Materiały do wykładu PMPiSt 3.12.2020

Temat. Zjawisko zbieżności w zbiorze rozkładów – TWIERDZENIA GRANICZNE.

Wstęp

Problem 1. Niech $\mathcal{R}$ oznacza zbiór rozkładów (dyskretnych i ciągłych).

Wtedy $r \in \mathcal{R}$ reprezentuje rozkład,

$(r_n)_{n \geq 1}$ – ciąg takich rozkładów.

Dokładniej, biorąc r, lub $(r_n)_{n \geq 1}$ mamy na myśli, że:

a) istnieje ZL

b) jego MPK (Ω, Σ, P)

c) albo (dla r) mamy zm. losową $X: \Omega \to \mathbb{R}$

albo (dla $(r_n)_{n \geq 1}$) mamy ciąg zm. losowych $(X_n)_{n \geq 1}$, $X_n: \Omega \to \mathbb{R}$.

Problem 2. Niech dany będzie $(X_n)_{n \geq 1}$ j.w.

Jak zdefiniować zbieżność takiego ciągu? Co ma być jego „granicą”?

W T.P. mówimy co najmniej dla rodzaju zbieżności:

① tzw. ZBIEŻNOŚCI PRAWIE WSZĘDZIE, co zapisujemy

$$(X_n)_{n\geq 1} \text{ z.p.w.} \equiv \exists_{X_0} \quad X_n \xrightarrow{p.w.} X_0$$

$$\equiv$$

$$P(\{\omega\in\Omega : X_n(\omega) \to X_0(\omega)\}) = 1$$

② tzw. ZBIEŻNOŚCI wg. ROZKŁADU, co zapisujemy

$$(X_n)_{n\geq 1} \text{ z. wg. } d \equiv \exists X_0$$

$$\forall_t \; F_{X_n}(t) \to F_{X_0}(t), \quad X_n \xrightarrow{d} X_0$$

TW.: $X_n \xrightarrow{d} X_0 \Rightarrow X_n \xrightarrow{p.w.} X_0$

Uwaga. Dla (X_n), $\omega \in \Omega$, mamy ciąg liczbowy $(X_n(\omega))_{n\geq 1}$, Zatem $X_n(\omega) \to X_0(\omega)$ oznacza zbieżność tego ciągu liczbowego.

Twierdzenia graniczne.

I MPWLB: MOCNE PRAWO WIELKICH LICZB BERNOULLIEGO.

Problem 1. Wiemy X: $\begin{array}{c|c} 0 & 1 \\ \hline q & p \end{array}$

Jak wyznaczyć p?

Problem 2. Wiemy, że w praktyce pomiarem X bierzemy EX.
Czy można wyznaczyć EX bez znajomości rozkładu X?

Uwaga. Zauważmy, że w PB, $p = EX$ (!).

Rozwiązanie P1 & P2: MPWLB.

Przypuśćmy, że obserwujemy ZL o pewnym (~~nieznanym~~) rozkładzie X, o którym wiemy tylko, że ma drugi moment a więc istnieje EX i $\operatorname{var} X$.
Obserwacji ZL dokonujemy w postaci nieskończonej serii powtórzeń: $X_1, X_2, \ldots, X_n, \ldots$, gdzie

$\forall_{n \geq 1}\ X_n \stackrel{d}{=} X$ oraz powtórzenia są stochastycznie niezależne.

Biezmy zdarzenie

$$A = \{ \omega \in \Omega : \text{ ciąg liczb } \frac{1}{n}(X_1(\omega) + X_2(\omega) + \ldots + X_n(\omega)) \text{ jest zbieżny} \}.$$

<u>Twierdzenie M.P.W.L.B.</u>

1) $A = \{ \omega \in \Omega : \frac{1}{n}(X_1(\omega) + \ldots + X_n(\omega)) \longrightarrow EX \}$

2) $P(A) = 1$.

//

<u>Uwagi:</u>

1) Warunek 2) i 1) M.P.W.L.B. oznacza, że dla $\omega_0 \in A$ z prawdop. 1 mamy

$$\frac{1}{n}(X_1(\omega_0) + X_2(\omega_0) + \ldots + X_n(\omega_0)) \longrightarrow EX,$$

a więc dla <u>odpowiednio dużych</u> n,

$$\frac{1}{n}(X_1(\omega_0) + X_2(\omega_0) + \ldots + X_n(\omega_0)) \approx EX$$

2) Zatem pomiar X (nie znamy rozkładu!) sprowadza się do wygenerowania <u>odpowiedniej</u> długości

serii powtórzeń obserwacji i na uśrednieniu tej serii.

o czym wszyscy doskonale wiemy i robimy to!

3) Weźmy schemat z P.?

$$X: \begin{array}{c|c} 0 & 1 \\ \hline q & p \end{array}$$

Niech $X_k \overset{d}{=} X$, $X_1, X_2, \dots, X_n$ niezależne.
$1 \le k \le n$

Zauważmy, że wtedy dla $\omega_0 \in \Omega$

$$\frac{1}{n}\left(X_1(\omega_0) + X_2(\omega_0) + \dots + X_n(\omega_0)\right) = \frac{\#\{i : X_i(\omega_0) = 1\}}{\#\text{powtórzeń}}$$

$$\| $$

$$f_n(\omega_0),$$

dlatego nazywamy to FREKWENCJA

MPWL B mówi, że $f_n(\omega_0) \xrightarrow{n\to\infty} p$, a więc

$$f_n(\omega_0) \approx p$$

PRAWDOPODOBIEŃSTWO EMPIRYCZNE

PRAWDOPODOBIEŃSTWO TEORETYCZNE

4| Ważny przykład: niech $B_n \in B(n, p)$, $n \geq 2$

Dla ε, $n \geq 2$ rozważmy zdarzenie

$$A_{\varepsilon, n} = \left\{ \omega \in \Omega : \left| \frac{1}{n} B_n(\omega) - p \right| \geq \varepsilon \right\}$$

Oszacujmy $P(A_{\varepsilon, n})$. W tym celu skorzystajmy z nierówności Czebyszewa, gdzie

$$X = \frac{1}{n} B_n, \quad \varepsilon = t\sigma \Rightarrow t = \frac{\varepsilon}{\sigma}$$

$$EX = \frac{1}{n} \cdot np = p, \quad \sigma^2 = Var(X) = \frac{1}{n^2} \cdot npq = \frac{pq}{n},$$

skąd $\varepsilon = t\sqrt{\frac{pq}{n}}$

oraz $P(\{\omega \in \Omega : |X - EX| \geq t\sigma\}) \leq \frac{1}{t^2}$

Mamy więc:

$$P(A_{\varepsilon, n}) \leq \frac{1}{t^2} = \frac{\sigma^2}{\varepsilon^2} = \frac{pq}{n\varepsilon^2},$$

albo w postaci odwrotnej

$$P(A^c_{\varepsilon, n}) \geq 1 - \frac{pq}{n\varepsilon^2}.$$

Wykonaj teraz symulacji (proszę oprogramować to!

Założenia: $p=q=0{,}5$, $\varepsilon=10^{-k}$, $k\geq 2$,

~~oraz niech~~ czyli

$$\frac{pq}{n\varepsilon^2}=\frac{1/4}{n10^{-2k}}=\frac{1}{4n\cdot 10^{-2k}} \text{ oraz}$$

dla danego $k\geq 2$ niech n będzie takie, że

$$1-\frac{1}{4n\cdot 10^{-2k}}\geq 1-10^{-3}, \text{ czyli}$$

$$\frac{1}{4n\cdot 10^{-2k}}\leq 10^{-3} \iff n>\frac{1}{4}10^{3+2k}$$

Weźmy $k=2$ (tylko!). Wtedy

$\varepsilon=0{,}01$

$$n>\frac{1}{4}\cdot 10^{7}=2.500.000.$$

Podsumujmy:

$$P(\{\omega\in\Omega: |\tfrac{1}{n}B_n(\omega)-p|<0{,}01\})\geq 0{,}999$$

o ile $n\geq 2.500.000$

Wtedy prawie na pewno (bo z pr. co najmniej 0,99)

$\frac{1}{n} B_n(\omega_0) \approx p$, gdzie odchyłka będzie rzędu $\underline{\underline{0{,}01}}$.

ZADANIE 1

Napisać procedurę pos. symulacji. Powtórzyć $k = 2, 3, 4, 5$!

II. C.T.W. – CENTRALNE TW. GRANICZNE

Zajmiemy się teraz mocniejszym rodzajem zbieżności w R – w.g. rozkładu.

Niech X spełnia założenia M.P.W.L.D., gdzie

$$m = EX, \quad \sigma^2 = var(X) > 0.$$

Jak wcześniej, niech dla $n \geq 1$, X_n oznacza kopię X, $d(X_n) = d(X)$, oraz $(X_n)_{n \geq 1}$ ci. niezależn.

Bierzemy:

1) $Y_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ – uśrednienie

Wtedy $EY_n = m$, $\operatorname{var}(Y_n) = \frac{1}{n^2} \cdot n \operatorname{var}(X) = \frac{\sigma^2}{n}$

2) Standaryzujemy Y_n:

$$Z_n = \frac{Y_n - EY_n}{\sqrt{\operatorname{var} Y_n}} = \frac{Y_n - m}{\frac{\sigma}{\sqrt{n}}}$$

Tw. (C.T.G.)

Przy założeniach j.w.

$$Z_n \xrightarrow{d} N, \quad (N \in N(0,1))$$

czyli

$$F_{Z_n}(t) = P(\{\omega \in \Omega : Z_n(\omega) < t\}) \longrightarrow \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-u^2/2} du$$

dla $t \in \mathbb{R}$.

Uwagi:

1°. Dla dostatecznie dużych $n \geq 2$, mamy przybliżenie

$$\mathcal{L}(Z_n) \approx \Phi$$

2°. Sprawdźmy, co to oznacza dla X:

$$F_{Z_n}(t) = P\left(\left\{\omega\in\Omega : \frac{Y_n(\omega)-m}{\frac{\sigma}{\sqrt{n}}} < t\right\}\right) =$$

$$\downarrow$$

$$\Phi(t) \qquad = P\left(\left\{\omega\in\Omega : Y_n(\omega) < m + \frac{t\sigma}{\sqrt{n}}\right\}\right)$$

$$\qquad\qquad\qquad\qquad\qquad \| $$

$$\qquad\qquad\qquad\qquad F_{Y_n}\left(m + \frac{t\sigma}{\sqrt{n}}\right)$$

Czyli $\underline{F_{Y_n}\left(m + \frac{t\sigma}{\sqrt{n}}\right) \approx \Phi(t)}$, (*)

Niech $u = m + \frac{t\sigma}{\sqrt{n}} \Rightarrow t = \frac{u-m}{\frac{\sigma}{n}}$, mamy

$$F_{Y_n}(u) \approx \Phi\left(\frac{u-m}{\frac{\sigma}{\sqrt{n}}}\right). \qquad (**)$$

Ale

$$\Phi\left(\frac{u-m}{\frac{\sigma}{\sqrt{n}}}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{u-m}{\frac{\sigma}{\sqrt{n}}}} e^{-t^2/2} \, dt =$$

podstawmy:

$$t = \frac{u-m}{\frac{\sigma}{\sqrt{n}}} \qquad dt = \frac{1}{\frac{\sigma}{\sqrt{n}}} du = \frac{\sqrt{n}}{\sigma} du$$

$$= \frac{1}{\sqrt{2\pi}} \frac{\sqrt{n}}{\sigma} \int_{-\infty}^{t} e^{-\frac{\left(\frac{u-m}{\frac{\sigma}{\sqrt{n}}}\right)^2}{2}} du =$$

$$= \underbrace{\frac{1}{\sqrt{2\pi \left(\frac{\sigma^2}{n}\right)}} \int_{-\infty}^{t} e^{-\frac{(u-m)^2}{2\frac{\sigma^2}{n}}} du}_{\text{dystrybuanta rozkładu } N\left(m, \frac{\sigma^2}{n}\right)}.$$

Zatem $F_{\bar{Y}_n} \approx N\left(m, \frac{\sigma^2}{n}\right)$

Przykład: I wersja C.T.G. – tw. Moivre'a–Laplace'a.

Dla $n \geq 2$, $B_n \in B(n,p)$.

Oszacujmy pr. zdarzenia

$$A_{a,b,n} = \{ \omega \in \Omega : a \leq B_n(\omega) < b \}, \quad a < b$$

Dostajemy kolejno:

uśrednienie

$$P(A_{a,b,n}) = P\left(\{\omega \in \Omega : \frac{a}{n} \leq \frac{B_n(\omega)}{n} < \frac{b}{n}\}\right)$$

$$E\left(\frac{B_n}{n}\right) = p, \quad \operatorname{var}\left(\frac{B_n}{n}\right) = \frac{p \cdot q}{n}$$

centrowanie i normowanie = standaryzacja

$$= P\left(\{\omega \in \Omega : \frac{\frac{a}{n} - p}{\sqrt{\frac{pq}{n}}} \leq \frac{\frac{B_n(\omega)}{n} - p}{\sqrt{\frac{pq}{n}}} < \frac{\frac{b}{n} - p}{\sqrt{\frac{pq}{n}}}\}\right)$$

$$\simeq \Phi\left(\frac{\frac{b}{n} - p}{\sqrt{\frac{pq}{n}}}\right) - \Phi\left(\frac{\frac{a}{n} - p}{\sqrt{\frac{pq}{n}}}\right)$$

— w takim sformułowaniu pojawiło się w XVIII w. I wersja C.T.G.

Symulacja:

Niech $n=100$, $a=30$, $b=70$, $p=0,25$

Wtedy

$$\mathbb{P}(A_{a,b,n}) \approx \Phi(10,3923) - \Phi(1,1547)$$

$$\approx 1-\Phi(1,15747) \approx$$

$$\underline{\underline{1-0,8943}} = \underline{\underline{0,1057}}.$$

Uwaga.

Oszacujmy teraz $P(\{\omega\in\Omega : B_n(\omega)=k\})$, np. $k=30$.

Zauważmy, że $(n=100)$

$$P(\{\omega\in\Omega: B_{100}(\omega)=30\})=$$

$$= P(\{\omega\in\Omega: 29,5 \le B_{100}(\omega) < 30,5\})$$

$$\approx \Phi(1,2701) - \Phi(1,03923)$$

$$= \underline{0,0495}$$

Z drugiej strony $P(\{\omega\in\Omega: B_{100}(\omega)=30\})=$

$$= \binom{100}{30}(0,25)^{30}(0,75)^{70}$$

Ale $\binom{100}{30} = 2{,}9372 \cdot 10^{25}$

$(0{,}25)^{30} = 8{,}6736 \cdot 10^{-19}$

$(0{,}75)^{70} = 1{,}7959 \cdot 10^{-9}$

Zatem $P(B_{100}=30) = \underline{\underline{0{,}045}}$

Odchyłka wynosi

$|0{,}0495 - 0{,}045| = \underline{\underline{0{,}004}}$

Zad 2.

Oprogramować te procedury!