BÀI4: MỘT SỐ ĐỊNH LÝ QUAN TRỌNG TRONG LÝ THUYẾT XÁC SUẤT

Mục tiêu

Giới thiệu những dạng đơn giản nhất (không chứng minh) của một số định lý cơ bản nhất của Lý thuyết Xác suất. Đây là những cơ sở quan trọng của lý thuyết Ước lượng và Lý thuyết Kiểm định.

Thời lượng

- 4 tiết

Các kiến thức cần có

- Định lý Poisson
- Luật số lớn
- Định lý giới hạn trung tâm

4.1. Định lý Poisson

Trong thực hành ta thường bắt gặp tình huống cần xác định khả năng xuất hiện k lần biến cố A trong n phép thử, được biết trước xác suất p của việc xảy ra biến cố A trong một phép thử. Lúc đó ta có thể dùng các công thức của phân phối nhị thức để tính toán. Tuy nhiên công thức đó chỉ thích hợp cho trường hợp số lượng n các phép thử tương đối nhỏ, còn khi số lượng phép thử lớn thì có thể áp dụng Định lý Poisson để tính gần đúng.

Định lý Poisson:

Xác suất của một biến cố xuất hiện k lần trong n phép thử (xác suất xuất hiện biến cố trong một phép thử là p) với n tương đối lớn, $p \ll 1$ *và* $np \approx \lambda$ *với* λ *là một số cố định nào đó, được tính xấp xỉ theo công thức:*

$$P_n(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \approx e^{-np} \frac{(np)^k}{k!} \approx e^{-\lambda} \frac{(\lambda)^k}{k!}.$$

Trong trường hợp cần tính xác suất biến cố A xuất hiện từ k_1 đến k_2 lần trong n phép thử, ký hiệu xác xuất đó là $P_n(k_1,k_2)$, áp dụng định lý Poisson để tính xấp xỉ cho giá trị $P_n(k_1,k_2)$ ta có công thức:

$$P_n(k_1,k_2) = \sum_{k=k_1}^{k_2} P_n(k) \approx \sum_{k=k_1}^{k_2} e^{-\lambda} \frac{(\lambda)^k}{k!}$$

Ví dụ 1:

Tổng sản phẩm của xí nghiệp A trong một quý là 800. Xác xuất để sản xuất ra một phế phẩm là 0,005. Tìm xác suất để cho

1. Có 3 sản phẩm là phế phẩm.
2. Có không quá 10 phế phẩm.

Giải:

Ta có $n = 800, p = 0,005$. Vậy $\lambda = np = 4$, từ đó

1. $P_{800}(3) = e^{-4} \frac{4^3}{3!} = 0,1954$.

2. $P_{800}(0,10) = \sum_{k=0}^{10} e^{-4} \frac{4^k}{k!} = 0,997$.

4.2. Luật Số lớn

Đối với mỗi một tham số của biến ngẫu nhiên (kỳ vọng, phương sai, xác suất, v.v.), người ta có thể dùng nhiều thống kê khác nhau để ước lượng. Do vậy người ta đã đưa ra một số tiêu chuẩn để đánh giá các ước lượng của tham số như tính vững, tính không chệch, tính hiệu quả, v.v. Luật Số lớn là một công cụ giúp đánh giá tính vững cho ước lượng của hai tham số thống kê là xác suất và kỳ vọng.

Định lý Bernoulli:

Nếu f *là tần suất xuất hiện biến cố* A *trong* n *phép thử độc lập và* p *là xác suất xuất hiện biến cố đó trong mỗi phép thử thì với mọi ε dương nhỏ tùy ý ta luôn có*

$$\lim_{n\to\infty} P(|f-p|<\varepsilon)=1.$$

Định lý trên còn được gọi là Luật số lớn Bernoulli. Định lý này cho thấy tần suất là một ước lượng vững của xác suất. Đối với kỳ vọng, ta có định lý dạng tổng quát, được phát biểu như sau:

Luật Số lớn: *Giả sử* $X_1, X_2, ..., X_n, ...$ *là dãy các biến ngẫu nhiên độc lập có cùng phân bố với kỳ vọng chung* μ *và phương sai* σ^2 *hữu hạn. Khi đó với mọi ε dương nhỏ tùy ý ta luôn có*

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

4.3. Định lý Giới hạn trung tâm

Trên đây ta thấy có thể tính xấp xỉ các xác suất của luật phân phối nhị thức với số lượng phép thử lớn thông qua luật phân phối Poisson. Các định lý Giới hạn trung tâm trình bày dưới đây sẽ cung cấp một công cụ khác để tính xấp xỉ các xác suất thông qua luật phân phối chuẩn tắc.

Định lý Moivre-Laplace:

Giả sử X_n *là biến ngẫu nhiên có phân bố nhị thức với tham số* (n,p). *Đặt*

$$S_n=\frac{X_n-np}{\sqrt{np(1-p)}}.$$

Khi đó với mọi $x\in(-\infty,+\infty)$ *ta có*

$$\lim_{n\to\infty} P(S_n<x)=P(Z<x)$$

trong đó Z *là biến ngẫu nhiên có phân bố chuẩn tắc.*

Định lý trên cho thấy có thể tính xấp xỉ xác suất $P_n(k)$ để một biến cố xuất hiện *k* lần trong n phép thử (p) là xác suất xuất hiện biến cố A trong một phép thử của lược đồ Bernoulli với n tương đối lớn theo công thức

$$P_n(k)=(2\pi)^{-1/2}e^{-(k-np)^2/2np(1-p)}=\varphi(x_k)$$

với φ là hàm mật độ của phân bố chuẩn tắc:

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

Còn:

$$x_k = \frac{k - np}{\sqrt{np(1-p)}}.$$

Ví dụ 2:

Xác suất để sản xuất ra một chi tiết loại tốt là 0,4. Tìm xác suất để trong 26 chi tiết sản xuất ra thì có 13 chi tiết loại tốt.

Giải:

Ta cần tìm $P_{26}(13)$ *với* $n = 26$, $p = 0{,}4$, $1 - p = 0{,}6$ và

$$x_k = \frac{(k - np)}{\sqrt{np(1-p)}} = 1{,}04,\ \varphi(x_k) = \varphi(1{,}04) = 0{,}2323,\ P_{26}(13) \approx \frac{0{,}2323}{2{,}5} = 0{,}093.$$

Khi áp dụng Định lý Moivre - Laplace để tính xấp xỉ cho giá trị $P_n(k_1, k_2)$ ta có công thức

$$P_n(k_1, k_2) = \Phi(\beta) - \Phi(\alpha).$$

Với:

$$\alpha = \frac{(k_1 - np)}{\sqrt{np(1-p)}},\ \beta = \frac{(k_2 - np)}{\sqrt{np(1-p)}}.$$

Và:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{\frac{-x^2}{2}} dx.$$

Ví dụ 3:

Một phân xưởng sản xuất bóng đèn đạt trung bình là 70% sản phẩm loại tốt. Tìm xác suất để trong 1000 bóng đèn có từ 652 đến 760 bóng đèn loại tốt.

Giải:

Ta có $n = 1000$, $p = 0{,}7$, $1\text{-}p = 0{,}3$, $k_1 = 652$, $k_2 = 700$. Xác suất phải tìm là $P_{1000}(652;760)$.

Như vậy

$$\alpha = \frac{(k_1 - np)}{\sqrt{np(1-p)}} = -3{,}31;\ \Phi(\alpha) = \Phi(-3{,}31) = -0{,}499520.$$

$$\beta = \frac{(k_2 - np)}{\sqrt{np(1-p)}} = 4{,}14;\ \Phi(\beta) = \Phi(4{,}14) = 0{,}499968$$

Từ đó $P_{1000}(652;760) = \Phi(\beta) - \Phi(\alpha) = 0{,}999488$.

Định lý Moivre – Laplace trên đây là một dạng đặc biệt của Định lý Giới hạn trung tâm, được áp dụng cho các biến ngẫu nhiên có phân phối 0–1. Đối biến ngẫu nhiên có phân phối dạng bất kỳ, ta có định lý tổng quát sau đây:

Định lý Giới hạn trung tâm:

Nếu $X_1, X_2, ..., X_n, ...$ *là một dãy các biến ngẫu nhiên độc lập cùng tuân theo một quy luật phân phối xác suất với kỳ vọng toán* μ *và phương sai hữu hạn* σ^2*, thì quy luật phân phối xác suất của biến ngẫu nhiên.*

$$U_n = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - n\mu}{\sqrt{n.\sigma^2}} \text{ với } S_n = \sum_{k=1}^{n} X_k$$

sẽ hội tụ tới quy luật chuẩn tắc $N(0,1)$ *khi* $n \to \infty$.

Các định lý Giới hạn trung tâm có ý nghĩa rất quan trọng đối với việc áp dụng thống kê toán học trong thực tế, không những chỉ với công dụng tính xấp xỉ các xác suất như đã trình bày ở trên mà còn cả trong quá trình tiến hành các phép kiểm định thống kê. Thật vậy, phần lớn các tiêu chuẩn kiểm định thống kê cổ điển như kiểm định so sánh các tần suất, so sánh các giá trị trung bình, so sánh phương sai, v.v. đều được xây dựng dựa trên cơ sở ban đầu của các biến ngẫu nhiên có phân phối chuẩn. Tuy nhiên trong các số liệu thực tế hầu như rất khi bắt gặp một biến ngẫu nhiên thực sự có phân phối chuẩn. Lúc đó phải dựa vào hiệu lực của các định lý Giới hạn trung tâm để áp dụng các tiêu chuẩn kiểm định thống kê một cách gần đúng cho các trường hợp số liệu có cỡ mẫu đủ lớn.

TÓM LƯỢC CUỐI BÀI

Bài này cung cấp cho học viên một số định lý cơ bản trong Lý thuyết Xác suất : Định lý Poisson, Luật Số lớn, Định lý Giới hạn trung tâm. Những định lý này sẽ là cơ sở quan trọng của Lý thuyết Ước lượng và Lý thuyết Kiểm định được trình bày trong phần 2 của giáo trình này, cũng như cung cấp cho học viên những công thức tính gần đúng với một số bài toán xác suất phổ biến.