

Metody probabilistyczne

14. Elementy statystyki matematycznej II

Wojciech Kotłowski

Instytut Informatyki PP
<http://www.cs.put.poznan.pl/wkotlowski/>

23.01.2020

Estymacja przedziałowa

Dotychczas rozważaliśmy tzw. **estymatory punktowe**, które szacują wartość parametru θ za pomocą jednej liczby $\hat{\theta}$.

Estymacja przedziałowa

Dotychczas rozważaliśmy tzw. **estymatory punktowe**, które szacują wartość parametru θ za pomocą jednej liczby $\hat{\theta}$.

Często chcemy mieć nie tylko oszacowanie parametru θ , ale również **przedział**, w którym (z dużym prawdopodobieństwem) θ się znajdzie:

$$[\hat{\theta}_L, \hat{\theta}_P]$$

Taki przedział nazywamy **przedziałem ufności**, a zadanie jego wyznaczenia nazywamy **estymacją przedziałową**.

Estymacja przedziałowa

Dotychczas rozważaliśmy tzw. **estymatory punktowe**, które szacują wartość parametru θ za pomocą jednej liczby $\hat{\theta}$.

Często chcemy mieć nie tylko oszacowanie parametru θ , ale również **przedział**, w którym (z dużym prawdopodobieństwem) θ się znajdzie:

$$[\hat{\theta}_L, \hat{\theta}_P]$$

Taki przedział nazywamy **przedziałem ufności**, a zadanie jego wyznaczenia nazywamy **estymacją przedziałową**.

Zaletą przedziału ufności jest **wskazanie stopnia dokładności** oszacowania parametru.

Przedział ufności

Przedziałem ufności dla parametru θ na **poziomie ufności** $1 - \alpha$ nazywamy przedział $[\hat{\theta}_L, \hat{\theta}_P]$ taki, że $\hat{\theta}_L$ i $\hat{\theta}_P$ są statystykami wyznaczonymi z próby, oraz:

$$P\left(\hat{\theta}_L \leq \theta \leq \hat{\theta}_P\right) = 1 - \alpha$$

Przedział ufności

Przedziałem ufności dla parametru θ na **poziomie ufności** $1 - \alpha$ nazywamy przedział $[\hat{\theta}_L, \hat{\theta}_P]$ taki, że $\hat{\theta}_L$ i $\hat{\theta}_P$ są statystykami wyznaczonymi z próby, oraz:

$$P\left(\hat{\theta}_L \leq \theta \leq \hat{\theta}_P\right) = 1 - \alpha$$

- Z prawdopodobieństwem $1 - \alpha$ prawdziwa wartość parametru znajduje się w przedziale

Przedział ufności

Przedziałem ufności dla parametru θ na **poziomie ufności** $1 - \alpha$ nazywamy przedział $[\hat{\theta}_L, \hat{\theta}_P]$ taki, że $\hat{\theta}_L$ i $\hat{\theta}_P$ są statystykami wyznaczonymi z próby, oraz:

$$P\left(\hat{\theta}_L \leq \theta \leq \hat{\theta}_P\right) = 1 - \alpha$$

- Z prawdopodobieństwem $1 - \alpha$ prawdziwa wartość parametru znajduje się w przedziale
- **Uwaga:** to przedział jest losowy, a nie wartość parametru!

Przedział ufności

Przedziałem ufności dla parametru θ na **poziomie ufności** $1 - \alpha$ nazywamy przedział $[\hat{\theta}_L, \hat{\theta}_P]$ taki, że $\hat{\theta}_L$ i $\hat{\theta}_P$ są statystykami wyznaczonymi z próby, oraz:

$$P\left(\hat{\theta}_L \leq \theta \leq \hat{\theta}_P\right) = 1 - \alpha$$

- Z prawdopodobieństwem $1 - \alpha$ prawdziwa wartość parametru znajduje się w przedziale
- **Uwaga:** to przedział jest losowy, a nie wartość parametru!
- Czasem α (zamiast $1 - \alpha$) nazywa się poziomem ufności

Przedział ufności

Przedziałem ufności dla parametru θ na **poziomie ufności** $1 - \alpha$ nazywamy przedział $[\hat{\theta}_L, \hat{\theta}_P]$ taki, że $\hat{\theta}_L$ i $\hat{\theta}_P$ są statystykami wyznaczonymi z próby, oraz:

$$P\left(\hat{\theta}_L \leq \theta \leq \hat{\theta}_P\right) = 1 - \alpha$$

- Z prawdopodobieństwem $1 - \alpha$ prawdziwa wartość parametru znajduje się w przedziale
- **Uwaga:** to przedział jest losowy, a nie wartość parametru!
- Czasem α (zamiast $1 - \alpha$) nazywa się poziomem ufności
- Czasem używa się przedziału otwartego $(\hat{\theta}_L, \hat{\theta}_P)$

Przedział ufności

Przedziałem ufności dla parametru θ na **poziomie ufności** $1 - \alpha$ nazywamy przedział $[\hat{\theta}_L, \hat{\theta}_P]$ taki, że $\hat{\theta}_L$ i $\hat{\theta}_P$ są statystykami wyznaczonymi z próby, oraz:

$$P\left(\hat{\theta}_L \leq \theta \leq \hat{\theta}_P\right) = 1 - \alpha$$

- Z prawdopodobieństwem $1 - \alpha$ prawdziwa wartość parametru znajduje się w przedziale
- **Uwaga:** to przedział jest losowy, a nie wartość parametru!
- Czasem α (zamiast $1 - \alpha$) nazywa się poziomem ufności
- Czasem używa się przedziału otwartego $(\hat{\theta}_L, \hat{\theta}_P)$
- **Przykład:** W problemie szacowania poparcia kandydata A, estymacja przedziałowa dała przedział ufności na poziomie 95% równy $[0.58, 0.62]$, tzn. z prawdopodobieństwem 95% prawdziwy parametr p znajduje się między 0.58 a 0.62.

Metody wyznaczania przedziałów ufności

Zwykle rozpinamy przedział symetrycznie wokół pewnego estymatora punktowego $\hat{\theta}$:

$$[\hat{\theta} - \Delta, \hat{\theta} + \Delta], \quad \text{tzn. } \hat{\theta}_L = \hat{\theta} - \Delta, \quad \hat{\theta}_P = \hat{\theta} + \Delta$$

Rozpiętość przedziału Δ wyznaczana jest w oparciu o docelowy poziom ufności, tzn. rozwiązując równanie (ze względu na Δ):

$$1 - \alpha = P(\hat{\theta} - \Delta \leq \theta \leq \hat{\theta} + \Delta)$$

Metody wyznaczania przedziałów ufności

Zwykle rozpinamy przedział symetrycznie wokół pewnego estymatora punktowego $\hat{\theta}$:

$$[\hat{\theta} - \Delta, \hat{\theta} + \Delta], \quad \text{tzn. } \hat{\theta}_L = \hat{\theta} - \Delta, \quad \hat{\theta}_P = \hat{\theta} + \Delta$$

Rozpiętość przedziału Δ wyznaczana jest w oparciu o docelowy poziom ufności, tzn. rozwiązując równanie (ze względu na Δ):

$$\begin{aligned} 1 - \alpha &= P(\hat{\theta} - \Delta \leq \theta \leq \hat{\theta} + \Delta) \\ &= P(-\Delta \leq \theta - \hat{\theta} \leq \Delta) \end{aligned}$$

Metody wyznaczania przedziałów ufności

Zwykle rozpinamy przedział symetrycznie wokół pewnego estymatora punktowego $\hat{\theta}$:

$$[\hat{\theta} - \Delta, \hat{\theta} + \Delta], \quad \text{tzn. } \hat{\theta}_L = \hat{\theta} - \Delta, \quad \hat{\theta}_P = \hat{\theta} + \Delta$$

Rozpiętość przedziału Δ wyznaczana jest w oparciu o docelowy poziom ufności, tzn. rozwiązując równanie (ze względu na Δ):

$$\begin{aligned} 1 - \alpha &= P(\hat{\theta} - \Delta \leq \theta \leq \hat{\theta} + \Delta) \\ &= P(-\Delta \leq \theta - \hat{\theta} \leq \Delta) \\ &= P(-\Delta \leq \hat{\theta} - \theta \leq \Delta) \end{aligned}$$

Przykład: rozkład normalny ze znaną wariancją

Niech $X \sim N(\mu, \sigma^2)$, gdzie σ^2 jest **znana**

Przykład: rozkład normalny ze znaną wariancją

Niech $X \sim N(\mu, \sigma^2)$, gdzie σ^2 jest **znana**

Konstruujemy estymator wartości oczekiwanej μ na podstawie próby $X_1, \dots, X_n$:

$$\hat{\mu} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Przykład: rozkład normalny ze znaną wariancją

Niech $X \sim N(\mu, \sigma^2)$, gdzie σ^2 jest **znana**

Konstruujemy estymator wartości oczekiwanej μ na podstawie próby $X_1, \dots, X_n$:

$$\hat{\mu} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Ponieważ $\bar{X}_n$ jest kombinacją liniową niezależnych zmiennych o rozkładzie normalnym, $\bar{X}_n$ **ma rozkład normalny** z parametrami:

$$E(\bar{X}_n) = \mu, \quad D^2(\bar{X}_n) = \frac{\sigma^2}{n}, \quad \text{czyli } \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Przykład: rozkład normalny ze znaną wariancją

Niech $X \sim N(\mu, \sigma^2)$, gdzie σ^2 jest **znana**

Konstruujemy estymator wartości oczekiwanej μ na podstawie próby $X_1, \dots, X_n$:

$$\hat{\mu} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

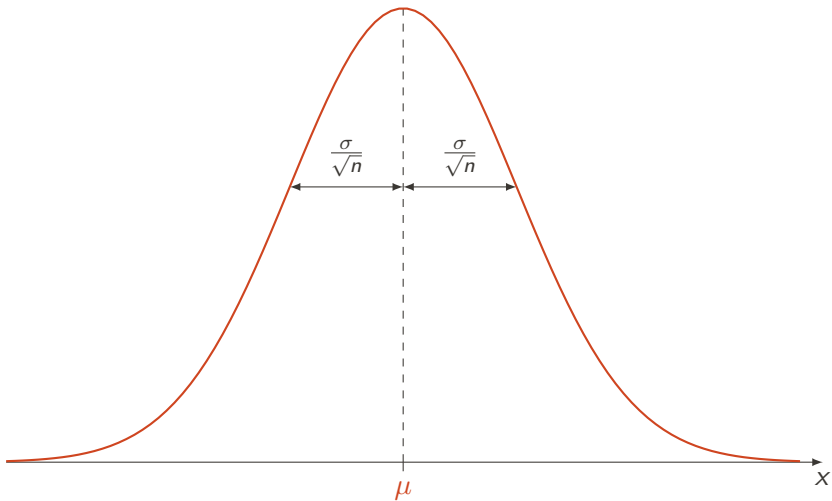
Ponieważ $\bar{X}_n$ jest kombinacją liniową niezależnych zmiennych o rozkładzie normalnym, $\bar{X}_n$ **ma rozkład normalny** z parametrami:

$$E(\bar{X}_n) = \mu, \quad D^2(\bar{X}_n) = \frac{\sigma^2}{n}, \quad \text{czyli } \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Konstruujemy następnie przedział ufności na poziomie ufności $1 - \alpha$:

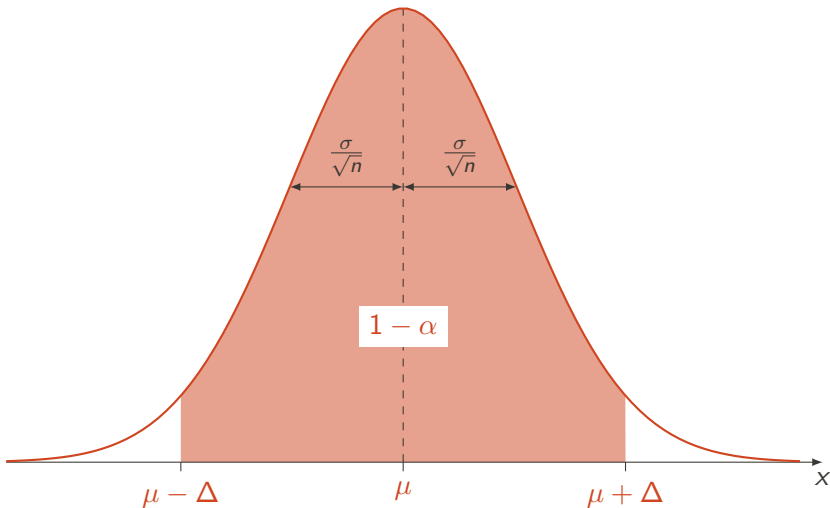
$$1 - \alpha = P\left(-\Delta \leq \bar{X}_n - \mu \leq \Delta\right)$$

Rozkład estymatora $\overline{X}_n$



Rozkład estymatora $\bar{X}_n$

$$P\left(\mu - \Delta \leq \bar{X}_n \leq \mu + \Delta\right) = 1 - \alpha$$



Przykład: rozkład normalny ze znaną wariancją

$$\text{Skoro } \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{to } Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Przykład: rozkład normalny ze znaną wariancją

$$\text{Skoro } \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{to } Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$1 - \alpha = P\left(-\Delta \leq \bar{X}_n - \mu \leq \Delta\right)$$

Przykład: rozkład normalny ze znaną wariancją

$$\text{Skoro } \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{to } Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$\begin{aligned} 1 - \alpha &= P\left(-\Delta \leq \bar{X}_n - \mu \leq \Delta\right) \\ &= P\left(-\frac{\Delta}{\sigma}\sqrt{n} \leq \frac{\bar{X}_n - \mu}{\sigma}\sqrt{n} \leq \frac{\Delta}{\sigma}\sqrt{n}\right) \end{aligned}$$

Przykład: rozkład normalny ze znaną wariancją

$$\text{Skoro } \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{to } Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$\begin{aligned} 1 - \alpha &= P\left(-\Delta \leq \bar{X}_n - \mu \leq \Delta\right) \\ &= P\left(-\frac{\Delta}{\sigma}\sqrt{n} \leq \frac{\bar{X}_n - \mu}{\sigma}\sqrt{n} \leq \frac{\Delta}{\sigma}\sqrt{n}\right) \\ &= P\left(-\frac{\Delta}{\sigma}\sqrt{n} \leq Z \leq \frac{\Delta}{\sigma}\sqrt{n}\right) \end{aligned}$$

Przykład: rozkład normalny ze znaną wariancją

$$\text{Skoro } \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{to } Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$\begin{aligned} 1 - \alpha &= P\left(-\Delta \leq \bar{X}_n - \mu \leq \Delta\right) \\ &= P\left(-\frac{\Delta}{\sigma}\sqrt{n} \leq \frac{\bar{X}_n - \mu}{\sigma}\sqrt{n} \leq \frac{\Delta}{\sigma}\sqrt{n}\right) \\ &= P\left(-\frac{\Delta}{\sigma}\sqrt{n} \leq Z \leq \frac{\Delta}{\sigma}\sqrt{n}\right) \\ &= \Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - \Phi\left(-\frac{\Delta}{\sigma}\sqrt{n}\right) \end{aligned}$$

Przykład: rozkład normalny ze znaną wariancją

$$\text{Skoro } \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{to } Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$\begin{aligned} 1 - \alpha &= P\left(-\Delta \leq \bar{X}_n - \mu \leq \Delta\right) \\ &= P\left(-\frac{\Delta}{\sigma}\sqrt{n} \leq \frac{\bar{X}_n - \mu}{\sigma}\sqrt{n} \leq \frac{\Delta}{\sigma}\sqrt{n}\right) \\ &= P\left(-\frac{\Delta}{\sigma}\sqrt{n} \leq Z \leq \frac{\Delta}{\sigma}\sqrt{n}\right) \\ &= \Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - \Phi\left(-\frac{\Delta}{\sigma}\sqrt{n}\right) \\ &= 2\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - 1 \end{aligned}$$


$$\Phi(-x) = 1 - \Phi(x)$$

Przykład: rozkład normalny ze znaną wariancją

$$\text{Skoro } \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{to } Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$1 - \alpha = 2\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - 1$$

Przykład: rozkład normalny ze znaną wariancją

$$\text{Skoro } \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{to } Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$1 - \alpha = 2\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - 1$$

$$\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) = 1 - \frac{\alpha}{2}$$

Przykład: rozkład normalny ze znaną wariancją

$$\text{Skoro } \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{to } Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$1 - \alpha = 2\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - 1$$

$$\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) = 1 - \frac{\alpha}{2}$$

$$\frac{\Delta}{\sigma}\sqrt{n} = \underbrace{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}_{z_{1-\alpha/2}}$$

Przykład: rozkład normalny ze znaną wariancją

$$\text{Skoro } \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{to } Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$1 - \alpha = 2\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - 1$$

$$\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) = 1 - \frac{\alpha}{2}$$

$$\frac{\Delta}{\sigma}\sqrt{n} = \underbrace{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}_{z_{1-\alpha/2}}$$

$$\Delta = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Przykład: rozkład normalny ze znaną wariancją

$$\text{Skoro } \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{to } Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$\begin{aligned} 1 - \alpha &= 2\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - 1 \\ \Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) &= 1 - \frac{\alpha}{2} \\ \frac{\Delta}{\sigma}\sqrt{n} &= \underbrace{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}_{z_{1-\alpha/2}} \\ \Delta &= z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \end{aligned}$$

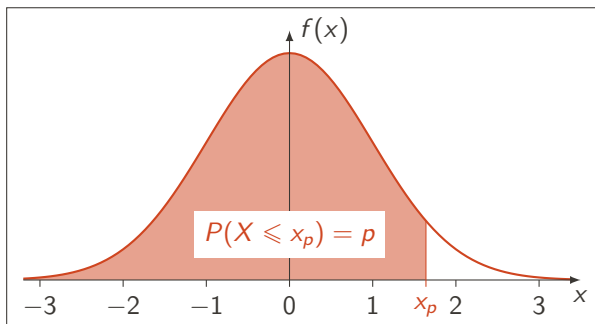
Uwaga: $z_{1-\alpha/2}$ jest **kwantylem rzędu $1 - \frac{\alpha}{2}$** rozkładu normalnego standardowego

Kwantyl (uproszczona definicja!)

Rozważmy zmienną losową X ze ściśle rosnącą dystrybuantą.

Kwantylem rzędu p zmiennej X nazywamy liczbę $x_p \in \mathbb{R}$ taką, że:

$$F_X(x_p) = P(X \leq x_p) = p, \quad \text{czyli } x_p = F_X^{-1}(p)$$



Kwantyl x_p otrzymujemy poprzez wyznaczenia wartości dystrybuanty odwrotnej (tzw. funkcji kwantylowej) w p

Kwantyl z_p dla $Z \sim N(0, 1)$ a tablica wartości $\Phi(z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5674	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.20	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Kwantyl z_p dla $Z \sim N(0, 1)$ a tablica wartości $\Phi(z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557					
0.20	0.5793	0.5832	0.5871	0.5910	0.5948					
0.30	0.6179	0.6217	0.6255	0.6293	0.6331					
0.40	0.6554	0.6591	0.6628	0.6664	0.6700					
0.50	0.6915	0.6950	0.6985	0.7019	0.7054					
0.60	0.7257	0.7291	0.7324	0.7357	0.7389					
0.70	0.7580	0.7611	0.7642	0.7673	0.7704					
0.80	0.7881	0.7910	0.7939	0.7967	0.7995					
0.90	0.8159	0.8186	0.8212	0.8238	0.8264					
1.00	0.8413	0.8438	0.8461	0.8485	0.8508					
1.10	0.8643	0.8665	0.8686	0.8708	0.8729					
1.20	0.8849	0.8869	0.8888	0.8907	0.8925					
1.30	0.9032	0.9049	0.9066	0.9082	0.9099					
1.40	0.9192	0.9207	0.9222	0.9236	0.9251					
1.50	0.9332	0.9345	0.9357	0.9370	0.9382					
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.20	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Przykładowe wartości:

p $z_p = \Phi^{-1}(p)$

0.95 1.64

0.975 1.96

0.99 2.33

0.995 2.58

Kwantyl z_p dla $Z \sim N(0, 1)$ a tablica wartości $\Phi(z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557					
0.20	0.5793	0.5832	0.5871	0.5910	0.5948					
0.30	0.6179	0.6217	0.6255	0.6293	0.6331					
0.40	0.6554	0.6591	0.6628	0.6664	0.6700					
0.50	0.6915	0.6950	0.6985	0.7019	0.7054					
0.60	0.7257	0.7291	0.7324	0.7357	0.7389					
0.70	0.7580	0.7611	0.7642	0.7673	0.7704					
0.80	0.7881	0.7910	0.7939	0.7967	0.7995					
0.90	0.8159	0.8186	0.8212	0.8238	0.8264					
1.00	0.8413	0.8438	0.8461	0.8485	0.8508					
1.10	0.8643	0.8665	0.8686	0.8708	0.8729					
1.20	0.8849	0.8869	0.8888	0.8907	0.8925					
1.30	0.9032	0.9049	0.9066	0.9082	0.9099					
1.40	0.9192	0.9207	0.9222	0.9236	0.9251					
1.50	0.9332	0.9345	0.9357	0.9370	0.9382					
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9858
2.20	0.9861	0.9864	0.9868	0.9871	0.9874	0.9877	0.9880	0.9883	0.9886	0.9889
2.30	0.9893	0.9896	0.9898	0.9901	0.9903	0.9905	0.9907	0.9909	0.9911	0.9913
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9933	0.9935	0.9937
2.50	0.9938	0.9940	0.9941	0.9943	0.9944	0.9945	0.9946	0.9947	0.9948	0.9949
2.60	0.9953	0.9955	0.9956	0.9957	0.9958	0.9959	0.9960	0.9961	0.9962	0.9963
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9978	0.9979	0.9980	0.9981	0.9982	0.9983
2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9985	0.9986	0.9987	0.9988	0.9989
3.00	0.9987	0.9987	0.9987	0.9988	0.9989	0.9990	0.9990	0.9991	0.9991	0.9992
3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9993	0.9993	0.9994	0.9994
3.20	0.9993	0.9993	0.9994	0.9994	0.9995	0.9995	0.9996	0.9996	0.9997	0.9997
3.30	0.9995	0.9995	0.9996	0.9996	0.9997	0.9997	0.9998	0.9998	0.9999	0.9999
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Przykładowe wartości:

p	$z_p = \Phi^{-1}(p)$
-----	----------------------

0.95	1.64
------	------

0.975	1.96
-------	------

0.99	2.33
------	------

0.995	2.58
-------	------

Przykładowe wartości:

α	$z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$
----------	--

0.1	1.64
-----	------

0.05	1.96
------	------

0.02	2.33
------	------

0.01	2.58
------	------

Przykład: rozkład normalny ze znaną wariancją

Próba o liczności $n = 16$ pomiarów pewnej cechy o rozkładzie normalnym ze znaną wariancją $\sigma^2 = 16$ dała wartość statystyki $\bar{x}_n = 2$. Wyznacz przedział ufności dla wartości oczekiwanej na poziomie ufności $1 - \alpha$ dla $\alpha = 0.05$.

Przykład: rozkład normalny ze znaną wariancją

Próba o liczności $n = 16$ pomiarów pewnej cechy o rozkładzie normalnym ze znaną wariancją $\sigma^2 = 16$ dała wartość statystyki $\bar{x}_n = 2$. Wyznacz przedział ufności dla wartości oczekiwanej na poziomie ufności $1 - \alpha$ dla $\alpha = 0.05$.

Rozwiązanie: Przedział ufności ma postać:

$$[\bar{x}_n - \Delta, \bar{x}_n + \Delta], \quad \text{gdzie } \Delta = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Przykład: rozkład normalny ze znaną wariancją

Próba o liczności $n = 16$ pomiarów pewnej cechy o rozkładzie normalnym ze znaną wariancją $\sigma^2 = 16$ dała wartość statystyki $\bar{x}_n = 2$. Wyznacz przedział ufności dla wartości oczekiwanej na poziomie ufności $1 - \alpha$ dla $\alpha = 0.05$.

Rozwiązanie: Przedział ufności ma postać:

$$[\bar{x}_n - \Delta, \bar{x}_n + \Delta], \quad \text{gdzie } \Delta = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$z_{1-\alpha/2} = z_{0.975} = \Phi^{-1}(0.975) = 1.96$$

Przykład: rozkład normalny ze znaną wariancją

Próba o liczności $n = 16$ pomiarów pewnej cechy o rozkładzie normalnym ze znaną wariancją $\sigma^2 = 16$ dała wartość statystyki $\bar{x}_n = 2$. Wyznacz przedział ufności dla wartości oczekiwanej na poziomie ufności $1 - \alpha$ dla $\alpha = 0.05$.

Rozwiązanie: Przedział ufności ma postać:

$$[\bar{x}_n - \Delta, \bar{x}_n + \Delta], \quad \text{gdzie } \Delta = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$z_{1-\alpha/2} = z_{0.975} = \Phi^{-1}(0.975) = 1.96$$

$$\Delta = 1.96 \frac{4}{\sqrt{16}} = 1.96$$

Przykład: rozkład normalny ze znaną wariancją

Próba o liczności $n = 16$ pomiarów pewnej cechy o rozkładzie normalnym ze znaną wariancją $\sigma^2 = 16$ dała wartość statystyki $\bar{x}_n = 2$. Wyznacz przedział ufności dla wartości oczekiwanej na poziomie ufności $1 - \alpha$ dla $\alpha = 0.05$.

Rozwiązanie: Przedział ufności ma postać:

$$[\bar{x}_n - \Delta, \bar{x}_n + \Delta], \quad \text{gdzie } \Delta = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$z_{1-\alpha/2} = z_{0.975} = \Phi^{-1}(0.975) = 1.96$$

$$\Delta = 1.96 \frac{4}{\sqrt{16}} = 1.96$$

$$[\bar{x}_n - \Delta, \bar{x}_n + \Delta] = [2 - 1.96, 2 + 1.96] = [0.04, 3.96]$$

Zadanie

Zadanie 1

W celu oszacowania średniego czasu przejazdu pociągów na odcinku łączącym pewne dwa miasta pobrano próbę o liczności $n = 100$ pomiarów (tzn. zmierzono czasy przejazdu 100 wybranych losowo pociągów na tym odcinku), uzyskując średnią z próby $\bar{x}_n = 54$ (minuty). Przyjmij, że czasy przejazdu można zamodelować rozkładem normalnym ze znanym odchyleniem standardowym $\sigma = 10$ (minut). Wyznacz przedział ufności dla oczekiwanej wartości czasu przejazdu na poziomie ufności $1 - \alpha$ dla $\alpha = 0.02$.

Przykład: rozkład dwupunktowy

Niech $X \sim B(p)$. Znajdziemy przedział ufności dla parametru p .

Przykład: rozkład dwupunktowy

Niech $X \sim B(p)$. Znajdziemy przedział ufności dla parametru p .

Estymator parametru p na podstawie próby $X_1, \dots, X_n$:

$$\hat{p} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{\# \text{ sukcesów}}{n}$$

Przykład: rozkład dwupunktowy

Niech $X \sim B(p)$. Znajdziemy przedział ufności dla parametru p .

Estymator parametru p na podstawie próby $X_1, \dots, X_n$:

$$\hat{p} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{\# \text{ sukcesów}}{n}$$

$$E(\hat{p}) = p, \quad D^2(\hat{p}) = \frac{D^2(X)}{n} = \frac{p(1-p)}{n},$$

Przykład: rozkład dwupunktowy

Niech $X \sim B(p)$. Znajdziemy przedział ufności dla parametru p .

Estymator parametru p na podstawie próby $X_1, \dots, X_n$:

$$\hat{p} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{\# \text{ sukcesów}}{n}$$

$$E(\hat{p}) = p, \quad D^2(\hat{p}) = \frac{D^2(X)}{n} = \frac{p(1-p)}{n},$$

Z Centralnego Twierdzenia Granicznego mamy:

$$\frac{\hat{p} - E(\hat{p})}{D(\hat{p})} = \frac{\hat{p} - p}{\sqrt{p(1-p)}} \sqrt{n} \xrightarrow{D} Z \sim N(0, 1)$$

Przykład: rozkład dwupunktowy

Niech $X \sim B(p)$. Znajdziemy przedział ufności dla parametru p .

Estymator parametru p na podstawie próby $X_1, \dots, X_n$:

$$\hat{p} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{\# \text{ sukcesów}}{n}$$

$$E(\hat{p}) = p, \quad D^2(\hat{p}) = \frac{D^2(X)}{n} = \frac{p(1-p)}{n},$$

Z Centralnego Twierdzenia Granicznego mamy:

$$\frac{\hat{p} - E(\hat{p})}{D(\hat{p})} = \frac{\hat{p} - p}{\sqrt{p(1-p)}} \sqrt{n} \xrightarrow{D} Z \sim N(0, 1)$$

Z kolei z Prawa Wielkich Liczb mamy $\hat{p} \xrightarrow{P} p$, a tym samym*:

$$\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}} \sqrt{n} \xrightarrow{D} Z \sim N(0, 1)$$

Załóżymy, że n jest wystarczająco duże aby skorzystać z tego przybliżenia.

* nietrywialne: jeśli $X_n \xrightarrow{D} X$ i $Y_n \xrightarrow{P} c$ to $X_n f(Y_n) \xrightarrow{D} X f(c)$ dla ciągłej f .

Przykład: rozkład dwupunktowy

Konstruujemy przedział ufności na poziomie ufności $1 - \alpha$:

$$1 - \alpha = P(-\Delta \leq \hat{p} - p \leq \Delta)$$

Przykład: rozkład dwupunktowy

Konstruujemy przedział ufności na poziomie ufności $1 - \alpha$:

$$\begin{aligned} 1 - \alpha &= P(-\Delta \leq \hat{p} - p \leq \Delta) \\ &= P\left(-\frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}} \leq \underbrace{\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}}\sqrt{n}}_{\simeq Z} \leq \frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}\right) \end{aligned}$$

Przykład: rozkład dwupunktowy

Konstruujemy przedział ufności na poziomie ufności $1 - \alpha$:

$$\begin{aligned}1 - \alpha &= P(-\Delta \leq \hat{p} - p \leq \Delta) \\&= P\left(-\frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}} \leq \underbrace{\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}}\sqrt{n}}_{\simeq Z} \leq \frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}\right) \\&\simeq \Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}\right) - \Phi\left(-\frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}\right) = 2\Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}\right) - 1\end{aligned}$$

Przykład: rozkład dwupunktowy

Konstruujemy przedział ufności na poziomie ufności $1 - \alpha$:

$$\begin{aligned}1 - \alpha &= P(-\Delta \leq \hat{p} - p \leq \Delta) \\&= P\left(-\frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}} \leq \underbrace{\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}}\sqrt{n}}_{\simeq Z} \leq \frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}\right) \\&\simeq \Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}\right) - \Phi\left(-\frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}\right) = 2\Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}\right) - 1\end{aligned}$$

Rozwiązując równanie otrzymujemy:

$$\Delta = z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Przykład: rozkład dwupunktowy

Konstruujemy przedział ufności na poziomie ufności $1 - \alpha$:

$$\begin{aligned}1 - \alpha &= P(-\Delta \leq \hat{p} - p \leq \Delta) \\&= P\left(-\frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}} \leq \underbrace{\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}}\sqrt{n}}_{\simeq Z} \leq \frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}\right) \\&\simeq \Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}\right) - \Phi\left(-\frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}\right) = 2\Phi\left(\frac{\Delta\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}\right) - 1\end{aligned}$$

Rozwiązując równanie otrzymujemy:

$$\Delta = z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Zadanie 2*

Rozwiąż powyższe zagadnienie krok po kroku

Przykład: rozkład dwupunktowy

W celu oszacowania poparcia kandydata A w drugiej turze wyborów prezydenckich pobrano losowo próbę o liczności $n = 1000$ osób, odnotowując czy dana osoba głosuje na kandydata A ($X = 1$), czy na jego konkurenta B ($X = 0$). Uzyskano 550 głosów za A i 450 głosów za B . Wyznacz przedział ufności dla poparcia kandydata A dla $\alpha = 0.05$.

Przykład: rozkład dwupunktowy

W celu oszacowania poparcia kandydata A w drugiej turze wyborów prezydenckich pobrano losowo próbę o liczności $n = 1000$ osób, odnotowując czy dana osoba głosuje na kandydata A ($X = 1$), czy na jego konkurenta B ($X = 0$). Uzyskano 550 głosów za A i 450 głosów za B . Wyznacz przedział ufności dla poparcia kandydata A dla $\alpha = 0.05$.

Rozwiązanie:

$$\hat{p} = \frac{550}{1000} = 0.55$$

Przykład: rozkład dwupunktowy

W celu oszacowania poparcia kandydata A w drugiej turze wyborów prezydenckich pobrano losowo próbę o liczności $n = 1000$ osób, odnotowując czy dana osoba głosuje na kandydata A ($X = 1$), czy na jego konkurenta B ($X = 0$). Uzyskano 550 głosów za A i 450 głosów za B . Wyznacz przedział ufności dla poparcia kandydata A dla $\alpha = 0.05$.

Rozwiązanie:

$$\hat{p} = \frac{550}{1000} = 0.55$$

$$z_{1-\alpha/2} = z_{0.975} = \Phi^{-1}(0.975) = 1.96$$

Przykład: rozkład dwupunktowy

W celu oszacowania poparcia kandydata A w drugiej turze wyborów prezydenckich pobrano losowo próbę o licznosci $n = 1000$ osób, odnotowując czy dana osoba głosuje na kandydata A ($X = 1$), czy na jego konkurenta B ($X = 0$). Uzyskano 550 głosów za A i 450 głosów za B . Wyznacz przedział ufności dla poparcia kandydata A dla $\alpha = 0.05$.

Rozwiązanie:

$$\hat{p} = \frac{550}{1000} = 0.55$$

$$z_{1-\alpha/2} = z_{0.975} = \Phi^{-1}(0.975) = 1.96$$

$$\Delta = z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96 \sqrt{\frac{0.55 \cdot 0.45}{1000}} \simeq 0.031$$

Przykład: rozkład dwupunktowy

W celu oszacowania poparcia kandydata A w drugiej turze wyborów prezydenckich pobrano losowo próbę o liczności $n = 1000$ osób, odnotowując czy dana osoba głosuje na kandydata A ($X = 1$), czy na jego konkurenta B ($X = 0$). Uzyskano 550 głosów za A i 450 głosów za B . Wyznacz przedział ufności dla poparcia kandydata A dla $\alpha = 0.05$.

Rozwiązanie:

$$\hat{p} = \frac{550}{1000} = 0.55$$

$$z_{1-\alpha/2} = z_{0.975} = \Phi^{-1}(0.975) = 1.96$$

$$\Delta = z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96 \sqrt{\frac{0.55 \cdot 0.45}{1000}} \simeq 0.031$$

$$[\hat{p} - \Delta, \hat{p} + \Delta] = [0.55 - 0.031, 0.55 + 0.031] = [0.519, 0.581]$$

Zadanie

Zadanie 3

Aby ocenić jakość nowego algorytmu rekomendacji muzyki na pewnym portalu, wybrano losowo próbę $n = 300$ użytkowników, którym nowy algorytm zarekomendował utwór muzyczny zgodnie z ich preferencjami, i sprawdzono ilu spośród użytkowników utwór ten odsłuchało. Okazało się, że utwór odsłuchało 75 osób. Jeśli jakość algorytmu będziemy mierzyć jako prawdopodobieństwo odsłuchania rekomendowanego utworu, wyznac przedział ufności dla jakości na poziomie ufności $1 - \alpha$, gdzie $\alpha = 0.1$.

Przykład: rozkład normalny z nieznaną wariancją

Niech $X \sim N(\mu, \sigma^2)$, ale teraz **nie znamy** σ^2

Przykład: rozkład normalny z nieznaną wariancją

Niech $X \sim N(\mu, \sigma^2)$, ale teraz **nie znamy** σ^2

Jak poprzednio, jako estymator wartości oczekiwanej μ wybieramy $\hat{\mu} = \bar{X}_n$

Przykład: rozkład normalny z nieznaną wariancją

Niech $X \sim N(\mu, \sigma^2)$, ale teraz **nie znamy** σ^2

Jak poprzednio, jako estymator wartości oczekiwanej μ wybieramy $\hat{\mu} = \bar{X}_n$

Ponieważ nie znamy wariancji σ^2 , **estymujemy ją z próby**:

$$\hat{\sigma}_*^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Przykład: rozkład normalny z nieznaną wariancją

Niech $X \sim N(\mu, \sigma^2)$, ale teraz **nie znamy** σ^2

Jak poprzednio, jako estymator wartości oczekiwanej μ wybieramy $\hat{\mu} = \bar{X}_n$

Ponieważ nie znamy wariancji σ^2 , **estymujemy ją z próby**:

$$\hat{\sigma}_*^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Wprowadzamy nową zmienną:

$$T = \frac{\bar{X}_n - \mu}{\hat{\sigma}_*} \sqrt{n}$$

Przykład: rozkład normalny z nieznaną wariancją

Niech $X \sim N(\mu, \sigma^2)$, ale teraz **nie znamy** σ^2

Jak poprzednio, jako estymator wartości oczekiwanej μ wybieramy $\hat{\mu} = \bar{X}_n$

Ponieważ nie znamy wariancji σ^2 , **estymujemy ją z próby**:

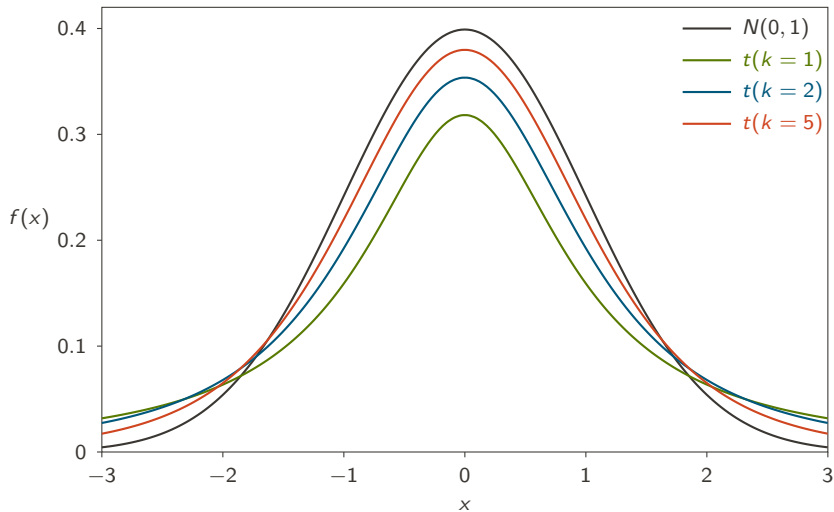
$$\hat{\sigma}_*^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Wprowadzamy nową zmienną:

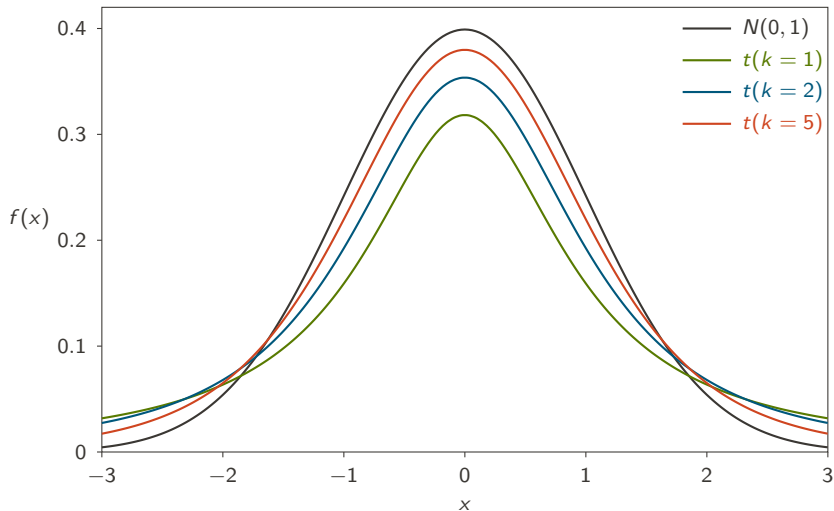
$$T = \frac{\bar{X}_n - \mu}{\hat{\sigma}_*} \sqrt{n}$$

Można pokazać, że zmienna T ma **rozkład t -Studenta** o $n - 1$ stopniach swobody

Rozkład t -Studenta o k stopniach swobody



Rozkład t -Studenta o k stopniach swobody



Dla $k \rightarrow \infty$ rozkład t zbiega do rozkładu $N(0, 1)$

Przykład: rozkład normalny z nieznaną wariancją

$$1 - \alpha = P\left(-\Delta \leq \bar{X}_n - \mu \leq \Delta\right)$$

Przykład: rozkład normalny z nieznaną wariancją

$$\begin{aligned}1 - \alpha &= P\left(-\Delta \leq \bar{X}_n - \mu \leq \Delta\right) \\&= P\left(-\frac{\Delta}{\hat{\sigma}_*} \sqrt{n} \leq \frac{\bar{X}_n - \mu}{\hat{\sigma}_*} \sqrt{n} \leq \frac{\Delta}{\hat{\sigma}_*} \sqrt{n}\right)\end{aligned}$$

Przykład: rozkład normalny z nieznaną wariancją

$$\begin{aligned}1 - \alpha &= P\left(-\Delta \leq \bar{X}_n - \mu \leq \Delta\right) \\&= P\left(-\frac{\Delta}{\hat{\sigma}_*} \sqrt{n} \leq \frac{\bar{X}_n - \mu}{\hat{\sigma}_*} \sqrt{n} \leq \frac{\Delta}{\hat{\sigma}_*} \sqrt{n}\right) \\&= P\left(-\frac{\Delta}{\hat{\sigma}_*} \sqrt{n} \leq T \leq \frac{\Delta}{\hat{\sigma}_*} \sqrt{n}\right)\end{aligned}$$

Przykład: rozkład normalny z nieznaną wariancją

$$\begin{aligned}1 - \alpha &= P\left(-\Delta \leq \bar{X}_n - \mu \leq \Delta\right) \\&= P\left(-\frac{\Delta}{\hat{\sigma}_*} \sqrt{n} \leq \frac{\bar{X}_n - \mu}{\hat{\sigma}_*} \sqrt{n} \leq \frac{\Delta}{\hat{\sigma}_*} \sqrt{n}\right) \\&= P\left(-\frac{\Delta}{\hat{\sigma}_*} \sqrt{n} \leq T \leq \frac{\Delta}{\hat{\sigma}_*} \sqrt{n}\right) \\&= F_T\left(\frac{\Delta}{\hat{\sigma}_*} \sqrt{n}\right) - F_T\left(-\frac{\Delta}{\hat{\sigma}_*} \sqrt{n}\right) = 2F_T\left(\frac{\Delta}{\hat{\sigma}_*} \sqrt{n}\right) - 1,\end{aligned}$$

gdzie użyliśmy symetrii rozkładu t -Studenta: $F_T(-x) = 1 - F_T(x)$.

Przykład: rozkład normalny z nieznaną wariancją

$$\begin{aligned}1 - \alpha &= P\left(-\Delta \leq \bar{X}_n - \mu \leq \Delta\right) \\&= P\left(-\frac{\Delta}{\hat{\sigma}_*} \sqrt{n} \leq \frac{\bar{X}_n - \mu}{\hat{\sigma}_*} \sqrt{n} \leq \frac{\Delta}{\hat{\sigma}_*} \sqrt{n}\right) \\&= P\left(-\frac{\Delta}{\hat{\sigma}_*} \sqrt{n} \leq T \leq \frac{\Delta}{\hat{\sigma}_*} \sqrt{n}\right) \\&= F_T\left(\frac{\Delta}{\hat{\sigma}_*} \sqrt{n}\right) - F_T\left(-\frac{\Delta}{\hat{\sigma}_*} \sqrt{n}\right) = 2F_T\left(\frac{\Delta}{\hat{\sigma}_*} \sqrt{n}\right) - 1,\end{aligned}$$

gdzie użyliśmy symetrii rozkładu t -Studenta: $F_T(-x) = 1 - F_T(x)$.

Postępując analogicznie do przypadku znanej wariancji dostajemy:

$$\Delta = t_{1-\alpha/2; n-1} \frac{\hat{\sigma}_*}{\sqrt{n}}, \quad \text{gdzie } t_{1-\alpha/2; n-1} = F_T^{-1}\left(1 - \frac{\alpha}{2}\right)$$

jest kwantylem rzędu $1 - \frac{\alpha}{2}$ rozkładu t -Studenta o $n - 1$ stopniach swobody

Przykładowa tablica kwantyli rozkładu t -Studenta

df	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
32	1.309	1.694	2.037	2.449	2.738
34	1.307	1.691	2.032	2.441	2.728
36	1.306	1.688	2.028	2.434	2.719
38	1.304	1.686	2.024	2.429	2.712
∞	1.282	1.645	1.960	2.326	2.576

Przykładowa tablica kwantyli rozkładu t -Studenta

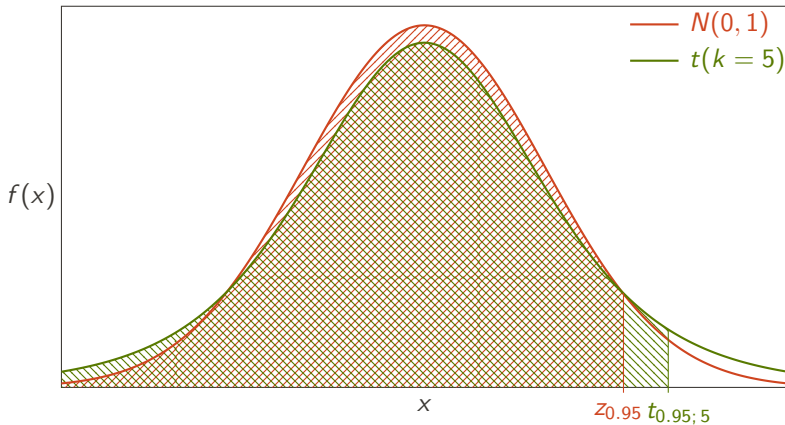
df	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	3.078	6.314	12.706	31.821	63.657
2				6.965	9.925
3				4.541	5.841
4				3.747	4.604
5				3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
32	1.309	1.694	2.037	2.449	2.738
34	1.307	1.691	2.032	2.441	2.728
36	1.306	1.688	2.028	2.434	2.719
38	1.304	1.686	2.024	2.429	2.712
∞	1.282	1.645	1.960	2.326	2.576

Liczba stopni swobody k
(df – degrees of freedom)

Rozkład t -Studenta



Rozkład t -Studenta



Kwantyle rozkładu t -Studenta są nieznacznie większe od odpowiadających im kwantyli rozkładu normalnego (rozkład t -Studenta ma grubsze ogony). Jest to spowodowane estymowaniem wariancji z danych (zwiększa niepewność przy konstrukcji przedziału ufności)

Przykład: rozkład normalny z nieznaną wariancją

Próba o liczności $n = 16$ pomiarów pewnej cechy o rozkładzie normalnym statystyki $\bar{x}_n = 1$ oraz $\hat{s}_*^2 = 9$. Wyznacz przedział ufności dla wartości oczekiwanej na poziomie ufności $1 - \alpha$ dla $\alpha = 0.05$.

Przykład: rozkład normalny z nieznaną wariancją

Próba o liczności $n = 16$ pomiarów pewnej cechy o rozkładzie normalnym statystyki $\bar{x}_n = 1$ oraz $\hat{s}_*^2 = 9$. Wyznacz przedział ufności dla wartości oczekiwanej na poziomie ufności $1 - \alpha$ dla $\alpha = 0.05$.

Rozwiązanie: Przedział ufności ma postać:

$$[\bar{x}_n - \Delta, \bar{x}_n + \Delta], \quad \text{gdzie } \Delta = t_{1-\alpha/2; n-1} \frac{\hat{s}_*}{\sqrt{n}}$$

Przykład: rozkład normalny z nieznaną wariancją

Próba o liczności $n = 16$ pomiarów pewnej cechy o rozkładzie normalnym statystyki $\bar{x}_n = 1$ oraz $\hat{s}_*^2 = 9$. Wyznacz przedział ufności dla wartości oczekiwanej na poziomie ufności $1 - \alpha$ dla $\alpha = 0.05$.

Rozwiązanie: Przedział ufności ma postać:

$$[\bar{x}_n - \Delta, \bar{x}_n + \Delta], \quad \text{gdzie } \Delta = t_{1-\alpha/2; n-1} \frac{\hat{s}_*}{\sqrt{n}}$$

$$t_{1-\alpha/2; n-1} = t_{0.975; 15} = 2.131$$

Przykład: rozkład normalny z nieznaną wariancją

Próba o liczności $n = 16$ pomiarów pewnej cechy o rozkładzie normalnym statystyki $\bar{x}_n = 1$ oraz $\hat{s}_*^2 = 9$. Wyznacz przedział ufności dla wartości oczekiwanej na poziomie ufności $1 - \alpha$ dla $\alpha = 0.05$.

Rozwiązanie: Przedział ufności ma postać:

$$[\bar{x}_n - \Delta, \bar{x}_n + \Delta], \quad \text{gdzie } \Delta = t_{1-\alpha/2; n-1} \frac{\hat{s}_*}{\sqrt{n}}$$

$$t_{1-\alpha/2; n-1} = t_{0.975; 15} = 2.131$$

$$\Delta = 2.131 \frac{3}{\sqrt{16}} = 1.598$$

Przykład: rozkład normalny z nieznaną wariancją

Próba o liczności $n = 16$ pomiarów pewnej cechy o rozkładzie normalnym statystyki $\bar{x}_n = 1$ oraz $\hat{s}_*^2 = 9$. Wyznacz przedział ufności dla wartości oczekiwanej na poziomie ufności $1 - \alpha$ dla $\alpha = 0.05$.

Rozwiązanie: Przedział ufności ma postać:

$$[\bar{x}_n - \Delta, \bar{x}_n + \Delta], \quad \text{gdzie } \Delta = t_{1-\alpha/2; n-1} \frac{\hat{s}_*}{\sqrt{n}}$$

$$t_{1-\alpha/2; n-1} = t_{0.975; 15} = 2.131$$

$$\Delta = 2.131 \frac{3}{\sqrt{16}} = 1.598$$

$$[\bar{x}_n - \Delta, \bar{x}_n + \Delta] = [1 - 1.598, 1 + 1.598] = [0.598, 2.598]$$

Zadanie

Zadanie 4

Postanowiono oszacować średnią wartość cen wynajmuj kawalerki w pewnej dzielnicy. W tym celu pobrano próbę o wielkości $n = 25$ uzyskując średnią cenę $\bar{x}_n = 1.4$ (w tys. zł) i wartość estymatora wariancji $\hat{s}_*^2 = 0.09$. Wyznacz przedział ufności dla wartości oczekiwanej ceny na poziomie ufności $1 - \alpha$ dla $\alpha = 0.05$. Załóż na potrzeby zadania, że rozkład cen jest rozkładem normalnym.