2.) $f(x;\vartheta) = \binom{100}{x} \cdot \vartheta^{x}(1-\vartheta)^{100-x}\, \mathbb{1}_{\{0,1,2,\dots,100\}}(x)$ $\qquad \vartheta \in (0,1) \subset \mathbb{R}$

$\Rightarrow f(x;\vartheta) = \begin{cases} \binom{100}{x} \cdot \vartheta^{x}(1-\vartheta)^{100-x} ; & x \in \{0,1,2,\dots,100\} \\ 0 ; & \text{sicer} \end{cases} \sim \mathrm{Bin}(n,\vartheta)$ za $n=100$

" $\mathrm{Bin}(100,\vartheta)$ $\qquad \vartheta \in (0,1)$

a) Kompletna zadostna statistika?

Vemo: Če $f(\vec{x},\vartheta)$ pripada eksponentni družini, torej če lahko f zapišemo kot

$f(\vec{x};\vartheta) = C(\vartheta)\, e^{\langle T(\vec{x}), Q(\vartheta)\rangle} h(\vec{x})$ in (je $Q(\vartheta)$ bijekcija) ima eksponentna družina poln rang, potem je $T(\vec{x})$ kompletna zadostna statistika.

$$f(\vec{x},\vartheta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \binom{100}{x_i} \cdot \vartheta^{x_i}(1-\vartheta)^{100-x_i} = \prod_{i=1}^{n} \binom{100}{x_i}(1-\vartheta)^{100} \cdot \left(\frac{\vartheta}{1-\vartheta}\right)^{x_i} =$$

$$= (1-\vartheta)^{100} \prod_{i=1}^{n} \binom{100}{x_i} \cdot e^{\sum_{i=1}^{n} x_i \ln\left(\frac{\vartheta}{1-\vartheta}\right)} =$$

$$= \underbrace{(1-\vartheta)^{100}}_{C(\vartheta)} e^{\langle \underbrace{\ln \frac{\vartheta}{1-\vartheta}}_{Q(\vartheta)}, \underbrace{\sum_{i=1}^{n} x_i}_{T(\vec{x})} \rangle} \cdot \underbrace{\prod_{i=1}^{n} \binom{100}{x_i}}_{h(\vec{x})}$$

NR $\rightarrow$ $= 6627$

Ker je $Q(\vartheta) = \ln\left(\frac{\vartheta}{1-\vartheta}\right)$ bijekcija, je $T(\vec{x}) = \sum_{i=1}^{n} x_i$ kompletna zadostna statistika.

b) Preizkusi domnev $\vartheta \le \vartheta_0$ proti $\vartheta > \vartheta_0$?

Imamo enoparametrični eksponentni model z gostoto $f(\vec{x},\vartheta) = C(\vartheta) e^{\langle Q(\vartheta), T(\vec{x})\rangle} h(\vec{x})$

Preizkusi, ki jih poznamo za preizkušanje naše domneve so:

nerandomizirani preizkusi, randomizirani preizkusi oblike: $\phi : \mathbb{R}^n \to [0,1]$ pri realizaciji $X = x$

$\phi(x) = \begin{cases} 1; & H \text{ zavrnemo} \\ 0; & H \text{ ne zavrnemo} \\ \gamma \in (0,1); & \text{vržemo kovanec z verjetnostjo cifre} = \gamma \end{cases}$ (H zavrnemo z verjetnostjo γ)

in H zavrnemo, če pade cifra oziroma ne zavrnemo, če pade grb.

preizkus na podlagi razmerja verjetij iz katerega sledi tudi preizkus po N-P lemi, ki nam pove, da obstaja enolični preizkus H proti A velikosti α, ki je najmočnejši med vsemi preizkusi H proti A st. znač. α in ima obliko $\phi(x) = \begin{cases} 1; & f_1(x) > D f_0(x) \\ \gamma; & f_1(x) = D f_0(x) \\ 0; & f_1(x) < D f_0(x) \end{cases}$

Ker imamo v našem primeru enoparametrični eksponentni model pa vemo (po STAT1), da za preizkus $H = \{\vartheta \mid \vartheta \le \vartheta_0\}$ proti $A = \{\vartheta \mid \vartheta > \vartheta_0\}$ obstaja preizkus velikosti α, ki je enakomerno najmočnejši med vsemi preizkusi stopnje značilnosti α in ima N-P obliko.

Ker je v našem primeru funkcija $Q(\vartheta) = \ln\left(\frac{\vartheta}{1-\vartheta}\right)$ za $\vartheta \in (0,1)$ naraščajoča, ima preizkus obliko:

$$\phi(\vec{x}) = \begin{cases} 1; & T(\vec{x}) > C \\ \gamma; & T(\vec{x}) = C \\ 0; & T(\vec{x}) < C \end{cases}$$

c) $H_0: \vartheta \leq 0.2$, najboljši preizkus z $\alpha = 0.05$, $n = 100$, $m = 300$, $N = n \cdot m = 30000$

Izberemo zadnji naveden preizkus iz točke b), saj imamo enoparametrični eksponentni model, preizkus pa bo tako enakomerno najmočnejši med vsemi preizkusi H_0 proti A st. znač. α. Ker je $Q(\vartheta)$ naraščajoča imamo torej preizkus oblike

$$\phi(\vec{x}) = \begin{cases} 1; & T(\vec{x}) > c \\ \gamma; & T(\vec{x}) = c \\ 0; & T(\vec{x}) < c \end{cases}, \text{ kjer je } T(\vec{x}) = \sum_{i=1}^{m} X_i = 6627$$

γ in c dobimo iz zahteve: $E_{\vartheta_0}(\phi) = \alpha = 0.05$

$$E_{\vartheta_0}(\phi) = 1 \cdot P(\{\sum_{i=1}^{m} X_i > c\}) + \gamma \cdot P(\{\sum_{i=1}^{m} X_i = c\}) = \alpha = 0.05$$

Imamo binomsko porazdelitev: $B(n,p) = B(100, \vartheta) \sim X_i$

Sledi $T(\vec{x}) = \sum_{i=1}^{m} X_i \sim Bin(m \cdot 100, \vartheta) = Bin(30000, \vartheta)$ (vsota Bin je spet Bin porazdelitev)

Določiti moramo funkcijo moči: $\beta(\vartheta)$

1. način: $\beta(\vartheta) = P_\vartheta(T > c) = \sum_{k=c+1}^{N} \binom{N}{k} p^k (1-p)^{N-k} \leq \alpha \Rightarrow \beta(\vartheta_0) \leq \alpha$

2. način: $\beta(\vartheta) = \sum_{k=c+1}^{N} P(\cancel{X}\sum X_i = k) + \gamma \cdot P(\cancel{X}\sum X_i = c) = \alpha$

Uporabimo 2. način, poračunamo v R s funkcijo.

$c = 6114$

$$\gamma = \frac{\alpha - \sum_{k=c+1}^{N} P(\cancel{X} = k)}{P(\cancel{X} = c)} = \frac{\alpha - pbinom(\cancel{X} = c)}{dbinom(\cancel{X} = c)} = 0.3689786$$

3. način: $Bin(N, \vartheta)$ aproksimiramo z normalno porazdelitvijo:

Vemo, da če je n dovolj velik, izkrivljenost (skewness) binomske porazdelitve ni prevelika. Zato lahko aproksimiramo $B(n,p) \approx N(np, np(1-p))$

Test, kaj pomeni "dovolj velik n":

$$\frac{1}{\sqrt{n}} \left| \sqrt{\frac{1-p}{p}} - \sqrt{\frac{p}{1-p}} \right| < 0.3 \Rightarrow n = 30.000, p = 0.2: \quad 0.0086 < 0.3 \checkmark \text{ je OK}$$

Ugotovimo, da za naše parametre ~~ta~~ aproksimacija ustreza.

Iščemo torej c, da bo $P(N(n \cdot \vartheta, n \cdot \vartheta(1-\vartheta)) > c) = \alpha$

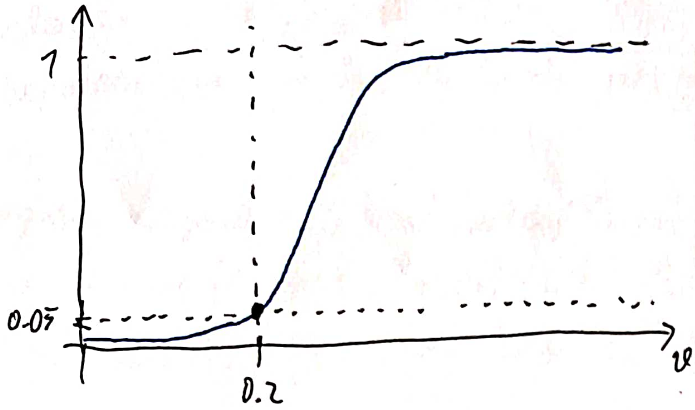
$$\Rightarrow P\left(N(0,1) > \frac{c - n \cdot \vartheta}{\sqrt{n \cdot \vartheta(1-\vartheta)}}\right) = \alpha \Rightarrow 1 - \phi\left(\frac{c - n\vartheta}{\sqrt{n \cdot \vartheta(1-\vartheta)}}\right) = \alpha \Rightarrow \phi^{-1}(1-\alpha) = \frac{c - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}$$

$\Rightarrow$ Sledi, da je $c = \phi^{-1}(1-\alpha)\sqrt{n\vartheta(1-\vartheta)} + n\vartheta$

Vstavimo $n = 30.000$, $\vartheta = 0.2$ in $\alpha = 0.05$, poračunamo v R in dobimo:

$c = 6113.9588 \approx 6114$ in $\gamma = 0.3689786$, kar je enak rezultat kot prej.

Skicirajmo še funkcijo moči (natančen graf je narisan v R):



d) Interval zaupanja za naš vzorec

Uporabimo izrek o pivotnih funkcijah v enoparametričnem eksponentnem modelu.

$X_i \sim Bin(100, \vartheta)$, $T(\vec{X}) = \sum_{i=1}^{m} X_i$, $m = 300$

$F_{T;\vartheta}$ je pridružena k.p.l. $T(\vec{X}) \Rightarrow T(\vec{X}) \sim Bin(30.000, \vartheta)$

$F_{T;\vartheta}(t)$ je nenaraščajoča v ϑ za fiksen t, vendar ni zvezna, torej uporabimo (ii) točko izreka iz predavanj.

Definiramo $\bar{\vartheta} = \sup\{\vartheta \mid F_{T;\vartheta}(x) \geq \alpha_1\}$ in $\underline{\vartheta} = \inf\{\vartheta \mid F_{T;\vartheta}(x^-) \leq 1-\alpha_2\}$, kjer je $\alpha = \alpha_1 + \alpha_2$.

Velja: $[\underline{\vartheta}(TX), \bar{\vartheta}(TX)]$ je I.Z. stopnje zaupanja α za ϑ.

~~Rešitev~~

$\bar{\vartheta}$ je največji tak ϑ, za katerega bo $F_{T,\vartheta}(x) \geq \alpha_1$

Vzemimo $\alpha_1 = \alpha_2 \rightarrow \alpha = 2\alpha_1 \Rightarrow \alpha_1 = \alpha_2 = \frac{\overset{0.05}{\alpha}}{2} = 0.025$

$\Rightarrow$ To bo I.Z. st. zaup 0.05

Potrebujemo I.Z. st. zaup. 0.95

$\Rightarrow \alpha = 0.95$

$\alpha_1 = \alpha_2 = 0.475$

Za zgornjo mejo potrebujemo torej tak ϑ, da bo $F_{T;\vartheta}(x) \geq 0.025 \Rightarrow \bar{\vartheta}$

0.475

Za spodnjo mejo pa prvi tak ϑ, da bo $F_{T_1;\vartheta}(x^-) \leq 1-0.025 = 0.975 \Rightarrow \underline{\vartheta}$

$1-0.475 = 0.525$

Izračunamo s funkcijo v R in dobimo:

I.Z.: $[0.2163, 0.2256]$ $\alpha=0.05$

I.Z. $\alpha=0.95$: $[0.2208, 0.2210]$

Ker je naša porazdelitev $Bin(n, \vartheta)$, mora biti I.Z. enak kot Clopper-Pearsonov I.Z.

C-P I.Z. izračunan v R z vgrajeno metodo: $[0.2162161, 0.225638]$. $\alpha=0.05$

$[0.2207363, 0.221107]$. $\alpha=0.95$

Vidimo, da se intervala dobro ujemata.

Interval z višjo stopnjo zaupanja (0.95) je seveda ožji.