Kurs : SSP ( SPC )

Wykład 4

Temat: Elementy t. prawdop. i st. matem. w zastosowaniach SPC c.d.

Kontynuacja opisu procedury

(4°). Na podstawie próby prostej dla danej P.G. i jej cechy $X$,

$$(x_1, x_2 \ldots x_n) = (X_1, X_2 \ldots, X_n)(\omega_0)$$

próbujemy ustalić rozkład $X$.

Na tym etapie stosuje się metody dystrybuanty empirycznej lub histogramu.

dystrybuanta empiryczna

$$\mathbb{R} \ni t \longrightarrow S(\omega_0, t) \overset{df}{=} \frac{1}{n} \left| \{ j : x_j < t \} \right|$$

(Patrz listy)

Można udowodnić, że

$$P(\{\omega \in \Omega : S(\omega, t) \longrightarrow F(t)\}) = 1,$$

zatem dla odp. dużych $n$,

$$\underbrace{S(\omega_0, t) \simeq F(t)}$$

## Histogram

rodzaj wykresu słupkowego: na osi poziomej przedziały liczbowe danych empirycznych (p.p.), na osi pionowej # obserwacji przypisanych do danej klasy.

Bardzo często przy konstrukcji histogram od razu konstruuje się dystrybuantę empiryczną (tzw. wartości skumulowane)

### Przykład.

Dla P.G. z cechy X pobrano p.p.:

24.975, 24.970, 25.014, 25.010, 25.010, 24.967
25.012, 25.014, 25.018, 24.993, 24.992, 25.072
24.991, 24.950, 25.042, 25.020, 24.925, 25.007
25.031, 25.090, 24.960, 24.968, 24.957, 25.031
24.985, 24.997, 24.995, 25.007, 24.991, 25.033

$n = 30$

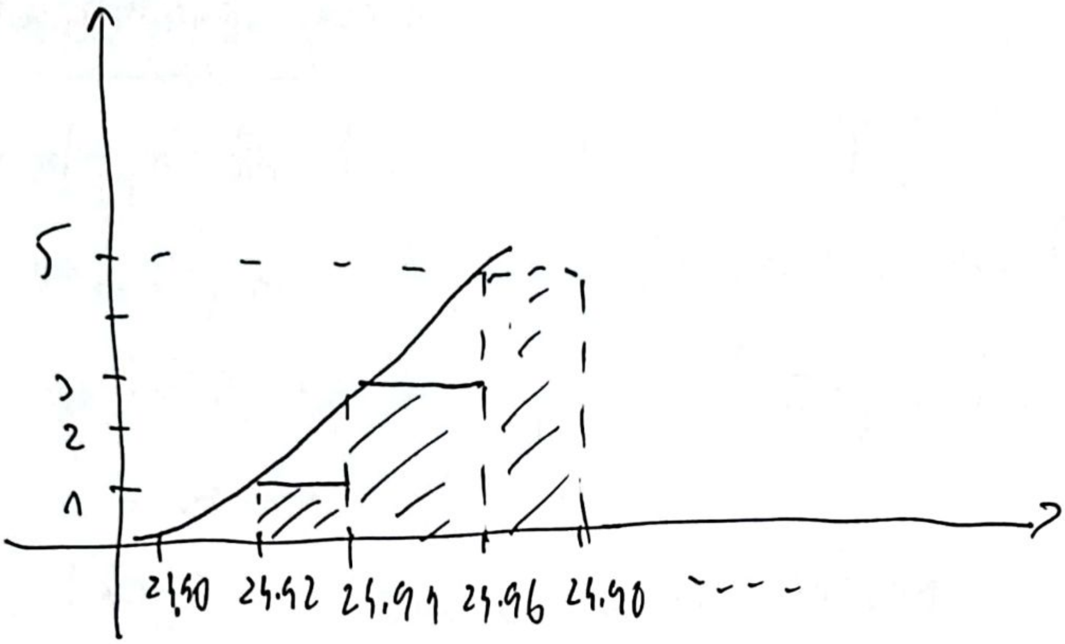
Wybieramy wartość najmniejszą - 24.900 jako początkową i największą - 25,100 i długość powstałego odcinka dzielimy na przedziały jednakowej długości, np. 0,02.

Dla każdego takiego przedziału zliczamy # $x_i$ z p.p.

Postawiemy:

| klasa | $\# x_j'$ | skumulowane # | $S(w_0, t)$ $n=30$ |
|---|---|---|---|
| (24,9000, 24,920> | 0 | 0 | 0 |
| (24,920, 24,940> | 1 | 1 | $\frac{1}{30}$ |
| (24,940, 24,960> | 3 | 4 | $\frac{4}{30}$ |
| (24,960, 24,980> | 5 | 9 | $\frac{9}{30}$ |
| (24.980, 25.000> | 6 | 15 | $\frac{15}{30}$ |
| (25.000, 25.020> | 9 | 24 | $\frac{24}{30}$ |
| (25.020, 25.040> | 3 | 27 | $\frac{27}{30}$ |
| (25.040, 25.060> | 1 | 28 | $\frac{28}{30}$ |
| (25.060, 25.080> | 1 | 29 | $\frac{29}{30}$ |
| 25.080, 25.100 | 1 | 30 | 1 |

Sporządź wykres histogramu i dystr. empirycznej.




<u>Obwiednia</u> histogramu daje przybliżoną postać f. gęstości (o ile $X$ ma rozkład ciągły)

<u>Podobnie dla</u> $S(w_0, t)$ . (Dokończyć).

(VI) Na ogół na etapie (V) ustalamy tylko:

- rodzaj rozkładu (dyskretny, ciągły)
- dla ustalonego rozkładu nie znamy parametrów,

czyli np. wiemy, że $X \in N(m, \sigma^2)$, ale $m = ?$, $\sigma^2 = ?$.

Ustalenie parametrów przebiega metodą ESTYMACJI (przedziałowej).

# Idea estymacji przedziałowej

Pamiętamy, że dla P.G. z cechą $X$ i P.P.

$(x_1, \dots, x_n) = (X_1, \dots X_n)(\omega_0)$, metodą (V)

ustaliliśmy typ rozkładu $X$, $F(~~t~~, \Theta)$, gdzie

$\Theta$ nieznana wartość parametru.

Należy: (i) skonstruować dwie statystyki

$$Z_1 = f_1(X_1, \dots, X_n), \quad Z_2 = f_2(X_1, \dots, X_n)$$

$$\omega \in \Omega, \quad Z_1(\omega) \leq Z_2(\omega)$$

(ii) ustalić $\alpha \in (0,1)$ – poziom <u>istotności</u>

(np. $\alpha = 0{,}05$)

(iii) w ten sposób, że

$$P(\{\omega \in \Omega : Z_1(\omega) < \Theta < Z_2(\omega)\}) = 1 - \alpha$$

poziom wiarygodności

Wtedy $\Theta \in (Z_1(\omega_0), Z_2(\omega_0))$ z prawd. $1-\alpha$

Przykład 1

Przypuśćmy, że $X \in N(m, \sigma^2)$, $\sigma^2$ znane, estymujemy $m$.

Wtedy: $$Z_1 = \overline{X}_n - u_\alpha \frac{\sigma}{\sqrt{n}}$$

$$Z_2 = \overline{X}_n + u_\alpha \frac{\sigma}{\sqrt{n}},$$

gdzie $\Phi(u_\alpha) = 1 - \alpha/2$

(Patrz Lista nr 3)

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

$$Z_1 = \overline{X}_n - t_\alpha \frac{S}{\sqrt{n-1}}; \quad Z_2 = \overline{X}_n + t_\alpha \frac{S}{\sqrt{n-1}},$$

gdzie $t_\alpha$ – wartość krytyczna r. t-Studenta $\overset{n-1 \text{ st. sw.}}{\frown}$ dla $\alpha$

$$P(\{\omega \in \Omega : |t_{n-1}(\omega)| > t_\alpha\}) = \alpha$$

(Patrz Lista nr. 7)

Jak zobaczymy dalej, założenie $X \in N(m, \sigma^2)$ będzie na ogół spełniane w metodzie SPC.

Uwaga. Pozostałe aspekty związane z T.P. i S.M. podamy przy okazji omawiania metody SPC.