Chương 4
CÁC QUY LUẬT PHÂN PHỐI

4.1. CÁC PHÂN PHỐI RỜI RẠC

1- Phân phối nhị thức

Đại lượng ngẫu nhiên rời rạc X = {0, 1, 2, ..., n} gọi là có phân phối nhị thức nếu tồn tại số $p \in (0,1)$ sao cho:

$$p_k = P(X = k) = C_n^k p^k q^{n-k}, \ q = 1 - p, \ k = \overline{0, n}$$

Trong trường hợp này ta ký hiệu X ~ B(n,p).

Định nghĩa trên là hợp lý vì:

$$\sum_{1=1}^{n} p_i = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p+q)^k = 1$$

Theo định lý 1.8 chương 1 ta có:

Định lý 4.1. Nếu X là số lần thành công trong dãy n phép thử Bernoulli với xác suất thành công p thì X ~ B(n,p).

Ta sẽ chứng minh định lý sau nói về các đặc trưng của phân phối nhị thức.

Định lý 4.2. Nếu X ~ B(n.p) thì E(X) = np và D(X) = npq.

Chứng minh. Từ định nghĩa ta có:

$$E(X) = \sum_{k=0}^{n} kC_n^k p^k q^{n-k}$$

Theo công thức nhị thức Newton:

$$\sum_{k=0}^{n} C_n^k x^k q^{n-k} = (x+q)^n$$

Lấy đạo hàm hai vế theo x, sau đó nhân cả hai vế với x ta có:

$$\sum_{k=0}^{n} kC_n^k x^k q^{n-k} = nx(x+q)^{n-1} \qquad (4.1)$$

Thay x = p vào (4.1) với chú ý p + q = 1 ta được:

Bây giờ lấy đạo hàm hai vế của (4.1) theo x, sau đó nhân hai vế với x:

$$\sum_{k=0}^{n} k^2 C_n^k x^k q^{n-k} = nx(x+q)^{n-1} + n(n-1)x^2(x+p)^{n-2}$$

Thay x = p vào đẳng thức này, ta được:

$$E(X^2) = \sum_{k=0}^{n} k^2 C_n^k x^k q^{n-k} = np + n(n-1)p^2$$

từ đó: $D(X) = E(X2) – (E(X))2 = np + n(n – 1)p2 – n2p2$

$= np(1 – p) = npq$

Nhận xét. Theo định lý 1.9 chương 1 ta có:

mod(X) = [np – q] hoặc mod(X) = [np – q] + 1

Ví dụ 4.1. Hàng đóng thành kiện mỗi kiện 10 sản phẩm, trong đó có ba phế phẩm. Khách hàng nhận kiện hàng nếu lấy ngẫu nhiên ra hai sản phẩm thì cả hai sản phẩm đều tốt. Khách hàng kiểm tra 100 kiện hàng. Gọi X là số kiện được khách hàng nhận. Tìm E(X), D(X), mod(X).

Giải. Gọi p là xác suất một kiện hàng kiểm tra được nhận, ta có:

$$p = \frac{C_7^2}{C_{10}^2} = \frac{7}{15}$$

Số kiện được nhận là số lần thành công trong dãy 100 phép thử Bernoulli với xác suất thành công $p = \frac{7}{15}$, do đó X ~ B(100; $\frac{7}{15}$). Từ đó:

$$E(X) = np = 100.\frac{7}{15} = \frac{140}{3}$$

$$D(X) = npq = 100.\frac{7}{15}.\frac{8}{15} = \frac{224}{9}$$

mod(X) = 46 hoặc mod(X) = 47 (np – q = 46,133)

2- Phân phối siêu bội

Đại lượng ngẫu nhiên X = {0, 1, 2, ..., n} gọi là có phân phối siêu bội nếu tồn tại các số tự nhiên N, M sao cho $n \le M \le N$ và:

$$p_k = P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \quad k = \overline{0, n}$$

Trong trường hợp này ta ký hiệu X ~ H(N,M,n).

Định nghĩa trên là hợp lý vì theo định nghĩa tổ hợp và quy tắc nhân:

$$\sum_{k=0}^{n} p_k = \frac{\sum_{k=0}^{n} C_M^k C_{N-M}^{n-k}}{C_N^n} = \frac{C_N^n}{C_N^n} = 1$$

Định lý 4.3. Nếu X ~ H(N,M,n) thì:

$$E(X) = np, \ D(X) = npq\frac{N-n}{N-1}$$

trong đó: $p = \dfrac{M}{N}$, $q = 1 - p$.

Chứng minh. Dựa vào công thức tính tổ hợp, dễ dàng kiểm tra rằng:

$$k.\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} = \frac{M}{N}.n.\frac{C_{M-1}^{k-1} C_{N-M}^{(n-1)-(k-1)}}{C_{N-1}^{n-1}}$$

Từ đó đặt $j = k - 1$, ta có:

$$E(X) = \sum_{k=0}^{n} kp_k = \frac{M}{N}.n.\sum_{j=0}^{n-1}\frac{C_{M-1}^{j} C_{N-M}^{(n-2)-j}}{C_{N-1}^{n-1}} = \frac{M}{N}.n$$

Tương tự, có thể kiểm tra rằng:

$$k(k-1)\frac{C_M^k.C_{N-M}^{n-k}}{C_M^n} = \frac{M(M-1)n(n-1)}{N(N-1)}.\frac{C_{M-2}^{k-2} C_{N-M}^{(n-2)-(k-2)}}{C_{N-2}^{n-2}}$$

Từ đó đặt $j = k - 2$, ta có:

$$\sum_{k=0}^{n} k(k-1)p_k = \frac{M(M-1)n(n-1)}{N(N-1)} \sum_{j=0}^{n-2} \frac{C_{M-2}^{j} C_{N-M}^{(n-2)-j}}{C_N^{n-2}}$$

Từ đó: $E(X^2) - E(X) = \dfrac{M(M-1)n(n-1)}{N(N-1)}$

$$D(X) = E(X^2) - (E(X))^2 = \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{M}{N}.n - \left(\frac{M}{N}n\right)^2$$

$$= np\left[\frac{(M-1)(n-1)}{N-1} + 1 - \frac{M}{N}n\right] = np\,\frac{(N-n)(N-M)}{N\,(N-1)}$$

$$= npq\,\frac{N-n}{N-1}$$

Ví dụ 4.2. Một lô hàng 30 sản phẩm, trong đó có 10 phế phẩm. Chọn ngẫu nhiên năm sản phẩm. Gọi X là số phế phẩm trong năm sản phẩm được chọn. Tìm phân phối xác suất, tính kỳ vọng, phương sai của X.

Giải. Ta có:

$$P(X=k) = \frac{C_{10}^{k}.C_{30-10}^{5-k}}{C_{30}^{5}};\ k=\overline{0,5}$$

Vì vậy X có phân phối siêu bội, X ~ H(30,10,5).

$$E(X) = n\frac{M}{N} = 5.\frac{10}{30} = \frac{5}{3}$$

$$D(X) = n.\frac{M}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1} = 5.\frac{1}{3}(1-\frac{1}{3}).\frac{25}{29} = \frac{250}{261}$$

3- Phân phối Poisson

Đại lượng ngẫu nhiên:

$X = \{0, 1, 2, ..., n, ...\}$

gọi là có phân phối Poisson nếu tồn tại số $a > 0$ sao cho:

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

Trong trường hợp này ta ký hiệu X ~ P(a). Số a gọi là tham số của phân phối Poisson.

Định nghĩa vừa nêu là hợp lý vì:

$$\sum_{k=0}^{\infty} p_k = e^{-a}\sum_{k=0}^{\infty}\frac{a^k}{k!} = e^{-a}.e^{a} = 1$$

Định lý 4.4. Nếu X ~ P(a) thì E(X) = D(X) = a.

Chứng minh. Theo công thức tính kỳ vọng ta có:

$$E(X) = \sum_{k=0}^{\infty} kp_k = \sum_{k=0}^{\infty} ke^{-a}.\frac{a^k}{k!} = e^{-a}.a\sum_{j=0}^{\infty}\frac{a^j}{j!} = e^{-a}.a.e^{a} = a$$

Sử dụng kết quả trên ta có:

$$E(X^2) = \sum_{k=0}^{\infty} k^2 p_k = \sum_{k=0}^{\infty} k^2.\frac{e^{-a}a^k}{k!} = a.e^{-a}\sum_{j=0}^{\infty}(j+1).\frac{a^j}{j!}$$

$$= ae^{-a}\left(a\sum_{j=0}^{\infty}\frac{a^j}{j!} + \sum_{j=0}^{\infty}\frac{a^j}{j!}\right) = a(a+1)$$

Từ đó: D(X) = E(X2) – (E(X))2 = a(a + 1) – a2 = a

Nhận xét. Trong thực tế người ta thấy rằng số người vào các trạm phục vụ công cộng trong một đơn vị thời gian có phân phối xấp xỉ Poisson. Dựa vào điều này có thể giải bài toán sau đây:

Ví dụ 4.3. Quan sát 5 phút thấy có 15 người ghé vào một đại lý bưu điện. Tính xác suất trong 1 phút có bốn người vào đại lý bưu điện đó.

Giải. Số người vào trung bình mỗi phút là Gọi X là số người vào trong 1 phút thì X ~ P(3). Từ đó:

$$P(X=4)=\frac{e^{-3}.3^4}{4!}=0{,}1680$$

4.2. CÁC PHÂN PHỐI LIÊN TỤC

1. Phân phối đều

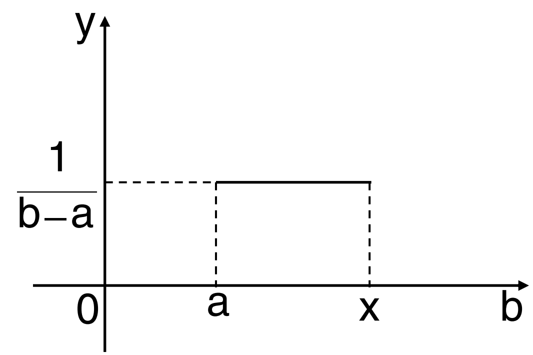
Đại lượng ngẫu nhiên X gọi là có phân phối đều trên đoạn [a,b] nếu hàm mật độ của X là:

$$\begin{cases} \dfrac{1}{b-a} & \text{nếu} \quad x\in[a,b] \\ 0 & \text{nếu} \quad x\notin[a,b] \end{cases}$$

Trong trường hợp này ta ký hiệu X ~ U(a,b).

Định nghĩa vừa nêu là hợp lý vì :

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



Định lý 4.5. Nếu X ~ U(a,b) thì:

$$E(X) = \frac{b+a}{2}, \; D(X) = \frac{(b-a)^2}{12}$$

Chứng minh. Ta có:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{a}^{b} \frac{x}{b-a}dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{b+a}{2}$$

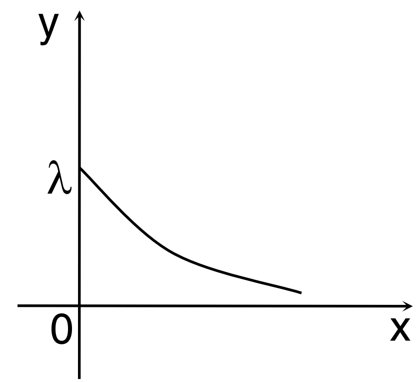
$$E(X^2) = \int_{a}^{b} \frac{x^2}{b-a}dx = \left.\frac{x^3}{3(b-a)}\right|_a^b = \frac{a^2+ab+b^2}{3}$$

$$D(X) = E(X^2) - (E(X))^2 = \frac{a^2+ab+b^2}{3} - \frac{(b+a)^2}{4} = \frac{(b-a)^2}{12}$$

2. Phân phối mũ

Đại lượng ngẫu nhiên X gọi là có phân phối mũ tham số λ ($\lambda > 0$) nếu hàm mật độ của X là:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu} \quad x \geq 0 \\ 0 & \text{nếu} \quad x < 0 \end{cases}$$



Trong trường hợp này ta ký hiệu $X \sim E(\lambda)$.

Định nghĩa vừa nêu là hợp lý vì:

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{+\infty} \lambda e^{-\lambda x}dx = -e^{-\lambda x}\Big|_{0}^{+\infty} = 1$$

Định lý 4.6. Nếu $X \sim E(\lambda)$ thì:

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

Chứng minh. Ta có:

$$E(X) = \int_{0}^{+\infty} x\lambda e^{-\lambda x}dx = -xe^{-\lambda x}\Big|_{0}^{+\infty} + \int_{0}^{+\infty} e^{-\lambda x}dx$$

$$= -\frac{1}{\lambda}e^{-\lambda x}\Big|_{0}^{+\infty} = \frac{1}{\lambda}$$

vì : $E(X^2) = \int_0^{+\infty} x^2 e^{-\lambda x} dx = \dfrac{2}{\lambda^2}$

nên: $D(X) = E(X^2) - (E(X))^2 = \dfrac{2}{\lambda^2} - \dfrac{1}{\lambda^2} = \dfrac{1}{\lambda^2}$

3. Phân phối chuẩn

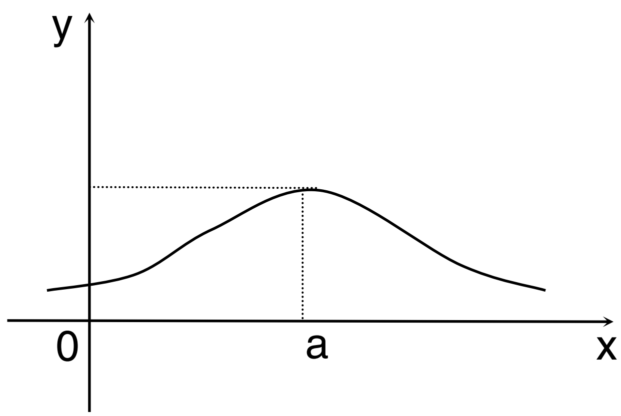
1- Phân phối chuẩn

Đại lượng ngẫu nhiên X gọi là có phân phối chuẩn nếu hàm mật độ của X có dạng:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \sigma > 0$$

Trong trường hợp này ta ký hiệu:

X ~ N(a,σ2)



Áp dụng tích phân Poisson:

$$\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$$

và bằng phép đổi biến:

$$t = \frac{x - a}{\sigma}$$

hay: $x = a + \sigma t$

ta có:

$$\int_{-\infty}^{+\infty} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma.e^{-\frac{t^2}{2}} dt = 1$$

Vậy định nghĩa trên là hợp lý.

Cũng áp dụng tích phân Poisson, dễ dàng nhận được:

Định lý 4.7. Nếu X ~ N(a,σ2) thì E(X) = a, D(X) = σ2.

Theo định lý 4.7, nói X có phân phối chuẩn với kỳ vọng a, phương sai σ2 có nghĩa là X ~ N(a, σ2).

2- Phân phối chuẩn chuẩn tắc

Đại lượng ngẫu nhiên X ~ N(0,1) gọi là có phân phối chuẩn chuẩn tắc.

Nếu X có phân phối chuẩn chuẩn tắc thì hàm mật độ của X là:

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

gọi là hàm mật độ Gauss. Hàm mật độ Gauss là hàm chẵn, ta có:

$$\max f(x) = f(0) = \frac{1}{\sqrt{2\pi}} = 0{,}3989$$

$$\int_{-\infty}^{0} f(x)dx = \int_{0}^{+\infty} f(x)dx = 0{,}5$$

Mọi phân phối chuẩn đều có thể chuẩn tắc hóa nhờ định lý sau đây:

Định lý 4.8. Nếu X ~ N(a, σ2) thì $Y = \dfrac{X-a}{\sigma} \sim N(0,1)$

Chứng minh. Ký hiệu FY(x) và fY(x) là hàm phân phối và hàm mật độ của Y.

Ta có: $$F_Y(x) = P(Y < x) = P\left(\frac{X-a}{\sigma} < x\right)$$

$$= P(X < \sigma x + a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\sigma x + a} e^{-\frac{(t-a)^2}{2\sigma^2}} dt$$

Từ đó: $f_Y(x) = F_Y'(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\sigma x + a - a)^2}{2\sigma^2}} . (\sigma x + a)' = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

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

nên: $Y \sim N(0,1)$

3- Tích phân Laplace

Cho f(x) là hàm mật độ Gauss. Khi đó ta có hàm phân phối Gauss:

$$F(u) = \int_{-\infty}^{u} f(x)dx$$

và tích phân Laplace:

$$\Phi(u) = \int_{0}^{u} f(x)dx = \frac{1}{\sqrt{2\pi}} \int_{0}^{u} e^{-\frac{x^2}{2}} dt$$

Giữa hàm phân phối Gauss và tích phân Laplace có mối liên hệ:

$$F(u) = \frac{1}{2} + \Phi(u)$$

hay: $\Phi(u) = F(u) - \frac{1}{2}$

Hàm $\Phi(u)$ là hàm số lẻ.

Định lý 4.9. Nếu $X \sim N(0,1)$ thì:

(i) $P(\alpha < X < \beta) = \Phi(\beta) - \Phi(\alpha)$

(ii) $P(|X| < \alpha) = 2\Phi(\alpha), \alpha > 0$

Nếu: $X \sim N(a, \sigma 2)$

thì: (iii) $P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - A}{\sigma}\right)$

(iv) $P(|X - a| < \alpha) = 2\Phi\left(\frac{\alpha}{\sigma}\right), \alpha > 0$

Chứng minh

(i) $P(\alpha < X < \beta) = F(\beta) - F(\alpha) = \left(F(\beta) - \frac{1}{2}\right) - \left(F(\alpha) - \frac{1}{2}\right) =$

$= \Phi(\beta) - \Phi(\alpha)$

(ii) $P(|X| < \alpha) = P(-\alpha < X < \alpha) = \Phi(\alpha) - \Phi(-\alpha) = 2\Phi(\alpha)$

(iii) và (iv) suy ra từ (i) và (ii) theo định lý 4.8.

Từ định lý 4.9 (iv) ta có định lý sau đây thường gọi là quy tắc k - sigma.

Định lý 4.10. Nếu X ~ N(a, σ2) thì:

$$P(|X - a| < k\sigma) = 2\Phi(k)$$

Với k = 3, ta có quy tắc 3 – sigma:

$$P(|X - a| < 3\sigma) = 2\Phi(3) = 0{,}9973$$

Quy tắc 3 - sigma có nghĩa là: sai số giữa X và a không quá 3σ là gần chắc chắn (xác suất gần bằng 1).

Ví dụ 4.4. Đường kính của một loại chi tiết do một máy sản xuất có phân phối chuẩn, kỳ vọng 20 mm, phương sai (0,2 mm)2. Tính xác suất lấy ngẫu nhiên một chi tiết.

a) Có đường kính trong khoảng 19,9 mm đến 20,3 mm

b) Có đường kính sai khác với kỳ vọng không quá 0,3 mm.

Giải. Gọi X là đường kính của một chi tiết. Ta có:

X ~ N(20, (0,2)2)

a) $$P(19{,}9 < X < 20{,}3) = \Phi\left(\frac{20{,}3 - 20}{0{,}2}\right) - \Phi\left(\frac{19{,}9 - 20}{0{,}2}\right) =$$

$$= \Phi(1{,}5) + \Phi(0{,}5)$$

$$= 0{,}4332 + 0{,}1915 = 0{,}6247$$

b) $$P(|X - a| \le 0{,}3) = 2\Phi\left(\frac{0{,}3}{0{,}2}\right) = 2\Phi(1{,}5) = 2.0{,}4332 = 0{,}8664$$

4. Phân phối "khi bình phương"

Đại lượng ngẫu nhiên *X* 2 gọi là có phân phối "khi bình phương" n bậc tự do nếu:

$$X\,2 = X_1^2 + X_2^2 + \ldots + X_n^2$$

trong đó: X1, X2, ..., Xn - là các đại lượng ngẫu nhiên độc lập có phân phối chuẩn chuẩn tắc.

Trong trường hợp này ta ký hiệu: *X* 2 ~ *X* 2(n)

Ký hiệu Γ(x) là hàm gama:

$$\Gamma(x) = \int_{o}^{+\infty} t^{x-t} e^{-t} dt$$

Định lý 4.11. Cho X ~ *X* 2(n). Khi đó:

(i) Hàm mật độ của *X*2 là:

$$f_n(x) = \begin{cases} \dfrac{e^{-\frac{x}{2}} x^{\frac{n}{2}-1}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} & \text{nếu } \quad x > 0 \\ 0 & \text{nếu } \quad x \leq 0 \end{cases}$$

(ii) E(*X*2)=n; D(*X*2) = 2N

5. Phân phối Student

Đại lượng ngẫu nhiên T gọi là có phân phối Student n bậc tự do nếu:

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

trong đó: U ~ N(0,1) và V ~ *X*2(n).

Trong trường hợp này ta ký hiệu: T ~ T(n).

Định lý 4.12. Cho T ~ T(n). Khi đó:

(i) Hàm mật độ của T là:

(ii) $E(T) = 0;\ D(T) = \dfrac{n}{n-2}$

4.3. CÁC ĐỊNH LÝ GIỚI HẠN

1- Định lý Chebyshev

Định lý 4.13. (Bất đẳng thức Chebyshev). Cho X là một đại lượng ngẫu nhiên. Khi đó, với mọi, ε > 0 ta có:

$$P\left(|X - E(X)| \geq \varepsilon\right) \leq \frac{D(X)}{\varepsilon^2}$$

Chứng minh. Gọi F(x) là hàm phân phối của X thì:

$$P\left(|X - E(X)| \geq \varepsilon\right) = \int\limits_{|x-E(X)|\geq\varepsilon} dF(x) \leq \frac{1}{\varepsilon^2} \int\limits_{|x-E(X)|\geq\varepsilon} \left(x - E(X)\right)^2 dF(x)$$

$$\leq \frac{1}{\varepsilon^2} \int_{-\infty}^{+\infty} \left(x - E(X)\right)^2 dF(x) = \frac{1}{\varepsilon^2} D(X)$$

Định lý 4.14 (Chebyshev). Cho X1, X2, ..., Xn, ... là dãy các đại lượng ngẫu nhiên đôi một độc lập, có phương sai bị chặn đều (tồn tại C > 0 để D(Xk) ≤ c với mọi k). Khi đó, với mọi, ε > 0 ta có:

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

Chứng minh. Do các Xk độc lập nên:

$$D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n} D(X_k) \le \frac{C}{n}$$

Từ đó theo định lý 4.13:

$$1 \ge P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| < \varepsilon\right) \ge 1 - \frac{D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)}{\varepsilon^2}$$

$$\ge 1 - \frac{C}{n\varepsilon^2}$$

Cho $n \to \infty$ ta được điều cần chứng minh.

Nhận xét. Với các giả thiết của định lý 4.14, khi n khá lớn, ta có:

$$\frac{1}{n}\sum_{k=1}^{n} X_k \approx \frac{1}{n}\sum_{k=1}^{n} E(X_k)$$

2- Định lý Bernoulli

Định lý 4.15 (Bernoulli). Nếu m là số lần thành công trong dãy n phép thử Bernoulli với xác suất thành công p thì:

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

Chứng minh. Gọi Xk là số lần thành công của phép thử thứ k thì phân phối xác suất của Xk là:

X_k	0	1
P	q	p

Do đó E(Xk) = p, D(Xk) = pq. Gọi X là số lần thành công trong n phép thử thì:

$$X = \sum_{k=1}^{n} X_k$$

Theo định lý 4.14:

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| < \varepsilon\right) = 1$$

Nhận xét. Theo định lý 4.15, với n khá lớn, ta có: $p \approx \dfrac{m}{n}$

Định lý 4.15 là cơ sở toán học của định nghĩa xác suất theo thống kê (phần 1 chương 1).

3- Định lý giới hạn trung tâm

Với các đại lượng ngẫu nhiên X1, X2, ..., Xk, ... ta đặt:

$$a_k = E(X_k), \sigma_k^2 = D(X_k), C_k = E|X_k - a_k|^3$$

$$Y_k = \sum_{k=1}^{n} X_k, e_n = \sum_{k=1}^{n} a_k, d_n^2 = \sum_{k=1}^{n} \sigma_k^2$$

Định lý 4.16 (Liapounov). Nếu các đại lượng ngẫu nhiên X1, X2, ..., Xn, ... đôi một độc lập và:

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} C_k}{\left(\sum_{k=1}^{n} \sigma_k^2\right)^{3/2}} = 0$$

thì:

$$\lim_{n\to\infty} P\left(\alpha < \frac{Y_n - e_n}{d_n} < \beta\right) = \Phi(\beta) - \Phi(\alpha)$$

với mọi $\alpha < \beta$, Φ là tích phân Laplace.

Theo định lý 4.16, nếu n khá lớn, ta có:

$$\frac{Y_n - e_n}{d_n} \sim N(0,1)$$

hay: $$Y_n \sim N(e_n, d_n^2)$$

Ta thường sử dụng trường hợp riêng sau đây của định lý 4.16.

Định lý 4.17. Nếu E(Xk) = a, D(Xk) = σ2 với mọi k thì, với n khá lớn, ta có:

$$\frac{\frac{1}{n}Y_n - a}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

hay:

$$\frac{1}{n}(X_1 + X_2 + \ldots + X_n) \sim N\left(a, \frac{\sigma^2}{n}\right)$$

Trong thống kê ta thường coi $n \geq 30$ là khá lớn.

4.4. CÁC CÔNG THỨC GẦN ĐÚNG

1- Phân phối siêu bội và phân phối nhị thức

Định lý 4.18. Cho X ~ H(N, M, n). Nếu n cố định và $\lim\limits_{N\to\infty} \frac{M}{N} = p$

thì với $k = \overline{0,n}$, ta có:

$$\lim_{N\to\infty} P(X = k) = C_n^k p^k q^{n-k}$$

Theo định lý 4.18, nếu N khá lớn so với n thì có thể coi:

$$X \sim B\left(n, \frac{M}{N}\right)$$

tức là ta có công thức gần đúng:

$$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \approx C_n^k \left(\frac{M}{M}\right)^k \left(1 - \frac{M}{N}\right)^{n-k}, k = \overline{0,n}$$

2- Phân phối nhị thức và phân phối Poisson

Định lý 4.19. Cho X ~ B(n,p). Nếu $p \to 0$ và $np \to \lambda$ khi $n \to \infty$ thì với ta có:

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

Ví dụ 4.5. Một máy dệt có 4000 ống sợi. Xác suất để mỗi ống sợi ấy bị đứt trong 1 phút là 0,0005. Tính xác suất trong 1 phút

a) Có ba ống sợi bị đứt

b) Có ít nhất hai ống sợi bị đứt.

Giải. Gọi X là số ống sợi bị đứt trong 1 phút thì X ~ B(4000; 0,0005). Có thể coi X ~ P(2).

a) $P(X = 3) = C_{4000}^{3}(0{,}0005)^{3}.(0{,}9995)^{3997}$

$$\approx e^{-2}.\frac{2^3}{3!} = 0{,}1804$$

Nhận xét. Giá trị đúng của P(X = 3) là 0,180492; sai số 10-5.

b) $P(X \geq 2) = 1 - (P(X = 0) + P(X = 1))$

$$\approx 1 - \left(e^{-2}.\frac{2^0}{0!} + e^{-1}.\frac{2^1}{1!}\right)$$

$$= 1 - (0{,}1353 + 0{,}2707)$$

$$= 0{,}594$$

3- Phân phối nhị thức và phân phối chuẩn

Định lý 4.20 (định lý Moivre – Laplace địa phương). Cho X ~ B(n,p). Nếu n, k ∈ N sao cho $x = \dfrac{k - np}{\sqrt{npq}}$ bị chặn khi n → ∞ thì:

$$P(X = k) = \frac{1}{\sqrt{npq}}.f\left(\frac{k - np}{\sqrt{npq}}\right).\lambda_n$$

trong đó $\lambda n \to 1$ khi $n \to \infty$ và f là hàm mật độ Gauss.

Theo định lý 4.20, khi n khá lớn ta có công thức gần đúng:

$$C_n^k p^k q^{n-k} \approx \frac{1}{\sqrt{npq}} f\left(\frac{k-np}{\sqrt{npq}}\right), k = \overline{0, n}$$

Định lý 4.21 (định lý Moivre – Laplace tích phân). Với giả thiết như trong định lý 4.20, ta có:

$$P(k_1 \le X \le k_2) = \lambda_n \left[\Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right)\right]$$

trong đó: $\lambda n \to 1$ khi $n \to \infty$ và Φ là tích phân Laplace.

Theo định lý 4.21, khi n khá lớn ta có công thức gần đúng:

$$P(k_1 \le X \le k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k}$$

$$\approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right)$$

Hai công thức gần đúng sau cùng này thường chỉ sử dụng khi p không quá gần 0 hoặc 1, vì trong trường hợp đó sai số là lớn.

Ví dụ 4.6. Một xạ thủ có xác suất bắn trúng của mỗi phát là 0,8. Xạ thủ này bắn 64 phát vào bia. Tính xác suất:

a) Có 50 phát trúng bia

b) Có từ 45 đến 52 phát trúng bia

c) Có không dưới 51 phát trúng bia.

Giải. Gọi X là số đạn trúng bia thì X ~ B(64; 0,8).

a) $$P_{64}(50) = C_{64}^{50}(0{,}8)^{50}(0{,}2)^{14}$$

$$\approx \frac{1}{3{,}2}F\left(\frac{50-51{,}2}{3{,}2}\right) = \frac{1}{3{,}2}F(0{,}375) = 0{,}1156$$

Nhận xét. Giá trị đúng của P64(50) là 0,1119; sai số 4.10-3. Nếu tính theo phân phối Poisson (định lý 2) kết quả là 0,0555; sai số khá lớn.

b) $$P_{64}(45{,}52) \approx \Phi\left(\frac{52-51{,}2}{3{,}2}\right) - \Phi\left(\frac{45-51{,}2}{3{,}2}\right)$$

$$= \Phi(0{,}25) + \Phi(1{,}94)$$

$$= 0{,}0987 + 0{,}4738 = 0{,}5725$$

c)
$$P_{64}(51,64) \approx \Phi\left(\frac{64-51{,}2}{3{,}2}\right) - \Phi\left(\frac{51-51{,}2}{3{,}2}\right)$$
$$= \Phi(4) + \Phi(0{,}06) = 0{,}5239$$

BÀI TẬP

4.1. Bài toán S.Pepys

Biến cố nào sau đây có xác suất lớn hơn:

a) Khi gieo sáu xúc xắc cân đối, đồng chất thì có ít nhất một mặt trên có sáu chấm

b) Khi gieo 12 xúc xắc cân đối, đồng chất thì có ít nhất hai mặt trên có sáu chấm

c) Khi gieo 18 xúc xắc cân đối, đồng chất thì có ít nhất ba mặt trên có sáu chấm.

4.2. Gieo 1000 hạt giống với xác suất nảy mầm của mỗi hạt là 0,8. Gọi X là biến ngẫu nhiên chỉ số hạt nảy mầm. X tuân theo quy luật gì? Tính kỳ vọng và phương sai của X.

4.3. Chiều cao của nam thanh niên khi trưởng thành ở một thành phố là một đại lượng ngẫu nhiên có phân phối chuẩn với kỳ vọng a = 165 cm và có độ lệch chuẩn σ = 5 cm. Một thanh niên bị coi là lùn nếu có chiều cao nhỏ hơn 160cm.

a) Tìm tỉ lệ thanh niên lùn ở thành phố

b) Tìm xác suất để lấy ngẫu nhiên ở thành phố đó bốn người thì có ít nhất một người không bị lùn.

4.4. Năng suất lúa ở đồng bằng sông Cửu Long là một đại lượng ngẫu nhiên có phân phối chuẩn với kỳ vọng a = 150 tạ/ha và độ lệch chuẩn σ = 10 tạ/ha. Tìm xác suất để gặt ngẫu nhiên ba thửa ruộng ở vùng đó thì có hai thửa ruộng có năng suất sai lệch so với năng suất trung bình không quá 2 tạ/ha.

4.5. Một xe tải vận chuyển 8000 chai rượu vào kho. Xác suất để khi vận chuyển mỗi chai bị vỡ là 0,001. Tìm xác suất để khi vận chuyển:

a) Có đúng sáu chai bị vỡ

b) Có không quá 12 chai bị vỡ.

4.6. Một tổng đài điện thoại nội bộ của một cơ quan phục vụ 1000 máy điện thoại. Xác suất để trong mỗi phút mỗi máy điện thoại gọi đến tổng đài là 0,02. Tìm số máy gọi đến tổng đài trung bình trong một phút.

4.7. Xác suất hỏng của các hạt thóc giống là 0,006. Tìm xác suất sao cho khi chọn 1000 hạt giống có:

a) Không ít hơn ba hạt bị hỏng

b) Có đúng sáu hạt bị hỏng

c) Có không quá 15 hạt bị hỏng.

4.8. Trong số 500 trang của một cuốn sách có 10 chỗ in nhầm. Tìm xác suất để khi lấy ngẫu nhiên một trang thì có không ít hơn hai chỗ in nhầm.

4.9. Một máy đếm để gần một nguồn phóng xạ sao cho xác suất để một hạt phát ra từ nguồn phóng xạ được ghi lại trong máy đếm là 10-4. Giả sử rằng trong thời gian quan sát có 40000 hạt được phóng ra từ nguồn phóng xạ. Tìm xác suất sao cho mấy đếm:

a) Ghi được trên sáu hạt

b) Không ghi được hạt nào cả

c) Tính số hạt ít nhất mà nguồn phóng xạ cần phát ra sao cho với xác suất lớn hơn 0,945 máy đếm ghi được không ít hơn bốn hạt

4.10. Gieo một xúc xắc cân đối, đồng chất 12000 lần. Tìm xác suất để số lần xuất hiện mặt sáu chấm trên con xúc xắc gần giữa 1900 và 2150.

4.11. Biến ngẫu nhiên X là trung bình cộng của n biến ngẫu nhiên độc lập có cùng phân phối: X1, X2, ..., Xn với phương sai:

$$D(Xk) = 5 \ (k = 1, 2, ..., n)$$

Xác định n sao cho với xác suất không bé hơn 0,9973.

a) Hiệu của X - E(X) không vượt quá 0,01

b) Trị tuyệt đối của X - E(X) không vượt quá 0,005.

4.12. Gieo 3200 lần một đồng xu cân đối, đồng chất. X là biến ngẫu nhiên chỉ số lần xuất hiện mặt sấp trong 3200 lần gieo đó. Tìm xác suất sao cho giá trị của X nằm trong khoảng .

4.13. Người ta muốn khảo sát thời gian cháy sáng trung bình của một lô bóng đèn bằng phương pháp chọn mẫu. Hỏi phải lấy mẫu cỡ bao nhiêu để cho, với xác suất không bé hơn 0,9876, có thể kết luận rằng trị số tuyệt đối của hiệu thời gian cháy sáng trung bình của bóng đèn của toàn lô và kỳ vọng của nó không vượt quá 10 giờ, biết rằng độ lệch tiêu chuẩn của thời gian cháy sáng của bóng đèn là 80 giờ.

4.14. Một cán bộ phòng thí nghiệm thực hiện việc chọn giống lúa. Anh ta kiểm tra 10000 hạt lúa giống, xác suất để mỗi hạt lúa đạt tiêu chuẩn là 0,2. Tìm xác suất sao cho độ lệch giữa tần suất các hạt lúa đạt tiêu chuẩn so với xác suất 0,2 không vượt quá 0,01.

4.15. Thời gian phục vụ mỗi khách hành tại một cửa hàng mậu dịch là một đại lượng ngẫu nhiên X tuân theo quy luật lũy thừa với hàm mật độ xác suất như sau:

$$\varphi(x) = \begin{cases} 5e^{-5x} & \text{khi } x > 0 \\ 0 & \text{khi } x \le 0 \end{cases}$$

với x được tính bằng phút/khách hàng.

a) Tìm xác suất để thời gian phục vụ một khách hàng nào đó sẽ nằm trong khoảng từ 0,4 đến 1 phút

b) Tìm kỳ vọng và phương sai của biến ngẫu nhiên X

4.16. Buffon đã gieo đồng tiền 4040 lần, thấy mặt sấp xuất hiện 2048 lần. Hãy tìm xác suất để khi gieo đồng tiền 4040 lần thì số lần mặt sấp xuất hiện dao động quanh 2020 không xa hơn kết quả mà Buffon nhận được.

4.17. Xác suất để xuất hiện biến cố A trong mỗi phép thử là 0,6. Cần phải làm bao nhiêu phép thử để xác suất có độ lệch của tần suất so với xác suất bé hơn 0,01 là 0,995.

4.18. 600 mẫu ngô đã gieo, xác suất nảy mầm là 0,9. Tìm sai số giới hạn của tần suất nảy mầm so với xác suất nảy mầm của từng mẫu với độ tin cậy 0,995.

4.19. Xác suất trúng đích của một viên đạn là 0,2. Tìm xác suất để khi bắn 400 viên thì có tất cả:

a) 70 viên trúng

b) Từ 60 đến 100 viên trúng