Prawdopodobieństwo sukcesu pewnego zdarzenia wynosi π_0. W n niezależnych próbach zanotowano s sukcesów. Estymatorem nieobciążonym i największej wiarygodności π_0 jest $p_0 = \frac{s}{n}$

Przedział ufności dla π_0 na poziomie $1-\alpha$ ma postać:

$$|\pi - p_0| \leqslant \lambda(\alpha)\sqrt{\frac{\pi_0(1-\pi_0)}{n}} \tag{1}$$

gdzie $\lambda(\alpha)$ jest takim współczynnikiem, aby prawdopodobieństwo spełnienia warunku (1) było równe $1-\alpha$.

Wartość $\lambda(\alpha)$ zależy od nieznanego parametru π_0. Zazwyczaj stosuje się przedział ufności

$$|\pi - p_0| \leqslant \lambda(\alpha)\sqrt{\frac{p_0(1-p_0)}{n}}$$

Standardowy przedział ufności

Z twierdzenia Moivre'a - Laplace'a, dla dużych n (przyjmuje się $n > 5 \setminus \min(p_0, 1-p_0)$), $\lambda(\alpha)$ jest kwantylem rzędu $1-\alpha/2$ rozkładu normalnego standardowego.

3.1 **Przedział ufności Wilsona**

Przedział ufności Wilsona jest rozwiązaniem nierówności

$$|\pi - p_0| \leqslant \lambda(\alpha)\sqrt{\frac{\pi(1-\pi)}{n}} \tag{2}$$

względem π. Dla dużych n, $\lambda(\alpha)$ jest kwantylem rzędu $1-\alpha/2$ rozkładu normalnego standardowego. Przedział ten jest dokładniejszy niż standardowy przedział ufności.

Rozwiąż nierówność (2) względem π. Pokaż, że rozwiązanie jest przedziałem o końcach

$$\widetilde{p_0} \pm \lambda(\alpha)\sqrt{\frac{\widetilde{p_0}(1-\widetilde{p_0})}{n}}$$

gdzie

$$\widetilde{p_0} = \frac{s+\lambda^2(\alpha)/2}{n+\lambda^2(\alpha)}$$

3.2 **Przedział ufności Agrestiego-Coulla**

Przedział ufności Agrestiego-Coulla jest modyfikacją przedziału Wilsona:

$$\widetilde{p_0} \pm \lambda(\alpha)\sqrt{\frac{\widetilde{p_0}(1-\widetilde{p_0})}{\widetilde{n}}}$$

gdzie

$$\widetilde{p_0} = \frac{s+\lambda^2(\alpha)/2}{n+\lambda^2(\alpha)},\ \widetilde{n} = n + \lambda^2(\alpha)$$

3.2.1 Oceń, który przedział jest szerszy: Wilsona, czy Agrestiego - Coulla? Jakiego rzędu jest ta różnica? Jakie są tego skutki? Który z nich jest bardziej "naturalny"?

3.2.2 W praktyce stosuje się przedział ufności dla π_0, nazywany *plus 2 sukcesy, plus 2 porażki*, postaci

$$\widetilde{p_0} \pm 2\sqrt{\frac{\widetilde{p_0}(1-\widetilde{p_0})}{\widetilde{n}}}$$

gdzie

$$\widetilde{p_0} = \frac{s+2}{n+4}, \ \widetilde{n} = n+4$$

Uzasadnij nazwę tego przedziału. Jakiemu, w przybliżeniu, poziomowi ufności odpowiada ten przedział?

3.3 **Ćwiczenia rachunkowe**

Wyznacz na poziomie ufności 95% przedziały: standardowy, Wilsona, Agrestiego-Coulla, *plus 2 sukcesy, plus 2 porażki* dla danych:

a) $s = 6$, $n = 20$,

b) $s = 0$, $n = 20$,

c) $s = 20$, $n = 20$,

d) $s = 30$, $n = 100$.