Kurs: SSP (SPC)

Wykład 3

Temat: Elementy teorii prawdop. i stat. matem. w zastosowaniach SPC c.d.

4. Parametry zmiennej losowej i ich znaczenie

4a Wartość oczekiwana, czyli średnia rozkładu.

Założenia. Dla ZL mamy (Ω, Σ, P) jego MPK oraz

$X : \Omega \to \mathbb{R}$ dana zmienna losowa o rozkładzie F.

(i) Przypadek r. dyskretnego.

Niech F skokowa, czyli $X(\Omega) = \{x_n, n \in \mathbb{N}_0\}$, $\mathbb{N}_0 \subset \mathbb{N}$

$$p_n = P(\{\omega \in \Omega : X(\omega) = x_n\}), \quad n \in \mathbb{N}_0$$

Bierzemy ciąg $(x_n p_n)_{n \in \mathbb{N}_0}$. Jeśli jest on sumowalny, czyli szereg $\sum_{n \in \mathbb{N}_0} x_n p_n$ jest zbieżny, to jego sumę oznaczamy przez EX, gdzie

(1) $$EX = \sum_{n \in \mathbb{N}_0} x_n p_n$$

i nazywamy wartością oczekiwaną zm. l. X (rozkładu)

Uwaga 1.

Gdy X ma rozkład skokowy, czyli

$X(\Omega) = \{x_1, x_2, \ldots, x_n\}$, to EX zawsze istnieje.

Wtedy $EX = \sum_{k=1}^{n} x_k p_k$ nazywamy jest średnią ważoną (p_k – wagi). Dlatego EX nazywany jest średnią i oznaczamy $m = EX$ (mean value).

W szczególności, gdy $p_1 = p_2 = \ldots = p_n$, czyli $p_k = \frac{1}{n}$

$$EX = m = \frac{1}{n}(x_1 + \ldots + x_n) \text{ – średnia arytmetyczna.}$$

Przykłady:

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

$X \in B(n, p) \Rightarrow EX = np$

$X \in P(\lambda) \Rightarrow EX = \lambda$

(ii) Przypadek rozkładu ciągłego

Wtedy $X(\Omega) = [a, b]$ i $F' = f$ – f. gęstości

Jeśli istnieje $\int_{-\infty}^{+\infty} t f(t)\,dt \in \mathbb{R}$, to liczbę tą oznaczamy

$EX = m$ i nazywamy średnią X.

<u>Przykłady:</u>.

$X \in N(m, \sigma) \Rightarrow EX = m$

Wtedy $E(X-m) = 0$ i operację $X \to X-m$ nazywamy „centrowaniem”

4.6 <u>Wariancja jako średnie kwadratowe odchylenie</u>.

Niech dana będzie zm. los. X o wartości oczekiwanej m. Bierzemy kolejno:

(i) $X \longrightarrow X - m = Y$

(ii) $Y \longrightarrow Y^2$

Jeśli Y^2 ma wartość oczekiwaną, tzn. istnieje

$$E(Y^2) = E((X-m)^2),$$

to mówimy, że X ma <u>wariancję</u> i piszemy

$$\mathrm{var}(X) = \sigma^2 = E(X-m)^2$$

Wtedy:

a) $\sigma \geq 0$, $\sigma = 0 \Leftrightarrow X = \text{const}$.

Czyli zawsze $\sigma > 0$!

b) $\operatorname{var}(X) = E(X-m)^2$

średnie (bo) kwadratowe (bo)

odchylenie (bo)

||

jako różnica $X - m$

c) $\sigma^2 = E(X^2) - (EX)^2 = E(X^2) - m^2$.

Przykłady

$X: \begin{array}{c|c} 0 & 1 \\ \hline q & p \end{array}$ $\quad X^2: \begin{array}{c|c} 0 & 1 \\ \hline q & p \end{array}$

$\sigma^2 = E(X^2) - (EX)^2 = p - p^2 = p(1-p) = pq$

$X \in B(n,p)$ $\quad \sigma^2 = npq$

$X \in P(\lambda)$ $\quad \sigma^2 = \lambda$

$X \in N(m,\sigma)$ $\quad \operatorname{var}(X) = \sigma^2$

Fakt (standaryzacja)

Niech X ma $\operatorname{var} X = \sigma^2 > 0$.

Wtedy $X \longrightarrow \frac{X-m}{\sigma} = Y$,

$EY = 0$ i $\operatorname{var}(Y) = 1$

Jasno $X \in N(m, \sigma) \Rightarrow \frac{X-m}{\sigma} \in N(0,1)$.

5°. Zjawisko niepewności pomiaru – stochastyczna zasada nieoznaczoności

Niech dane będzie $\mathcal{Z}$, jako element przyrody.

Wtedy dla obserwacji X dla zjawiska, gdy

$\Omega \ni \omega \longrightarrow X(\omega) \in \mathbb{R}$

(i) Istnieje $\operatorname{var}(X) = \sigma^2$

(ii) $\sigma > 0$

(iii) istnieje $m = EX$.

Problem. Co jest wynikiem obserwacji dla $\mathcal{Z}$?

Z teoretycznego punktu widzenia liczba $X(\omega_0)$ dla pewnego $\omega_0 \in \Omega$.

Ale, jeśli np. ZL ma rozkład dyskretny,

czyli

$$X: \begin{array}{c|c|c|c|c|c} x_1 & x_2 & \dots & x_j & \dots & x_n \\ \hline p_1 & p_2 & & p_j & & p_n \end{array},$$

to $X(\omega_0) = x_j$ dla pewnego j i

$$P(\{\omega \in \Omega : X(\omega) = x_j\}) = \underline{\underline{p_j < 1}}.$$

<u>Zatem nie można stwierdzić, że na pewno</u> $X(\omega_0) = x_j$.

Jeszcze „gorzej” jest gdy <u>rozkład jest ciągły</u>, bowiem wtedy $\forall x_0 \in \mathbb{R}$

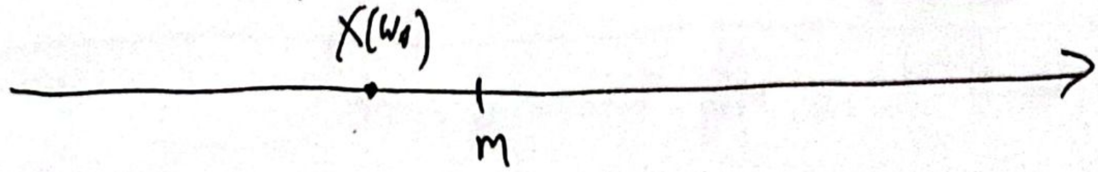
$$P(\{\omega \in \Omega : X(\omega) = x_0\}) = 0 \quad (!!!)$$

Zatem „teoretyczna wartość $X(\omega_0)$” <u>jest nieakceptowalna</u>.

Akceptowalnym wynikiem obserwacji jest $m = EX$,

Tym razem mamy sytuację, gdy na ogół

$$X(\omega_0) \neq m \quad (!!)$$

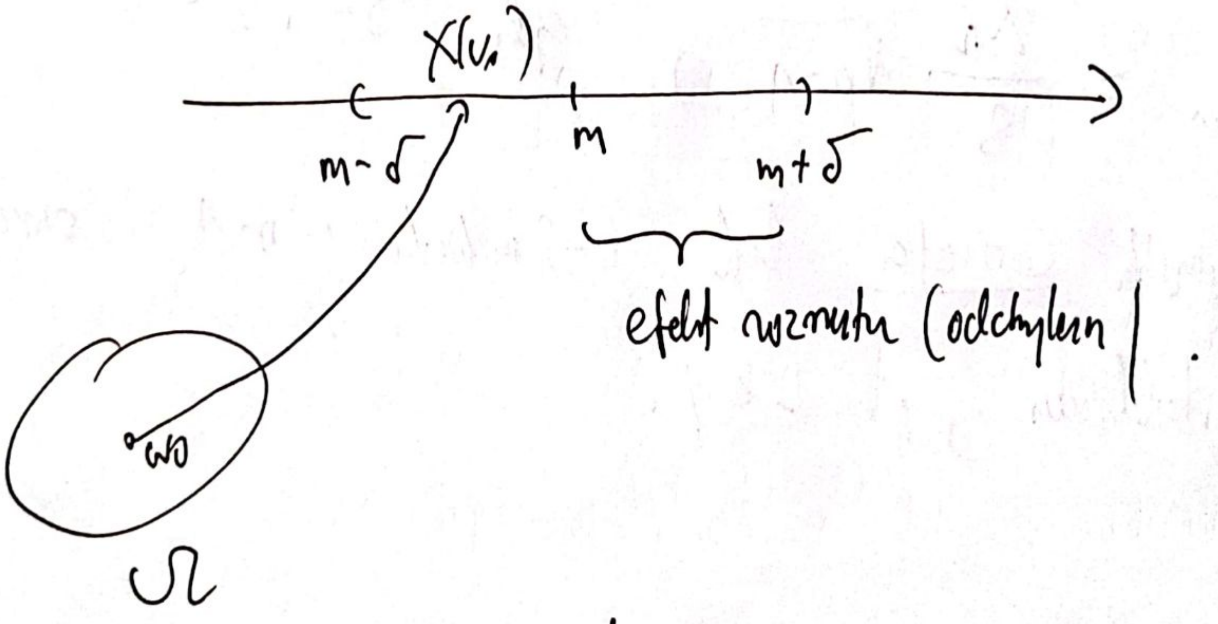


Wymagana jest zatem kontrola wzajemnego położenia m, $X(\omega_0)$.

Uzyskuje się ją z SZN, która mówi, że

(SZN)

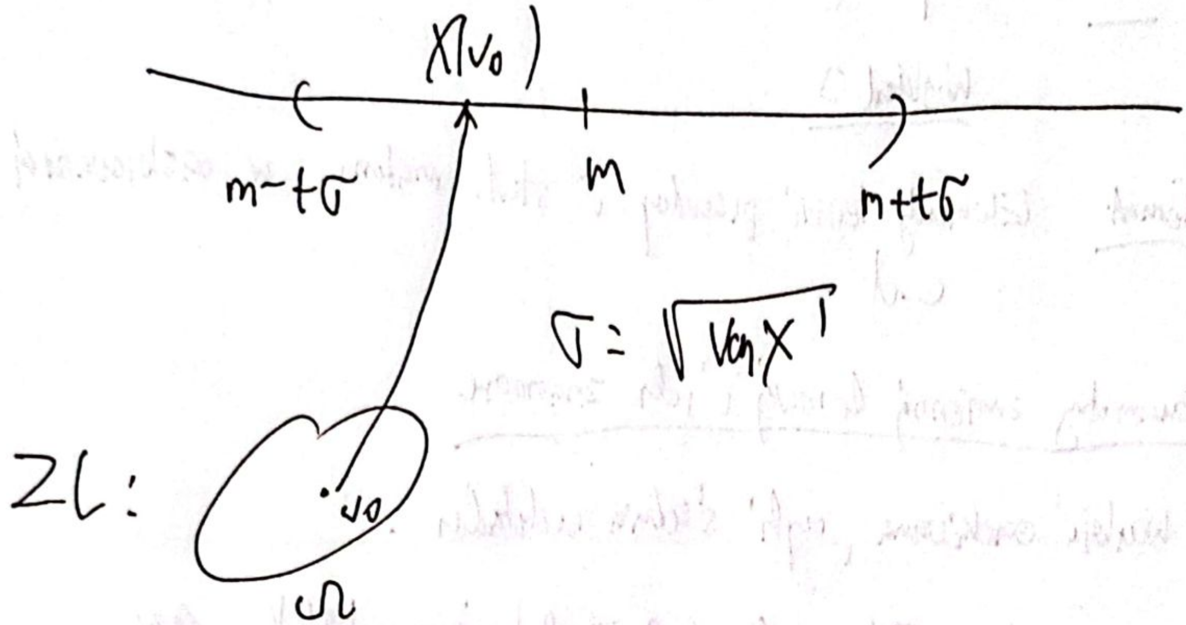
$$\forall_{\varepsilon>0} \exists_{\delta>0} \; P(\{\omega\in\Omega : |X(\omega)-m| < \delta\}) \geq 1-\varepsilon$$



(SZN) można też zapisać odmiennie:

$$\underbrace{\forall_{t>0} \; \underbrace{P(\{\omega\in\Omega : |X(\omega)-m| < t\sigma\})} \geq 1-\frac{1}{t^2}}_{\text{miara niepewności pomiaru}}$$

$\sigma = \sqrt{\operatorname{var} X}$ → dyspersja

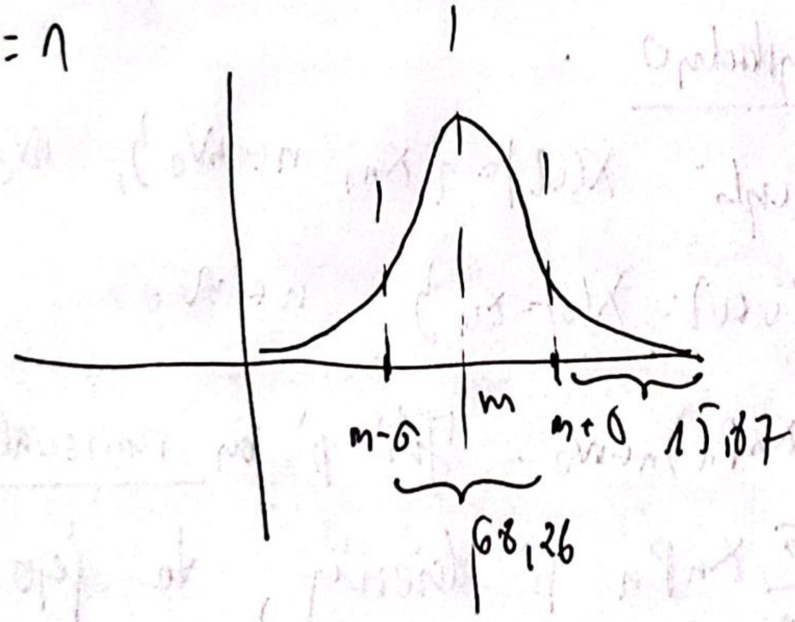


$\sigma = \sqrt{VarX}$

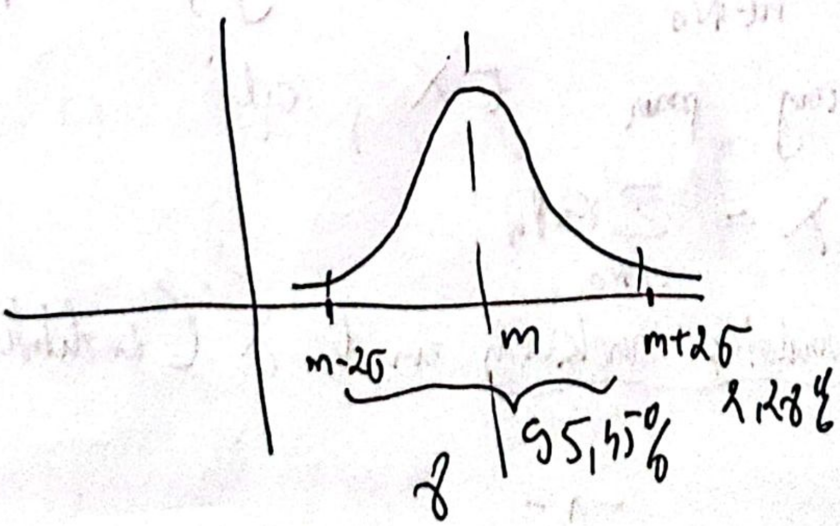
Przykład. Niech $X \in N(m, \sigma^2)$.

Dokonaj oceny sytuacji dla: $t = 1, 2, 1$.

$t = 1$

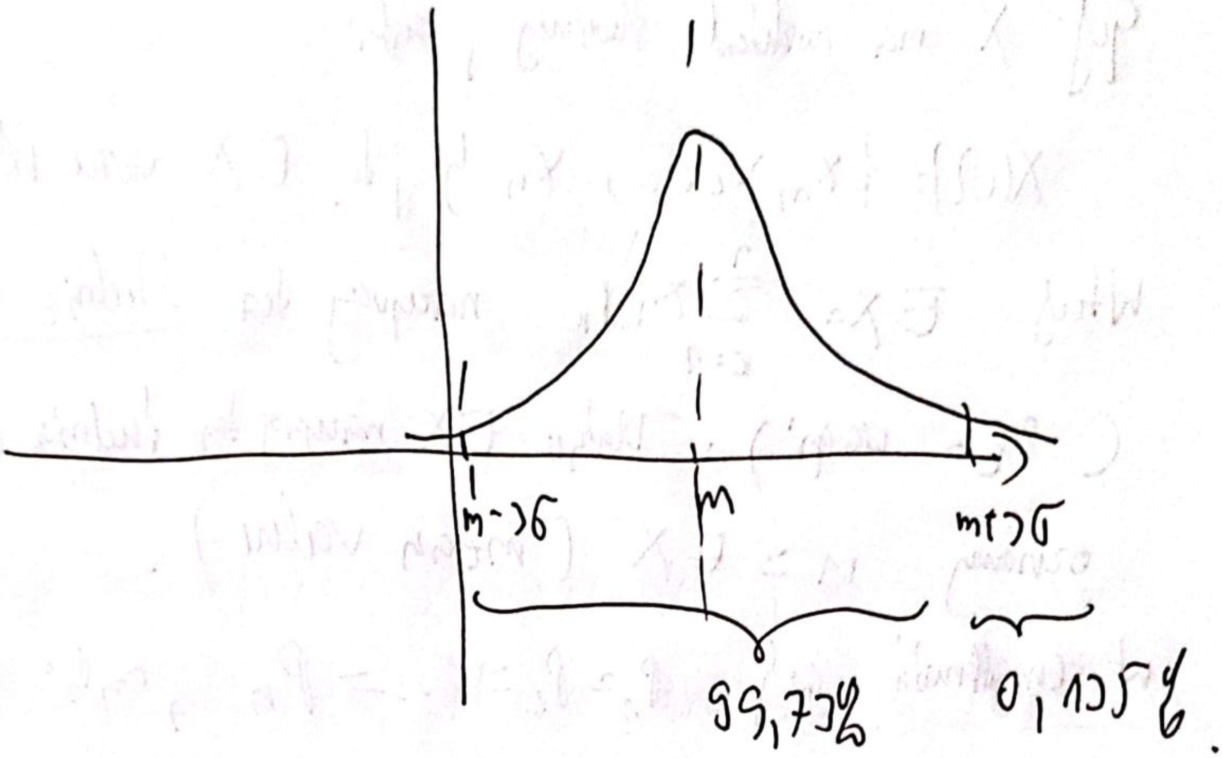


$t = 2$



Zeskanowane w CamScanner

$t=3$ („3σ")



6°. Pojęcie wektora losowego

Niech dane będzie ZL oraz jej model (Ω, Σ, P).

Do tej pory obserwowaliśmy tylko jedną cechę (atrybut) ZL używając pojedynczej zm. losowej X.

W praktyce istnieje potrzeba obserwacji $n \geq 2$ takich atrybutów. Wtedy taki proces ma opis

$$\Omega \ni \omega \longrightarrow (X_1, X_2, \dots, X_n)(\omega) =$$

$$= (X_1(\omega), X_2(\omega), \dots, X_n(\omega)) =$$

$$= (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$$

gdzie $X_1, X_2, .., X_n$ są zm. losowymi.

$(X_1, X_2, .., X_n)$ – wektor losowy.

Dalej będą potrzebować, aby kolejne obserwacje X_j były od siebie „odseparowane", a więc

X_1 nie wpływało na X_j $j \geq 2$

X_2 ~~nie~~ „ – X_j $j \geq 3$ itd.

Będą więc mówić, że składowe wektora losowego $(X_1, .., X_n)$ są stochastycznie niezależne.

Przykład.

Niech ZL$_0$ z obserwacją X_0 $\begin{array}{c|c} 0 & 1 \\ \hline q & p \end{array}$

Niech $X_1, X_2, .., X_n$ kopie X_0 i dodatkowo stochastycznie niezależne.

Bierzemy wektor $(X_1, X_2, .., X_n)$. Wtedy dla $\omega_0 \in \Omega$,

$(X_1, X_2, .., X_n)(\omega_0) = (b_1, b_2, .., b_n)$,

gdzie $b_j \in \{0, 1\}$

Niech $A_k = \{\omega \in \Omega : \#$ „1" w ciągu $(b_1, \dots, b_n)$ wynosi $k\}$.

Wtedy $P(A_k) = \binom{n}{k} p^k (1-p)^{n-k}$, czyli

tak jak w $B(n,p)$.

7°. Wprowadzenie do statystyki mat. – pojęcie populacji generalnej, jej cechy i próby prostej.

Więc, że każde ZL ma swój model $(\Omega, \mathcal{F}, P)$ i można obserwować dany atrybut ZL używając zm. losowej X. W T.P. zakładamy, że wtedy znamy jej rozkład X, umiemy zatem policzyć m i σ^2, zatem znamy wyniki obserwacji na mocy SZN!

Z perspektywy statystyki matematycznej rozkład X nie jest znany!

Metody statystyki matematycznej pozwalają ten rozkład odtworzyć, czyli wrócić do warunków T.P.

Opis procedury

1°. Ω nazywa się POPULACJĄ GENERALNĄ.

Dokładniej, jest to zbiór, dla którego

$$\omega \in \Omega \iff \omega \text{ ma (tę samą) cechę},$$

którą oznacza się przez X (jest to zm. losowa)

Cel: wyznaczenie rozkładu X

2°. Ω może być nieskończony. Z Ω wybieramy podzbiór skończony, odpowiednio liczebny $\Omega_0 \subset \Omega$

$\Omega_0 = \{\omega_1, \omega_2, \ldots, \omega_n\}$ i obserwujemy wartości X na podstawie Ω_0, tzn. mamy

$$(X(\omega_1), X(\omega_2), \ldots, X(\omega_n)) = (x_1, x_2, \ldots, x_n)$$

Jeśli Ω_0 wybrany jest właściwie, to ciąg można skonstruować następująco:

$$(x_1, x_2, \ldots, x_n) = (X_1, X_2, \ldots, X_n)(\omega_0), \text{ gdzie}$$

$\omega_0 \in \Omega$, $X_1, \ldots, X_n$ niezależne kopie X.

Możemy uważać, że $(x_1, x_2, \ldots, x_n)$ jest próbą prostą z P.G. dla cechy X.

3°. Mając próbę prostą

$$(x_1, \ldots, x_n) = (X_1, \ldots, X_n)(\omega_0) \text{ z P.G. z cechy } X$$

definiujemy nowe zmienne losowe nazywane statystykami:

(i) $\overline{X}_n \stackrel{df}{=} \frac{1}{n}(X_1 + \ldots + X_n)$.

Wtedy $\overline{X}(\omega_0) = \frac{1}{n}(x_1 + \ldots + x_n)$ - wartości zaobserwowane

$\overline{X}_n$ - średnia z próby

Dowodzi się, że gdy n odp. duże, to

$$\overline{X}(\omega_0) \cong m = EX \text{ (zjawisko ESTYMACJI).}$$

(ii) $S^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2$ - wariancja z próby

$S^2(\omega_0)$ - w. zaobs. S^2

(iii) Przypadek, w szczególności wtedy, gdy $X \in N(m, \sigma^2)$ (m i σ^2 nieznane).

Wtedy $\frac{nS^2}{\sigma^2} = \chi^2_{n-1}$ – chi-kwadrat Pearsona o $n-1$ st. swobody

(stablicowany, patrz. zadanie (2))

(iv) X j.w.

$$t_{n-1} = \frac{\overline{X}_n}{S}\sqrt{n-1}, \quad \text{gdzie } S = \sqrt{S^2}$$

statystyka Goseta lub t-Studenta o $n-1$ st. swobody

(stablicowana, p. (2)).