Rachunek prawdopodobieństwa MAT1332

Wydział Matematyki, Matematyka Stosowana

Wykładowca: dr hab. A. Jurlewicz

Wykład 10: Twierdzenie de Moivre'a-Laplace'a. Centralne twierdzenie graniczne.

Szacowanie jakości przybliżenia prawdopodobieństwa sukcesu przez częstość występowania sukcesów w schemacie Bernoulliego:

Niech S_n to ilość sukcesów w n próbach Bernoulliego z prawdopodobieństwem sukcesu p. Z PWL Bernoulliego wiemy, że $P\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) \underset{n\to\infty}{\longrightarrow} 0$ dla dowolnego $\epsilon > 0$. Ustalmy małe $\epsilon > 0$ i $\delta > 0$. Chcemy wyznaczyć n, dla jakiego pojedyncza wylosowana częstość występowania sukcesów w n próbach różnić się będzie od prawdopodobieństwa sukcesu p nie więcej niż o ϵ z prawdopodobieństwem $1-\delta$ (bliskim 1). Inaczej mówiąc, dla jakiego n mamy

$$P\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) < \delta? \tag{1}$$

Wyznaczanie n z nierówności Czebyszewa:

Dla $X = \frac{S_n}{n}$ mamy $\mathrm{E}X = \frac{np}{n} = p$, $\mathrm{D}^2X = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$, a zatem z nierówności Czebyszewa, dla każdego $\epsilon > 0$

$$P\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) = P(|X - \mathrm{E}X| > \epsilon) \leqslant \frac{\mathrm{D}^2X}{\epsilon^2} = \frac{\frac{p(1-p)}{n}}{\epsilon^2} = \frac{p(1-p)}{\epsilon^2}\cdot\frac{1}{n}.$$

Zatem warunek (1) jest spełniony, gdy

$$\frac{p(1-p)}{\epsilon^2}\cdot\frac{1}{n} < \delta \iff n > \frac{p(1-p)}{\delta\epsilon^2}.$$

Jak widać, gdy ϵ i δ są małe, oszacowana tą metodą wartość n jest bardzo duża i na ogół dużo większa niż potrzeba.

Uwaga: Nierówność Czebyszewa odkryto dość późno. Wcześniej szacowano wartość n korzystając z **twierdzenia de Moivre'a-Laplace'a** (XVIII w.). Twierdzenie to dało początek tematyce **twierdzeń granicznych**.

Opracowanie: dr hab. Agnieszka Jurlewicz

TWIERDZENIE DE MOIVRE'A-LAPLACE'A:

Niech S_n będzie liczbą sukcesów w n próbach Bernoulliego z prawdopodobieństwem sukcesu p. Wtedy

$$\frac{S_n - np}{\sqrt{np(1-p)}} = \frac{S_n - \mathrm{E}S_n}{\sqrt{\mathrm{D}^2 S_n}} \xrightarrow[n\to\infty]{d} Y$$

gdzie Y ma standardowy rozkład normalny $\mathcal{N}(0,1)$.

Inaczej mówiąc, dla dowolnego $x \in \mathbb{R}$

$$P\left(\frac{S_n - np}{\sqrt{np(1-p)}} < x\right) \xrightarrow[n\to\infty]{} P(Y < x) = \Phi(x)$$

gdzie $\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ to dystrybuanta standardowego rozkładu normalnego $\mathcal{N}(0,1)$.

UWAGA:

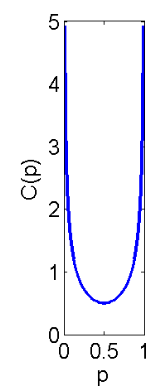
Twierdzenie de Moivre'a-Laplace'a mówi o tym, że liczba sukcesów w n próbach Bernoulliego z prawdopodobieństwem sukcesu p po standardyzacji (tzn. unormowaniu do zmiennej losowej o średniej 0 i wariancji 1) dąży według rozkładu do standardowego rozkładu normalnego, gdy $n \to \infty$.

Zatem dla dużych n liczba sukcesów w n próbach Bernoulliego z prawdopodobieństwem sukcesu p ma asymptotycznie rozkład normalny $\mathcal{N}(np, \sqrt{np(1-p)})$. Równoważnie, częstość występowania sukcesów $\frac{S_n}{n}$ ma asymptotycznie rozkład normalny $\mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$.

OSZACOWANIE DOKŁADNOŚCI PRZYBLIŻENIA W TWIERDZENIU DE MOIVRE'A-LAPLACE'A:

$$\sup_{x\in\mathbb{R}} \left| P\left(\frac{S_n - np}{\sqrt{np(1-p)}} < x\right) - \Phi(x) \right| \leqslant \frac{C(p)}{\sqrt{n}},$$

gdzie $C(p) = \frac{p^2 + (1-p)^2}{2\sqrt{p(1-p)}}$.



Widać, że dokładność przybliżenia rośnie wraz ze wzrostem n.

Dla $p \in (0,1)$ funkcja $C(p)$ przyjmuje wartość najmniejszą dla $p = 0.5$ i rośnie do ∞ przy $p \to 0+$ oraz przy $p \to 1-$. Zatem jakość przybliżenia dla danego n jest lepsza, gdy p jest z wnętrza przedziału $(0,1)$. Dla skrajnych wartości p, bliskich 0 lub bliskich 1, jakość tego przybliżenia mocno się pogarsza.

Opracowanie: dr hab. Agnieszka Jurlewicz

Wyznaczanie n spełniającego (1) z twierdzenia de Moivre'a-Laplace'a:

$$P\left(\left|\frac{S_n}{n}-p\right|>\epsilon\right)=P\left(\left|\frac{S_n-np}{\sqrt{np(1-p)}}\right|>\frac{n\epsilon}{\sqrt{np(1-p)}}\right)\approx 2\left(1-\Phi\left(\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)\right)$$

dla n dostatecznie dużych, przy czym błąd przybliżenia nie przekracza $\frac{2C(p)}{\sqrt{n}}$.

Zatem warunek (1) jest spełniony, gdy

$$2\left(1-\Phi\left(\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)\right)+\frac{2C(p)}{\sqrt{n}}<\delta.$$

Aproksymacja rozkładu Bernoulliego rozkładem normalnym.

Z twierdzenia de Moivre'a-Laplace'a można w przybliżony sposób szybko obliczyć prawdopodobieństwa wystąpienia k sukcesów w n próbach Bernoulliego dla dużych n i p z wnętrza przedziału $(0,1)$, odległych od 0 i od 1.

Niech S_n oznacza liczbę sukcesów w n próbach Bernoulliego z prawdopodobieństwem sukcesu p. Wtedy dla dowolnego $k=0,1,\ldots,n$

$$\begin{aligned}
P(S_n<k)=P(S_n<k-0.5) &\approx \Phi\left(\frac{k-0.5-np}{\sqrt{np(1-p)}}\right)\\
P(S_n\leqslant k)=P(S_n<k+0.5) &\approx \Phi\left(\frac{k+0.5-np}{\sqrt{np(1-p)}}\right)\\
P(S_n\geqslant k)=P(S_n>k-0.5) &\approx 1-\Phi\left(\frac{k-0.5-np}{\sqrt{np(1-p)}}\right)\\
P(S_n>k)=P(S_n>k+0.5) &\approx 1-\Phi\left(\frac{k+0.5-np}{\sqrt{np(1-p)}}\right)
\end{aligned}$$

z błędem, który nie przekracza $\frac{C(p)}{\sqrt{n}}$.

$$P(S_n=k)=P(k-0.5<S_n<k+0.5)\approx\Phi\left(\frac{k+0.5-np}{\sqrt{np(1-p)}}\right)-\Phi\left(\frac{k-0.5-np}{\sqrt{np(1-p)}}\right)$$

z błędem, który nie przekracza $\frac{2C(p)}{\sqrt{n}}$.

Uwaga: PWL Bernoulliego (tw. Borela) to szczególny przypadek PWL dla ciągu niezależnych zmiennych losowych o jednakowym rozkładzie o skończonej średniej (gdy rozkład ten jest zerojedynkowy $\mathcal{B}(1,p)$). Podobnie, twierdzenie de Moivre'a-Laplace'a jest szczególnym przypadkiem ogólniejszego **Centralnego Twierdzenia Granicznego (CTG) Lindeberga-Lévy'ego**

Opracowanie: dr hab. Agnieszka Jurlewicz

Centralne Twierdzenie Graniczne Lindeberga-Lévy'ego

Niech (X_n) będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie, przy czym $0 < \mathrm{D}^2 X_n = \sigma^2 < \infty$. Oznaczmy $m = \mathrm{E}X_n$ (ta wartość oczekiwana istnieje na mocy założenia, że istnieje wariancja). Wówczas

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} Y,$$

gdzie Y ma standardowy rozkład normalny $\mathcal{N}(0,1)$.

Inaczej mówiąc, dla dowolnego $x \in \mathbb{R}$

$$P\left(\frac{S_n - nm}{\sigma\sqrt{n}} < x\right) \xrightarrow[n\to\infty]{} P(Y < x) = \Phi(x)$$

gdzie $\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$ to dystrybuanta standardowego rozkładu normalnego $\mathcal{N}(0,1)$.

Oszacowanie dokładności przybliżenia w CTG Lindeberga-Lévy'ego - nierówność Berry-Esseena:

Niech (X_n) będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie, przy czym $\mathrm{E}|X_n|^3 < \infty$. Ponadto niech $m = \mathrm{E}X_n$, $\sigma^2 = \mathrm{D}^2 X_n > 0$ (ta wartość oczekiwana i wariancja istnieją na mocy założenia o rozkładzie). Wówczas

$$\sup_{x\in\mathbb{R}} \left| P\left(\frac{S_n - nm}{\sigma\sqrt{n}} < x\right) - \Phi(x)\right| \leqslant C\frac{\mathrm{E}|X_1 - m|^3}{\sigma^3\sqrt{n}},$$

dla pewnej stałej C, która spełnia nierówność $\frac{1}{\sqrt{2\pi}} \leqslant C < 0.5$.

Uwagi:

1. Jeżeli zmienne losowe X_n maja niezerową skończoną wariancję, to do szacowania szybkości zbieżności w PWL Kołmogorowa można stosować metody analogiczne do tych dotyczacych PWL Bernoulliego, opartych na twierdzeniu de Moivre'a-Laplace'a.

2. Twierdzenie de Moivre'a-Laplace'a, CTG Lindeberga-Lévy'ego to przykłady centralnych twierdzeń granicznych. CTG dla zmiennych o różnych rozkładach to np. twierdzenie Lindeberga-Fellera, twierdzenie Lapunowa.

3. CTG Lindeberga-Lévy'ego to wynikanie:
istnieje niezerowa wariancja $\Longrightarrow$ zbieżność unormowanych sum do standardowego rozkładu normalnego.

Interpretacja CTG Lindeberga-Lévy'ego:

Jeżeli wielkość fizyczna jest opisana zmienną losową i jest wynikiem sumowania wielu niezależnych jednakowych statystycznie efektów, przy czym rozkład efektu ma skończoną wariancję, to wielkość ta ma asymptotycznie rozkład normalny.

Opracowanie: dr hab. Agnieszka Jurlewicz