Za exaktní testy a konfidenční intervaly pro parametr binomického rozdělení logičtější!

Jan Klaschka
klaschka@cs.cas.cz
ÚI AV ČR, Praha

Robust 2016, Rejhotice 11.–16. 9. 2016

Osnova

- Úvod: Oboustranné exaktní testy a konfidenční intervaly (CI) pro parametr binomického rozdělení

Osnova

- Úvod: Oboustranné exaktní testy a konfidenční intervaly (CI) pro parametr binomického rozdělení
- Stať: 5 typů rozporu, s nimiž se můžeme setkat – a co s nimi.

Osnova

- Úvod: Oboustranné exaktní testy a konfidenční intervaly (CI) pro parametr binomického rozdělení
- Stať: 5 typů rozporu, s nimiž se můžeme setkat – a co s nimi.
- Závěr.

Úvod

- Úloha: k . . . realizace $X \sim Binom(n, p)$, n pevné, známé $\rightarrow$ exaktní (konzervativní) dvoustranný

Úvod

- Úloha: k ... realizace $X \sim Binom(n, p)$, n pevné, známé $\rightarrow$ exaktní (konzervativní) dvoustranný
  - test hypotézy $p = p_0$ na hladině významnosti α,

Úvod

- Úloha: k ... realizace $X \sim Binom(n, p)$, n pevné, známé $\rightarrow$ exaktní (konzervativní) dvoustranný
  - test hypotézy $p = p_0$ na hladině významnosti α,
  - konfidenční interval (CI) pro p na hladině spolehlivosti $1 - \alpha$.

Úvod

- Úloha: k … realizace $X \sim Binom(n, p)$, n pevné, známé $\rightarrow$ exaktní (konzervativní) dvoustranný
  - test hypotézy $p = p_0$ na hladině významnosti α,
  - konfidenční interval (CI) pro p na hladině spolehlivosti $1 - \alpha$.

- Máme bohatý „zvěřinec“ typů testů/CI – není všeobecná shoda na kritériích optimality, která by navíc byla splnitelná (pokud nechceme sáhnout po znáhodněných či fuzzy testech/CI).

Úvod

- Úloha: k ... realizace $X \sim Binom(n, p)$, n pevné, známé $\rightarrow$ exaktní (konzervativní) dvoustranný
  - test hypotézy $p = p_0$ na hladině významnosti α,
  - konfidenční interval (CI) pro p na hladině spolehlivosti $1 - \alpha$.

- Máme bohatý „zvěřinec“ typů testů/CI – není všeobecná shoda na kritériích optimality, která by navíc byla splnitelná (pokud nechceme sáhnout po znáhodněných či fuzzy testech/CI).

- Některá optimalizační kritéria (i návrhy, které nic neoptimalizují) vedou k nemilým důsledkům – viz dále.

Úvod

- Úloha: $k \ldots$ realizace $X \sim Binom(n, p)$, n pevné, známé $\rightarrow$ exaktní (konzervativní) dvoustranný
  - test hypotézy $p = p_0$ na hladině významnosti α,
  - konfidenční interval (CI) pro p na hladině spolehlivosti $1 - \alpha$.

- Máme bohatý „zvěřinec“ typů testů/CI – není všeobecná shoda na kritériích optimality, která by navíc byla splnitelná (pokud nechceme sáhnout po znáhodněných či fuzzy testech/CI).

- Některá optimalizační kritéria (i návrhy, které nic neoptimalizují) vedou k nemilým důsledkům – viz dále.

- Mnohé lze zobecnit na jiná rozdělení, popř. i čtyřpolní tabulky.

Úvod

- $\beta_{k,n}(p_0)$ ... p-hodnota testu hypotézy $p = p_0$ při k úspěších z n pokusů.

Úvod

- $\beta_{k,n}(p_0)$ ... p-hodnota testu hypotézy $p = p_0$ při k úspěších z n pokusů.
- Funkci $\beta_{k,n}(p)$ budeme říkat p-funkce, zkratkou pvf (z p-value function). V literatuře též evidence function, confidence interval function aj.

Úvod

- $\beta_{k,n}(p_0)$ ... p-hodnota testu hypotézy $p = p_0$ při k úspěších z n pokusů.
- Funkci $\beta_{k,n}(p)$ budeme říkat p-funkce, zkratkou pvf (z p-value function). V literatuře též evidence function, confidence interval function aj.
- Hypotéza $p = p_0$ se zamítá na hladině α, když $\beta_{k,n}(p_0) \leq \alpha$.

Úvod

- $\beta_{k,n}(p_0)$ ... p-hodnota testu hypotézy $p = p_0$ při k úspěších z n pokusů.

- Funkci $\beta_{k,n}(p)$ budeme říkat p-funkce, zkratkou pvf (z p-value function). V literatuře též evidence function, confidence interval function aj.

- Hypotéza $p = p_0$ se zamítá na hladině α, když $\beta_{k,n}(p_0) \leq \alpha$.

- Konfidenční interval (CI) na hladině spolehlivosti $1 - \alpha$ pro k úspěchů z n pokusů:

$$CI(k, n, \alpha) = \{p; \beta_{k,n}(p) > \alpha\}.$$

Úvod

- $\beta_{k,n}(p_0)$ ... p-hodnota testu hypotézy $p = p_0$ při k úspěších z n pokusů.
- Funkci $\beta_{k,n}(p)$ budeme říkat p-funkce, zkratkou pvf (z p-value function). V literatuře též evidence function, confidence interval function aj.
- Hypotéza $p = p_0$ se zamítá na hladině α, když $\beta_{k,n}(p_0) \leq \alpha$.
- Konfidenční interval (CI) na hladině spolehlivosti $1 - \alpha$ pro k úspěchů z n pokusů:

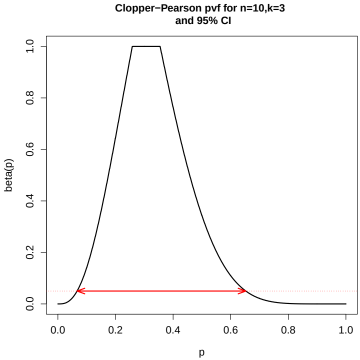
$$CI(k, n, \alpha) = \{p; \beta_{k,n}(p) > \alpha\}.$$

- Takto to ovšem funguje jen v lepším případě (jak uvidíme).

Úvod

- Příklad: Clopper-Pearson (1934):

$$\beta_{k,n}^{CP}(p) = \min\{2P_p(X \leq k), 2P_p(X \geq k), 1\}.$$



Úvod

- Clopper-Pearsonův CI/test je příliš konzervativní: současně

  P_p(celý CI je pod skutečným p) $\leq \alpha/2$,

  P_p(celý CI je nad skutečným p) $\leq \alpha/2$

Úvod

- Clopper-Pearsonův CI/test je příliš konzervativní: současně

  $P_p(\text{celý CI je pod skutečným } p) \leq \alpha/2,$

  $P_p(\text{celý CI je nad skutečným } p) \leq \alpha/2$

- Méně konzervativní exaktní alternativy: Hlídá se jen součet

  $P_p(\ldots\text{pod }\ldots) + P_p(\ldots\text{nad }\ldots) \leq \alpha$

Úvod

- Clopper-Pearsonův CI/test je příliš konzervativní: současně

  P_p(celý CI je pod skutečným p) $\leq \alpha/2$,

  P_p(celý CI je nad skutečným p) $\leq \alpha/2$

- Méně konzervativní exaktní alternativy: Hlídá se jen součet

  $P_p(\ldots \text{pod} \ldots) + P_p(\ldots \text{nad} \ldots) \leq \alpha$

- Nejvýznamnější návrhy:

Úvod

- Clopper-Pearsonův CI/test je příliš konzervativní: současně

  $$P_p(\text{celý CI je pod skutečným } p) \leq \alpha/2,$$

  $$P_p(\text{celý CI je nad skutečným } p) \leq \alpha/2$$

- Méně konzervativní exaktní alternativy: Hlídá se jen součet

  $$P_p(\ldots\text{pod}\ \ldots) + P_p(\ldots\text{nad}\ \ldots) \leq \alpha$$

- Nejvýznamnější návrhy:

  - Sterne (1954), Crow (1956)

Úvod

- Clopper-Pearsonův CI/test je příliš konzervativní: současně

  P_p(celý CI je pod skutečným p) $\leq \alpha/2$,

  P_p(celý CI je nad skutečným p) $\leq \alpha/2$

- Méně konzervativní exaktní alternativy: Hlídá se jen součet

  $P_p(\ldots\text{pod}\ \ldots) + P_p(\ldots\text{nad}\ \ldots) \leq \alpha$

- Nejvýznamnější návrhy:

  - Sterne (1954), Crow (1956)
  - Blyth a Still (1983), Casella (1986)

Úvod

- Clopper-Pearsonův CI/test je příliš konzervativní: současně

$$P_p(\text{celý CI je pod skutečným } p) \leq \alpha/2,$$
$$P_p(\text{celý CI je nad skutečným } p) \leq \alpha/2$$

- Méně konzervativní exaktní alternativy: Hlídá se jen součet

$$P_p(\ldots \text{pod} \ldots) + P_p(\ldots \text{nad} \ldots) \leq \alpha$$

- Nejvýznamnější návrhy:
  - Sterne (1954), Crow (1956)
  - Blyth a Still (1983), Casella (1986)
  - Blaker (2000)

Úvod

- Clopper-Pearsonův CI/test je příliš konzervativní: současně

$$P_p(\text{celý CI je pod skutečným } p) \leq \alpha/2,$$

$$P_p(\text{celý CI je nad skutečným } p) \leq \alpha/2$$

- Méně konzervativní exaktní alternativy: Hlídá se jen součet

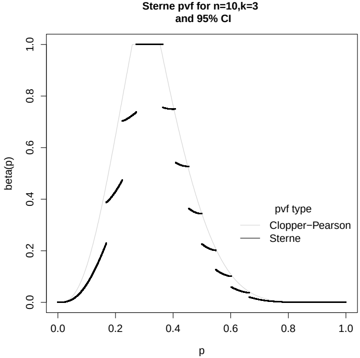
$$P_p(\ldots\text{pod } \ldots) + P_p(\ldots\text{nad } \ldots) \leq \alpha$$

- Nejvýznamnější návrhy:
  - Sterne (1954), Crow (1956)
  - Blyth a Still (1983), Casella (1986)
  - Blaker (2000)
- S nimi ale přicházejí také potíže. . .

Rozpory: I - konfidenční ne-intervaly

- Sterne (1954):

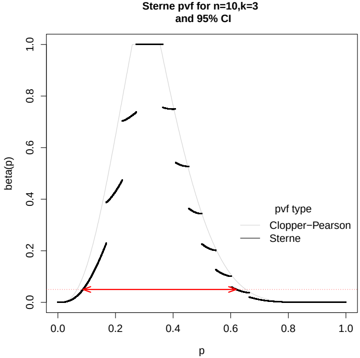
$$\beta^S_{k,n}(p) = \sum_{j \in A} P_p(X = j), \quad A = \{j; P_p(X = j) \leq P_p(X = k)\}.$$



Rozpory: I - konfidenční ne-intervaly

- Sterne (1954):

$$\beta^S_{k,n}(p) = \sum_{j \in A} P_p(X = j), \quad A = \{j; P_p(X = j) \le P_p(X = k)\}.$$



Rozpory: I - konfidenční ne-intervaly

- Crow (1956): $\{p; \beta^S_{k,n}(p) > \alpha\}$ může být nesouvislá množina (funkce $\beta^S_{k,n}(p)$ nejsou unimodální).

Rozpory: I - konfidenční ne-intervaly

- Crow (1956): $\{p; \beta^S_{k,n}(p) > \alpha\}$ může být nesouvislá množina (funkce $\beta^S_{k,n}(p)$ nejsou unimodální).

- Tedy (obecně) přesněji (S $\sim$ set, I $\sim$ interval):

$$
\begin{aligned}
CS(k,n,\alpha) &= \{p; \beta_{k,n}(p) > \alpha\}, \\
CI(k,n,\alpha) &= conv(CS(k,n,\alpha))
\end{aligned}
$$

Rozpory: I - konfidenční ne-intervaly

- Crow (1956): $\{p; \beta^S_{k,n}(p) > \alpha\}$ může být nesouvislá množina (funkce $\beta^S_{k,n}(p)$ nejsou unimodální).

- Tedy (obecně) přesněji (S $\sim$ set, I $\sim$ interval):

$$\begin{aligned} CS(k,n,\alpha) &= \{p; \beta_{k,n}(p) > \alpha\}, \\ CI(k,n,\alpha) &= conv(CS(k,n,\alpha)) \end{aligned}$$

- $CS^S(k,n,\alpha) = \{p; \beta^S_{k,n}(p) > \alpha\}$ mají v součtu přes $k = 0, 1, \ldots, n$ minimální délku (Leb. míru), ale po doplnění na intervalty $CI^S(\ldots)$ optimality pozbývají.

Rozpory: I - konfidenční ne-intervaly

- Blaker (2000):

$$\beta^{B}_{k,n}(p) = \min\{P_p(X \geq k) + P_p(X \leq k^*_p), \\ P_p(X \leq k) + P_p(X \geq k^{**}_p), 1\},$$

$$k^*_p = \max\{j; P_p(X \leq j) \leq P_p(X \geq k\},$$
$$k^{**}_p = \min\{j; P_p(X \geq j) \leq P_p(X \leq k\}.$$

Rozpory: I - konfidenční ne-intervaly

- Blaker (2000):

$$\beta^B_{k,n}(p) = \min\{P_p(X \geq k) + P_p(X \leq k^*_p), \\ P_p(X \leq k) + P_p(X \geq k^{**}_p), 1\},$$

$$k^*_p = \max\{j; P_p(X \leq j) \leq P_p(X \geq k\},$$
$$k^{**}_p = \min\{j; P_p(X \geq j) \leq P_p(X \leq k\}.$$

- Nic se neoptimalizuje, ale $\beta^B(k,n,\alpha) \leq \beta^{CP}(k,n,\alpha)$, takže $CI^B(\dots) \subseteq CI^{CP}(\dots)$.

Rozpory: I - konfidenční ne-intervaly

- Blaker (2000):

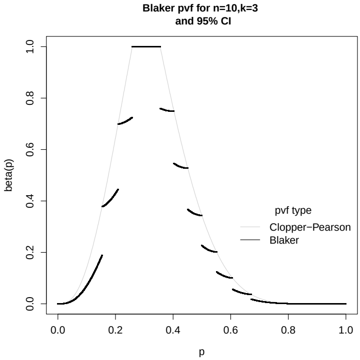
$$\beta^B_{k,n}(p) = \min\{P_p(X \geq k) + P_p(X \leq k^*_p), \\ P_p(X \leq k) + P_p(X \geq k^{**}_p), 1\},$$

$$\begin{aligned} k^*_p &= \max\{j; P_p(X \leq j) \leq P_p(X \geq k\}, \\ k^{**}_p &= \min\{j; P_p(X \geq j) \leq P_p(X \leq k\}. \end{aligned}$$

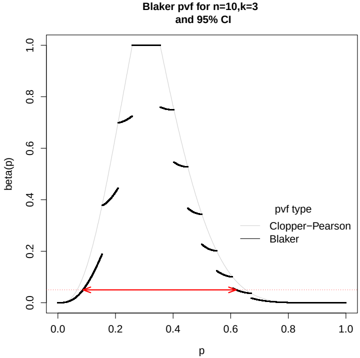
- Nic se neoptimalizuje, ale $\beta^B(k,n,\alpha) \leq \beta^{CP}(k,n,\alpha)$, takže $CI^B(\dots) \subseteq CI^{CP}(\dots)$.

- Funkce $\beta^B_{k,n}(p)$ ovšem také nejsou unimodální – tedy $CS^B(k,n,\alpha) = \{p; \beta^B_{k,n}(p) > \alpha\}$ mohou být nesouvislé a krok $CI^B(\dots) = conv(CS^B(\dots))$ je nutný.

Rozpory: I - konfidenční ne-intervaly



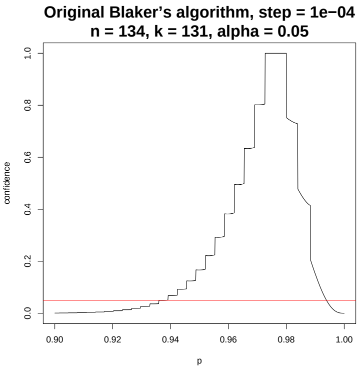
Rozpory: I - konfidenční ne-intervaly



Rozpory: I - konfidenční ne-intervaly

- Ne-unimodální pvf „zlobí“ při příliš jednoduchém numerickém výpočtu konfidenčních mezí.

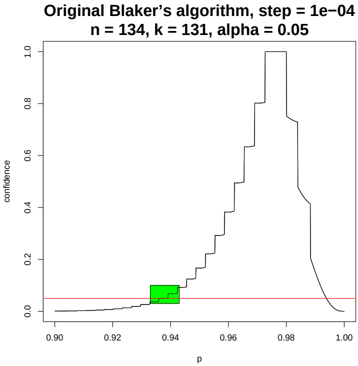
- Hledání s pevným krokem – Blaker (2000) $\sim CI^B$, Reiczigel (199?) $\sim CI^S$.



Rozpory: I - konfidenční ne-intervaly

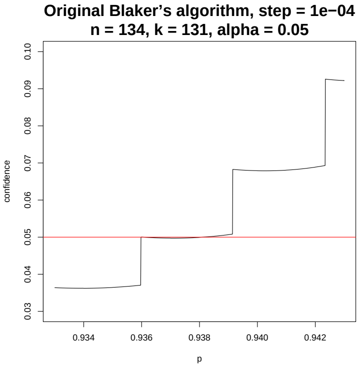
- Ne-unimodální pvf „zlobí“ při příliš jednoduchém numerickém výpočtu konfidenčních mezí.

- Hledání s pevným krokem – Blaker (2000) $\sim CI^B$, Reiczigel (199?) $\sim CI^S$.



Rozpory: I - konfidenční ne-intervaly

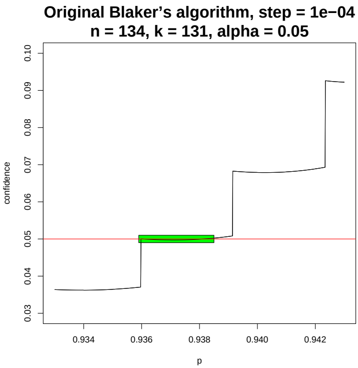
- Ne-unimodální pvf „zlobí“ při příliš jednoduchém numerickém výpočtu konfidenčních mezí.
- Hledání s pevným krokem – Blaker (2000) $\sim CI^B$, Reiczigel (199?) $\sim CI^S$.



Rozpory: I - konfidenční ne-intervaly

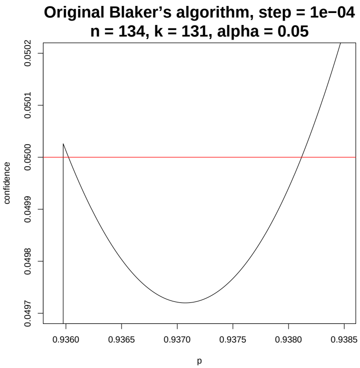
- Ne-unimodální pvf „zlobí“ při příliš jednoduchém numerickém výpočtu konfidenčních mezí.

- Hledání s pevným krokem – Blaker (2000) $\sim CI^B$, Reiczigel (199?) $\sim CI^S$.



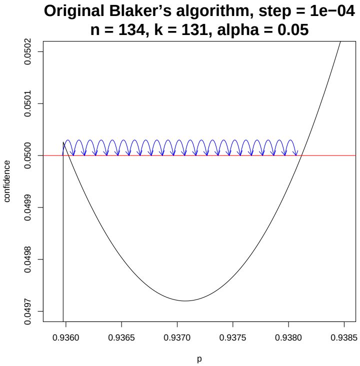
Rozpory: I - konfidenční ne-intervaly

- Ne-unimodální pvf „zlobí“ při příliš jednoduchém numerickém výpočtu konfidenčních mezí.
- Hledání s pevným krokem – Blaker (2000) $\sim CI^B$, Reiczigel (199?) $\sim CI^S$.



Rozpory: I - konfidenční ne-intervaly

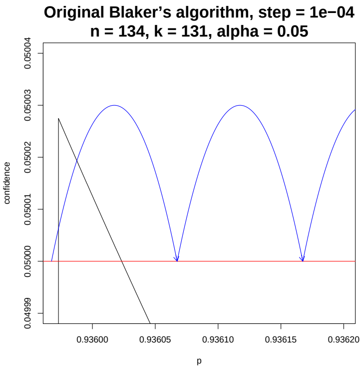
- Ne-unimodální pvf „zlobí“ při příliš jednoduchém numerickém výpočtu konfidenčních mezí.
- Hledání s pevným krokem – Blaker (2000) $\sim CI^B$, Reiczigel (199?) $\sim CI^S$.



Rozpory: I - konfidenční ne-intervaly

- Ne-unimodální pvf „zlobí“ při příliš jednoduchém numerickém výpočtu konfidenčních mezí.

- Hledání s pevným krokem – Blaker (2000) $\sim CI^B$, Reiczigel (199?) $\sim CI^S$.



Rozpory: I - konfidenční ne-intervaly

- Moje parketa: Algoritmus pro výpočet Blakerova CI založený na teoretické analýze průběhu pvf.

Rozpory: I - konfidenční ne-intervaly

- Moje parketa: Algoritmus pro výpočet Blakerova CI založený na teoretické analýze průběhu pvf.
- „Nepřehlédne“ ani velmi krátký interval, kde $\beta^B_{k,n}(p) > \alpha$ (a je rychlý i při vysoké přesnosti).

Rozpory: I - konfidenční ne-intervaly

- Moje parketa: Algoritmus pro výpočet Blakerova CI založený na teoretické analýze průběhu pvf.

- „Nepřehlédne“ ani velmi krátký interval, kde $\beta^{B}_{k,n}(p) > \alpha$ (a je rychlý i při vysoké přesnosti).

- Referoval jsem na ROBUSTu 2010 a 2012.

Rozpory: I - konfidenční ne-intervaly

- Moje parketa: Algoritmus pro výpočet Blakerova CI založený na teoretické analýze průběhu pvf.
- „Nepřehlédne“ ani velmi krátký interval, kde $\beta^B_{k,n}(p) > \alpha$ (a je rychlý i při vysoké přesnosti).
- Referoval jsem na ROBUSTu 2010 a 2012.
- Tech. zpráva č. 1099 ÚI AV (dostupná online).

Rozpory: I - konfidenční ne-intervaly

- Moje parketa: Algoritmus pro výpočet Blakerova CI založený na teoretické analýze průběhu pvf.
- „Nepřehlédne“ ani velmi krátký interval, kde $\beta^B_{k,n}(p) > \alpha$ (a je rychlý i při vysoké přesnosti).
- Referoval jsem na ROBUSTu 2010 a 2012.
- Tech. zpráva č. 1099 ÚI AV (dostupná online).
- Implementace: `BlakerCI` – balíček v R (1. verze 2010).

Rozpory: I - konfidenční ne-intervaly

- Moje parketa: Algoritmus pro výpočet Blakerova CI založený na teoretické analýze průběhu pvf.
- „Nepřehlédne“ ani velmi krátký interval, kde $\beta^B_{k,n}(p) > \alpha$ (a je rychlý i při vysoké přesnosti).
- Referoval jsem na ROBUSTu 2010 a 2012.
- Tech. zpráva č. 1099 ÚI AV (dostupná online).
- Implementace: `BlakerCI` – balíček v R (1. verze 2010).
- Podobný algoritmus pro Sterneho CI navrhl (pracovně) Reiczigel – snad o něm v připravované práci Reiczigel, Klaschka & Thulin.

Rozpory: II - ne-vnořenost CI

- Blyth a Still (1983), Casella (1986): Konfidenční intervaly minimální délky (v součtu přes $k = 0, 1, \ldots, n$).

Rozpory: II - ne-vnořenost CI

- Blyth a Still (1983), Casella (1986): Konfidenční intervaly minimální délky (v součtu přes $k = 0, 1, \ldots, n$).
- Intervaly bohužel nejsou „vnořené“ (nested): Neplatí

$$\alpha_1 \geq \alpha_2 \Rightarrow CI^{BSC}(k, n, \alpha_1) \subseteq CI^{BSC}(k, n, \alpha_2).$$

Rozpory: II - ne-vnořenost CI

- Blyth a Still (1983), Casella (1986): Konfidenční intervaly minimální délky (v součtu přes $k = 0, 1, \ldots, n$).

- Intervaly bohužel nejsou „vnořené“ (nested): Neplatí

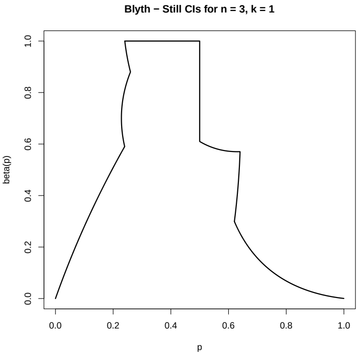
$$\alpha_1 \geq \alpha_2 \Rightarrow CI^{BSC}(k, n, \alpha_1) \subseteq CI^{BSC}(k, n, \alpha_2).$$

- Takové intervaly nelze odvodit z pvf.

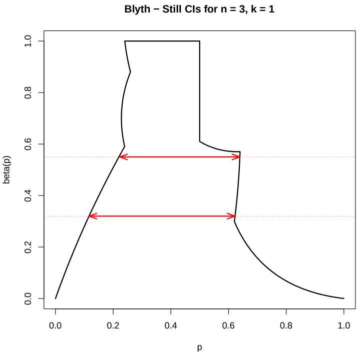
Rozpory: II - ne-vnořenost CI

- Blyth a Still (1983), Casella (1986): Konfidenční intervaly minimální délky (v součtu přes $k = 0, 1, \ldots, n$).
- Intervaly bohužel nejsou „vnořené“ (nested): Neplatí
$$\alpha_1 \geq \alpha_2 \Rightarrow CI^{BSC}(k, n, \alpha_1) \subseteq CI^{BSC}(k, n, \alpha_2).$$
- Takové intervaly nelze odvodit z pvf.
- Co s tím? Nepoužívat? „Na smetišti dějin“ tyto intervaly stále nejsou – viz např. SW `StatXact` C. Mehty.

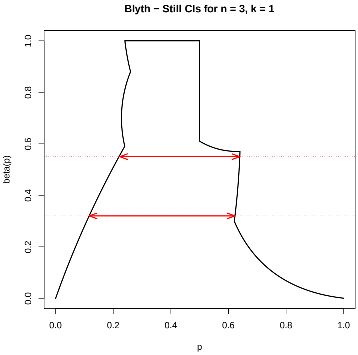
Rozpory: II - ne-vnořenost CI



Rozpory: II - ne-vnořenost CI



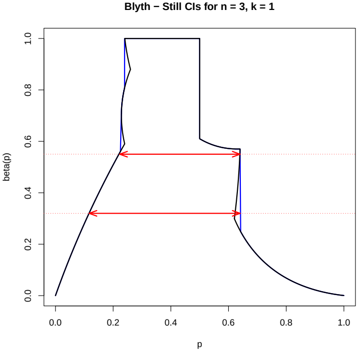
Rozpory: II - ne-vnořenost CI



- Lze zavést „opravu“

$$CI'^{BSC}(k,n,\alpha) = \bigcup_{\alpha' \geq \alpha} CI^{BSC}(k,n,\alpha')$$

Rozpory: II - ne-vnořenost CI



- Lze zavést „opravu“

$$CI'^{BSC}(k,n,\alpha) = \bigcup_{\alpha' \geq \alpha} CI^{BSC}(k,n,\alpha')$$

Rozpory: III - špatně spárované typy testu a CI

- Fay (2009): Mnohý standardní statistický SW (včetně renomovaného) dává vzájemně si odporující inference – např. frekvence jevu ve dvou skupinách se významně liší, ale CI pro poměr šancí (odds ratio) obsahuje 1.

Rozpory: III - špatně spárované typy testu a CI

- Fay (2009): Mnohý standardní statistický SW (včetně renomovaného) dává vzájemně si odporující inference – např. frekvence jevu ve dvou skupinách se významně liší, ale CI pro poměr šancí (odds ratio) obsahuje 1.
- Způsobeno je to použitím vzájemně si neodpovídajících typů testu a CI.

Rozpory: III - špatně spárované typy testu a CI

- Fay (2009): Mnohý standardní statistický SW (včetně renomovaného) dává vzájemně si odporující inference – např. frekvence jevu ve dvou skupinách se významně liší, ale CI pro poměr šancí (odds ratio) obsahuje 1.
- Způsobeno je to použitím vzájemně si neodpovídajících typů testu a CI.
- Fay vytvořil balíky `exactci` a `exact2x2` v R – nabízí v nich pro 1- a 2-výběrovou úlohu správně spárované testy a CI typů à la Clopper-Pearson, Sterne a Blaker.

Rozpory: IV - „nevyhnutelné“ nekonsistence

- Fay (2009, 2010) nabízí nástroj k odstranění většiny rozporů mezi výsledky testů a CI, ale zbývají podle něj (byť řidší) „nevyhnutelné“ nekonsistence (unavoidable inconsistencies).

Rozpory: IV - „nevyhnutelné“ nekonsistence

- Fay (2009, 2010) nabízí nástroj k odstranění většiny rozporů mezi výsledky testů a CI, ale zbývají podle něj (byť řidší) „nevyhnutelné“ nekonsistence (unavoidable inconsistencies).
- I u správně spárovaného testu a CI se může stát, že test hypotézy $p = p_0$ vyjde na hladině α významně (tj. $\beta_{k,n}(p_0) \leq \alpha$), ale $p_0 \in CI(k, n, \alpha)$.

Rozpory: IV - „nevyhnutelné“ nekonsistence

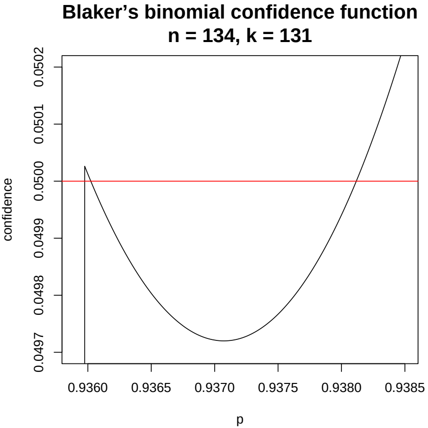
- Fay (2009, 2010) nabízí nástroj k odstranění většiny rozporů mezi výsledky testů a CI, ale zbývají podle něj (byť řidší) „nevyhnutelné“ nekonsistence (unavoidable inconsistencies).

- I u správně spárovaného testu a CI se může stát, že test hypotézy $p = p_0$ vyjde na hladině α významně (tj. $\beta_{k,n}(p_0) \leq \alpha$), ale $p_0 \in CI(k, n, \alpha)$.

- Čím to: $p_0 \in CI(k, n, \alpha) \setminus CS(k, n, \alpha)$.

Rozpory: IV - „nevyhnutelné“ nekonsistence

- Příklad „nevyhnutelné“ nekonsistence:



- Bod $p = 0{,}937$ patří do Blakerova CI, ale „odpovídající“ test hypotézu $H_0 : p = 0{,}937$ zamítá

Rozpory: IV - „nevyhnutelné“ nekonsistence

- „Nevyhnutelnému“ se lze vyhnout tzv. unimodalizací pvf – nahrazením nejmenší větší unimodální funkcí:

$$\widetilde{\beta}_{k,n}(p) = \begin{cases} \sup_{p' \leq p} \beta_{k,n}(p'), & p \leq k/n \\ \sup_{p' \geq p} \beta_{k,n}(p'), & p > k/n \end{cases}$$

Rozpory: IV - „nevyhnutelné“ nekonsistence

- „Nevyhnutelnému“ se lze vyhnout tzv. unimodalizací pvf – nahrazením nejmenší větší unimodální funkcí:

$$\widetilde{\beta}_{k,n}(p) = \begin{cases} \sup_{p' \leq p} \beta_{k,n}(p'), & p \leq k/n \\ \sup_{p' \geq p} \beta_{k,n}(p'), & p > k/n \end{cases}$$

- Pro Blakerův test implementováno v balíčku `BlakerCI`. Pro Sterneho test máme pracovní návrh algoritmu od J. Reiczigela.

Rozpory: IV - „nevyhnutelné“ nekonsistence

- „Nevyhnutelnému“ se lze vyhnout tzv. unimodalizací pvf – nahrazením nejmenší větší unimodální funkcí:

$$\widetilde{\beta}_{k,n}(p) = \begin{cases} \sup_{p' \leq p} \beta_{k,n}(p'), & p \leq k/n \\ \sup_{p' \geq p} \beta_{k,n}(p'), & p > k/n \end{cases}$$

- Pro Blakerův test implementováno v balíčku `BlakerCI`. Pro Sterneho test máme pracovní návrh algoritmu od J. Reiczigela.

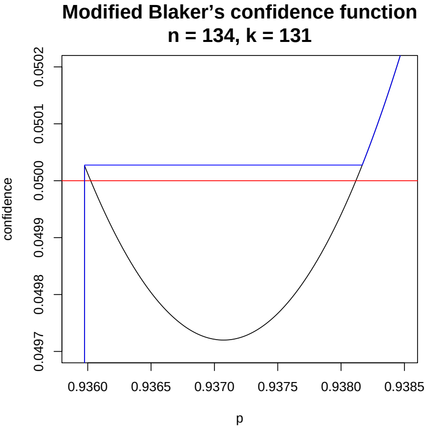
- V obou případech

$$\widetilde{\beta}_{k,n}(p) = \max(\beta_{k,n}(p), \beta_{k,n}(p^*)),$$

kde p^* je pro $p \leq x/n$ ($p > x/n$) nejbližší bod nespojitosti pvf vlevo (vpravo) od p.
(Pro Blakera důkaz triviální, se Sternem jsme se zapotili.)

Rozpory: IV - „nevyhnutelné“ nekonsistence

- Příklad unimodalizace:



- Modifikovaný test v souhlasu s Blakerovým CI hypotézu $H_0 : p = 0{,}937$ nezamítá

Rozpory: V - nekonsistentní inference při různém n

- Přednášel jsem o tom na ROBUSTu 2014, teď jen nejstručněji.

Rozpory: V - nekonsistentní inference při různém n

- Přednášel jsem o tom na ROBUSTu 2014, teď jen nejstručněji.
- Vos a Hudson (2009) upozornili, že u některých testů / CI (včetně Blakerova a Sterneho) si výsledky pro různá n vzájemně odporují.

Rozpory: V - nekonsistentní inference při různém n

- Přednášel jsem o tom na ROBUSTu 2014, teď jen nejstručněji.

- Vos a Hudson (2009) upozornili, že u některých testů / CI (včetně Blakerova a Sterneho) si výsledky pro různá n vzájemně odporují.

- Např. test při k úspěších v n pokusech zamítne hypotézu $p = p_0$, ale nezamítne ji při K úspěších v $N > n$ pokusech, přestože $K/N \geq k/n > p_0$ (popř. $K/N \leq k/n < p_0$).

Rozpory: V - nekonsistentní inference při různém n

- Přednášel jsem o tom na ROBUSTu 2014, teď jen nejstručněji.
- Vos a Hudson (2009) upozornili, že u některých testů / CI (včetně Blakerova a Sterneho) si výsledky pro různá n vzájemně odporují.
- Např. test při k úspěších v n pokusech zamítne hypotézu $p = p_0$, ale nezamítne ji při K úspěších v $N > n$ pokusech, přestože $K/N \geq k/n > p_0$ (popř. $K/N \leq k/n < p_0$).
- Pro Blakerův CI / test mám opravu („Vos – Hudson adjustment“), která toto nelogické chování potlačí – je implementovaná v balíku `BlakerCI` od verze ze srpna 2015.

Rozpory: V - nekonsistentní inference při různém n

- Přednášel jsem o tom na ROBUSTu 2014, teď jen nejstručněji.
- Vos a Hudson (2009) upozornili, že u některých testů / CI (včetně Blakerova a Sterneho) si výsledky pro různá n vzájemně odporují.
- Např. test při k úspěších v n pokusech zamítne hypotézu $p = p_0$, ale nezamítne ji při K úspěších v $N > n$ pokusech, přestože $K/N \geq k/n > p_0$ (popř. $K/N \leq k/n < p_0$).
- Pro Blakerův CI / test mám opravu („Vos – Hudson adjustment"), která toto nelogické chování potlačí – je implementovaná v balíku `BlakerCI` od verze ze srpna 2015.
- Obdobnou opravu Pro Sterneho CI / test nemám ještě promyšlenou po výpočetní stránce.

Rozpory: V - nekonsistentní inference při různém *n*

- Poznámka #1: S J. Reiczigelem na tomto poli nespolupracujeme – na článek Vos & Hudson (2009) má vyhraněně negativní názor.

Rozpory: V - nekonsistentní inference při různém n

- Poznámka #1: S J. Reiczigelem na tomto poli nespolupracujeme – na článek Vos & Hudson (2009) má vyhraněně negativní názor.
- Poznámka #2: U Clopper-Pearsonova CI / testu se daný typ rozporu nevyskytuje. Intuitivně je to zřejmé a mám to dokázané (stěží jako první), ale dalo to zabrat a klíčovou ingrediencí důkazu je netriviální výsledek W. Hoeffdinga z r. 1956.

Závěr: I – nadšenecký

- Rozpory všech 5 uvedených typů jsou řešitelné.
  (Záleží jen, za jakou cenu. . .)

Závěr: I – nadšenecký

- Rozpory všech 5 uvedených typů jsou řešitelné.
  (Záleží jen, za jakou cenu. . .)

- Vždy je třeba a) přesvědčit, že oprava / náprava je žádoucí,
  b) poskytnout k tomu nástroj (SW).

Závěr: I – nadšenecký

- Rozpory všech 5 uvedených typů jsou řešitelné.
  (Záleží jen, za jakou cenu. . .)

- Vždy je třeba a) přesvědčit, že oprava / náprava je žádoucí,
  b) poskytnout k tomu nástroj (SW).

- Všeobecně uznávaná je potřeba opravy / nápravy asi jen
  v případě nesouvislých konfidenčních množin – doplnění
  na interval se bere jako samozřejmost.

Závěr: I – nadšenecký

- Rozpory všech 5 uvedených typů jsou řešitelné.
(Záleží jen, za jakou cenu. . .)

- Vždy je třeba a) přesvědčit, že oprava / náprava je žádoucí,
b) poskytnout k tomu nástroj (SW).

- Všeobecně uznávaná je potřeba opravy / nápravy asi jen v případě nesouvislých konfidenčních množin – doplnění na interval se bere jako samozřejmost.

- Pak by ale stejnou samozřejmostí mělo být také upravovat („unimodalizovat“) pvf, aby test souhlasil s CI. Používat pvf bez opravy je ale jednodušší. . . (Např. Fay zřejmě o úpravě pvf vůbec neuvažoval, když rozpor mezi testem a CI prohlásil za nevyhnutelný.)

Závěr: II – skeptický

- Jestliže důsledná aplikace nějakého principu (zde konkrétně principu konstrukce CI / testu, ale platí to i obecněji) má důsledky odporující zdravému rozumu a vyvolává potřebu dodatečných oprav / úprav, může to vést k pochybnostem o smysluplnosti principu jako takového.

Závěr: II – skeptický

- Jestliže důsledná aplikace nějakého principu (zde konkrétně principu konstrukce CI / testu, ale platí to i obecněji) má důsledky odporující zdravému rozumu a vyvolává potřebu dodatečných oprav / úprav, může to vést k pochybnostem o smysluplnosti principu jako takového.

Děkuji za pozornost.