Przypuśćmy, że cecha $\mathbb{X}$ populacji generalnej na podstawie metody (2) ma określony typ rozkładu $F(x,\tau)$, gdzie τ jest nieznaną wartością jednego z parametrów znanego typu rozkładu F. Lub ogólniej, być może nie znamy nawet typu tego rozkładu, ale dysponujemy informacjami na temat istnienia jego momentów.

Jeśli uda się skonstruować dwie statystyki

$$\mathbb{Z}_j = f_j(\mathbb{X}_1 \dots \mathbb{X}_n), \quad j = 1, 2,$$

gdzie $(x_1 \dots x_n) = (\mathbb{X}_1 \dots \mathbb{X}_n)(\omega_o)$ jest próbą prostą i takie, że dla dostatecznej małej liczby α (np. 0,05) wiemy, że

$$P(\{\omega \in \Omega : \ \tau \in (\mathbb{Z}_1(\omega), \mathbb{Z}_2(\omega))\}) = 1 - \alpha,$$

to będziemy mówili, że otrzymaliśmy *przedział losowy* przybliżający nieznaną wartość parametru τ z prawdopodobieństwem $1 - \alpha$.

Wtedy, wiedząc, że zdarzenie elementarne ω_o, które wyznacza próbę prostą należy do zdarzenia, które opisuje przedział losowy, otrzymalibyśmy oszacowanie na wartość tego parametru

$$\tau \in (\mathbb{Z}_1(\omega_o), \ \mathbb{Z}_2(\omega_o)).$$

Ponieważ weryfikacja tego faktu może być niemożliwa, lepiej jest powiedzieć, że na *poziomie ufności* $1 - \alpha$ oszacowaliśmy przedział, z którego można wybrać wartość liczbową parametru τ.

Metodę tę omówimy dokładniej w poniższych przykładach. Ograniczymy się tylko do kilku sytuacji, zakładając, że cecha ma drugi moment (patrz też [10]).

Przykład 6.4.2 *Załóżmy, że z obserwacji histogramu dla próby prostej*

$$(x_1 \dots x_n) = (\mathbb{X}_1 \dots \mathbb{X}_n)(\omega_o)$$

zauważyliśmy, że

$$\mathbb{X} \in \mathcal{N}(m, \sigma^2), \ \textit{gdzie nieznany jest tylko parametr } m.$$

Pokażemy, jak w tym przypadku można zdefiniować przedział losowy.

Bierzemy statystyki

$$\mathbb{Z}_1 = \overline{\mathbb{X}}_n - n_\alpha \frac{\sigma}{\sqrt{n}} \quad \mathbb{Z}_2 = \overline{\mathbb{X}}_n + n_\alpha \frac{\sigma}{\sqrt{n},}$$

gdzie

$$P(\{\omega \in \Omega : \ | \ \mathbb{N}(\omega) \ | > n_\alpha\}) = \alpha,$$

i n_α jest takie, że

$$\Phi(n_\alpha) = 1 - \frac{\alpha}{2}, \quad \textit{co odczytujemy z tablic.}$$

Z twierdzenia 6.3.1 (3) wynika, że uzyskane statystyki poprawnie definiują przedział losowy na zadanym poziomie ufności.

Niech teraz $(2{,}23\ , 2{,}12\ , 1{,}97\ , 2{,}01) = (\mathbb{X}_1 \ldots \mathbb{X}_4)(\omega_o)$ *będzie próbą prostą i* $\mathbb{X}$ *rozkładem typu* $\mathcal{N}(m,\ (0{,}25)^2)$.

Dla $\alpha = 0{,}05$ *mamy:*

$$x_4 = \overline{\mathbb{X}_4}(\omega_o) \simeq 2{,}083,\ \Phi(n_\alpha) = 1 - \frac{\alpha}{2}\ \textit{implikuje, że}\ n_\alpha = 1{,}96,$$

więc z prawdopodobieństwem 0,95

$$m \in \Big(2{,}083 - 1{,}96\frac{0{,}25}{2},\ 2{,}083 + 1{,}96\frac{0{,}25}{2}\Big) = (1{,}838\ , 2{,}328).$$

Przykład 6.4.3 *Załóżmy, że z obserwacji histogramu dla próby prostej* $(x_1 \ldots x_n) = (\mathbb{X}_1 \ldots \mathbb{X}_n)(\omega_o)$ *zauważyliśmy, że*

$$\mathbb{X} \in \mathcal{N}(m, \sigma^2),\ \textit{gdzie nieznane są oba parametry oraz estymujemy } m.$$

W tej sytuacji bierzemy:

$$\mathbb{Z}_1 = \overline{\mathbb{X}}_n - t_\alpha \frac{\mathbb{S}}{\sqrt{n-1}}, \quad \mathbb{Z}_2 = \overline{\mathbb{X}}_n + t_\alpha \frac{\mathbb{S}}{\sqrt{n-1},}$$

gdzie dla danej wartości α

$$P(\{\omega \in \Omega :\ |\ t_{n-1}(\omega)\ | > t_\alpha\}) = \alpha,\ i\ \mathbb{S} = \sqrt{\mathbb{S}^2}.$$

Korzystając z Twierdzenia 6.3.1 (2) można pokzać, że uzyskane statystyki poprawnie definiują przedział losowy na zadanym poziomie ufności.

Odpowiednią symulację liczbową przeprowadzimy na danych liczbowych pochodzących z wcześniejszego przykładu. Korzystając z rozkładu t-Studenta o trzech stopniach swobody, dostaniemy

$$P(\{\omega \in \Omega :\ |\ t_3(\omega)\ | > t_\alpha\}) = 0{,}05\ \textit{implikuje, że}\ t_\alpha = 3{,}182.$$

Ponieważ z poprzedniego przykładu $\overline{\mathbb{X}_4}(\omega_o) = 2{,}083$, *więc*

$$s = \mathbb{S}(\omega_o) =$$

$$\sqrt{1/4((2{,}23-2{,}083)^2+(2{,}12-2{,}083)^2+(1{,}97-2{,}083)^2+(2{,}01-2{,}083)^2)}$$
$$=\sqrt{0{,}01}=0{,}1.$$

Stąd z postaci przedziału losowego dostaniemy

$$m\in\Big(2{,}083-3{,}182\frac{0,1}{1{,}73}\ ,\ 2{,}083+3{,}182\frac{0,1}{1{,}73}\Big)=(1{,}899\ ,2{,}267)$$

z prawdopodobieństwem 0,95.

Przykład 6.4.4 *W tym przykładzie obserwacja dystrybuanty empirycznej albo histogramu powstałego na podstawie zaobserwowanej próby prostej nie wiele dałanie znamy typu rozkładu cechy $\mathbb{X}$. Skądinąd wiemy natomiast, że ma on drugi moment. W dalszym ciągu będziemy estymowali parametr m. Weźmy teraz n dostatecznie duże, co najmniej równe 100.*

Następujące 3 fakty:

1.

$$\frac{\overline{\mathbb{X}_n}-m}{\frac{\sigma}{\sqrt{n}}}\xrightarrow{d}\mathcal{N}(0,1),\quad (z\quad CTG),$$

2.

$$E\widehat{\mathbb{S}}^2=var(\mathbb{X}),$$

3.

$$\widehat{\mathbb{S}}^2\overset{p.w.}{\to}\sigma^2,\quad (z\ MPWL)$$

uzasadniają, że w tym przypadku następujące statystyki

$$\mathbb{Z}_1=\overline{\mathbb{X}}_n-n_\alpha\frac{\widehat{\mathbb{S}}}{\sqrt{n}},\quad \mathbb{Z}_2=\overline{\mathbb{X}}_n+n_\alpha\frac{\widehat{\mathbb{S}}}{\sqrt{n},}$$

gdzie n_α ma takie samo znaczenie jak w przykładzie 6.4.2 wyznaczają przedział losowy dla parametru m z prawdopodobieństwem $1-\alpha$.

Ze względu na potrzebę korzystania w tej sytuacji z twierdzeń granicznych wielkość próby prostej jest tutaj bardzo istotna. Aby uprościć symulację liczbową załóżmy, że:

$$\overline{x}_{100}=2{,}0031,\ \widehat{s}=0,1967,\ \alpha=0{,}05.$$

Wtedy z prawdopodobieństwem 0,95 wartość średnia cechy $\mathbb{X}$ należy do przedziału

$$\Big(2{,}0031-1{,}96\frac{0{,}1967}{10}\ ,2{,}0031+1{,}96\frac{0{,}1967}{10}\Big)=(1{,}965\ ,2{,}042).$$