Prawa wielkich liczb
Centralne Twierdzenie Graniczne

Dr hab. Marzena Nowakowska, prof. PŚk
Wydział Zarządzania i Modelowania Komputerowego

Obszar odniesienia PWL

Prawa wielkich liczb (PWL) to zbiór twierdzeń z matematycznych, które opisują związek między liczbą wykonywanych doświadczeń, a faktycznym prawdopodobieństwem wystąpienia zdarzenia, którego te doświadczenia dotyczą.

Najwcześniejsza postać PWL to prawo Bernoulliego. Opublikowane pośmiertnie w 1713 w książce „Ars conjectandi” (Sztuka zgadywania) przez siostrzeńca Jakoba, Nicolasa, jako „złote twierdzenie” (https://www.tomaszgrebski.pl/blog/matematycy-znaczki-historie/jacob-bernouuli):

> „Z prawdopodobieństwem dowolnie bliskim 1 można się spodziewać, iż przy dostatecznie wielkiej liczbie prób *n* częstość *k* danego zdarzenia losowego będzie się dowolnie mało różniła od jego prawdopodobieństwa *p*.”

$$\forall \varepsilon > 0 \; \lim_{n\to\infty} P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = 1$$

Jacob Bernoulli, szwajcarski matematyk i fizyk, 1654-1705

Inne (niektóre) prawa wielkich liczb:

Prawo wielkich liczb Markowa
Prawo wielkich liczb Czebyszewa
Mocne prawo wielkich liczb Bernouliego
Prawo wielkich liczb Chińczyna
Pierwsze prawo wielkich liczb Kołmogorowa
Drugie prawo wielkich liczb Kołmogorowa
CTG – Centralne Twierdzenie Graniczne

CTG → Centralne Twierdzenie Graniczne

Jedno z najważniejszych twierdzeń rachunku prawdopodobieństwa, uzasadniające powszechność występowania rozkładów zbliżonych do rozkładu normalnego.

CTG Lindberga-Levy'ego

Dany jest ciąg X_k niezależnych zmiennych losowych o takim samym rozkładzie, takich że $E(X_k) = m$, $V(X_k) = \sigma^2$ dla każdego k (wartość oczekiwana m i wariancja σ^2 istnieją i są skończone). Wtedy dla dowolnego b:

$$\lim_{n\to\infty} P\left(\frac{X_1 + \ldots + X_n - n\cdot m}{\sigma\cdot\sqrt{n}} \le b\right) = \Phi(b)$$

gdzie Φ jest dystrybuantą rozkładu N(0, 1).

Twierdzenie mówi, że rozkład zmiennej losowej $\frac{X_1 + \ldots + X_n - n\cdot m}{\sigma\cdot\sqrt{n}}$

(zestandaryzowanej sumy zmiennych losowych $X_1, \ldots, X_n$)

jest asymptotycznie równy standardowemu rozkładowi normalnemu.

$\color{blue}{\text{CTG}}$ → Twierdzenie Graniczne Lapunowa

Dany jest ciąg X_k niezależnych zmiennych losowych o dowolnych rozkładach.

Przyjmując, że dla każdego k: $Z_k^3 = E(|X_k - m_k|^3) < \infty$

oraz, że dla oznaczeń:

$$\sigma_k^2 = V(X_k) \qquad \sigma_n = \sqrt[2]{\sigma_1^2 + \cdots + \sigma_n^2} \qquad Z_n = \sqrt[3]{Z_1^3 + \cdots + Z_n^3}$$

zachodzi warunek: $\lim\limits_{n\to\infty} \frac{Z_n}{\sigma_n} = 0$

to
$$\lim_{n\to\infty} P\left(\frac{X_1 + \ ... + \ X_n - (m_1 + \cdots + m_n)}{\sqrt[2]{\sigma_1^2 + \cdots + \sigma_n^2}} \le b\right) = \Phi(b)$$

gdzie Φ jest dystrybuantą rozkładu N(0, 1).

Twierdzenie mówi, że rozkład zmiennej losowej $\frac{X_1 + \ ... + \ X_n - (m_1 + \cdots + m_n)}{\sqrt[2]{\sigma_1^2 + \cdots + \sigma_n^2}}$
jest asymptotycznie równy standardowemu rozkładowi normalnemu.

CTG - idea

Nazwa **CTG** podkreśla doniosłą rolę, jaką twierdzenie (wspólna nazwa dla twierdzeń wielu autorów) odgrywa w rachunku prawdopodobieństwa i statystyce matematycznej.

Jeżeli zmienne losowe X_1, X_2,...,X_n, są niezależne i żadna z nich nie przyjmuje zbyt dużych wartości, to suma zmiennych losowych $X_1+X_2+...+X_n$ (i przeciętna wartość tych zmiennych) ma rozkład normalny lub zbliżony do rozkładu normalnego, gdy liczba zmiennych jest odpowiednio duża.
To oznacza, że jeżeli liczba składowych zmiennych losowych tworzących sumę (nową zmienną losową) jest dostatecznie duża, to rozkład tej zmiennej można aproksymować rozkładem normalnym.

Twierdzenie **Lapunowa** wyjaśnia z teoretycznego punktu widzenia dobrze znany z doświadczenia fakt, że empiryczne zmienne losowe mają na ogół rozkład zbliżony do normalnego. Spotykane w praktyce zmienne losowe mogą być przeważnie traktowane jako sumy znacznej liczby zmiennych losowych, z których żadna nie ma dominującego wpływu na wielkość tej sumy.

CTG - znaczenie

Nawet gdy liczba rozpatrywanych zmiennych jest tylko umiarkowanie duża, to jeżeli żadna z nich nie dominuje nad pozostałymi i o ile nie są w wysokim stopniu zależne, rozkład ich sumy będzie bliski normalnemu.

CTG zachodzi dla większości zmiennych losowych opisujących zjawiska fizyczne:
(1) jeżeli są niezależne i mają jednakowe rozkłady,
(2) jeżeli zmienne są niezależne ale nie mają jednakowych rozkładów,
(3) jeżeli nie są niezależne ale korelacja między jedną zmienną a każdą skończoną liczbą pozostałych zmiennych jest równa zero.

Olbrzymie znaczenie praktyczne rozkładu normalnego polega na tym, że CTG można sformułować bez dokładnej znajomości:
(1) rozkładów brzegowych składowych zmiennych losowych,
(2) ich liczby,
(3) ich rozkładu łącznego.