XỬ LÝ THỐNG KÊ SỐ LIỆU THỰC NGHIỆM TRONG PHÒNG THÍ NGHIỆM

Tác giả: Nguyễn Văn Lân, PGS/TS

CHƯƠNG 2

Đại lượng ngẫu nhiên và các phân bố xác suất

1. ĐẠI LƯỢNG NGẪU NHIÊN

Trong đo lường kiểm tra, *đại lượng đo* sẽ được thể hiện bằng một hay nhiều giá trị bất kỳ không thể đoán trước chính xác là bao nhiên. Chúng xảy ra một cách ngẫu nhiên nên đại lượng đo còn được gọi là *đại lượng ngẫu nhiên.*

Những kết quả đạt được x_i khi đo một đại lượng ngẫu nhiên có thể biểu thị bằng $x_i = \mu + e_r + e_s$

trong đó: μ - giá trị thực của đại lượng đo,

e_r - sai số ngẫu nhiên,

e_s - sai số hệ thống.

Trong thực tế, giá trị thực không bao giờ biết được và người ta thường thay nó bằng *giá trị thực quy ước*.

2. PHÂN BỐ XÁC SUẤT VÀ HÀM PHÂN BỐ

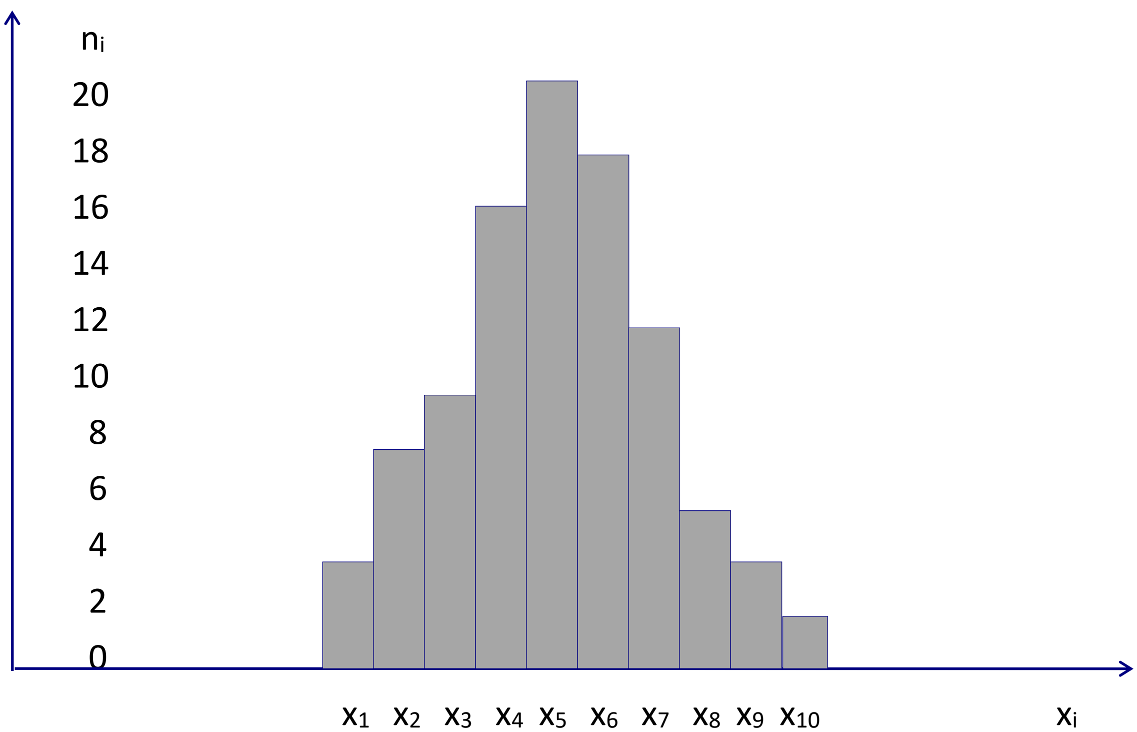
Khi thực hiện nhiều phép đo lặp để có *n* kết quả đo x_i , nếu thống kê số lần xuất hiện (còn gọi là *tần số*) n_i của những giá trị x_i và biểu diễn các cặp giá trị (x_i, n_i) lên hệ trục tọa độ, ta sẽ được dạng *phân bố thực nghiệm* của các kết quả đo. Có thể thay n_i bằng y_i là tỷ số n_i/n mà không làm dạng phân bố đó thay đổi. Ví dụ : Với $n = 100$ kết quả quan trắc, phân bố các giá trị x_i như sau:

x_i	1	2	3	4	5	6	7	8	9	10
n_i	4	8	10	16	20	18	12	6	4	2

và được biểu diễn theo đồ thị chữ nhật trên hình 2.1 dưới đây.

Khi tăng số kết quả quan trắc *n* đến một số vô cùng lớn thì đỉnh của các hình chữ nhật sẽ nằm trên một đường cong nào đó. Tùy theo bản chất của đại lượng đo, đường cong này đặc trưng cho một hàm phân bố xác suất lý thuyết khi tỷ số y_i trở thành p_i , xác suất xuất hiện giá trị x_i tương ứng.

Trong thực tế đo lường, có hai loại đại lượng ngẫu nhiên.



Hình 2.1 Phân bố thực nghiệm của đại lượng ngẫu nhiên X

2.1 Đại lượng ngẫu nhiên rời rạc

Đó là những đại lượng thể hiện các *kết quả đếm* (x_i = 0, 1, 2, . . .), ví dụ số sản phẩm hỏng hoặc số sản phẩm có khuyết tật, số khuyết tật trên một đơn vị đo của sản phẩm, số vi khuẩn trong một mẫu quan sát, v.v.. Hàm phân bố có dạng chung là:

$$f(x) = P(X = x_i) = p_i$$

Hàm mật độ xác suất có dạng :

$$F(x) = \Sigma P(X < x) = \sum_{xi<x} p_i$$

Nếu *n* hữu hạn: $\sum_{i=1}^{n} p_i = 1$ và nếu *n* vô cùng lớn: $\sum_{i=1}^{\infty} p_i = 1$

Hai tham số cơ bản của phân bố các đại lượng ngẫu nhiên rời rạc là :

Số trung bình : $\mu = EX = \sum_{i=1}^{n} p_i x_i$

Phương sai : $\sigma^2 = E(X - EX)^2 = EX^2 - \mu^2 = \sum_{i=1}^{n} p_i x_i^2 - \mu^2$

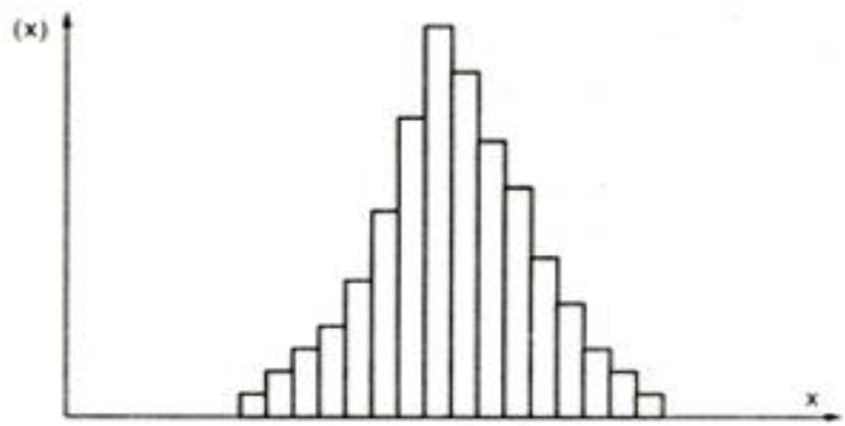
Phân bố nhị thức

Khi kiểm tra một lô sản phẩm nào đó theo hai chỉ tiêu hỏng và tốt, gọi sản phẩm *hỏng* tìm thấy là biến cố A và sản phẩm tốt là biến cố $\overline{A}$. Giả sử lô có p % sản phẩm hỏng thì xác suất để lấy từ lô đúng một sản phẩm hỏng là :

$$P(A) = p$$

và do vậy, xác suất để lấy từ lô một sản phẩm tốt là:

$$P(\overline{A}) = 1 - p = q$$



Hình 2.2 Phân bố nhị thức

Nếu từ lô lần lượt lấy ra đến sản phẩm thứ *n* để kiểm tra thì xác suất để trong số *n* sản phẩm đó có đúng *x* sản phẩm hỏng là :

$$P(\underbrace{A.\overline{A}.\overline{A}......A}_{n\,chu}) = \underbrace{p.q.q.......p}_{n\,chu}$$

Nếu trong *n* chữ, có *x* chữ A và (*n*–*x*) chữ $\overline{A}$ thì

$$P(A.\overline{A}.\overline{A}......A) = p^x.q^{n-x}$$

Dù vị trí của A và $\overline{A}$ như thế nào, vế bên phải của đẳng thức vẫn không thay đổi. Vậy nếu không kể thứ tự xuất hiện của A và $\overline{A}$ thì xác suất f(x) sao cho trong *n* phép thử có đúng *x* lần xuất hiện A là

$$f(x) = P(X = x) = C_n^x p^x q^{n-x}$$

với $$C_n^x = \frac{n(n-1)....(n-x+1)}{1.2.3...x} = \frac{n!}{x!(n-x)!}$$

trong đó *x* lấy các giá trị 0, 1, 2, ...,n.

Tập hợp các f(x) khi *x* lấy các giá trị từ 0 đến n lập nên phân bố nhị thức. Gọi phân bố nhị thức là vì các f(x) chính là các thành phần được khai triển của nhị thức $(p + q)^n$. Đó là

$$\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} C_n^x p^x q^{n-x} = C_n^0 q^n + C_n^1 p^1 q^{n-1} + ... + C_n^n p^n = (p+q)^n = 1$$

với quy ước 0! = 1 và $C_n^0 = C_n^n = 1$, ngoài ra. $C_n^x = C_n^{n-x}$

Để việc tính f(x) được dễ dàng với các giá trị khác nhau của *x*, thay *x* bằng (*x*+1) vào công thức tính f(x) trên sẽ được công thức truy chứng :

$$f(x+1) = \frac{p}{q}.\frac{n-x}{x+1}.f(x)$$

Trong thực hành, để tìm nhanh giá trị của các hàm phân bố nhị thức, ta áp dụng các hàm Excel như sau:

f(x) = binomdist(x,n,p,false) và F(x) = binomdist(x,n,p,true)

Phân bố nhị thức có một đỉnh, và cực đại của nó là một số nguyên x nằm trong khoảng từ (np−q) đến (np+q). Nếu p = q, phân bố nhị thức là đối xứng.

Hai tham số cơ bản là :

Số trung bình μ = np

Phương sai σ^2 = npq

Nếu thay đại lượng X bằng đại lượng Y = X/n (tức số sản phẩm hỏng bằng tỷ lệ sản phẩm hỏng) thì Y cũng thuộc phân bố nhị thức. Các tham số cơ bản trở thành :

Số trung bình μ = p

Phương sai $\sigma^2 = \frac{pq}{n}$

Phân bố nhị thức, còn gọi là phân bố Bernoulli, được áp dụng cho trường hợp kiểm tra chọn mẫu có hoàn lại và kiểm tra chọn mẫu có thay sản phẩm hỏng. Khi đó với lô có N sản phẩm , sản phẩm bất kỳ của lần chọn nào trong *n* sản phẩm của mẫu được chọn cũng đều có cùng xác suất rơi vào sản phẩm hỏng *p* như sản phẩm hỏng thứ nhất được phát hiện trong *n*.

Ví dụ: *Chất lượng sản phẩm của một nhà máy đạt như sau: sản phẩm hợp chuẩn chiếm p = 80 %, không hợp chuẩn chiếm q = 20 %. Nếu lấy một mẫu gồm n = 5 sản phẩm để kiểm tra, hãy tính xác suất xuất hiện x sản phẩm hợp chuẩn và không quá x sản phẩm hợp chuẩn*.

Tính toán	*f(x)*		*F(x)*
$f(0) = C_n^0.p^0q^{5-0} = C_5^0.0{,}2^5 =$	0,00032	$\rightarrow$	0,00032
$f(1) = \frac{0{,}8}{0{,}2}.\frac{5-0}{0+1}.0{,}00032 =$	0,0064	$\rightarrow$	0,00672
$f(2) = \frac{0{,}8}{0{,}2}.\frac{5-1}{1+1}.0{,}0064 =$	0,0512	$\rightarrow$	0,05792
$f(3) = \frac{0{,}8}{0{,}2}.\frac{5-2}{2+1}.0{,}0512 =$	0,2048	$\rightarrow$	0,26272
$f(4) = \frac{0{,}8}{0{,}2}.\frac{5-3}{3+1}.0{,}2048 =$	0,4096	$\rightarrow$	0,67232

$$f(5) = \frac{0{,}8}{0{,}2}.\frac{5-4}{4+1}.0{,}4096 \quad = \quad 0{,}3277 \quad \rightarrow \quad 1{,}00000$$

Số sản phẩm hợp chuẩn trung bình : $\mu = 5 \times 0{,}8 = 4$

Phương sai của sản phẩm hợp chuẩn : $\sigma^2 = 5 \times 0{,}8 \times 0{,}2 = 0{,}8$

Bài tập 2.1 *: Trên địa bàn một huyện, 30 % số giếng bị ô nhiễm. Từ rất nhiều giếng, lấy ngẫu nhiên một mẫu gồm n = 5 giếng . Hãy xác định trong số giếng mẫu có:*

a) Đúng 3 giếng bị ô nhiễm;

b) Tối thiểu 3 giếng bị ô nhiễm;

c) Ít hơn 3 giếng bị ô nhiễm.

Phân bố siêu bội

Phân bố nhị thức được xây dựng trên hai biến cố A và $\overline{A}$ với điều kiện quan trọng là chúng độc lập với nhau, trong đó p không thay đổi trong suốt quá trình thử. Điều kiện này chỉ có được khi mẫu kiểm tra có hoàn lại hoặc khi số sản phẩm trong lô N khá lớn. Nhưng trong thực tế, người ta thường chọn mẫu không hoàn lại, nghĩa là p sẽ thay đổi mỗi khi kiểm tra xong một sản phẩm. Ví dụ sau khi kiểm tra đến sản phẩm thứ i và đã phát hiện được k sản phẩm hỏng thì lần kiểm tra thứ (i+1) tiếp theo, p sẽ không còn bằng M/N (M là số sản phẩm hỏng thực có trong lô) mà lúc bấy giờ bằng (M−k)/ (N−i).

Đại lượng X trong trường hợp chọn mẫu không hoàn lại có phân bố xác suất là phân bố siêu bội. Với N là cỡ lô (tức số sản phẩm có trong lô), n là cỡ mẫu (số sản phẩm được lấy ra để kiểm tra), giả sử lô chứa M sản phẩm hỏng và tỷ lệ sản phẩm hỏng do đó bằng p = M/N. Xác suất để phát hiện đúng x sản phẩm hỏng trong mẫu kiểm tra sẽ bằng :

$$f(x) = P(X = x) = \frac{C_M^x . C_{N-M}^{n-x}}{C_N^n}$$

trong đó x = 0, 1, 2,....., min(M;n).

Phân bố siêu bội đóng vai trò quan trọng trong lĩnh vực kiểm tra thống kê chất lượng khi so sánh các kế hoạch chọn mẫu. Nó là trường hợp giới hạn của phân bố nhị thức khi $N \rightarrow \infty$ và khi thay p bằng M/N.

Với P(X = 0) : $$f(0) = \frac{C_{N-M}^n}{C_N^n}$$

và cũng để tiện việc tính toán, nên dùng công thức truy chứng :

$$f(x+1) = \frac{(n-x)(M-x)}{(x+1)(N-M-n+x+1)}.f(x)$$

Các tham số cơ bản của phân bố siêu bội là :

Số trung bình: $\mu = n\frac{M}{N} = np$

Phương sai: $\sigma^2 = n\dfrac{N-n}{N-1}p(1-p)$

Phân bố siêu bội áp dụng cho mẫu kiểm tra không hoàn lại và mẫu kiểm tra không thay sản phẩm hỏng , kể cả trường hợp cỡ lô không lớn và n/N vượt quá 0,10.

Ví dụ: *Một lô có N = 500 sản phẩm với tỷ lệ sản phẩm hỏng là p = 2 %. Hãy tính xác suất để trong n = 50 sản phẩm được chọn không có một sản phẩm hỏng nào và xác suất để không có quá p.n = 1 sản phẩm hỏng để từ kết quả kiểm tra của mẫu, lô sản phẩm được chấp nhận*?

Vì p = 0,02 = M/500 nên M = 10 và với x = 0 :

$$f(x = 0) = \frac{C_{10}^{0}.C_{490}^{50}}{C_{500}^{50}} = \frac{490!50!450!}{50!440!500!} = 0,345$$

Có thể dùng hàm Excel:

f(x = 0) = combin(10,0)*combin(490,50)/combin(500,50)

Vậy P(X=0) trong trường hợp này là 34,5%

$$f(x = 1) = \frac{50 \times 10}{500-10-50+1} \times 0,345 = \frac{500}{441} \times 0,345 = 0,391$$

$$F(X \le 1) = P(X = 0) + P(X = 1) = 0,345 + 0,391 = 0,736$$

Với xác suất chắc chắn 73,6 % để cho lô có thể được chấp nhận, rủi ro của phương án chọn mẫu này quá lớn!

Bài tập 2.2 *: Lập kế hoạch kiểm tra 20 hãng về sự vi phạm quy định về ô nhiễm. Trong một thời gian hạn chế, chỉ có thể kiểm tra 3 hãng. Nếu biết rằng trong số 20 hãng có 5 hãng vi phạm, hãy tìm xác suất sao cho khi kiểm tra:*

a) Không có hãng nào trong 3 hãng kiểm tra vi phạm;

b) Toàn bộ 3 hãng kiểm tra đều vi phạm;

c) Ít nhất 1 trong 3 hãng kiểm tra bị vi phạm.

Phân bố Poisson

Phân bố Poisson là một trường hợp giới hạn khác của phân bố nhị thức khi $n \to \infty$ và p xác suất của một trong hai biến cố A và $\overline{A} \to 0$ nhưng tích số np = μ vẫn là một số hữu hạn. Vì vậy phân bố này còn được gọi là phân bố của các "biến cố hiếm", tức là những biến cố ít xảy ra trong một đơn vị tính (như thời gian, chiều dài, diện tích, khối lượng...), ví dụ như số khuyết tật trên một km đường ống dẫn khí đốt; số lần gọi của máy điện thoại trong một khoảng thời gian nhất định; số nguyên tử phân rã trong thời gian ngắn của một chất phóng xạ v.v..

Hàm mật độ xác suất có dạng như sau :

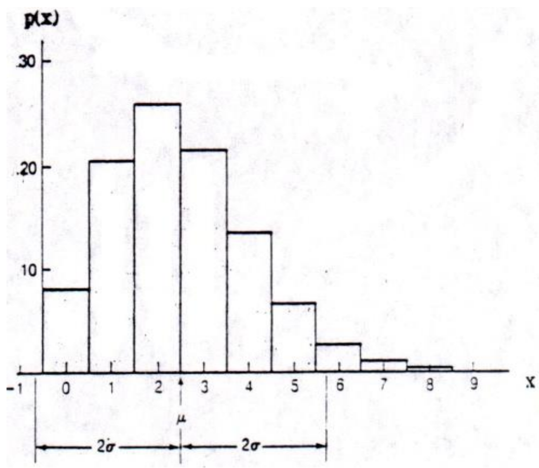
$$f(x) = \frac{\lambda^x \times e^{-\lambda}}{x!}$$

với x = 0, 1, 2,....

trong đó λ vừa là số trung bình vừa là phương sai của phân bố.

Số trung bình: $\mu = \lambda$

Phương sai: $\sigma^2 = \lambda$

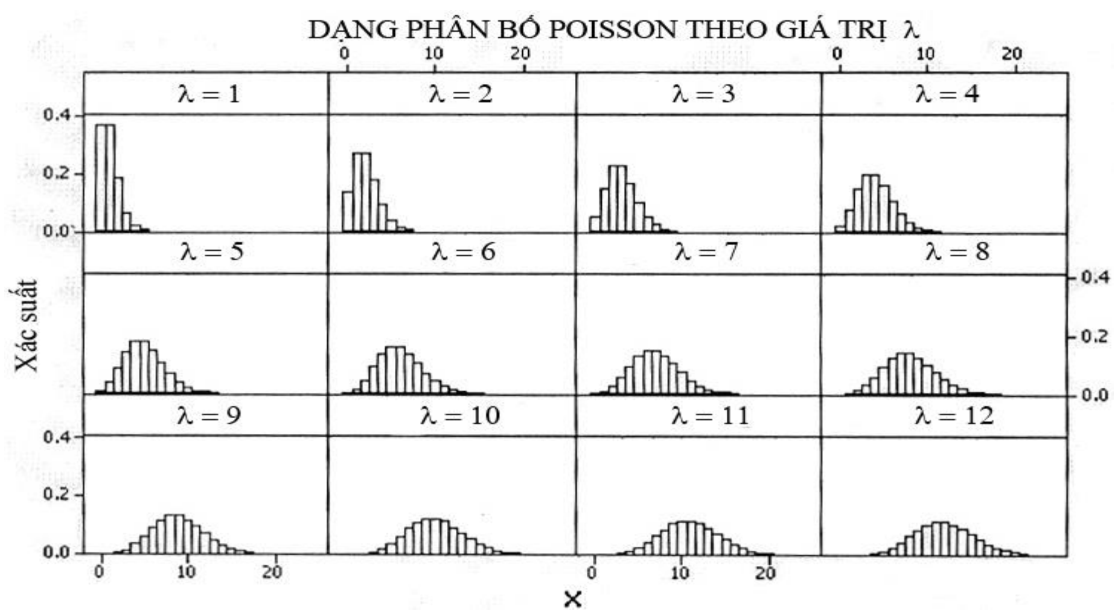


Hình 2.3. Phân bố Poisson

Công thức truy chứng dùng để tính các f(x) với những giá trị x khác nhau :

$$f(x+1) = \frac{\lambda}{x+1}.f(x)$$

và bắt đầu từ $f(0) = e^{-\lambda}$.



Hình 2.4. Các dạng phân bố Poisson theo giá trị λ

Theo độ lớn của λ, đường cong nối liền các điểm của phân bố Poisson có những dạng khác nhau. Giá trị λ càng lớn , đường cong càng trở nên đối xứng nhất là từ khi λ = 10 trở lên.

Trong thực hành, để tìm nhanh giá trị của các hàm phân bố Poisson, ta áp dụng các hàm Excel như sau:

$$f(x) = \text{Poisson}(x,\lambda,\text{false}) \text{ và } F(x) = \text{Poisson}(x,\lambda,\text{true})$$

Ví dụ: Kiểm tra một mẫu n = 500 cọc của máy kéo sợi, thấy trung bình có 0,6 lần đứt sợi. Áp dụng phân bố Poisson, hãy xác định xác suất và số cọc lý thuyết có số lần đứt sợi x = 0, 1, 2, 3, 4.

Với $f(x) = e^{-0,6} = 0,5488$ dùng công thức truy cập chứng minh tiếp f(1), f(2), v.v.. và làm tròn số cọc $n_i = n.f(x_i)$, ta có bảng sau:

x_i	*$f(x_i)$*	*n_i xấp xỉ*
0	0,5488	274
1	0,3293	165
2	0,0988	49
3	0,0197	10
4	0,0030	2

Bài tập 2.3*: Số xe ca x đến tại giao lộ trong một đơn vị thời gian quy định thuộc phân bố Poisson. Nếu biết số xe ca trung bình đến đúng giao lộ trong thời gian đó là λ, người kỹ sư giao thông có thể thiết kế hệ thống kiểm soát đường. Giả sử λ trong một phút bằng 1. Hãy xác định:*

1) Xác suất để trong 1 phút số xe bằng hoặc lớn hơn 3 ;

2) Có chắc là số xe đến giao lộ vượt quá 3 trong 1 phút ít khi xảy ra ?

2.2 Đại lượng ngẫu nhiên liên tục

Đại lượng ngẫu nhiên liên tục lấy các giá trị liên tục trên trục số thực được hiểu một cách quy ước bởi vì thực tế không có một dụng cụ đo nào cho độ chính xác với số chữ số có nghĩa vô hạn. Nếu làm trơn đường gấp khúc nối liền các điểm thể hiện tọa độ nằm sít nhau (x_i, $f(x_i)$) của đại lượng ngẫu nhiên đó, ta được một đường cong liên tục biểu diễn hàm phân bố xác suất của đại lương ngẫu nhiên liên tục.

Đại lượng thể hiện các *kết quả đo* thuộc loại đại lượng ngẫu nhiên liên tục. Có thể kể ra rất nhiều phân bố của các đại lượng ngẫu nhiên liên tục như phân bố Weibull, phân bố mũ, phân bố chuẩn, phân bố Student, v.v. .Hàm phân bố có dạng chung là :

$$f(x) = P(X = x_i) = p_i$$

Hàm mật độ xác suất có dạng :

$$F(x) = \Sigma P(X < x) = \int_{-\infty}^{x} f(t)dt$$

Tương tự như trên, cũng có : $\int_{-\infty}^{+\infty} f(t)d(t) = 1$

Xác suất để X nằm giữa x_1 và x_2 :

$$P(x_1 \le X \le x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t)d(t)$$

Hai tham số cơ bản của phân bố là :

Số trung bình : $\mu = EX = \int_{-\infty}^{+\infty} x.f(x)dx$

Phương sai : $\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2$

Ví dụ: *Một phân bố chữ nhật (hình 2.5) của đại lượng ngẫu nhiên liên tục có hàm phân bố xác suất là :*

$f(x) = 0$ khi $x < a_1$

$f(x) = C$ khi $a_1 \le x \le a_2$

$f(x) = 0$ khi $x > a_2$

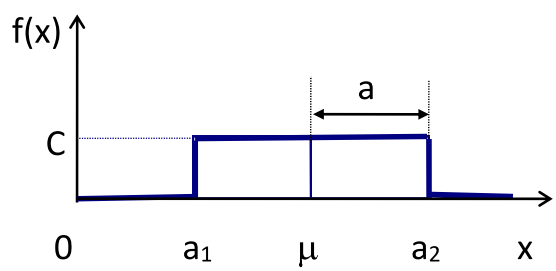
Hãy tính số trung bình và phương sai.

Theo công thức trên :

$$\int_{a_1}^{a_2} f(x)dx = \int_{a_1}^{a_2} Cdx = 1 \text{ từ đó } C = \frac{1}{a_2 - a_1}$$

$$\mu = \int_{a_1}^{a_2} x.f(x)dx = \frac{1}{a_2 - a_1}\int_{a_1}^{a_2} xdx = \frac{a_1 + a_2}{2};$$

$$\sigma^2 = \int_{a_1}^{a_2} x^2 f(x)dx - \mu^2 = \frac{(a_2 - a_1)^2}{12} = \frac{a^2}{3} \quad \text{tức là} \quad \sigma = \frac{a}{\sqrt{3}}$$

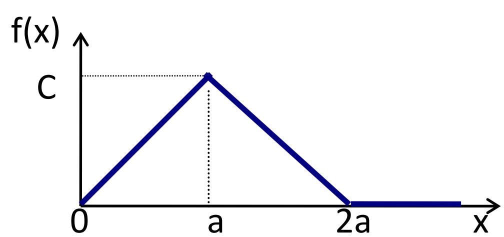


Hình 2.5. Phân bố chữ nhật

Ví dụ : *Một phân bố tam giác cân (Hình 2.6) của đại lượng ngẫu nhiên liên tục có hàm phân bố xác suất*:

$$f_1(x) = \frac{x}{a^2} \qquad \text{khi } 0 \le x \le a$$

$$f_2(x) = -\frac{x}{a^2} + \frac{2}{a} \quad \text{khi } a \le x \le 2a$$



Hình 2.6. Phân bố tam giác cân

Như vậy :

$$\mu = a$$

$$\sigma^2 = \int_0^a x^2\left(\frac{x}{a^2}\right)dx + \int_a^{2a} x^2\left(-\frac{x}{a^2}+\frac{2}{a}\right)dx - a^2 = \frac{a^2}{6} \quad \text{do đó} \quad \sigma = \frac{a}{\sqrt{6}}$$

Bài tập 2.4

1. Giả sử sai số đọc với độ phân giải của một pipét là 0,1 mL tuân theo phân bố chữ nhật, hãy tính độ lệch chuẩn của thể tích dung dịch mỗi lần hút.

2. Nhiệt độ không chế của một máy điều nhiệt là 15°C $\pm$ 1°C. Giả sử dao động của sai số nhiệt độ tuân theo phân bố tam giác cân, hãy tính độ lệch chuẩn của sai số nhiệt độ.

Các phân bố thường gặp trong kiểm tra chất lượng là:

Phân bố Weibull

Đó là phân bố của các đại lượng biểu thị thời gian xuất hiện hỏng hóc của sản phẩm, tuổi thọ của thiết bị, v.v... Biểu thức giải tích của hàm phân bố là

$$f(x) = \frac{\alpha}{\beta}.x^{\alpha-1}\exp\left\{-\frac{x^\alpha}{\beta}\right\}$$

và
$$F(x) = \int_0^x f(t)dt = \frac{\alpha}{\beta}\int_0^x t^{\alpha-1}.\exp\left\{-\frac{t^\alpha}{\beta}\right\}dt$$

với $0 \le x < \infty$ và $\alpha > 0$, $\beta > 0$.

Hai tham số cơ bản là:

Số trung bình: $\mu = \beta^{\frac{1}{\alpha}}\Gamma\left(\frac{\alpha+1}{\alpha}\right)$

Phương sai: $\sigma^2 = \beta^{\frac{2}{\alpha}}\left[\Gamma\left(\frac{\alpha+2}{\alpha}\right) - \Gamma^2\left(\frac{\alpha+1}{\alpha}\right)\right]$

trong đó Γ(Z) gọi là *hàm gamma* của Z tính theo công thức truy chứng :

$$\Gamma(Z+1) = Z.\Gamma(Z) \quad \text{với} \quad Z > 0$$

Khi Z = n là một số nguyên thì :

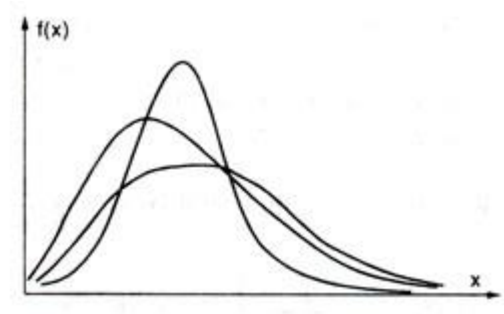
$$\Gamma(n) = (n-1)!$$

Khi Z = 0,5 thì :

$$\Gamma(0,5) = \sqrt{\pi}$$

Khi cho $1 \le Z \le 2$, Γ(Z) được tính theo hàm Excel:

Γ(Z) = exp(gammaln(Z))



Hình 2.7. Các dạng phân bố Weibull (phụ thuộc α và β)

Trong thực hành, để tìm nhanh giá trị của các hàm phân bố Weibull, ta áp dụng các hàm Excel như sau:

$$f(x) = weibull(x,\alpha,\beta,false) \text{ và } F(x) = weibull(x,\alpha,\beta,true)$$

Barella qua nghiên cứu nhận thấy khi bị kéo lặp đi lặp lại nhiều lần trên máy dệt, độ mỏi sợi dọc tuân theo phân bố Weibull và đại lượng thời gian được thay bằng số chu trình kéo sau đó xuất hiện sợi đứt.

Ví dụ : *Tuổi thọ (tính bằng giờ) của một loại mũi khoan trong một quá trình gia công cơ khí thuộc phân bố Weibull với $\alpha = 2$ và $\beta = 100$. Tìm xác suất mà mũi khoan bị hỏng trước 8 giờ sử dụng và tính tuổi thọ trung bình của loại mũi khoan này*.

Đặt $z = x^{\alpha}$, do đó $dz = \alpha.x^{\alpha-1}.dx$ và tính được

$$F(x) = 1 - \exp\left\{-\frac{z}{\beta}\right\} = 1 - \exp\left\{-\frac{x^{\alpha}}{\beta}\right\}$$

$$P(x < 8) = F(8) = 1 - \exp\left\{-\frac{8^2}{100}\right\} = 1 - \exp\{-0{,}64\} = 0{,}473$$

$$\mu = 100^{1/2}\Gamma\left(\frac{2+1}{2}\right) = 10\Gamma(1{,}5) = 10\Gamma(0{,}5+1) = 10\times 0{,}5\Gamma(0{,}5) = 5\sqrt{\pi} \approx 8{,}86 \text{ giờ}$$

Bài tập 2.5 *: Thời gian tính bằng số tháng x sau khi bảo trì thiết bị camera tại một ngân hàng tuân theo phân bố Weibull có $\alpha = 2$ và $\beta = 60$. Nếu ngân hàng muốn xác suất mà thiết bị camera có thể hư hỏng trước thời hạn bảo trì tiếp theo là 0,05 thì chu kỳ bảo trì sẽ là bao nhiêu?*

Phân bố mũ

Phân bố mũ cũng là một dạng của phân bố Weibull nếu cho $\alpha = 1$. Nó là phân bố của các đại lượng như thời gian sửa chữa máy, thời gian giữa hai lần dừng máy, thời gian giữa hai lần phục vụ khách hàng, v.v...Biểu thức giải tích của phân bố này là :

$$f(x) = \frac{1}{\beta}\exp\left\{-\frac{x}{\beta}\right\}$$

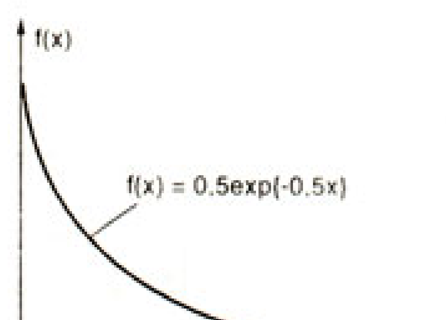
và

$$F(x) = \frac{1}{\beta}\int_0^x \exp\left\{-\frac{t}{\beta}\right\}dt$$

với $0 \le x < \infty$.

Hai tham số cơ bản là:

Số trung bình: $\mu = \beta$; Phương sai: $\sigma^2 = \beta^2$



Hình 2.8. Phân bố mũ

Trong thực hành, để tìm nhanh giá trị của các hàm phân bố mũ, ta áp dụng các hàm Excel như sau:

$$f(x) = \text{expondist}(x,1/\beta,\text{false}) \text{ và } F(x) = \text{expondist}(x,1/\beta,\text{true})$$

Ví dụ : *Trong một xí nghiệp có một loại máy mà thời gian giữa hai lần dừng trung bình do sự cố là 2 giờ. Thời gian giữa hai lần dừng máy thuộc phân bố mũ. Hãy xác định xác suất để thời gian này lâu hơn 2 giờ.*

Biểu thức giải tích của phân bố là

$$f(x) = \frac{1}{2}\exp\left\{-\frac{x}{2}\right\}$$

$P(x > 2) = 1 - P(x \leq 2) = 1 - F(2)$

$$= 1 - \frac{1}{2}\int_0^2 \exp\left\{-\frac{x}{2}\right\}dx = 1 - \left[-\exp\left\{-\frac{x}{2}\right\}\right]_0^2 = 1 - \left(-\frac{1}{e}+1\right) = \frac{1}{e} = 0,367879$$

Bài tập 2.6 *: Tốc độ gió (km/s) thuộc phân bố mũ. Ở một vùng cao nguyên có tốc độ gió trung bình trong năm là 5 km/s. Hãy tính:*

a) Xác suất để tốc độ gió dưới 5 km/s;

b) Xác suất để tốc độ gió lớn hơn 10 km/s.

Phân bố chuẩn

Phân bố chuẩn còn được gọi là *phân bố Gauss*, có biểu thức giải tích đường cong như sau:

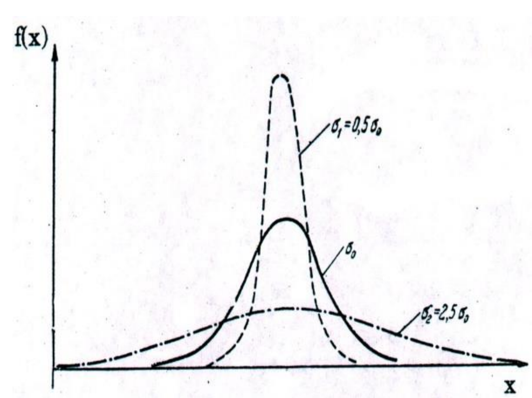
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

và hàm phân bố tích lũy :

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

Nếu tính từ đẳng thức:

$$P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t)dt$$



Hình 2.9. Đường cong phân bố chuẩn

trong đó $x_1 = \mu - z.\sigma$ và $x_2 = \mu + z.\sigma$, ta sẽ có:

$z = 1,00 \Rightarrow P(\mu - 1,00\sigma \le X \le \mu + 1,00\sigma) = 0,68269$

$z = 1,65 \Rightarrow P(\mu - 1,65\sigma \le X \le \mu + 1,65\sigma) = 0,90106$

$z = 1,96 \Rightarrow P(\mu - 1,96\sigma \le X \le \mu + 1,96\sigma) = 0,95000$

$z = 2,00 \Rightarrow P(\mu - 2,00\sigma \le X \le \mu + 2,00\sigma) = 0,95450$

$z = 2,58 \Rightarrow P(\mu - 2,58\sigma \le X \le \mu + 2,58\sigma) = 0,99012$

$z = 3,00 \Rightarrow P(\mu - 3,00\sigma \le X \le \mu + 3,00\sigma) = 0,99730$

$z = 3,29 \Rightarrow P(\mu - 3,29\sigma \le X \le \mu + 3,29\sigma) = 0,99900$

$z = 4,00 \Rightarrow P(\mu - 4,00\sigma \le X \le \mu + 4,00\sigma) = 0,99994$

Trong thực hành, để tìm nhanh giá trị của các hàm phân bố chuẩn, ta áp dụng các hàm Excel như sau:

f(x) = normdist(z,0,1,false) và F(x) = normdist(z,0,1,true)

Trong đo lường thử nghiệm, phân bố chuẩn được sử dụng khá phổ biến, nhất là khi số quan trắc lặp *n* tương đối lớn.

Ví dụ: *Một sản phẩm được ghi nhận có chất lượng trung bình là 6,5 đơn vị và độ lệch chuẩn là 0,4 đơn vị. Áp dụng phân bố chuẩn, những giá trị nào của chất lượng nằm trong phạm vi 95 % so với chất lượng trung bình và ngoài 95 % so với chất lượng trung bình*?

Theo phép tính trên, ứng với P = 95 % thì z = 1,96. Vậy:

- Những giá trị nằm trong phạm vi 95 % là:

  $6,5 - 1,96.0,4 \le x \le 6,5 + 1,96.0,4$ tức là $5,716 \le x \le 7,284$
- Những giá trị nằm ngoài phạm vi 95 % là:

  $x < 5,716$ và $x > 7,284$.

***Bài tập 2.7** : Các kết quả đo x_i của một mẫu thuộc phân bố Gauss. Giả sử số trung bình mẫu bằng 12,5 và độ lệch chuẩn mẫu bằng 0,2 . Nếu quy định số lạc phải loại bỏ ra khỏi phép tính với xác suất rủi ro 0,1% hãy xác định những x_i nào là số lạc?*

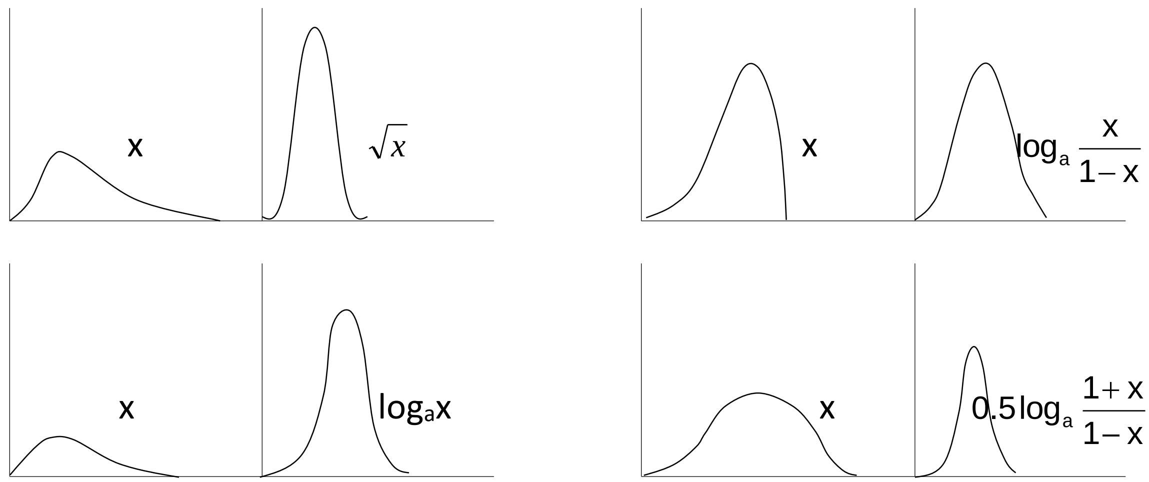
Có những đại lượng ngẫu nhiên X không thuộc phân bố chuẩn nhưng nếu qua một phép biến đổi giá trị của nó, đại lượng mới Z sẽ thuộc phân bố chuẩn, ví dụ hình 2.11

$Z = \sqrt{X}$ với X là số đếm ; $Z = \arcsin\sqrt{X}$ với X là tần suất ; $Z = \log_a X$ với X là độ bền mỏi ;

$Z = \log_a \frac{X}{1-X}$ với $0 \le X \le 1$ hoặc $Z = \frac{1}{2}\log_a \frac{1+X}{1-X}$ với $-1 \le X \le +1$

Phân bố nhị thức có thể thay bằng phân bố chuẩn khi cỡ mẫu n lớn. Trong thực tế, có thể thực hiện việc chuyển đổi này khi $n \ge 20$ và $0,3 \le p \le 0,7$.

Phân bố Poisson có thể thay bằng phân bố chuẩn khi số trung bình μ lớn. Thực tế chấp nhận việc chuyển đổi khi $\mu \geq 3$ với P = 0,95 , khi $\mu \geq 6$ với P = 0,99 và khi $\mu \geq 10$ với P = 0,999.



Hình 2.11. Đường cong phân bố chuyển đổi gần với phân bố chuẩn

Phân bố Student

Phân bố Student thường được áp dụng cho những mẫu có số quan trắc *n* bé.

Những đại lượng là số trung bình của tập hợp các kết quả đo $\{x_i\}$ thuộc phân bố bất kỳ đều thuộc phân bố Student.

Biểu thức giải tích của đường cong có dạng bên cạnh, trong đó ν là số bậc tự do. Hàm phân bố tích lũy có dạng:

$$F(x) = \int_{-\infty}^{x} f(z)dz$$

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

Hàm Excel để xác định phân bố Student như sau:

f(x) = tdist(x,ν,m) với m = 1 khi xét một phía và m = 2 khi xét hai phía.

Việc tính xác suất P trong phân bố Student phức tạp hơn nhiều so với phân bố chuẩn do có số bậc tự do ν. Trong thực tế , người ta thường dùng xác suất:

$$P(-t \leq X \leq +t) = \int_{-t}^{+t} f(x)dx$$

Giá trị t được gọi là *hệ số Student*. Khi $t_{P;\nu} = t_{0,95;3} = 3,182$ điều đó có nghĩa là:

$$P = \int_{-3,182}^{+3,182} f(x)dx = 0,95$$

Không cần tra bảng, có thể tìm phân vị t theo hàm Excel sau:

t =tinν(2*α,ν) cho trường hợp một phía

và t =tinν(2*α/2,ν) cho trường hợp hai phía

***Bài tập 2.8** : Hãy tìm t với P = 0,95 và P = 0,99 trong trường hợp một phía và hai phía.*

3. MỨC TIN CẬY

Mức tin cậy là xác suất khẳng định một kết luận nào đó. Mức tin cậy P = 95% tương đương với trong 100 trường hợp có 95 trường hợp đúng như kết luận còn 5 trường hợp khác với kết luận. Ngược với mức tin cậy là mức rủi ro thường ký hiệu bằng $\alpha = 100 - P$. Trong công tác điều tra người ta thường dùng P = 90-95 %, công tác nghiên cứu công nghệ và thiết bị, P = 95-98 %, công tác kiểm tra chất lượng sản phẩm, P = 95-99 %.

Mức tin cậy có hai loại:

- ***Mức tin cậy một phía**:* khi kết quả kiểm tra không được vượt quá một giới hạn nào đó (trên hay dưới), nằm dưới giới hạn trên thì chất lượng đạt mà vượt quá giới hạn trên thì chất lượng không đạt. Ví dụ khi xét chất lượng nước thải, hàm lượng các chất độc hại không được vượt quá chỉ tiêu cho phép;

- ***Mức tin cậy hai phía**:* khi điều kết luận quan tâm đến hai giới hạn trên và dưới, vượt quá giới hạn trên và nằm dưới giới hạn dưới thì chất lượng sản phẩm không đạt. Ví dụ khi xét đường kính các chi tiết cơ khí lắp ghép, đường kính không được lớn quá giới hạn trên và cũng không được nhỏ quá giới hạn dưới.

Bình thường trong đo lường thử nghiệm, người ta áp dụng mức tin cậy 95 % và khi cần có sự khẳng định quan trọng mới dùng mức tin cậy 99 % , thậm chí 99,9 %.