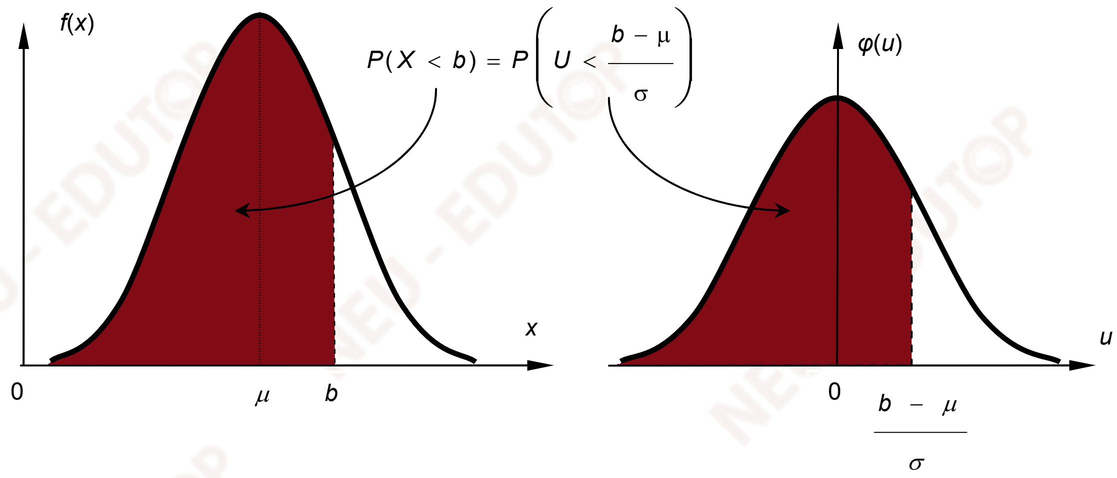
Nếu X phân phối Chuẩn trung bình là μ, phương sai σ^2 thì biến ngẫu nhiên $U = \dfrac{X - \mu}{\sigma}$ sẽ phân phối Chuẩn hóa.

Hay $X \sim N(\mu; \sigma^2)$ thì $U = \dfrac{X - \mu}{\sigma} \sim N(0;1)$

Từ đó ta có công thức quan trọng để tính xác suất của biến ngẫu nhiên phân phối Chuẩn:

$$X \sim N(\mu; \sigma^2) \text{ thì } P(X < b) = P\left(U < \frac{b - \mu}{\sigma}\right)$$

Ý nghĩa hình học của công thức được minh họa qua hình 4.12.

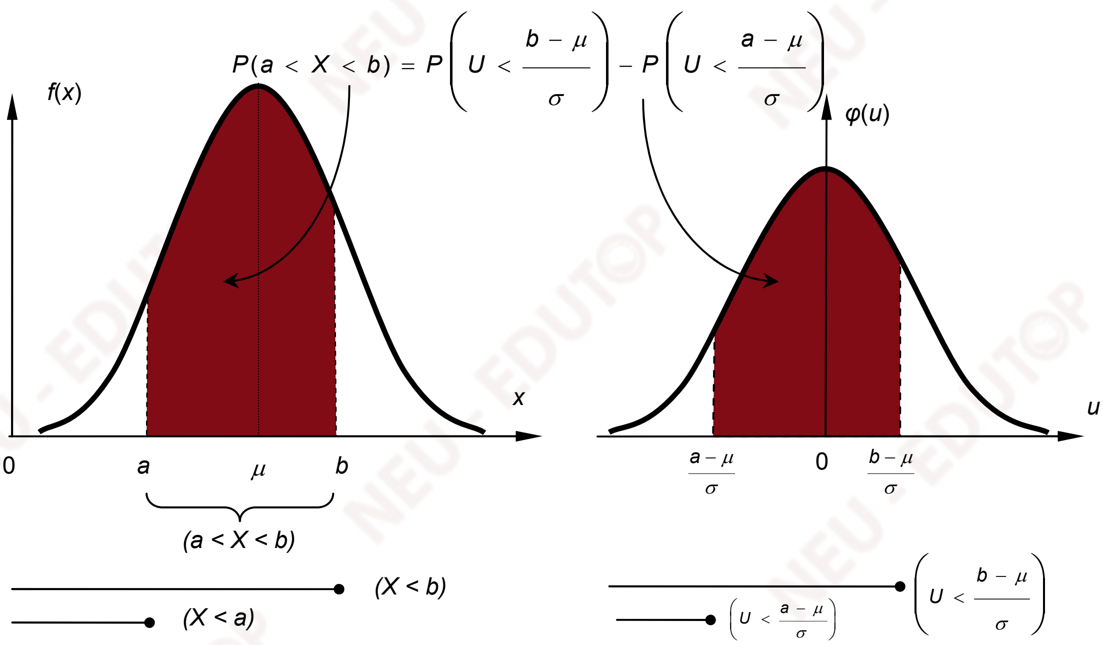


Hình 4.12. ***Tính xác suất qua biến ngẫu nhiên Chuẩn hóa***

Từ công thức trên, dễ dàng suy ra:

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

Có thể minh họa công thức qua hình 4.13.



Hình 4.13. ***Minh họa công thức tính xác suất trong khoảng***

Nhắc lại là với biến ngẫu nhiên phân phối Chuẩn là biến ngẫu nhiên liên tục nên việc xét dấu bằng đối với các khoảng là không cần thiết.

Ví dụ 4.6.

Thời gian một khách chờ đợi ở một quầy dịch vụ công (đơn vị: phút) là biến ngẫu nhiên phân phối Chuẩn với trung bình 15 phút và phương sai là 16 phút^2.

(a) Tính xác suất một khách đến quầy chờ ít hơn 17 phút.

(b) Tính tỷ lệ khách đến quầy phải chờ từ 10 đến 16 phút.

(c) Tính tỷ lệ khách phải chờ nhiều hơn 12 phút.

(d) Xác định một mức thời gian mà 15% số khách phải chờ lâu hơn mức thời gian đó.

Giải:

Ký hiệu thời gian chờ đợi là X (phút), X phân phối Chuẩn $X \sim N(\mu; \sigma^2)$. Theo đề bài thì trung bình của X là 15 cũng chính là kỳ vọng của X bằng 15 hay $\mu = 15$; phương sai của X bằng 16 hay $\sigma^2 = 16$, suy ra độ lệch chuẩn bằng $\sqrt{16}$ bằng 4 hay $\sigma = 4$.

Có thể viết $X \sim N(15;16)$ hoặc $N(15; 4^2)$

(a) Xác suất một khách hàng đến quầy chờ ít hơn 17 phút là $P(X < 17)$

Theo công thức ta có:

$$P(X<17)=P\left(U<\frac{17-\mu}{\sigma}\right)=P\left(U<\frac{17-15}{4}\right)=P(U<0,5)$$

Tra bảng phụ lục ta có: $P(U < 0,5) = 0,6915$.

Vậy xác suất một khách đến quầy chờ ít hơn 17 phút là 0,6915.

(b) Để tính tỷ lệ khách đến quầy chờ từ 10 đến 16 phút ta tính xác suất một khách hàng chờ trong khoảng 10 đến 16 phút, hay $P(10 < X < 16)$.

Theo công thức ta có:

$$P(10<X<16)=P\left(U<\frac{16-\mu}{\sigma}\right)-P\left(U<\frac{10-\mu}{\sigma}\right)$$
$$=P\left(U<\frac{16-15}{4}\right)-P\left(U<\frac{10-15}{4}\right)=P(U<0,25)-P(U<-1,25)$$

Tra bảng ta được: $P(U < 0,25) = 0,5987$ và $P(U < -1,25) = 0,1056$

Suy ra: $P(10 < X < 16) = 0,5987 - 0,1056 = 0,4931$

Vậy tỷ lệ khách chờ từ 10 đến 16 phút là 0,4931 hay 49,31%.

(c) Để tính tỷ lệ khách đến chờ nhiều hơn 12 phút ta tính $P(X > 12)$.

Để tính xác suất X lớn hơn 12, trong công thức chỉ có xác suất nhỏ hơn một số, do đó ta tính xác suất biến cố đối lập:

$$P(X>12)=1-P(X<12)=1-P\left(U<\frac{12-15}{4}\right)=1-P(U<-0,75)$$

Tra bảng ta được: $P(U < -0,75) = 0,2266$

Suy ra: $P(X > 12) = 1 - 0,2266 = 0,7734$

Vậy tỷ lệ khách chờ nhiều hơn 12 phút là 0,7734 hay 77,34%

(d) Câu hỏi này yêu cầu một suy luận ngược. Đề bài không cho trước con số để tính xác suất, mà lại cho trước xác suất để tìm con số. Do đó chúng ta phải lập luận theo cách sau:

Gọi mức cần tìm là a, ta phải tìm a sao cho: $P(X > a) = 0,15$

Theo công thức ta có:

$$P(X>a)=1-P(X<a)=1-P\left(U<\frac{a-15}{4}\right)$$

Suy ra: $1-P\left(U<\dfrac{a-15}{4}\right)=0,15$ hay $P\left(U<\dfrac{a-15}{4}\right)=0,85$

Tra trong bảng số xem có giá trị xác suất nào trong lòng bảng gần với giá trị 0,85 nhất, ta thấy giá trị gần nhất là 0,8508 ứng với con số 1,04.

u	…	0.03	0.04	0.05	…
…	…	…	…	…	…
0.8	…	0.7967	0.7995	0.8023	…
0.9	…	0.8238	0.8264	0.8289	…
1.0	…	0.8485	0.8508	0.8531	…
1.1	…	0.8708	0.8729	0.8749	…
…	…	…	…	…	…

Như vậy: $P(U < 1{,}04) \approx 0{,}85$, do đó $\dfrac{a-15}{4} = 1{,}04$

Hay: $a = 15 + 1{,}04 \times 4 = 19{,}15$

Vậy mức thời gian mà 15% khách hàng chờ lâu hơn là 19,16 phút.

Ví dụ 4.7.

Đây chính là cụ thể hóa của tình huống dẫn nhập.

Doanh nghiệp sản xuất sản phẩm điện tử với tuổi thọ của sản phẩm là biến ngẫu nhiên phân phối Chuẩn với trung bình bằng 3160 giờ, phương sai là 6400 giờ^2. Thời gian bảo hành của sản phẩm là 3000 giờ. Khi bán một sản phẩm thì doanh nghiệp lãi là 2 triệu đồng, nhưng nếu sản phẩm dừng hoạt động trong thời hạn bảo hành thì doanh nghiệp phải đền bù và khi đó sẽ lỗ 10 triệu.

(a) Tính tỷ lệ sản phẩm dừng hoạt động trong thời hạn bảo hành.

(b) Tính kỳ vọng, phương sai và độ lệch chuẩn của lợi nhuận khi bán một sản phẩm.

(c) Nếu giảm thời hạn bảo hành xuống còn 2940 giờ thì kỳ vọng tiền lãi thay đổi thế nào?

(d) Nếu thời hạn bảo hành 3000 giờ là không thể thay đổi và chi phí để tăng tuổi thọ trung bình của sản phẩm lên 3200 giờ là 100 nghìn đồng thì xét về mặt lợi nhuận, doanh nghiệp có nên bỏ chi phí đó để tăng tuổi thọ sản phẩm lên hay không?

(e) Doanh nghiệp có thể chi phí để giảm phương sai của tuổi thọ xuống 2500 giờ^2. Mức chi phí như thế nào thì được coi là có hiệu quả kinh tế với doanh nghiệp?

Giải:

Đặt X là tuổi thọ sản phẩm, đơn vị là giờ, do giả thiết phân phối Chuẩn: $X \sim N(\mu;\, \sigma^2)$. Theo đề bài $\mu = 3200$ (giờ), $\sigma^2 = 6400$ (giờ^2), suy ra $\sigma = 80$ (giờ).

(a) Để tính tỷ lệ sản phẩm dừng hoạt động trong thời hạn bảo hành, ta tính xác suất để tuổi thọ của sản phẩm là ngắn hơn thời hạn bảo hành, hay $P(X < 3000)$.

Theo công thức:

$$P(X < 3000) = P\left(U < \frac{3000-3160}{80}\right) = P(U < -2)$$

Tra bảng số ta có: $P(U < -2) = 0{,}0228$

Vậy tỷ lệ sản phẩm dừng hoạt động trong thời hạn bảo hành là 0,0228 hay 2,28%.

(b) Đặt Y là lợi nhuận khi bán một sản phẩm, đơn vị là triệu đồng. Nhận thấy Y không phải là một số cố định, mà còn tùy thuộc vào sản phẩm có bị dừng hoạt động trong thời gian bảo hành hay không, nên Y là biến ngẫu nhiên. Y có hai giá trị có thể xảy ra là bằng 2 khi sản phẩm không bị hỏng trong thời gian bảo hành, và Y bằng âm 10 khi sản phẩm bị hỏng trong thời gian bảo hành.

Để tính kỳ vọng và phương sai của Y, lập bảng phân phối xác suất.

Xác suất để Y bằng âm 10 chính là xác suất đã tính trong câu (a), và bằng 0,0228.

Xác suất để Y bằng 2 là xác suất đối lập với câu (a), bằng 1 – 0,0228 = 0,9772.

Ta có bảng phân phối xác suất của Y:

Y	–10	2
P(y)	0,0228	0,9772

Kỳ vọng của Y là: $E(Y) = -10 \times 0{,}0228 + 2 \times 0{,}9772 = 1{,}7264$

$$E(Y^2) = (-10)^2 \times 0{,}0228 + 2^2 \times 0{,}9772 = 6{,}1888$$

Phương sai của Y là: $V(Y) = E(Y^2) - (E(Y))^2 = 6{,}1888 - 1{,}7264^2 \approx 3{,}2$

Độ lệch chuẩn của Y là: $\sigma(Y) = \sqrt{3{,}2} \approx 1{,}79$

Vậy lợi nhuận khi bán một sản phẩm có kỳ vọng là 1,7264 (triệu đồng), phương sai là 3,2 (triệu đồng)^2 và độ lệch chuẩn là 1,79 (triệu đồng).

(c) Khi giảm thời hạn bảo hành thì xác suất sản phẩm bị hỏng trong thời gian bảo hành giảm xuống, do đó giá trị xác suất trong bảng phân phối xác suất của Y cũng thay đổi. Để tính kỳ vọng lợi nhuận, cần tính lại các xác suất và lập bảng phân phối xác suất mới.

Xác suất một sản phẩm bị hỏng trước thời hạn bảo hành 2940 giờ là:

$$P(X < 2950) = P\left(U < \frac{2940 - 3160}{80}\right) = P(U < -2{,}75)$$

Tra bảng số ta có: P(U < –2,75) = 0,003

Do đó xác suất sản phẩm hỏng trong thời gian bảo hành là 0,003 và xác suất sản phẩm không hỏng là 0,997.

Bảng phân phối xác suất của lợi nhuận là:

Y	–10	2
P(y)	0,003	0,997

Kỳ vọng của Y là:

$$E(Y) = -10 \times 0{,}003 + 2 \times 0{,}997 = 1{,}964$$

Vậy kỳ vọng lợi nhuận khi thời hạn bảo hành giảm xuống 2940 giờ là 1,964 triệu đồng, tăng thêm 0,2376 triệu đồng so với khi thời hạn bảo hành là 3000 giờ.

(d) Khi thời hạn bảo hành không đổi nhưng tuổi thọ trung bình của sản phẩm tăng thêm thì xác suất sản phẩm hỏng trong thời gian bảo hành sẽ giảm đi, do đó làm tăng lợi nhuận kỳ vọng. Nếu mức tăng của lợi nhuận kỳ vọng là vượt quá chi phí bỏ ra thì việc chi phí là có hiệu quả về kinh tế, nếu ngược lại thì không có hiệu quả về kinh tế. Ta sẽ thực hiện tính toán lần lượt các kết quả.

Với thời hạn bảo hành 3000 giờ, tuổi thọ trung bình mới bằng 3200 giờ, hay $\mu = 3200$ thì xác suất sản phẩm hỏng trong thời gian bảo hành là:

$$P(X < 3000) = P\left(U < \frac{3000-3200}{80}\right) = P(U < -2,5)$$

Tra bảng số ta có: $P(U < -2,5) = 0,0062$

Do đó xác suất sản phẩm hỏng trong thời gian bảo hành là 0,0062 và xác suất sản phẩm không hỏng là 0,9938.

Bảng phân phối xác suất của lợi nhuận là:

Y	–10	2
***P*(*y*)**	0,0062	0,9938

Kỳ vọng của *Y* là:

$$E(Y) = -10 \times 0,0062 + 2 \times 0,9938 = 1,9256$$

Lợi nhuận kỳ vọng là 1,9256 triệu đồng, so với lợi nhuận kỳ vọng khi không tăng tuổi thọ trung bình đã tính trong câu (a) là 1,7264 triệu đồng thì tăng là 0,1992 triệu đồng.

Chi phí bỏ ra để tăng tuổi thọ trung bình của sản phẩm là 100 nghìn đồng hay 0,1 triệu đồng, mức này nhỏ hơn mức 0,1992 triệu đồng ở trên, do đó bỏ chi phí là có hiệu quả về kinh tế. Khi đó mức kỳ vọng lợi nhuận ròng mới sẽ tăng thêm là 0,0992 triệu đồng.

(e) Khi thời hạn bảo hành, tuổi thọ kỳ vọng không đổi, phương sai tuổi thọ giảm thì xác suất sản phẩm hỏng trong thời gian bảo hành sẽ giảm đi, do đó lợi nhuận kỳ vọng có thể tăng lên. Mức lợi nhuận kỳ vọng tăng lên sẽ cho biết doanh nghiệp có thể chấp nhận chi phí tối đa bao nhiêu.

Với thời hạn bảo hành 3000 giờ, tuổi thọ trung bình $\mu = 3260$, phương sai là 2500 giờ^2 hay độ lệch chuẩn $\sigma = 50$ giờ thì xác suất sản phẩm hỏng trong thời gian bảo hành là:

$$P(X < 3000) = P\left(U < \frac{3000-3160}{50}\right) = P(U < -3,2)$$

Tra bảng số ta có: $P(U < -3,2) = 0,0007$

Do đó xác suất sản phẩm hỏng trong thời gian bảo hành là 0,0007 và xác suất sản phẩm không hỏng là 0,9993.

Bảng phân phối xác suất của lợi nhuận là:

Y	–10	2
***P*(*y*)**	0,0007	0,9993

Kỳ vọng của *Y* là:

$$E(Y) = -10 \times 0,0007 + 2 \times 0,9993 = 1,9916$$

Lợi nhuận kỳ vọng mới bằng 1,9916 triệu đồng, so với khi không giảm phương sai đã tính trong câu (a) là 1,7264 triệu đồng thì tăng lên 0,2652 triệu đồng.

Do đó nếu chi phí để giảm phương sai tuổi thọ là nhỏ hơn 0,2652 triệu đồng thì chi phí đó có hiệu quả kinh tế.

4.2.5. Giá trị tới hạn Chuẩn

Trong ví dụ ở trên, ta thấy bên cạnh bài toán xuôi là cho trước một con số, tính xác suất để biến ngẫu nhiên phân phối Chuẩn hoặc Chuẩn hóa lớn hơn con số đó. Trong thực tế có những bài toán ngược như câu (d) của ví dụ 4.6, cho trước mức xác suất, cần tìm con số để xác suất biến ngẫu nhiên lớn hơn con số đó bằng xác suất cho trước. Bài toán tìm con số đó gọi là bài toán tìm giá trị tới hạn.

Với ví dụ 4.6 câu (d), $X \sim N(\mu = 15; \sigma^2 = 16)$, cho trước xác suất bằng 0,15 cần tìm con số a sao cho xác suất X lớn hơn a bằng 0,15. Con số tìm được là 19,16. Khi đó con số 19,16 được gọi là giá trị tới hạn mức 0,15 của biến X.

Trong các tính toán về sau, không cần tính giá trị tới hạn của biến ngẫu nhiên phân phối Chuẩn bất kỳ, mà chỉ cần tính giá trị tới hạn của biến ngẫu nhiên Chuẩn hóa, do đó có khái niệm giá trị tới hạn Chuẩn với mức xác suất α tổng quát như sau:

Định nghĩa 4.4 – Giá trị tới hạn Chuẩn: *Giá trị tới hạn Chuẩn mức α là một con số, ký hiệu là u_α, sao cho với U phân phối Chuẩn hóa thì xác suất để U lớn hơn u_α bằng đúng α.*

Nghĩa là: $P(U > u_\alpha) = \alpha$. Tất nhiên theo khái niệm xác suất thì $0 \le \alpha \le 1$.

Ví dụ 4.8.

Trong ví dụ 4.5 ở trên ta đã tính được: $P(U < 0{,}52) = 0{,}6985$ suy ra $P(U > 0{,}52) = 0{,}3015$. Do đó có thể nói giá trị tới hạn Chuẩn mức 0,3015 là 0,52, hay ký hiệu là: $u_{0,3015} = 0{,}52$.

Tương tự: $P(U < 1{,}25) = 0{,}8944$ nên $P(U > 1{,}25) = 0{,}1056$, suy ra: $u_{0,1056} = 1{,}25$.

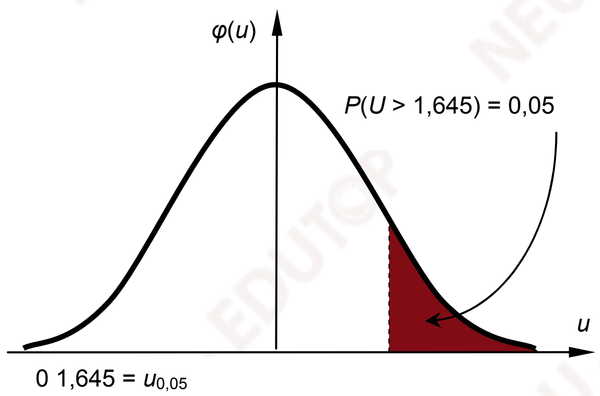
Trong thực tế thường dùng một số giá trị tới hạn Chuẩn quan trọng, đó là giá trị tới hạn Chuẩn với mức $\alpha = 0{,}05$ (hay 5%) và mức $\alpha = 0{,}025$ (hay 2,5%). Các giá trị quan trọng này được cho trong dòng cuối của bảng phụ lục số 4 hoặc ngay tại bảng sau:

Giá trị tới hạn chuẩn một số mức quan trọng:

α	0,1	0,05	0,025	0,01	0,005
u_α	1,282	1,645	1,960	2,326	2,576

Trong đó hai giá trị $u_{0,05} = 1{,}645$ và $u_{0,025} = 1{,}96$ là thường xuyên được sử dụng.

Hình 4.14 minh họa cho giá trị tới hạn Chuẩn mức 0,05.

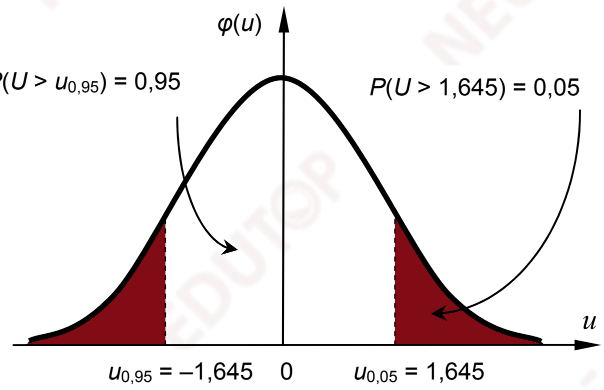


Hình 4.14. ***Giá trị tới hạn Chuẩn mức 0,05***

Từ định nghĩa của giá trị tới hạn Chuẩn và tính đối xứng của hàm mật độ Chuẩn hóa, có một số tính chất sau:

- Mức α càng lớn thì giá trị tới hạn Chuẩn u_α càng nhỏ.
- $u_{0,5} = 0$: do đồ thị đối xứng qua trục tung, $P(U > 0) = 0,5$ nên $u_{0,5} = 0$.
- $u_0 = +\infty$: do đồ thị tiệm cận với trục hoành nên $P(U > +\infty) = 0$.
- $u_1 = -\infty$: do đồ thị tiệm cận với trục hoành nên $P(U > -\infty) = 1$.
- $u_{1-\alpha} = -u_\alpha$: do tính đối xứng của đồ thị.

Hình 4.15 minh họa tính chất đối xứng của giá trị tới hạn với α = 5%



Hình 4.15. ***Tính chất đối xứng của giá trị tới hạn***

4.2.6. Sự hội tụ về quy luật Chuẩn

Chứng minh được một số tính chất như sau:

- Tổ hợp bậc nhất của các biến ngẫu nhiên phân phối Chuẩn cũng sẽ phân phối Chuẩn.
- Biến ngẫu nhiên phân phối Nhị thức $B(n; p)$ với n lớn hơn hoặc bằng 100 cũng hội tụ xấp xỉ phân phối Chuẩn.
- Tổng của một số lượng đủ lớn biến ngẫu nhiên có cùng quy luật phân phối, dù quy luật đó là như thế nào, thì cũng sẽ phân phối Chuẩn. Số lượng đủ lớn đó thường được lấy là trên 30.

Ví dụ 4.9.

Giả sử với người lao động lương và thu nhập khác là độc lập nhau, có đơn vị là USD. Lương phân phối Chuẩn với trung bình là 400 USD, độ lệch chuẩn là 80 USD; Thu nhập khác phân phối Chuẩn với trung bình là 200 USD, độ lệch chuẩn là 60 USD.

(a) Tính xác suất để với một người lao động bất kỳ lương cao hơn 460 USD và thu nhập khác cao hơn 240 USD.

(b) Tính xác suất để một người lao động bất kỳ có tổng thu nhập cao hơn 700 USD.

Giải:

Ký hiệu Lương là X và Thu nhập khác là Y, theo đề bài ta có: $X \sim N(\mu_X = 400; \sigma_X^2 = 80^2)$ và $Y \sim N(\mu_Y = 200; \sigma_Y^2 = 60^2)$; có $\sigma_X = 80$ và $\sigma_Y = 60$.

(a) Để tính xác suất xảy ra đồng thời biến cố lương cao hơn 450 và thu nhập khác cao hơn 250, do hai biến cố này độc lập nhau nên xác suất biến cố tích bằng tích hai xác suất.

$P(X > 460$ và $Y > 240) = P(X > 460) \times P(Y > 240)$

Theo công thức tính riêng các xác suất ta có:

$$P(X > 460) = 1 - P(X < 460) = 1 - P\left(U < \frac{460 - \mu_X}{\sigma_X}\right)$$

$$= 1 - P\left(U < \frac{460 - 400}{80}\right) = 1 - P(U < 0,75) = 1 - 0,7734 = 0,2266$$

$$P(Y > 240) = 1 - P(Y < 240) = 1 - P\left(U < \frac{240 - \mu_Y}{\sigma_Y}\right)$$

$$= 1 - P\left(U < \frac{240 - 200}{60}\right) = 1 - P(U < 0,67) = 1 - 0,7486 = 0,2514$$

Suy ra $P(X > 460$ và $Y > 240) = 0,2266 \times 0,2514 \approx 0,057$

Vậy xác suất để một người lao động bất kỳ có lương cao hơn 460 và thu nhập khác cao hơn 240 là 0,057.

(b) Tổng thu nhập bằng lương cộng với thu nhập khác. Đặt T là tổng thu nhập thì $T = X + Y$. Trong câu (a) có một trường hợp tổng thu nhập lớn hơn 700 (lương cao hơn 460 và thu nhập khác cao hơn 240), tuy nhiên còn vô số trường hợp khác tổng thu nhập cũng lớn hơn 700, do đó không thể dùng kết quả câu (a) để tính cho câu (b) được.

Do tính chất "tổ hợp bậc nhất của các biến ngẫu nhiên phân phối Chuẩn cũng phân phối Chuẩn, do đó T bằng tổng của X và Y thì T cũng sẽ phân phối Chuẩn.

$T \sim N(\mu_T; \sigma_T^2)$

Để tính xác suất tổng thu nhập lớn hơn 700 hay $P(T > 700)$, cần phải biết μ_T và σ_T.

Theo tính chất của kỳ vọng và phương sai, do $T = X + Y$ mà X và Y độc lập nên:

$E(T) = E(X + Y) = E(X) + E(Y)$ và $V(T) = V(X + Y) = V(X) + V(Y)$.

Như vậy: $\mu_T = \mu_X + \mu_Y = 400 + 200 = 600$ (USD)

Và: $\sigma_T^2 = \sigma_X^2 + \sigma_Y^2 = 80^2 + 60^2 = 6400 + 3600 = 10000$ (USD^2)

Suy ra: $\sigma_T = 100$ (USD)

Theo công thức tính xác suất, ta có:

$$P(T > 700) = 1 - P(X < 700) = 1 - P\left(U < \frac{460 - \mu_T}{\sigma_T}\right)$$

$$= 1 - P\left(U < \frac{700 - 600}{100}\right) = 1 - P(U < 1) = 1 - 0,8413 = 0,1587$$

Vậy xác suất để tổng thu nhập lớn hơn 700 là 0,1587.

4.3. Biến ngẫu nhiên phân phối Khi – bình phương

Biến ngẫu nhiên phân phối Chuẩn là loại biến ngẫu nhiên liên tục cơ bản và có ứng dụng trong các trường hợp thực tế. Bên cạnh đó có một số biến ngẫu nhiên sẽ dành cho tính toán các bài toán thống kê sẽ được xét trong các bài sau. Với các biến ngẫu nhiên này ta chỉ cần nắm được cách tra bảng số để sử dụng ở phần sau.

Khái niệm

Nếu có n biến ngẫu nhiên Chuẩn hóa độc lập, khi bình phương các biến đó rồi lấy tổng, thì tổng đó sẽ phân phối theo một quy luật gọi là quy luật "Khi – bình phương", ký hiệu là $\chi^2(n)$, đọc là quy luật "Khi – bình phương bậc tự do n".

Nếu $U_1, U_2, ..., U_n$ đều phân phối Chuẩn hóa: $U_i \sim N(0;1)$ và độc lập nhau thì:

$V = U_1^2 + U_2^2 + ... + U_n^2 = \sum_{i=1}^{n} U_i^2$ phân phối theo quy luật Khi – bình phương bậc tự do n, ký hiệu: $V \sim \chi^2(n)$

Lưu ý:

- Chữ "Khi" χ viết khác chữ *X*, và ký hiệu χ^2 là ký hiệu viết liền, luôn gắn liền với nhau. Trong nội dung môn học không dùng đại lượng "Khi" χ đứng riêng, mà chỉ có "Khi – bình phương" χ^2 viết liền.
- Giá trị n được gọi là bậc tự do, thể hiện biến ngẫu nhiên đó được tập hợp từ n biến ngẫu nhiên tự do.

Giá trị tới hạn

Với quy luật Khi – bình phương bậc tự do *n*, ta cần quan tâm giá trị tới hạn mức *α*, là một con số sao cho biến ngẫu nhiên lớn hơn con số đó bằng đúng *α*. Con số đó ký hiệu là $\chi_\alpha^2(n)$, đọc là *giá trị tới hạn mức α bậc tự do n.*

$$P\left(V > \chi_\alpha^2(n)\right) = \alpha$$

Để có giá trị tới hạn, tra bảng phụ lục.

Ví dụ 4.10.

Muốn tra giá trị tới hạn Khi – bình phương mức 0,05 bậc tự do 10, tìm cột 0,05 và dòng 10, đối chiếu được giá trị 18,31. Ta viết: $\chi_{0,05}^2(10) = 18,31$.

Muốn tra giá trị tới hạn Khi – bình phương mức 0,975 bậc tự do 24, tìm cột 0,975 và dòng 24, đối chiếu được giá trị 12,4. Ta viết $\chi_{0,975}^2(24) = 12,4$.

4.4. Biến ngẫu nhiên phân phối Student

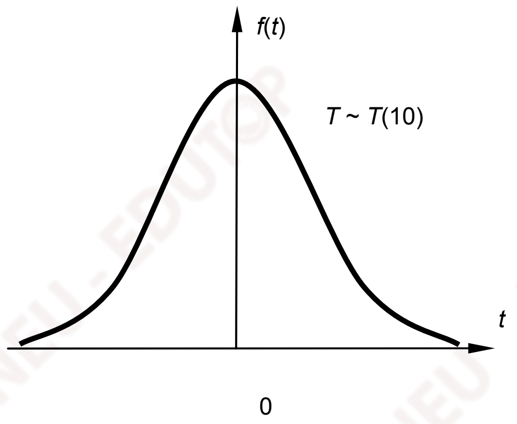
Khái niệm

Cho *U*, *V* là các biến ngẫu nhiên độc lập, *U* có phân phối Chuẩn hóa $U \sim N(0,1)$ và *V* có phân phối khi bình phương bậc tự do n: $V \sim \chi^2(n)$, khi đó

$$T = \frac{U}{\sqrt{V/n}}$$

là một biến ngẫu nhiên, quy luật phân phối xác suất của *T* gọi là quy luật Student với bậc tự do *n*, ký hiệu là *T*(*n*).

Đồ thị hàm mật độ của phân phối *T*(*n*) đối xứng qua trục tung, nó có dạng hình chuông gần giống hàm mật độ Chuẩn hóa *N*(0; 1). Khi n tăng lên đến 30 trở đi thì hàm mật độ sẽ gần như trùng với hàm mật độ Chuẩn hóa.



Hình 4.16. ***Đồ thị hàm mật độ biến ngẫu nhiên phân phối T(10)***

Giá trị tới hạn Student

Đối với biến ngẫu nhiên phân phối Student, ta cần quan tâm giá trị tới hạn mức α. Giá trị tới hạn mức α bậc tự do n, ký hiệu là $t_\alpha^{(n)}$ là một con số sao cho xác suất để T lớn hơn con số đó bằng đúng α.

$$P\left(T > t_\alpha^{(n)}\right) = \alpha$$

Các giá trị này có thể tra trong bảng Phụ lục.

Ví dụ 4.11:

Tra giá trị tới hạn mức 0,05 bậc tự do 10, ta tìm cột 0,05 và dòng 10, đối chiếu ô tương ứng được con số 1,812. Ta viết: $t_{0,05}^{(10)} = 1,812$

Tra giá trị tới hạn mức 0,025 bậc tự do 24, tà tìm cột 0,025 và dòng 24, đối chiếu ô tương ứng được con số 2,064. Ta viết: $t_{0,025}^{(24)} = 2,064$

Tính chất

- Do phân phối Student có hàm mật độ gần giống phân phối Chuẩn hóa nên tính chất đối xứng cũng tương tự: $t_{1-\alpha}^{(n)} = -t_\alpha^{(n)}$.

  Như vậy $t_{0,95}^{(10)} = -t_{0,05}^{(10)} = -1,812$; $t_{0,975}^{(24)} = -t_{0,025}^{(24)} = -2,064$.

- Do hàm mật độ phân phối Student hội tụ về Chuẩn hóa khi bậc tự do đủ lớn, n > 30, nên với bậc tự do lớn hơn 30 có thể dùng xấp xỉ giá trị tới hạn Chuẩn để tính toán. Như vậy $t_{0,05}^{(40)} \approx u_{0,05} = 1,645$, $t_{0,025}^{(50)} \approx u_{0,025} = 1,96$.

Tóm lược cuối bài

- Biến ngẫu nhiên liên tục là biến ngẫu nhiên có giá trị lấp đầy một khoảng, khi đó đặc trưng bởi hàm mật độ xác suất. Hàm mật độ càng cao ở chỗ nào thì xác suất tập trung quanh đó càng nhiều. Với biến ngẫu nhiên liên tục, xác suất để nhận giá trị tại một điểm duy nhất bằng 0 nên không xét đến cận trong các bài toán tính xác suất.
- Biến ngẫu nhiên phân phối Chuẩn có đồ thị hàm mật độ dạng quả chuông, đối xứng qua giá trị trung bình. Khi phương sai càng tăng thì quả chuông càng thấp và bẹt, phương sai càng nhỏ thì quả chuông càng cao và hẹp. Phân phối Chuẩn là phân phối thông dụng và phổ biến nhất trong kinh tế – xã hội và cả tự nhiên.
- Để tính xác suất biến ngẫu nhiên phân phối Chuẩn, cần tính qua biến ngẫu nhiên phân phối Chuẩn hóa. Các giá trị xác suất của biến ngẫu nhiên Chuẩn hóa được tra trong bảng phụ lục, từ đó qua công thức chuyển đổi xác suất để tính các xác suất.
- Bên cạnh các bài toán xác suất xuôi, còn có bài toán tính xác suất ngược, là cơ sở cho bài toán tìm giá trị tới hạn.
- Giá trị tới hạn Chuẩn, tới hạn Khi – bình phương và tới hạn Student được cho trong bảng phụ lục, sẽ được sử dụng cho các bài sau.

Câu hỏi ôn tập

1. Thế nào là hàm mật độ xác suất của biến ngẫu nhiên liên tục?
2. Biến phân phối Chuẩn có đồ thị hàm mật độ như thế nào?
3. Với biến phân phối Chuẩn khi trung bình và phương sai thay đổi thì đồ thị hàm mật độ thay đổi như thế nào?
4. Thế nào là biến ngẫu nhiên phân phối Chuẩn hóa?
5. Tính xác suất trên biến ngẫu nhiên phân phối Chuẩn hóa như thế nào?
6. Tính xác suất trên biến ngẫu nhiên phân phối Chuẩn như thế nào?
7. Giá trị tới hạn Chuẩn được lấy ở đâu?
8. Tìm giá trị tới hạn Khi – bình phương và Student như thế nào?