1 Laboratorium 5. Przedziały ufności dla prawdopodobieństwa sukcesu.

Niech $X_1, \ldots, X_n$ i.i.d. $B(1,p)$, $X = X_1 + \ldots + X_n$, $\widehat{p}_n = \dfrac{X}{n}$. Dla ustalonego $\alpha \in (0,1)$, $z_{\alpha/2}$ oznacza kwantyl rzędu $1-\alpha/2$ rozkładu $N(0,1)$.

Asymptotyczne przedziały ufności na poziomie ufności $1-\alpha$ dla prawdopodobieństwa sukcesu (frakcji) p.

1. **Przedział ufności Walda:**

$$CI_S = \widehat{p}_n \pm z_{\alpha/2} \cdot \sqrt{\frac{\widehat{p}_n(1-\widehat{p}_n)}{n}}.$$

2. **Przedział ufności Wilsona:**

$$CI_W = \left\{ p \,:\, \frac{\sqrt{n}|\,\widehat{p}_n - p\,|}{\sqrt{p(1-p)}} \le k \right\}.$$

Ten przedział można zapisać w równoważnej postaci

$$CI_W = \tilde{p} \pm \frac{k\sqrt{n}}{n+k^2}\sqrt{\tilde{p}\tilde{q} + \frac{k^2}{4n}},$$

gdzie

$$k = z_{\alpha/2},\ \tilde{X} = X + k^2/2,\ \tilde{n} = n + k^2,\ \tilde{p} = \frac{\tilde{X}}{\tilde{n}},\ \tilde{q} = 1 - \tilde{p}.$$

3. **Przedział ufności Agresti–Coulla:**

$$CI_{AC} = \tilde{p} \pm k\frac{\sqrt{\tilde{p}\tilde{q}}}{\sqrt{\tilde{n}}}.$$

4. **Przedział ufności oparty na ilorazie wiarogodności:**

$$CI_{LR} = \left\{ p \,:\, -2\ln\Lambda_n(p) \le z^2_{\alpha/2} \right\}.$$

W tym wzorze Λ_n jest ilorazem wiarogodności

$$\Lambda_n(p_0) = \frac{L(p_0)}{\sup_p L(p)} = \frac{p_0^X(1-p_0)^{n-X}}{(X/n)^X(1-X/n)^{n-X}}$$

a L oznacza funkcję wiarogodności, tzn. $L(p) = p^X(1-p)^{n-X}$.

****

Przedział ufności Walda, pojawiający się w prawie każdej książce ze statystyki, jest zaimplementowany we wszystkich pakietach statystycznych. Niestety, ten przedział, przy którego konstrukcji wykorzystano to, że

$$\frac{\sqrt{n}(\,\widehat{p}_n - p\,)}{\sqrt{\widehat{p}_n(1-\widehat{p}_n)}} \xrightarrow{D} N(0,1),$$

nie ma zbyt dobrych własności. Dla ustalonego n, nawet przekraczającego 100, oraz dla ustalonego p, nawet bliskiego $1/2$, prawdopodobieństwo pokrycia dla CI_S jest dalekie od $1-\alpha$. Istnieją przedziały lepsze od CI_S.

****

Niech $p = 1/2$ i $\alpha = 0.05$.

1. Dla każdego $n = 10, 11, 12, \ldots, 100$ wyznacz **dokładną** wartość* prawdopodobieństwa pokrycia dla CI_S, to znaczy liczbę

$$p_n := P_p\left(\frac{\sqrt{n}|\,\widehat{p}_n - p\,|}{\sqrt{\widehat{p}_n(1-\widehat{p}_n)}} \leq z_{\alpha/2}\right).$$

   W sprawozdaniu podaj wartości p_n, $n = 10, 11, \ldots, 100$, z dokładnością do trzech cyfr po przecinku.

2. Narysuj łamaną łączącą punkty (n, p_n), $n = 10, 11, \ldots, 100$.

3. Powtórz poprzednią analizę dla pozostałych trzech przedziałów ufności.

4. Dla którego z tych czterech przedziałów ufności oscylacje prawdopodobieństwa pokrycia względem $1 - \alpha$ wydają się być najmniejsze?

5. Dla CI_S i dla CI_{AC} oraz dla każdego $p \in \{1/10, 2/10, 3/10, 4/10; 5/10\}$ wyznacz najmniejszą liczbę $n(p)$, taką że $p_n \geq 0.93$ dla $n \geq n(p)$. Sporządź tabelkę z wartościami $n(p)$ dla obu przedziałów i dla $p \in \{1/10, 2/10, 3/10, 4/10; 5/10\}$. Co wynika z ostatniej analizy?

Uwaga*. By wyznaczyć tę **dokładną** wartość należy skorzystać z tego, że $n\widehat{p}_n \overset{D}{=} B(n, p)$. W obliczeniach można posłużyć się pakietem *Mathematica*.