confidence intevral

Подсећање:

Када смо тражили интервал поверења за m на основу узорка из $\mathcal{N}(m, \sigma^2)$, где је σ^2 познато, користили смо стожерну величину:

$$\boxed{\frac{\overline{X}_n - m}{\sigma}\sqrt{n}}$$

и показали смо да она има $\mathcal{N}(0,1)$ расподелу.
Како бисмо нашли интервал поверења за m да нам σ^2 није познато?
Не можемо користити исти стожер, јер се у њему појављује σ, па је потребно да осмислимо неки други.

деф. Случајна величина чија ф-ја густине гласи:

$$f(x) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}, \quad x>0$$

има χ^2 расподелу са n степени слободе

Тврђење: Ако су $X_1, \dots, X_n$ независне сл. вел. са $\mathcal{N}(0,1)$ расподелом, тада сл. вел. $\chi^2_n := X_1^2 + \dots + X_n^2$ има χ^2 расподелу са n степени слободе.

Својства: $E\chi^2_n = n, \quad D\chi^2_n = 2n$

Доказ: Видимо да је χ^2 расподела заправо $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ расподела, па знамо да је $E\chi^2_n = \frac{\frac{n}{2}}{\frac{1}{2}} = n$, $D\chi^2_n = \frac{\frac{n}{2}}{\left(\frac{1}{2}\right)^2} = 2n$

Други начин да то покажемо био би следећи:

$$E\chi^2_n = E(X_1^2 + \dots + X_n^2) = nEX_1^2 = n(DX_1 + \underbrace{(EX_1)^2}_{0}) = nDX_1 = n$$

Теорема 1: $(X_1, \dots, X_n)$ ПСУ из $\mathcal{N}(m, \sigma^2)$. Тада

$$\sum_{i=1}^{n} \frac{(X_i - \overline{X}_n)^2}{\sigma^2} \sim \chi^2_{n-1}$$

Теорема 2: $(X_1, \dots, X_n)$ ПСУ из $\mathcal{N}(m, \sigma^2)$. Тада су $\overline{X}_n$ и $\tilde{S}^2$ независне.

Овај резултат се добија применом Басуове теореме, о којој ће бити више речи на математичкој статистици.

последица: $\frac{\overline{X}_n - m}{\sigma}\sqrt{n}$ и $\frac{n-1}{\sigma^2} \cdot \tilde{S}^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X}_n)^2}{\sigma^2}$

су независне сл. величине.

Тврђење: Ако $X \sim \mathcal{N}(0,1)$, $Y \sim \chi^2_n$ и независне су, тада $\frac{X}{\sqrt{\frac{Y}{n}}} \sim t_n$ (студентова расподела са n степени слободе)

последица:

$$\frac{\frac{\overline{X}_n - m}{\cancel{\sigma}}\sqrt{n}}{\sqrt{\cancel{\frac{(n-1)}{\sigma^2}}\frac{\tilde{S}_n^2}{\cancel{n-1}}}} = \frac{\overline{X}_n - m}{\tilde{S}_n}\sqrt{n} \sim t_{n-1}$$

Сада видимо да бисмо могли $\boxed{\frac{\overline{X}_n - m}{\tilde{S}_n}\sqrt{n}}$ да користимо као стожерну величину.

25. $\overline{x}_n = 2{,}6$, $\tilde{s}_n^2 = 1{,}79$, $1-\alpha = 0{,}95$

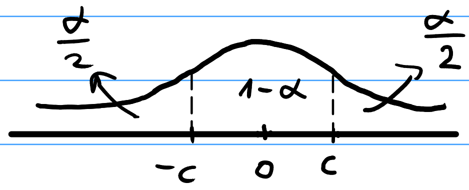
Тражимо U_n и V_n т.д.

$$P\{U_n \le m \le V_n\} = 1-\alpha$$

Рекли смо да је згодно поћи од стожера. Како је студентова расподела попут нормалне симетрична,

U_n и V_n тежимо наћи на исти начин као и у 24. зад. Дакле, тражимо $c \in \mathbb{R}$ т.д.

$$(*) \qquad P\left\{-c \le \frac{\overline{X}_n - m}{\widetilde{S}_n}\sqrt{n} \le c\right\} = 1-\alpha$$



$F_{t_{n-1}}$ - ф-ја расподеле за t_{n-1} расподелу

За c важи: $F_{t_{n-1}}(c) = 1-\frac{\alpha}{2} \Rightarrow c = F^{-1}_{t_{n-1}}\left(1-\frac{\alpha}{2}\right)$

$$= qt\left(1-\frac{\alpha}{2}, n-1\right)$$

у R-у

Код нас $n = 25$, $\alpha = 1 - 0{,}95 = 0{,}05$

$$\Rightarrow c = 2{,}063899$$

Када смо нашли c, преостаје нам да трансформишемо израз $(*)$:

$$(*) \qquad P\left\{-c \le \frac{\overline{X}_n - m}{\widetilde{S}_n}\sqrt{n} \le c\right\} = 1-\alpha$$

$$\Leftrightarrow P\left\{\overline{X}_n - c\frac{\widetilde{S}_n}{\sqrt{n}} \le m \le \overline{X}_n + c\cdot\frac{\widetilde{S}_n}{\sqrt{n}}\right\} = 1-\alpha$$

$$\Leftrightarrow U_n = \overline{X}_n - c\frac{\widetilde{S}_n}{\sqrt{n}}, \quad V_n = \overline{X}_n + c\frac{\widetilde{S}_n}{\sqrt{n}}$$

Преостаје нам још да заменимо $\overline{x}_n$, $\widetilde{s}_n$, c и n и добијамо тражени интервал.

Шта бисмо радили да нам није дато $\widetilde{s}_n^2$, него $\overline{s}_n^2$?

$\overline{S}_n^2 = \frac{n-1}{n}\widetilde{S}_n^2$, па је стожер:

$$\frac{\overline{X}_n - m}{\frac{\cancel{\sqrt{n}}}{\sqrt{n-1}}\overline{S}_n}\cancel{\sqrt{n}} = \boxed{\frac{\overline{X}_n - m}{\overline{S}_n}\sqrt{n-1}}$$

26. ПСУ из $\mathcal{N}(m,\sigma^2)$, m непознато, дато n и $\tilde{S}_n^2$
99% интервал поверења за σ^2?

стожер: $\frac{(n-1)}{\sigma^2}\tilde{S}_n^2 \sim \chi^2_{n-1}$

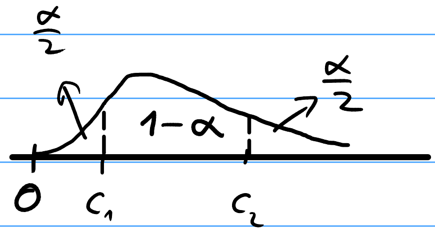
Тражимо c_1 и c_2 т.д.:

$$P\left\{c_1 \le \frac{(n-1)\tilde{S}_n}{\sigma^2} \le c_2\right\} = 1-\alpha = 0,99$$

$$\iff P\left\{\frac{(n-1)\tilde{S}_n^2}{c_2} \le \sigma^2 \le (n-1)\frac{\tilde{S}_n^2}{c_1}\right\} = 1-\alpha$$

$$\iff \left(\frac{(n-1)\tilde{S}_n^2}{c_2}, \frac{(n-1)\tilde{S}_n^2}{c_1}\right) \text{ је 99\% инт. пов. за } \sigma^2$$

χ^2_{n-1} расподела није симетрична, како изабрати c_1 и c_2?
Не можемо изабрати нпр $c_2=-c_1$, јер је носач расподеле $(0,+\infty)$:



обично узимамо c_1 и c_2 т.д. важи

$F_{\chi^2_{n-1}}(c_1)=\frac{\alpha}{2}$

$F_{\chi^2_{n-1}}(c_2)=1-\frac{\alpha}{2}$

густина χ^2 расподеле са бар 3 степена слободе (код нас је n углавном велики број, па ће бити баш овај облик)

$$c_1 = F^{-1}_{\chi^2_{19}}\left(\frac{\alpha}{2}\right) = qchisq\left(\frac{\alpha}{2}, 19\right) = 6,843971$$

$$c_2 = qchisq\left(1-\frac{\alpha}{2}, 19\right) = 38,58226$$

Сада нам преостаје да заменимо познате вредности и добијемо тражени интервал.

27. X - обележје које нам каже да ли је јединка против гушења у затвореном простору

$$X : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

1 – против
0 – за

Имамо ПСУ $(X_1, \dots, X_{500})$ из $Ber(p)$ и дато нам је да је 285 сл. вел. из овог узорка узело вр. 1, а остале 0.

Дакле, $\sum_{i=1}^{n} x_i = 285$, односно $\bar{x}_n = \frac{285}{500}$

Нас интересује интервал поверења за p, па је потребно да смислимо који ћемо стожер користити.

Како је обим узорка велики $(n=500)$, позваћемо се на централну граничну теорему:

једна верзија

ЦГТ: $(X_1, \dots, X_n)$ ПСУ, n велико $EX_i = m$, $DX_i = \sigma^2 < +\infty$, $S_n := X_1 + \dots + X_n$, тада:

$$\frac{S_n - ES_n}{\sqrt{DS_n}} \approx \mathcal{N}(0,1)$$

често користимо и: $$\frac{S_n - ES_n}{\sqrt{DS_n}} = \frac{\sum_{i=1}^{n} X_i - nm}{\sqrt{n\sigma^2}} \overset{(:n)}{=} \frac{\bar{X}_n - m}{\sigma}\sqrt{n}$$

У нашем примеру за стожер можемо узети:

$$\boxed{\frac{\bar{X}_n - p}{\sqrt{p(1-p)}}\sqrt{n}} \qquad EX_i = p, \ DX_i = p(1-p)$$

јер знамо да ова сл. вел. по ЦГТ има приближно $\mathcal{N}(0,1)$ расподелу.

Још један трик који можемо применити да себи олакшамо посао је да се позовемо на закон великих бројева:

ЗВБ: $(X_1, X_2, \dots, X_n)$ псу, n велико, $EX_i = m$
$n \geq 100$

тада је $\frac{X_1 + \dots + X_n}{n} \approx m$

Дакле, у нашем примеру је $p \approx \overline{X}_n$, односно
$$\sqrt{p(1-p)} \approx \sqrt{\overline{X}_n(1-\overline{X}_n)}.$$

Пошто се у стожеру $\sqrt{p(1-p)}$ појављује у имениоцу, идеја је да се ова апроксимација врши само када $\overline{X}_n$ није близу 0 или 1.
Код нас је $\overline{x}_n = 0{,}57$, па можемо применити ову апроксимацију и наш стожер изгледа:

$$\boxed{\frac{\overline{X}_n - p}{\sqrt{\overline{X}_n(1-\overline{X}_n)}}\sqrt{n}}$$ и има приближно $N(0,1)$ расподелу

Интервал пов. за p добијамо исто као и до сада:
Тражимо c т.д.:

$$P\left\{-c \leq \frac{\overline{X}_n - p}{\sqrt{\overline{X}_n(1-\overline{X}_n)}}\sqrt{n} \leq c\right\} = 1-\alpha = 0{,}95$$

$$\Rightarrow c = \Phi^{-1}\left(1-\frac{\alpha}{2}\right) = \text{qnorm}\left(1-\frac{\alpha}{2}\right) = 1{,}96$$

$$P\left\{\overline{X}_n - c\frac{\sqrt{\overline{X}_n(1-\overline{X}_n)}}{\sqrt{n}} \leq p \leq \overline{X}_n + c\frac{\sqrt{\overline{X}_n(1-\overline{X}_n)}}{\sqrt{n}}\right\} = 1-\alpha$$

$$\Rightarrow (U_n, V_n) = \left(\overline{X}_n - c\frac{\sqrt{\overline{X}_n(1-\overline{X}_n)}}{\sqrt{n}}, \overline{X}_n + c\frac{\sqrt{\overline{X}_n(1-\overline{X}_n)}}{\sqrt{n}}\right)$$ је

$(1-\alpha)\cdot 100-\%$ инт. пов. за p

28. $X \sim Ber(p)$ p - веров. да особа има висок притисак

$n = 25$, $\bar{x}_n = \frac{20}{25} = 0{,}8$

ЦГТ се може користити за $n \geq 30$, али толерисаћемо и 25 у овом примеру. За ЗВБ је потребно да $n \geq 100$, па се на њега не можемо позвати.

Дакле, стожерна величина је:

$$\frac{\bar{X}_n - p}{\sqrt{p(1-p)}}\sqrt{n} \sim \mathcal{N}(0,1)$$

Тражимо c т.д.

$$P\left\{-c \leq \frac{\bar{X}_n - p}{\sqrt{p(1-p)}}\sqrt{n} \leq c\right\} = 1-\alpha = 0{,}95$$

Како је $P\left\{-c \leq \frac{\bar{X}_n - p}{\sqrt{p(1-p)}}\sqrt{n} \leq c\right\} = P\left\{\left|\frac{\bar{X}_n - p}{\sqrt{p(1-p)}}\sqrt{n}\right| \leq c\right\}$

$$= P\left\{\frac{(\bar{X}_n - p)^2 \cdot n}{p(1-p)} \leq c^2\right\},$$ U_n и V_n ћемо наћи трансформисањем ове вероватноће:

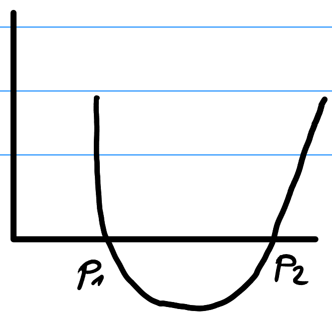
$$P\left\{\frac{(\bar{X}_n - p)^2 n}{p(1-p)} \leq c^2\right\} = P\left\{\bar{X}_n^2 - 2p\bar{X}_n + p^2 \leq \frac{c^2 p(1-p)}{n}\right\}$$

$$= P\left\{p^2(n + c^2) - p(2\bar{X}_n \cdot n + c^2) + n\bar{X}_n^2 \leq 0\right\} = (*)$$

Посматрајмо на тренутак ф-ју од p:

$$g(p) = p^2(n+c^2) - p(2n\bar{X}_n + c^2) + n\bar{X}_n^2$$

и нађимо тачке које задовољавају $g(p) \leq 0$



за p између p_1 и p_2 важи $g(p) \leq 0$, па је потребно да решимо кв. једн. $g(p) = 0$

$$p_{1,2} = \frac{2n\overline{X}_n + c^2 \pm \sqrt{(2n\overline{X}_n + c^2)^2 - 4(n+c^2)n\overline{X}_n^2}}{2(n+c^2)}$$

$$= \frac{2n\overline{X}_n + c^2 \pm \sqrt{\cancel{4n^2\overline{X}_n^2} + c^4 + 4n\overline{X}_n c^2 - \cancel{4n^2\overline{X}_n^2} - 4nc^2\overline{X}_n^2}}{2(n+c^2)}$$

$$= \frac{2n\overline{X}_n + c^2 \pm \sqrt{c^2(c^2 + 4n\overline{X}_n - 4n\overline{X}_n^2)}}{2(n+c^2)}$$

Ако узмемо за стат. U_n и V_n да буду: $U_n = p_1$ (мање решење)
$V_n = p_2$ (веће решење),

$(*) = P\{U_n \leq p \leq V_n\}$, па је наш интервал поверења баш интервал (U_n, V_n).

Мало прецизније је могуће записати тражени интервал поверења као: $(\max\{0, U_n\}, \min\{1, V_n\})$, јер p мора бити у $(0,1)$

$$c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1{,}96$$

$$u_n = \frac{2 \cdot 25 \cdot 0{,}8 + 1{,}96^2 - 1{,}96\sqrt{1{,}96^2 + 4 \cdot 25 \cdot 0{,}8 - 4 \cdot 25 \cdot 0{,}8^2}}{2(25 + 1{,}96^2)} = 0{,}61$$

$$v_n = 0{,}91$$

$$\Rightarrow (u_n, v_n) = (0{,}61,\ 0{,}91)$$

У R-у ф-ја polyroot() враћа корене полинома

Пример: $p(x) = a_0 + a_1 x + a_2 x^2$

polyroot(c(a_0, a_1, a_2)) враћа корене овог полинома

29. $X \sim Ber(p)$ p – верв. да особа буде леворука

дато је: $n = 100 \rightarrow$ може ЦГТ

$\bar{x}_n = 0{,}05 \rightarrow$ не може апроксимација $\sqrt{p(1-p)} \approx \sqrt{\bar{x}_n(1-\bar{x}_n)}$,

јер је $\bar{x}_n$ блиско 0

код у R-у:

```
n <- 100
x_sr <- 0.05
alpha <- 1-0.9
c <- qnorm(1-alpha/2)
polyroot(c(n*x_sr^2, -(2*x_sr*n+c^2),n+c^2))
```

враћа u_n и v_n

31. X - број новорођенчади у једном дану

$X \sim \mathcal{P}(\lambda)$

- Треба нам инт. поверења за λ, $1-\alpha = 0{,}95$
- $n = 365$ (1 година), $\bar{x}_n = 175{,}34$

n је велико, па можемо применити ЦГТ:

ЦГТ: $\dfrac{\bar{X}_n - \lambda}{\sqrt{\lambda}}\sqrt{n} \sim \mathcal{N}(0,1)$ $EX = \lambda$, $DX = \lambda$

ЗВБ: $\bar{X}_n \approx \lambda$

стожер: $\boxed{\dfrac{\bar{X}_n - \lambda}{\sqrt{\bar{X}_n}}\sqrt{n}}$

Нађи c т.д. $P\left\{-c \leq \dfrac{\bar{X}_n - \lambda}{\sqrt{\bar{X}_n}}\sqrt{n} \leq c\right\} = 1-\alpha$

$\Longleftrightarrow P\left\{\bar{X}_n - c\dfrac{\sqrt{\bar{X}_n}}{\sqrt{n}} \leq \lambda \leq \bar{X}_n + c\dfrac{\sqrt{\bar{X}_n}}{\sqrt{n}}\right\} = 1-\alpha \Rightarrow$

$U_n = \bar{X}_n - c\dfrac{\sqrt{\bar{X}_n}}{\sqrt{n}}$

$V_n = \bar{X}_n + c\dfrac{\sqrt{\bar{X}_n}}{\sqrt{n}}$