Matematyka dla biologów — Zajęcia nr 13.

Dariusz Wrzosek

15 stycznia 2024

Plan:

1. Rachunek prawdopodobienstwa-zmienne losowe o rozkładzie ciągłym
2. Prawo wielkich liczb
3. Rozkład jednostajny
4. Rozkład wykładniczy
5. Rozkład normalny
6. Centralne twierdzenie graniczne

Przypomnienie – zmienna losowa o rozkładzie dyskretnym

Niech $(\Omega\,, P)$ będzie przestrzenią probabilistyczną, a zmienna losowa $X : \Omega \to \mathbb{R}$ przyjmuje wartości dyskretne (zbiór wartości jest skończony)

$$x_1\,, x_2\,, \ldots x_n\,.$$

Rozkładem zmiennej losowej X nazywamy zbiór R_X par , z których każda określa z jakim prawdopodobieństwem zmienna losowa przyjmuje daną wartość

$$R_X = \{(x_1\,, p_1)\,, (x_2\,, p_2) \ldots (x_n\,, p_n)\}$$

gdzie

$$p_i = P(\{\omega : X(\omega) = x_i\}) \text{ lub w skróconym zapisie } p_i = P(X = x_i).$$

Dla uproszczenia zapisu pisze się zwykle

$$P(a \leqslant X \leqslant b)$$

zamiast

$$P(\{\omega : a \leqslant X(\omega) \leqslant b\})$$

dla określenia prawdopodobieństwa tego, że zmienna losowa X przyjmie wartości z przedziału $[a\,, b]$

Wtedy dla odcinka $[a, b] \subset \mathbb{R}$

$$P(a \leqslant X \leqslant b) = \Sigma_{\{i : x_i \in [a,b]\}}\, p_i$$

czyli sumujemy tylko te prawdopodobieństwa p_i dla których $x_i \in [a, b]$

Zmienne losowe o rozkładzie ciągłym w biologii

- W biologii w naturalny sposób natrafiamy na zmienne losowe o rozkładzie ciągłym rozpatrując takie cechy osobników w populacji jak masa lub średnica ciała czy ciśnienie krwi itp.
- Wtedy zmierzoną wartość danej cechy u losowo wybranego osobnika, będącą liczbą rzeczywistą, uznaje się jako realizację zmiennej losowej o rozkładzie ciągłym.
- Tym w jaki sposób na podstawie kilku-kilkunastu losowych próbkowań wyciągnąć wiarygodny wniosek o rozkładzie dla całej populacji zajmuje się statystyka.

Rozkład ciągły zmiennej losowej

Definicja

Zmienna losowa $X : \Omega \to \mathbb{R}$ *ma* **rozkład ciągły** *jeśli istnieje funkcja nieujemna* $f : \mathbb{R} \mapsto [0\,, +\infty)$ *taka, że dla dowolnych* $a\,, b \in \mathbb{R}$

$$P(a \leqslant X \leqslant b) = \int_a^b f(x)dx$$

oraz

$$\int_{-\infty}^{+\infty} f(x)dx = 1\,.$$

Funkcję f *nazywa się wtedy* **gęstością zmiennej losowej**.

Całka powyżej to całka niewłaściwa.

Dystrybuantą zmiennej losowej o gęstości f nazywa się funkcję

$$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(x)dx\,,$$

określa ona prawdopodobieństwo tego, że zmienna losowa X przyjmie wartość mniejszą niż x. Z definicji wynika, że dystrybuanta zmiennej losowej o rozkładzie ciągłym jest funkcją funkcją ciągłą i

$$\lim_{x \to -\infty} F(x) = 0\,, \quad \lim_{x \to +\infty} F(x) = 1\,.$$

Zauważmy, że dystrybuanta jest funkcją pierwotną funkcji gęstości i że dystrybuanta jednoznacznie wyznacza rozkład zmiennej losowej (bo F'(x)=f(x)).

Prawdopodobieństwo tego, że zmienna losowa o rozkładzie ciągłym przyjmuje punktową wartość $x_0 \in \mathbb{R}$ jest równa zeru bo można ją przedstawić jako granicę

$$\lim_{n\to+\infty} P(x_0 - \frac{1}{n} \leqslant X \leqslant x_0 + \frac{1}{n}) = \lim_{n\to+\infty} \int_{x_0-\frac{1}{n}}^{x_0+\frac{1}{n}} f(x)dx$$
$$= \lim_{n\to+\infty} \left[F(x_0 + \frac{1}{n}) - F(x_0 - \frac{1}{n}) \right] = F(x_0) - F(x_0) = 0\,.$$

Przedostatnia równość wynika z faktu, że skoro F ma pochodną, to jest funkcją ciągłą

Zdefiniujemy teraz wartość oczekiwaną i wariancję zmiennej losowej o rozkładzie ciągłym

Definicja

Wartością oczekiwaną zmiennej losowej X o gęstości f nazywamy liczbę

$$EX = \int_{-\infty}^{+\infty} xf(x)dx\,.$$

Wariancję definiuje się jako

$$D^2X = E((X - EX)^2) = \int_{-\infty}^{+\infty} (x - EX)^2 f(x)dx\,.$$

Zwróćmy uwagę, że zarówno wartość oczekiwana jaki wariancja mogą nie istnieć jeśli nie istnieją odpowiednie całki. Można podać takie przykłady ale nie są one elementarne.

Powstaje naturalne pytanie– jakie jest prawdopodobieństwo tego, że wartości zmiennej losowej będą odbiegać znacznie od wartości oczekiwanej. Ogólną odpowiedź na to pytanie daje słynna **nierówność Czebyszewa** (P. Czebyszew (1821-1894)), której dotyczy następujące twierdzenie

Twierdzenie

Dla dowolnej zmiennej losowej X posiadającej wartość oczekiwaną i wariancję i dowolnej liczby $t > 0$ zachodzi nierówność

$$P(|X - EX| \geqslant t) \leqslant \frac{D^2X}{t^2}. \tag{1}$$

Podstawiając np. $t = 3\sqrt{D^2X}$ otrzymujemy wniosek, że prawdopodobieństwo, tego, iż odległość pomiędzy wartością, którą przyjmie zmienna losowa X i jej wartością oczekiwaną trzykrotnie przekracza wielkość dyspersji jest mniejsze niż $\frac{1}{9}$, czyli stosunkowo małe. To oszacowanie nie wykorzystuje żadnej konkretnej informacji o zmiennej losowej X i dlatego nie jest zbyt precyzyjne.

Dowód twierdzenia Czebyszewa jest bardzo prosty i dlatego prezentujemy go poniżej.

Dowód twierdzenia Czebyszewa. Przyjmijmy, że pewna zmienna losowa Y, która przyjmuje tylko wartości nieujemne, ma gęstość f . Zatem $f(x) = 0$ dla $x \leqslant 0$. Udowodnimy najpierw, że wtedy dla dowolnej liczby $\varepsilon > 0$

$$P(Y \geqslant \varepsilon) \leqslant \frac{EY}{\varepsilon} .$$

Jest tak dlatego, że

$$EY = \int_0^{+\infty} xf(x)dx \geqslant \int_\varepsilon^{+\infty} xf(x)dx \geqslant \varepsilon \int_\varepsilon^{+\infty} f(x)dx = \varepsilon P(Y \geqslant \varepsilon) .$$

Wstawiając teraz $Y = (EX - X)^2$ i $\varepsilon = t^2$ otrzymujemy (1). Dowód w przypadku ogólnym, gdy nie zakłada się istnienia gęstości zmiennej losowej jest bardzo podobny.

Zastosowanie nierówności Czebyszewa

Zadanie. Rzucamy 10000 razy monetą. Jakie jest prawdopodobieństwo, że liczba sukcesów S_{10000} będzie różna od 5000 o więcej niż 150?

Odp. Interesuje nas prawdopodobieństwo zdarzenia, że $(|S_{10000} - 5000| \geqslant 150)$ czyli $(|\frac{S_{10000}}{10000} - \frac{1}{2}| \geqslant 0,015)$, zatem stosując nierówność Czebyszewa dostajemy

$$P\left(\left|\frac{S_{10000}}{10000} - \frac{1}{2}\right| \geqslant 0,015\right) \leqslant \frac{0,25}{(0,015)^2 \cdot 10000} = \frac{1}{9}\,.$$

Prawo wielkich liczb (PWL)

Z twierdzenia Czebyszewa wynika jedno z najbardziej znanych twierdzeń rachunku prawdopodobieństwa sformułowane w najprostszej wersji już przez J. Bernoulliego (1655-1705). W praktyce mówi ono, że w dostatecznie długiej serii powtórzeń doświadczenia w schemacie Bernoulliego o prawdopodobieństwie sukcesu p, z prawdopodobieństwem dowolnie bliskim 1 częstość uzyskania sukcesu $\frac{S_n}{n}$ równa jest p z dowolnie zadaną dokładnością.

Przejdźmy teraz do precyzyjnego sformułowania.

PWL

Twierdzenie

Niech $X_1, X_2, \ldots X_n$ *będzie ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie. Oznaczmy przez* $S_n = \sum_{i=1}^{n} X_i$. *Załóżmy ,że dla każdego* $i = 1, \ldots, n$, $EX_i = m$ *oraz* $D^2 X_i = \sigma^2$. *Wtedy dla dowolnego* $\varepsilon > 0$

$$\lim_{n \to +\infty} P\left(\left|\frac{S_n}{n} - m\right| > \varepsilon\right) = 0 \,. \tag{2}$$

Co więcej,

$$P\left(\left|\frac{S_n}{n} - m\right| \leqslant \varepsilon\right) > 1 - \frac{\sigma^2}{\varepsilon^2 n} \tag{3}$$

Pierwsza część tego twierdzenia zwana jest **prawem wiekich liczb**.

Dowód. Skorzystamy z twierdzenia Czebyszewa. W tym celu połóżmy $X = S_n$. Wtedy $EX = nm$ oraz $D^2X = n\sigma^2$. Stosując (1) otrzymujemy

$$P\left(|S_n - nm| > t\right) \leqslant \frac{n\sigma^2}{t^2}\,.$$

Wstawiając $t = \varepsilon n$ otrzymujemy

$$P\left(\left|\frac{S_n}{n} - m\right| > \varepsilon\right) \leqslant \frac{\sigma^2}{\varepsilon^2 n}\,.$$

Przechodząc do granicy z $n \to \infty$ otrzymujemy (2). Korzystając z tego, że $P(A) = 1 - P(\Omega \setminus A)$ gdzie $A \subset \Omega$ to dowolne zdarzenie w Ω. możemy zapisać, że

$$P\left(\left|\frac{S_n}{n} - m\right| > \varepsilon\right) = 1 - P\left(\left|\frac{S_n}{n} - m\right| \leqslant \varepsilon\right)\,,$$

a stąd już wynika (3).

Przykładowe rozkłady ciągłe- rozkład jednostajny

Rozważmy zmienną losową X_J, która przyjmuje wartości jedynie w odcinku $[a\,,b]$ o tej własności, że prawdopodobieństwo że przyjmie wartości w odcinku $A=[x_1\,,x_2]\subset[a\,,b]$ nie zależy od jego położenia a jedynie od jego długości i wynosi $\frac{|x_2-x_1|}{b-a}$. Gęstość tej zmiennej losowej wyznacza funkcja

$$f_J(x)=\begin{cases}\frac{1}{b-a} & \text{gdy} \quad x\in[a\,,b]\,,\\ 0 & \text{gdy} \quad x\notin[a\,,b]\,.\end{cases}$$

Jeżeli $x_1\,,x_2\in[a\,,b]$ to

$$P(x_1\leqslant X_J\leqslant x_2)=\int_{x_1}^{x_2}f_J(x)dx=\frac{|x_2-x_1|}{b-a}\,.$$

Jako proste ćwiczenie pozostawiamy sprawdzenie, że

$$EX_J=\frac{a+b}{2}\,,\quad D^2X_J=\frac{1}{12}(b-a)^2\,.$$

Przykład zastosowania rozkładu jednostajnego

Otrzymujemy informację, że przy prawym brzegu rzeki ukryto skarb, który znajduje się gdzieś pomiędzy dwoma mostami. Poszukiwania odbywają się stopniowo , odcinkami. Prawdopodobieństwo znalezienia skarbu na danym odcinku rzeki jest wtedy równe stosunkowi długości przeszukiwanego odcinka rzeki do odległości pomiędzy mostami.

Rozkład wykładniczy

Zmienna losowa X_w przyjmująca wartości nieujemne ma rozkład wykładniczy gdy jej gęstość zadana jest przez funkcję

$$f_w(x) = \lambda e^{-\lambda x}, \quad x \geqslant 0 .$$

Wtedy

$$EX_w = \frac{1}{\lambda}, \quad D^2 X_w = \frac{1}{\lambda^2} .$$

Łatwo sprawdzić, że

$$P(X_w > a + b | X_w > b) = P(X_w > a),$$

bo

$$\frac{P(X_w > a + b)}{P(X_w > b)} = \frac{\int_{a+b}^{+\infty} \lambda e^{-\lambda x} dx}{\int_{b}^{+\infty} \lambda e^{-\lambda x} dx} = \frac{e^{-\lambda(a+b)}}{e^{-\lambda b}} = e^{-\lambda a} = P(X_w > a) .$$

Czas oczekiwania na zdarzenie

Ta ostatnia własność powoduje, że najczęściej zmienna losowa o rozkładzie wykładniczym bywa interpretowana jako czas oczekiwania na jakieś zdarzenie jeśli można przyjąć, że może ono zajść w każdej chwili i fakt, że czeka się już jakiś czas na zajście zdarzenia nie wpływa na to jak długo jeszcze będzie się czekać.(**tzw. efekt braku pamięci**).

Rozkład wykładniczy stosuje się do szacowania czasu oczekiwania na

– emisję cząstki z materiału promieniotwórczego (rozpad promieniotwórczy),

– autobus miejski (było to dobre oszacowanie w czasach PRL),

– zajście mutacji w danym odcinku DNA.

Ale także w pewnych sytuacjach do opisu czasu przeżycia organizmu (lub niezawodnej pracy urządzenia) , jeśli średni czas życia wynosi $\frac{1}{\lambda}$, a śmierć może przyjść w każdej chwili.

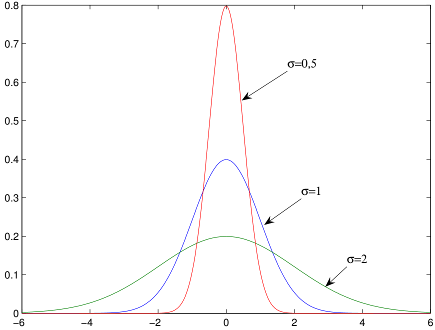
Rozkład normalny

Istotną rolę w rachunku prawdopodobieństwa pełni **rozkład normalny zwany także rozkładem Gaussa**. Oznacza się go zwyczajowo przez $N(\mu\,,\sigma)$ i jest on zadany przez funkcję gęstości, którą oznaczymy przez $f_{N(\mu\,,\sigma)}(x)$:

$$f_{N(\mu\,,\sigma)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Jej wykres to charakterystyczna krzywa o dzwonowatym kształcie, symetryczna względem $x = \mu$. Wiadomo, że jeśli zmienna losowa X ma rozkład $N(\mu\,,\sigma)$ to

$$EX = \mu\,, \qquad \text{a dyspersja} \qquad \sqrt{VarX} = \sigma\,.$$

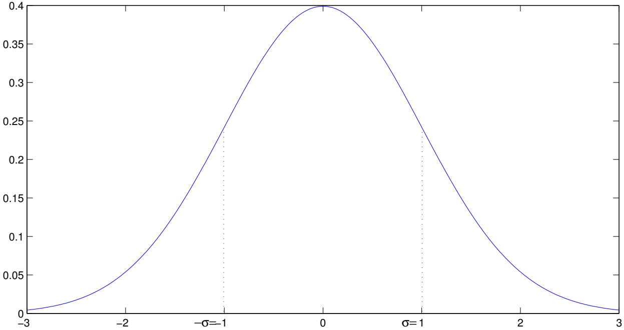


Krzywe Gaussa, gęstości rozkładu normalnego dla $\mu = 0$ różnych σ .

Funkcja będąca gęstością rozkładu normalnego $N(\mu\,,\sigma)$ ma dwa punkty przegięcia o współrzędnych $\mu - \sigma$ i $\mu + \sigma$. Można się o tym przekonać przyrównując do zera drugą pochodną funkcji $f_{N(\mu\,,\sigma)}(x)$.

Rozkład zmiennej losowej nazywa się rozkładem standardowym normalnym jeśli zadany jest przez $N(0\,,1)$. Zauważmy, że

$$f_{N(\mu\,,\sigma)}(x) = \frac{1}{\sigma} f_{N(0\,,1)}\left(\frac{x-\mu}{\sigma}\right) .$$



Gęstość rozkładu normalnego standardowego $N(0,1)$ z zaznaczonymi punktami przegięcia w punktach -1 i 1.

Oznaczmy przez $\bar{X}$ zmienną losową o rozkładzie $N(0\,,1)$. W tablicach matematycznych można znaleźć wartości dystrybuanty standardowego rozkładu normalnego $N(0\,,1)$ oznaczanej zwykle przez $\Phi(x)$

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=P(\bar{X}\leqslant x).$$

Stąd wynika, że

$$P(a\leqslant \bar{X}\leqslant b)=\int_{a}^{b}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\Phi(b)-\Phi(a)\,.$$

Tę dystrybuantę wykorzystuje się do obliczania w praktyce wartości prawdopodobieństw zdarzeń takich, że zmienna X o rozkładzie normalnym $N(\mu\,,\sigma)$ przyjmuje wartości w jakimś przedziale np. $[a\,,b]$.

Korzystając z deficji całki oznaczonej i definicji dystrybuanty obliczamy, że

$$
\begin{aligned}
P(a \leqslant X \leqslant b) = \int_a^b f_{N(\mu\,,\sigma)}(x)dx = \int_a^b \frac{1}{\sigma} f_{N(0\,,1)}\left(\frac{x-\mu}{\sigma}\right)dx \\
= \quad \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) = P\left(\bar{X} \leqslant \frac{b-\mu}{\sigma}\right) - P\left(\bar{X} \leqslant \frac{a-\mu}{\sigma}\right) \\
= \quad P\left(\frac{a-\mu}{\sigma} \leqslant \bar{X} \leqslant \frac{b-\mu}{\sigma}\right).
\end{aligned}
$$

Podobnie można sprawdzić, że jeśli X jest zmienną losową o rozkładzie normalnym $N(\mu\,,\sigma)$ to zmienna losowa

$$\frac{X-\mu}{\sigma} \quad \text{ma rozkład normalny standardowy} \quad N(0\,,1)$$

Tę procedurę nazywa się **standaryzacją zmiennej losowej**. Dla przykładu obliczmy korzystając z tablic wartości funkcji Φ

$$P(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) = \Phi(2) - \Phi(-2) = 0,954\,. \tag{4}$$

Wynika stąd ważny wniosek: zdarzenie, że zmienna losowa o rozkładzie normalnym $N(\mu\,,\sigma)$ przyjmuje wartości w przedziale $[\mu - 2\sigma\,,\mu + 2\sigma]$ jest niemal pewne bo wynosi $0,954$.

W przypadku zmiennych o rozkładzie ciągłym definicja niezależności zmiennych losowych jest nieco bardziej skomplikowana ale wyraża w istocie to samo co w przypadku zmiennych o rozkładzie dyskretnym.

Jeśli X_1 i X_2 są niezależnymi zmiennymi losowymi o rozkładzie normalnym odpowiednio $N(\mu_1, \sigma_1)$ i $N(\mu_2, \sigma_2)$ to można udowodnić, że kombinacja liniowa $\alpha_1 X_1 + \alpha_2 X_2$ jest zmienną losową również o rozkładzie normalnym $N(\mu, \sigma)$ gdzie $\mu = \alpha_1 \mu_1 + \alpha_2 \mu_2$ i $\sigma = \sqrt{\alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2}$.

Jest wiele innych ciągłych rozkładów zmiennych losowych ważnych w statystyce przy weryfikacji hipotez. Należą do nich rozkład χ^2 i rozkład t-studenta.

Centralne twierdzenie graniczne

Poniższe twierdzenie zwane centralnym twierdzeniem granicznym jest kluczowe w rachunku prawdopodobieństwa i w statystyce. Mówi ono, że rozkład sum wartości niezależnych doświadczeń losowych po standaryzacji jest zbieżny, przy liczbie powtórzeń n dążącej do $+\infty$, do standardowego rozkładu normalnego. Warto tu podkreślić, że zmienne losowe określające wyniki kolejnych doświadczeń mają wszystkie ten sam rozkład, który jest **dowolny !** byleby miał on skończoną wartość oczekiwaną i wariancję.

Centralne twierdzenie graniczne CTG

Twierdzenie

Niech X_k *będzie ciągiem niezależnych zmiennych losowych o takim samym rozkładzie, przy czym* $EX_k = \mu$ *oraz* $VarX_k = \sigma^2$ *i niech* $S_n = \sum_{i=1}^{n} X_i$*. Wtedy dla dowolnych ustalonych* $a, b \in \mathbb{R}$ *zachodzi wzór*

$$P\left(a \leqslant \frac{S_n - n\mu}{\sigma\sqrt{n}} \leqslant b\right) \xrightarrow{n \to +\infty} \Phi(b) - \Phi(a)\,.$$

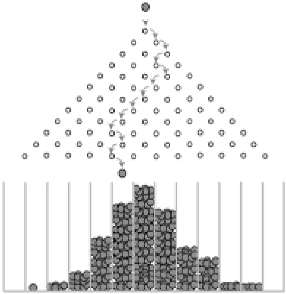
.

Dowodu nie przedstawiamy bo wymaga zastosowania zaawansowanych narzędzi matematycznych.

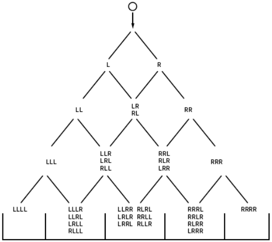
Jeśli X_k reprezentuje sukces w k-tym doświadczeniu Bernoulliego $P(X_k = 1) = p$ to wtedy

$$P\left(\left|\frac{S_n - np}{\sqrt{npq}}\right| \leqslant \gamma\right) \xrightarrow{n \to +\infty} \Phi(\gamma) - \Phi(-\gamma) = \Phi(\gamma) - (1 - \Phi(\gamma)) = 2\Phi(\gamma) - 1\,.$$

Zamiast dowodu CTG deska Galtona



Zamiast dowodu CTG deska Galtona



Położenie kulki na n-tym poziomie określone jest przez $\sum_{k=1}^{n} X_k$ gdzie

$$R_{X_k} = \{(-1, 1/2), (1, 1/2)\}, \quad EX_k = 0, VarX_k = 1$$

Aby zastosować to twierdzenie wybieramy korzystając z tablic wartość γ tak aby wyrażenie $\Phi(\gamma) - \Phi(-\gamma)$ miało wartość bliską 1 np. 0.95. Zauważmy, że

$$P\left(\left|\frac{S_n - np}{\sqrt{npq}}\right| \leqslant \gamma\right) = P\left(\left|\frac{S_n}{n} - p\right| \leqslant \frac{\gamma\sqrt{pq}}{\sqrt{n}}\right)$$

Po dokonaniu tylu prób n, że $\sqrt{n} > \frac{\gamma\sqrt{pq}}{0.001}$ prawdopodobieństwo tego, że średnia z n prób różni się od p o mniej niż 0.001 czyli

$$\left|\frac{S_n}{n} - p\right| \leqslant 0.001$$

wynosi 0.95.

Przykłady i zadania: Częstość względna a prawdopodobieństwo

Jeżeli nie mamy dostępu do tablic z wartościami dystrybuanty rozkładu normalnego (brak tablic lub komputera) to pozostaje do dyspozycji PWL. Ogólnie CTG daje dokładniejsze oszacowania liczby prób, które trzeba wykonać aby z dużym prawdopodobieństwem $\frac{S_n}{n}$ było blisko p .

Powiedzmy, że ciąg niezależnych zmiennych losowych $X_1 , X_2 , \ldots X_n$ jest ciągiem prób Bernoulliego takim, że każda ze zmiennych losowych określona jest na przestrzeni Ω i przyjmuje wartość 1 gdy zaszło zdarzenie $A \subset \Omega$ z prawdopodobieństwem $P(A)$ i wartość 0 w przeciwnym przypadku. Wtedy oczywiście dla każdego k mamy $EX_k = P(A)$ oraz $D^2 X_k = P(A) - P(A)^2$.

Zastosowanie PWL

Zauważmy,że częstość względną występowania zdarzenia A reprezentuje zmienna $\frac{S_n}{n}$,

1.

gdyż w liczniku jedynka występuje dokładnie tyle razy ile razy zaszło zdarzenie A w trakcie n prób. Można zadać pytanie

Jak wielkie powinno być n aby z prawdopodobieństwem większym niż $0,95$ częstość względna wystąpienia zdarzenia A różniła się od $P(A)$ o mniej niż $,01$?.

Skorzystamy z PWL podstawiając $\varepsilon = 0,01$, $\sigma^2 = P(A)(1 - P(A))$ i z nierówności $1 - \frac{\sigma^2}{\varepsilon^2 n} > 0,095$, po przekształceniu i podstawieniu otrzymujemy, że

$$n > \frac{P(A)(1 - P(A))}{0,00005}.$$

Jeśli $P(A) = 0,5$ to $n > 5000$. Warto tu zaznaczyć, że to oszacowanie

Zadanie: Czas przeżycia

Przyjmujemy, że średni czas życia organizmu (nie stosuje się do populacji ludzkich) wynosi β jednostek czasu. Innymi słowy jest to średni czas oczekiwania na śmierć. Znaleźć prawdopodobieństwo przeżycia do czasu β posługując się rozkładem wykładniczym.

Przyjmujemy zatem, że zmienna losowa T ma rozkłada wykładniczy i określa moment śmierci, $\lambda = \frac{1}{\beta}$, i obliczamy prawdopodobieństwo tego, że śmierć nastąpi po czasie s

$$P(T \geqslant s) = \int_s^{+\infty} \frac{1}{\beta} e^{-\frac{t}{\beta}} dt = e^{-\frac{s}{\beta}}.$$

czyli $P(T \geqslant \beta) = e^{-1} \approx 0,37$.

To jest najprostszy model przeżywalności, jest wiele modeli bardziej realistycznych i zaawansowanych.