CENTRALNE TWIERDZENIE GRANICZNE
CTG

Założenie

$X_1, \dots, X_n$ – z. l. niezależne, o takim samym rozkładzie

$S_n = \sum_{i=1}^{n} X_i$, $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

Teza

a) $S_n \underset{n\to\infty}{\sim} N(n \cdot \mu ; \sqrt{n} \cdot \delta)$ $\quad \mu = EX_1$

b) $\overline{X}_n \underset{n\to\infty}{\sim} N(\mu , \frac{\delta}{\sqrt{n}})$ $\quad \delta^2 = V(X_1)$

Zadanie 1

Rzucono 100 razy sześcienną symetryczną kością. Wyznacz przybliżenie p-stwo, że suma oczek jest zawarta między 300 a 400.

Niech X_i – liczba oczek w i-tym rzucie, $i = 1 \dots 100$

x	1	2	3	4	5	6
P(x)	1/6	1/6	1/6	1/6	1/6	1/6

$EX_i = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3,5$

$E(X_i^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \dots + 6^2 \cdot \frac{1}{6} = \frac{91}{6} = 15,2$

$V(X_i) = \frac{91}{6} - \left(\frac{21}{6}\right)^2 \approx 15,2$

$S_{100} = \sum_{i=1}^{100} X_i \overset{CTG}{\sim} N(100 \cdot 3,5 ; \sqrt{100} \cdot \sqrt{2,92})$

$S_{100} \sim N(350; 17,09)$

$P(300 \leq S_{100} \leq 400) \overset{\text{standaryzacja}}{=\!=} P\left(\frac{300-350}{17,09} \leq \frac{S_{100}-350}{17,09} \leq \frac{400-350}{17,09}\right) =$

$= P\left(-2,93 \leq \frac{S_{100}-350}{17,09} \leq 2,93\right) = \Phi(2,93) - \Phi(-2,93) = 0,9966$

Jeśli z. l. $Z \approx N(0,1)$ to:

$P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$ $\quad \Phi(-x) = 1 - \Phi(x)$

$P(Z \leq b) = \Phi(b)$ $\quad \Phi(x) \approx 1$ dla $x > 3$

$P(a \leq Z) = 1 - \Phi(a)$ $\quad \Phi(x) \approx 0$ dla $x < -3$

Zadanie 2

Liczba awarii sprzętu komputerowego w ciągu dnia w firmie A jest zmienną losową o rozkładzie Poissona ze średnią 1. Wyznacz przybliżone p-stwo, że w ciągu 20 dni wystąpi więcej niż 15 awarii.

Niech X_i - liczba awarii w ciągu i-tego dnia, $i=1 \dots 20$

$X_i \sim P(\lambda)$ $\quad EX_i = 1 = \lambda$ $\quad V(X_i)=1$ $\quad \mu=1$ $\quad \delta=1$

$S_{20} = \sum_{i=1}^{20} X_i$ - liczba awarii w ciągu 20 dni

$P(S_{20} > 15) \overset{\text{standaryzacja}}{=} P\left(\frac{S_{20}-20}{4.47} > \frac{15-20}{4.47}\right) = P\left(\frac{S_{20}-20}{4.47} > -1.12\right) =$

$= 1 - \Phi(-1.12) = \Phi(1.12) = 0.8686$

Zadanie 3

Mamy 100 żarówek, których czas działania jest wykładniczy o średniej 5 dni. Używamy jednocześnie tylko jednej żarówki, a gdy się zepsuje, ustawiamy natychmiast na jej miejsce nową. Wyznacz p-stwo, że:

a) po 525 dniach będzie działała jeszcze jakaś żarówka

b) wśród tych 100 żarówek ich średni czas działania będzie mniejszy niż 4 dni.

Niech X_i - czas działania i-tej żarówki

$X_i \sim \exp(\lambda)$ $\quad EX_i = 5 = \frac{1}{\lambda} \rightarrow \lambda = \frac{1}{5}$ $\quad V(X_i) = \frac{1}{\lambda^2} = 25 = \delta^2$ $\quad \mu = 5$

$S_{100} = \sum_{i=1}^{100} X_i$ - łączny czas działania 100 żarówek

$S_{100} \overset{CTG}{\sim} N(100 \cdot 5, \sqrt{100} \cdot \sqrt{25})$

$S_{100} \sim N(500, 50)$

a) $P(S_{100} > 525) \overset{\text{standaryzacja}}{=} P\left(\frac{S_{100}-500}{50} > \frac{525-500}{50}\right) = P\left(\frac{S_{100}-500}{50} > 0.5\right) =$

$= 1 - \Phi(0.5) = 1 - 0.691 = 0.3085$

b) $\overline{X}_{100} = \frac{1}{100} \sum_{i=1}^{100} X_i \overset{CTG}{\sim} N\left(5, \frac{\sqrt{25}}{\sqrt{100}}\right)$ $\quad \overline{X}_{100} \sim N(5; 0.5)$

$P(\overline{X} < 4) \overset{STD}{=} P\left(\frac{X_{100}-5}{0.5} < \frac{4-5}{0.5}\right) = P\left(\frac{X_{100}-5}{0.5} < -2\right) =$

$= \Phi(-2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$

ZMIENNA LOSOWA DYSKRETNA DWUWYMIAROWA

Y \ X	0	1
1	0,5	0,1
2	0.1	c

Znajdź stałą c, $P(X>0 \mid Y=1)$, EX, V(X), EY, V(Y).
Współczynnik kowariancji $c(X,Y)$, korelacji $\varrho(X,Y)$
Sprawdź czy X i Y są niezależne

$0,5+0.1+0.1+c=1 \Rightarrow c=0.3$

$P(X>0 / Y=1) = \frac{P(X=1 \wedge Y=1)}{P(Y=1)} = \frac{0.1}{0.6} = \frac{1}{6}$

$EX = 0 \cdot 0,6 + 1 \cdot 0,4 = 0,4$ $\qquad$ $EY = 1 \cdot 0,6 + 2 \cdot 0,4 = 1,4$

$E(X^2) = 0^2 \cdot 0,6 + 1^2 \cdot 0,4 = 0,4$ $\qquad$ $E(Y^2) = 1^2 \cdot 0,6 + 2^2 \cdot 0,4 = 2,4$

$V(X) = 0,4 - (0,4)^2 = 0,24$ $\qquad$ $V(Y) = 2,4 - (1,4)^2 = 0,44$

Współczynnik kowariancji $CON(X,Y) = E(XY) - (EX)(EY)$

$$E(XY) = \sum_{k,l} k \cdot l \cdot P(X=k, Y=l)$$

$E(XY) = \overset{x}{0} \cdot \overset{y}{1} \cdot \overset{P(x,y)}{0,5} + 0 \cdot 2 \cdot 0,1 + 1 \cdot 1 \cdot 0,1 + 1 \cdot 2 \cdot 0,3 = 0,1 + 0,6 = 0,7$

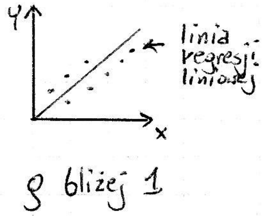
Współczynnik korelacji $\varrho(X,Y) = \frac{CON(X,Y)}{\sqrt{V(X)} \cdot \sqrt{V(Y)}}$

$\varrho(X,Y) = \frac{0,14}{\sqrt{0,24} \cdot \sqrt{0,44}} = 0,43$

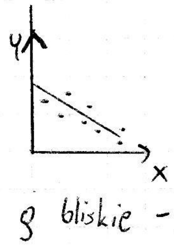
$\varrho(X,Y) \in \langle -1, 1 \rangle$

Jeśli $|\varrho| = 1$ to $\exists\, a,b \in \mathbb{R}$ $Y = aX + b$ (czyli X i Y są zależne liniowo)

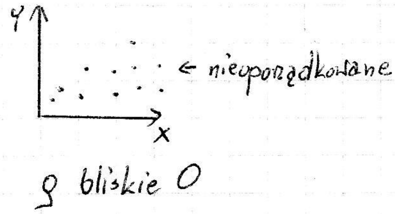
Im $|\varrho|$ bliżej 1, tym silniejsza zależność liniowa między X i Y



ϱ bliżej 1



ϱ bliskie −1



ϱ bliskie 0

Zależność liniowa $\forall\, k,l$ $P(X=k, Y=l) = P(X=k) \cdot P(Y=l)$

$P(X=0, Y=1) = 0.5$

$P(X=0) \cdot P(Y=1) = 0.6 \cdot 0,6$

z.l. X i Y są zależne

PRZEDZIAŁY UFNOŚCI - ESTYMACJA PRZEDZIAŁOWA

Model 1 Przedział ufności dla wartości średniej μ. (model całej populacji)

$X \sim N(\mu, \delta)$ δ znane

$$[\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\delta}{\sqrt{n}} ; \bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\delta}{\sqrt{n}}]$$

z - kwartyl rzędu α rozkładu $N(0,1)$

Model 2 Przedział ufności dla wartości średniej μ (model grupy próbkowej)

Cecha $X \sim N(\mu, \delta)$ δ nieznane

$$[\bar{x} - t_{1-\frac{\alpha}{2};n-1} \cdot \frac{s}{\sqrt{n}} ; \bar{x} + t_{1-\frac{\alpha}{2};n-1} \cdot \frac{s}{\sqrt{n}}]$$

t - kwartyl rzędu α rozkładu Studenta o n stopniach swobody

Model 3 Przedział ufności dla wariancji δ^2.

Cecha $X \sim N(\mu, \delta)$

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2};n-1}} , \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2};n-1}}\right]$$

$\chi^2_{\alpha,n}$ - kwartyl rzędu α z rozkładu χ^2 (chi kwadrat) o n stopniach swobody

Model 4 Przedział ufności dla proporcji p.

$$[\hat{p} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} ; \hat{p} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$$

$\hat{p}$ - proporcja (frakcja) próbkowa = liczba elementów próby o danej cesze/n.

Zad 1. W teście psychotechnicznym dla kierowców zmierzono czas reakcji 9 losowo wybranych kierowców. Otrzymano średnią próbkową 7 sekund i wariancję próbkową 1 sek^2 (bo wariancja jest do ^{2}). Wyznacz 95% przedział ufności dla wartości średniej reakcji zakładając, że ma rozkład normalny.

Model 2, bo nieznane jest sigma (δ), znane jest natomiast $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - x)^2$

DANE: $n = 9$ $\alpha = 0.05$ $\bar{x} = 7$ $s^2 = 1 \Rightarrow s = 1$

SZUKANE: $t_{1-\frac{\alpha}{2};n-1} = t_{1-\frac{0.05}{2};9-1} = t_{0.975;8}$ ← sprawdzamy w rozkładzie studenta

$$[7 - 2.306 \cdot \frac{1}{\sqrt{9}} ; 7 + 2.306 \cdot \frac{1}{\sqrt{9}}] = [6.23 ; 7.77]$$

im ~~przedział~~ ufności większy, tym przedział szerszy.

w tym przypadku dla 100%, będzie przedział $<0; +\infty)$

A gdyby przy tych samych warunkach dana była prawdziwa wariancja $\delta^2 = 1$

Model 1: $n = 9$ $\alpha = 0.05$ $\bar{x} = 7$ $\delta = 1$

$z_{1-\frac{0.05}{2}} = z_{0.975} = 1.96$ ← szukamy w rozkładzie normalnym

$$[7 - 1.96 \cdot \frac{1}{\sqrt{9}} ; 7 + 1.96 \cdot \frac{1}{\sqrt{9}}] = [6.35 ; 7.65]$$

Zad 2 Wagi 5 losowo wybranych chińskich noworodków wyniosły: 3.75, 3.45, 3.50, 3.90, 3.25.

Zakładając rozkład normalny noworodka wyznacz 99% przedział ufności dla wariancji wagi noworodka.

Model 3, bo mamy wyznaczyć przedział ufności dla wariancji.

DANE: $n = 5$ $\quad \bar{x} = 3.57$ $\quad s^2 = 0.06575$

$$\bar{x} = \frac{3.75 + \ldots + 3.25}{5} = 3.57 \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s^2 = \frac{1}{4}\left[(3,75-3.57)^2 + \ldots + (3.25-3.57)^2\right] = 0.06575 \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

SZUKANE: $\chi^2_{\cancel{1-}\frac{\alpha}{2};n-1} = \chi^2_{0.005;4} = 0.207$ ← sprawdzamy w rozkładzie χ^2 $\quad \chi^2_{1-\frac{\alpha}{2};n-1} = \chi^2_{0.995;4} = 14.86$

$$\left[\frac{(5-1)\cdot 0.06575}{14.86} ; \frac{4 \cdot 0.06575}{0.207}\right] = [0.018 ; 1.27]$$

Zad 3 Na 150 losowo wybranych Polaków 100 podaje wątpliwość rzetelności raportu MAK.

Wyznacz 95% przedział ufności dla proporcji Polaków podających wątpliwość tego raportu.

Model 4, bo mamy proporcję

DANE: $n = 150$ $\quad \hat{p} = \frac{100}{150} \approx \frac{1}{5}$ $\quad \alpha = 0.05$

SZUKANE: $z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.05}{2}} = z_{0.975} = 1.96$

$$\left[\frac{2}{3} - 1.96 \cdot \sqrt{\frac{\frac{2}{3}\cdot\frac{1}{3}}{150}} ; \frac{2}{3} + 1.96\sqrt{\frac{\frac{2}{3}\cdot\frac{1}{3}}{150}}\right] = [0.59 ; 0.74]$$