Intervaly spolehlivosti

Tomáš Hons

25. března 2021

1 Úvod

Minulý týden jsme počítali pouze s asymptotickými odhady následujícího typu

$$\left(\bar{X} - u_{\frac{\alpha}{2}}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \bar{X} + u_{\frac{\alpha}{2}}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right),$$

kde $u_{\cdot}$ jsou kvantily normálního rozdělení a $\hat{p}$ je odhad parametru alternativního rozdělení $\hat{p} = \bar{X}$.

Nyní budeme počítat i s těmito intervaly

$$\left(\bar{X} - t_{n-1;\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1;\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right),$$

které jsou založeny na Studentově rozdělení a používají výběrový rozptyl S^2 jako odhad rozptylu. Tento typ intervalu je přesný pro odhad střední hodnoty normálního rozdělení. V našem případu ovšem pracujeme s alternativním rozdělením a také tedy půjde pouze o asymptotický interval spolehlivosti.

Pro přehled budeme interval z minulého týdne označovat jako interval *prvního typu*, nově zavedený interval bude *druhého typu*.

2 Interval spolehlivosti druhého typu

Dle postupu ve videu jsem spočítal interval druhého typu pro svá data z minulého týdne. Pro připomenutí: šlo o házení víčkem, u kterého jsme sledoval, zda dopadlo dnem nahoru, či dolů.

Interval s hladinou významnosti 0.05 vychází po zaokrouhlení $[0.26; 0.40]$, tedy naprosto stejně jako interval prvního typu. Kvantily Studentova rozdělení jsou v absolutní hodnotě o něco větší (a tedy interval prodlužují), stejně tak je i větší hodnota výběrového rozptylu oproti odhadu $\hat{p}(1-\hat{p})$. Změny jsou nicméně tak malé, že se ve dvou desetinných místech neprojevily. Projeví se až při zaokrouhlení na tři desetinná místa, kdy první interval je kratší o jednu setinu na obě strany.

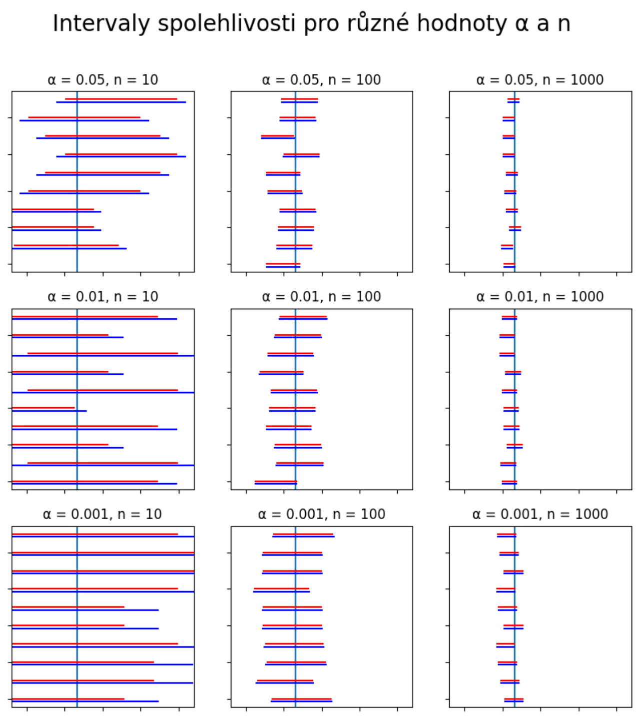
3 Popis simulace

Dalším bodem úkolu bylo experimentálně porovnat intervaly prvního a druhého typu generováním dat z binomického rozdělení, hledáním příslušných intervalů a porovnáním jejich kvalit. Sledujeme jednak délku intervalu a jednak zda pokrývá skutečnou hodnotu parametru. Parametr jsem volil $p = 0.33$, což byla moje hodnota odhadu na parametr z minulého týdne.

Simulace jsem provedl pro různé hodnoty spolehlivosti $1 - \alpha$ a pro různé počty pozorování n. Speciálně je zajímavé zkoumat chování intervalů pro malé n. Pro $n \to \infty$ totiž transformované veličiny, na kterých jsme intervaly postavili, konvergují v distribucí k $N(0, 1)$ a intervaly jsou tedy pro velké n prakticky stejné.

4 Ilustrace simulace

Pro představu, jak intervaly závisí na α a n, jsem pro všechny dvojice $\alpha \in \{0.05, 0.01, 0.001\}$ a $n \in \{10, 100, 1000\}$ vytvořit deset výběrů z $\mathrm{Bi}(n, 0.33)$ a vytvořil příslušné intervaly.



Červené intervaly jsou intervaly prvního typu a vidíme, že jsou kratší než modré intervaly druhého typu. To je patrné samozřejmě hlavně pro malé n. Pro větší n už jsou ovšem intervaly velmi podobné, téměř identické.

Dále si všimněme, že pro malé n a velké α jsou intervaly vysloveně nesmyslné a sahají mimo logické meze, tj. meze intervalu $[0, 1]$. V některých případech dokonce pokrývají celý tento interval. Takové chování není překvapivé, protože intervaly jsou pouze asymptotické a pro malý vzorek selhávají.

5 Srovnání intervalů

Pro lepší porovnání obou intervalů jsem pro každou z výše popsaných dvojic α, n udělal 10000 výběrů a počítal průměrnou délku intervalů a podíl intervalů, které minuly cíl, tj. nepokryly skutečnou hodnotu parametru.

```
##### alpha = 0.05, n = 10 #####
Type 0: Average length: 0.544; Miss rate: 0.1282
Type 1: Average length: 0.662; Miss rate: 0.1282

##### alpha = 0.05, n = 100 #####
Type 0: Average length: 0.183; Miss rate: 0.0581
Type 1: Average length: 0.186; Miss rate: 0.0581

##### alpha = 0.05, n = 1000 #####
Type 0: Average length: 0.058; Miss rate: 0.0495
Type 1: Average length: 0.058; Miss rate: 0.0495

##### alpha = 0.01, n = 10 #####
Type 0: Average length: 0.714; Miss rate: 0.0235
Type 1: Average length: 0.949; Miss rate: 0.0235

##### alpha = 0.01, n = 100 #####
Type 0: Average length: 0.241; Miss rate: 0.0143
Type 1: Average length: 0.247; Miss rate: 0.0126

##### alpha = 0.01, n = 1000 #####
Type 0: Average length: 0.077; Miss rate: 0.0094
Type 1: Average length: 0.077; Miss rate: 0.0094

##### alpha = 0.001, n = 10 #####
Type 0: Average length: 0.913; Miss rate: 0.0205
Type 1: Average length: 1.398; Miss rate: 0.0178

##### alpha = 0.001, n = 100 #####
Type 0: Average length: 0.308; Miss rate: 0.0022
Type 1: Average length: 0.319; Miss rate: 0.0019
```

```
##### alpha = 0.001, n = 1000 #####
Type 0: Average length: 0.098; Miss rate: 0.0014
Type 1: Average length: 0.098; Miss rate: 0.0013
```

Z výpisu vidíme přesně dle očekávání, že pro $n = 1000$ je úspěšnost intervalů téměř úplně stejná, jen pro $\alpha = 0.001$ se liší o desetitisícinu. Délky jsou identické.

Pro $n = 100$ jsou už intervaly druhého typu nepatrně delší, na úspěšnosti se to mírně projevilo jen pro $\alpha = 0.01$ a $\alpha = 0.001$.

Pro $n = 10$ je už rozdíl délek dobře patrný (i dle obrázku výše), ovšem podle simulace se výrazně nezměnila úspěšnost zachycení hodnoty parametru. Ta se tentokrát liší jen pro $\alpha = 0.001$. To je ovšem degenerovaný případ, protože průměrná délka intervalů druhého typu je přes 1.

Z nasimulovaných dat považuji za lepší první typ intervalu, protože větší délka druhého intervalu je jen chabě kompenzována jeho nepatrně větší přesností.