

# Rachunek prawdopodobo- bieństwa i statystyka

Wykład 12

# P. ufności dla wartości oczekiwanej

Szukamy przedziału ufności dla nieznannej wartości oczekiwanej  $\mu$  populacji w której badana cecha ma rozkład normalny  $\mathcal{N}(x; \mu, \sigma)$ , w przypadku gdy znana jest dyspersja  $\sigma$ , na podstawie  $n$ -elementowej próby prostej.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \rightarrow \quad \mathcal{N}\left(\bar{x}; \mu, \frac{\sigma}{\sqrt{n}}\right)$$

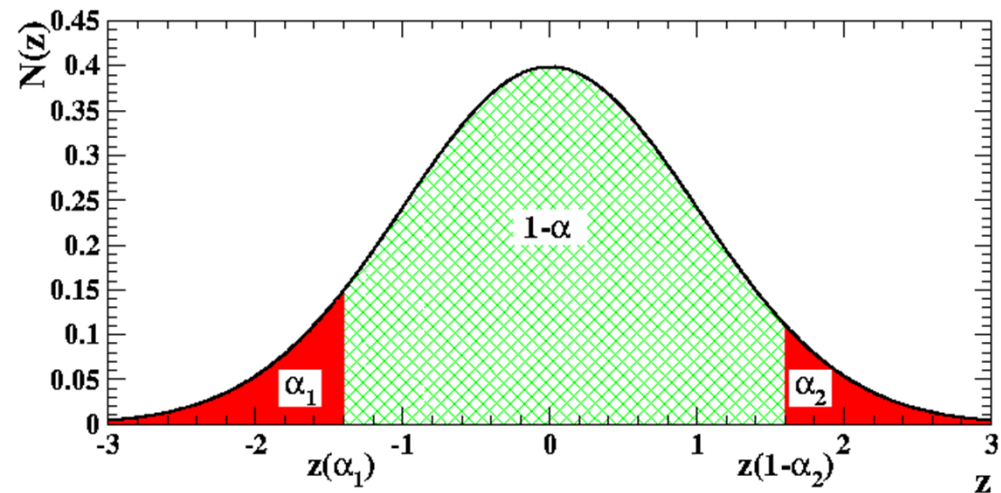
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad \rightarrow \quad \mathcal{N}(z; 0, 1)$$

Dla ustalonego  $\alpha = \alpha_1 + \alpha_2$  możemy znaleźć takie wartości  $z_1$  i  $z_2$  aby:

$$P(z_1 < z < z_2) = F(z_2) - F(z_1) = 1 - \alpha$$

Wystarczy założyć, że  $z_1 = z(\alpha_1)$  oraz  $z_2 = z(1 - \alpha_2)$  są kwantylami rozkładu zmiennej losowej z rzędów, odpowiednio,  $\alpha_1$  oraz  $1 - \alpha_2$ :

$$P\left(z(\alpha_1) < \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} < z(1 - \alpha_2)\right) = F(z(1 - \alpha_2)) - F(z(\alpha_1)) = 1 - \alpha_2 - \alpha_1 = 1 - \alpha$$



Math  
Player

# P. ufności dla wartości oczekiwanej

Rozwiązując nierówność w nawiasie względem parametru  $\mu$  dostajemy poszukiwany przedział ufności:

$$\bar{x} - z(1 - \alpha_2) \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} - z(\alpha_1) \frac{\sigma}{\sqrt{n}}$$

Przypadki szczególne:

- $\alpha_1 = 0, \alpha_2 = \alpha: z(0) = -\infty, z(1 - \alpha_2) = z(1 - \alpha)$

$$P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z(1 - \alpha)\right) = 1 - \alpha \quad \mu > \bar{x} - z(1 - \alpha) \frac{\sigma}{\sqrt{n}}$$

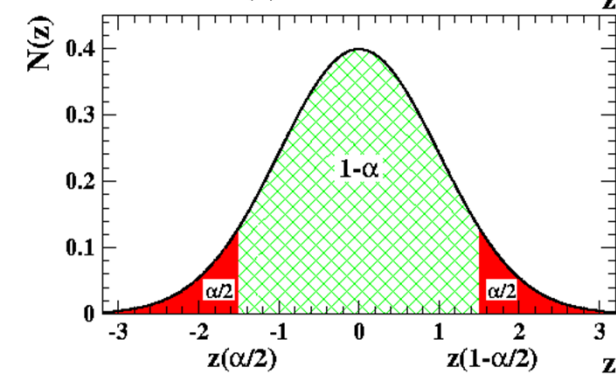
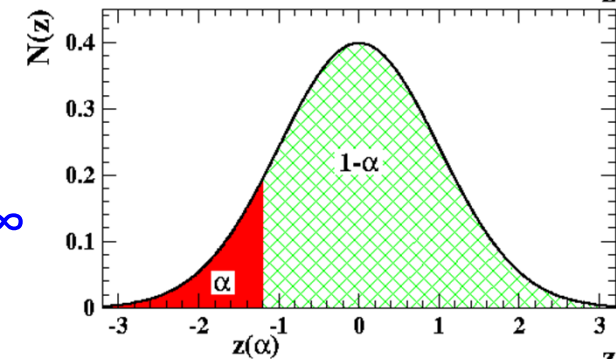
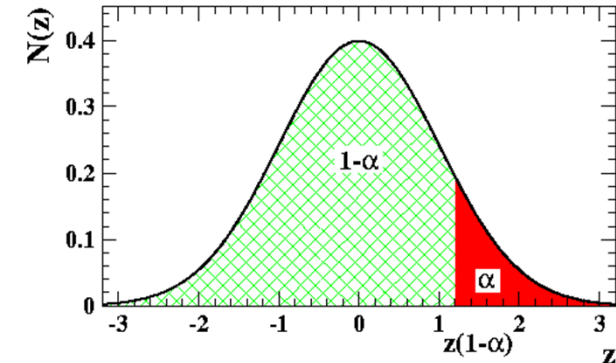
- $\alpha_1 = \alpha, \alpha_2 = 0: z(\alpha_1) = z(\alpha), z(1 - \alpha_2) = z(1) = +\infty$

$$P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > z(\alpha)\right) = 1 - \alpha$$

$$\mu < \bar{x} - z(\alpha) \frac{\sigma}{\sqrt{n}} = \bar{x} + z(1 - \alpha) \frac{\sigma}{\sqrt{n}}$$

- $\alpha_1 = \alpha_2 = \frac{1}{2} \alpha: \bar{x} - z\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} - z\left(\frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}}$

$$\bar{x} - z\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}}$$



# P. ufności dla parametru $\mu$ - przykład

**Przykład:** W losowo wybranej grupie 10 samochodów pewnej marki przeprowadzono badanie zużycia benzyny na tej samej trasie o długości 100 km. Okazało się, że średnie zużycie benzyny wynosiło 8.1 litra. Skądinąd wiadomo, że badana cecha ma rozkład normalny o dyspersji 0.8 litra. Wyznaczyć 99%-ową realizację przedziału ufności dla wartości przeciętnej zużycia paliwa na danej trasie przez samochody tej marki.

Szukamy przedziału ufności dla wartości oczekiwanej  $\mu$  w sytuacji gdy znamy dyspersję  $\sigma$  rozkładu normalnego z którego pochodzą pomiary.

$$P\left(z\left(\frac{\alpha}{2}\right) < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z\left(1 - \frac{\alpha}{2}\right)\right) = F\left(z\left(1 - \frac{\alpha}{2}\right)\right) - F\left(z\left(\frac{\alpha}{2}\right)\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$$

$$\bar{x} - z\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}}$$

**Wielkości dane:**  $\bar{x} = 8.1$        $\sigma = 0.8$        $n = 10$        $\alpha = 0.01$

**Z tablic odczytujemy:**  $z\left(1 - \frac{\alpha}{2}\right) = z(0.995) = 2.58$

**Szukany przedział ufności:**  $\mu \in (7.45, 8.75) \Rightarrow \Delta = 1.3$

# Przedział ufności dla $\mu$ - nieznaną $\sigma$

Szukamy przedziału ufności dla nieznaną wartości oczekiwanej  $\mu$  populacji w której badana cecha ma rozkład normalny  $\mathcal{N}(x; \mu, \sigma)$ , w przypadku nieznaną dyspersji  $\sigma$ , na podstawie  $n$ -elementowej próby prostej.

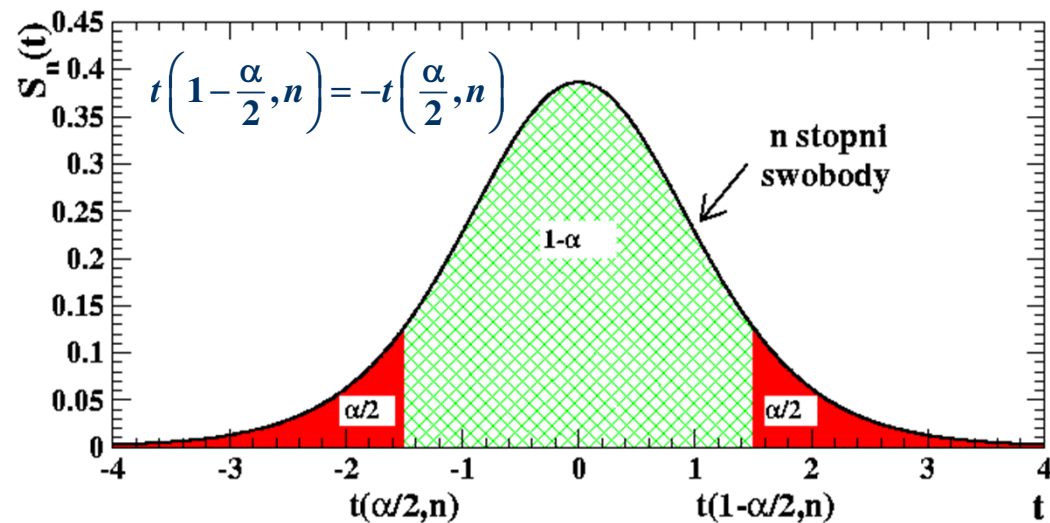
Math  
Player

Wiemy, że statystyka:

$$t = \frac{\bar{x} - \mu}{s_x / \sqrt{n}}$$

podlega rozkładowi Studenta o  $n-1$  stopniach swobody.

Niech  $t(\alpha, n)$  będzie kwantylem rzędu  $\alpha$  rozkładu Studenta o  $n$  stopniach swobody.



$$P\left(\left|\frac{\bar{x} - \mu}{s_x / \sqrt{n}}\right| < t\left(1 - \frac{\alpha}{2}, n - 1\right)\right) = F\left(t\left(1 - \frac{\alpha}{2}, n - 1\right)\right) - F\left(t\left(\frac{\alpha}{2}, n - 1\right)\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$$

$$\bar{x} - t\left(1 - \frac{\alpha}{2}, n - 1\right) \frac{s_x}{\sqrt{n}} < \mu < \bar{x} + t\left(1 - \frac{\alpha}{2}, n - 1\right) \frac{s_x}{\sqrt{n}} \quad \Rightarrow \quad \Delta = 2t\left(1 - \frac{\alpha}{2}, n - 1\right) \frac{s_x}{\sqrt{n}}$$

# P. ufności dla parametru $\mu$ - przykład

**Przykład:** W losowo wybranej grupie 10 samochodów pewnej marki przeprowadzono badanie zużycia benzyny na tej samej trasie o długości 100 km. Okazało się, że średnie zużycie benzyny wynosiło 8.1 litra, natomiast odchylenie standardowe wynosiło 0.8 litra. Zakładając, że badana cecha podlega rozkładowi normalnemu, wyznaczyć 99%-ową realizację przedziału ufności dla wartości przeciętnej zużycia paliwa na danej trasie przez samochody tej marki.

Szukamy przedziału ufności dla wartości oczekiwanej  $\mu$  w sytuacji gdy nie znamy dyspersji  $\sigma$  rozkładu normalnego z którego pochodzą pomiary, a jedynie jej estymatę  $s_x$ .

$$\bar{x} - t\left(1 - \frac{\alpha}{2}, n - 1\right) \frac{s_x}{\sqrt{n}} < \mu < \bar{x} + t\left(1 - \frac{\alpha}{2}, n - 1\right) \frac{s_x}{\sqrt{n}}$$

Wielkości dane:  $\bar{x} = 8.1$        $s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 0.8$        $n = 10$        $\alpha = 0.01$

Z tablic odczytujemy:  $t\left(1 - \frac{\alpha}{2}, n - 1\right) = t(0.995, 9) = 3.25$

Szukany przedział ufności:  $\mu \in (7.23, 8.97) \Rightarrow \Delta = 1.74$

# Przedział ufności dla $\mu$ - duża próbka

Założmy, że badana cecha populacji ma dowolny rozkład o nieznanym wartości oczekiwanej  $\mu$  i skończonej wariancji  $\sigma^2$ , oraz że mamy liczną ( $n \geq 100$ ) próbkę losową.

Na podstawie CTG Lindeberga-Levy'ego wiemy, że statystyka  $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$  ma asymptotyczny rozkład normalny  $\mathcal{N}(z; 0, 1)$ .

Ze względu na dużą próbkę, nieznaną wartość  $\sigma$  zastępujemy jej estymatą  $s_x$ , a szukany przedział ufności ma postać:

$$\bar{x} - z \left(1 - \frac{\alpha}{2}\right) \frac{s_x}{\sqrt{n}} < \mu < \bar{x} + z \left(1 - \frac{\alpha}{2}\right) \frac{s_x}{\sqrt{n}}$$

**Przykład:** W celu oszacowania czasu absencji chorobowej pracowników pewnego zakładu wybrano losową grupę 100 pracowników i zanotowano liczby dni opuszczonych z powodu choroby w ciągu ostatniego roku. Na tej podstawie znaleźć 95% realizację przedziału ufności dla nieznanego wartości przeciętnej czasu absencji chorobowej wszystkich pracowników tego zakładu.

Nieobecności [dni]	0	1-3	4-6	7-9	10-15	16-24	25-30
Liczba pracowników	13	37	22	17	8	2	1

# Przedział ufności dla $\mu$ - duża próbka

Wielkości dane:  $n = 100$   $\alpha = 0.05$

Obliczamy:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^7 n_i x_i = 4.875 \quad s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^7 n_i (x_i - \bar{x})^2} = 4.731$$

Ponieważ mamy do czynienia z liczną próbką, więc szukany przedział ufności możemy wyznaczyć na podstawie:

$$\bar{x} - z \left(1 - \frac{\alpha}{2}\right) \frac{s_x}{\sqrt{n}} < \mu < \bar{x} + z \left(1 - \frac{\alpha}{2}\right) \frac{s_x}{\sqrt{n}}$$

Z tablic odczytujemy:  $z \left(1 - \frac{\alpha}{2}\right) = z(0.975) = 1.96$

95%-wa realizacja przedziału ufności dla nieznannej wartości przeciętnej czasu absencji chorobowej pracowników tego zakładu to  $3.95 < \mu < 5.80$

Gdybyśmy zastosowali rozkład Studenta  $t \left(1 - \frac{\alpha}{2}, n - 1\right) = t(0.975, 99) = 1.98$  to otrzymalibyśmy:

$$\bar{x} - t \left(1 - \frac{\alpha}{2}, n - 1\right) \frac{s_x}{\sqrt{n}} < \mu < \bar{x} + t \left(1 - \frac{\alpha}{2}, n - 1\right) \frac{s_x}{\sqrt{n}} \quad \Rightarrow \quad 3.94 < \mu < 5.81$$



# Przedział ufności dla wariancji

Założmy, że badana cecha populacji ma rozkład normalny  $\mathcal{N}(x; \mu, \sigma)$  o nieznanym parametrach  $\mu$  i  $\sigma$ , a próbka nie jest liczna ( $n < 50$ ).

Wiemy, że statystyka  $u = \frac{(n-1)s_x^2}{\sigma^2}$  podlega rozkładowi  $\chi^2$  o  $n-1$  stopniach swobody.

Korzystając z kwantyli rozkładu  $\chi^2$  znajdujemy:

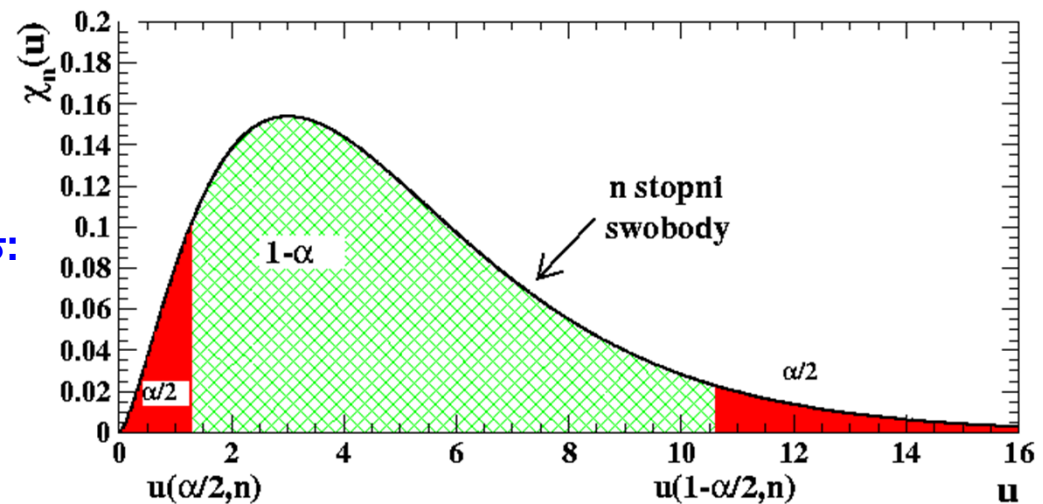
$$P\left(u\left(\frac{\alpha}{2}, n-1\right) < \frac{(n-1)s_x^2}{\sigma^2} < u\left(1-\frac{\alpha}{2}, n-1\right)\right) = 1-\alpha$$

a stąd przedziały ufności dla  $\sigma^2$ :

$$\frac{(n-1)s_x^2}{u\left(1-\frac{\alpha}{2}, n-1\right)} < \sigma^2 < \frac{(n-1)s_x^2}{u\left(\frac{\alpha}{2}, n-1\right)}$$

oraz dla odchylenia standardowego  $\sigma$ :

$$\sqrt{\frac{n-1}{u\left(1-\frac{\alpha}{2}, n-1\right)}} s_x < \sigma < \sqrt{\frac{n-1}{u\left(\frac{\alpha}{2}, n-1\right)}} s_x$$



# Przedział ufności dla wariancji

**Przykład:** Wykonano pomiary pewnej wielkości uzyskując następujące wyniki: 87, 102, 119, 81, 97, 93, 100, 114, 99, 100, 113, 93, 95, 85, 123, 99. Zakładając, że pomiary tej wielkości podlegają rozkładowi normalnemu, znajdź 90% realizacje przedziałów ufności dla wariancji i odchylenia standardowego.

Wielkości dane:  $n = 16$     $\alpha = 0.10$

Obliczamy:  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 100$     $s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \cong 12$

Z tablic odczytujemy:  $u\left(\frac{\alpha}{2}, n-1\right) = u(0.05, 15) = 7.261$

$$u\left(1 - \frac{\alpha}{2}, n-1\right) = u(0.95, 15) = 24.996$$

90%-ową realizacją przedziału ufności dla wariancji jest:

$$\frac{15 \times 12^2}{24.996} < \sigma^2 < \frac{15 \times 12^2}{7.261}$$

A stąd:  $86.41 < \sigma^2 < 297.48$    oraz    $9.3 < \sigma < 17.2$

# P. ufności dla wariancji – duża próbka

Założmy, że badana cecha populacji ma rozkład normalny  $\mathcal{N}(x; \mu, \sigma)$  o nieznanym parametrach  $\mu$  i  $\sigma$ , a próbka jest liczna ( $n \geq 50$ ).

W tym przypadku korzystamy z faktu, że statystyka:

$$\sqrt{2\chi^2} = \sqrt{2 \frac{(n-1)s_x^2}{\sigma^2}} = \frac{s_x}{\sigma} \sqrt{2(n-1)}$$

ma w przybliżeniu rozkład normalny  $\mathcal{N}\left(x; \sqrt{2n-3}, 1\right)$

Przedział ufności dla nieznanego odchylenia standardowego znajdujemy z warunku:

$$P\left(\sqrt{2n-3} - z\left(1 - \frac{\alpha}{2}\right) < \frac{s_x}{\sigma} \sqrt{2(n-1)} < \sqrt{2n-3} + z\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha$$

Rozwiązując powyższą nierówność otrzymujemy:

$$\frac{s_x \sqrt{2(n-1)}}{\sqrt{2n-3} + z\left(1 - \frac{\alpha}{2}\right)} < \sigma < \frac{s_x \sqrt{2(n-1)}}{\sqrt{2n-3} - z\left(1 - \frac{\alpha}{2}\right)}$$

# Obszar ufności dla parametrów $\mu$ i $\sigma$

Założmy, że badana cecha populacji ma rozkład normalny  $\mathcal{N}(x; \mu, \sigma)$  o nieznanymi parametrach  $\mu$  i  $\sigma$ . Chcemy przeprowadzić jednoczesną estymację przedziałową obu parametrów.

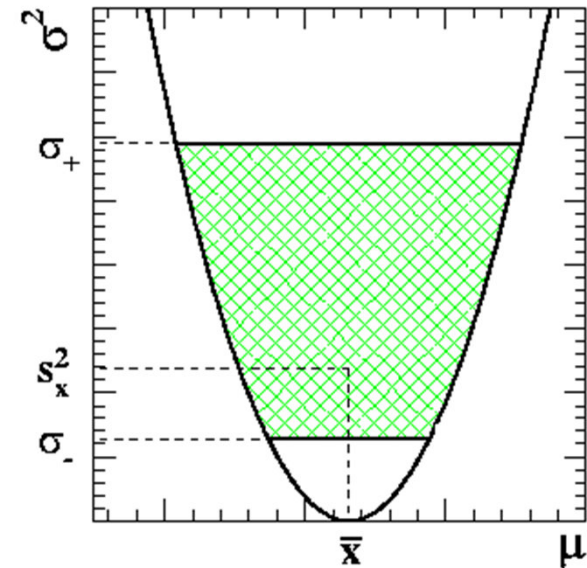
$$\mathbf{P}\left(z_- \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq z_+, u_- \leq \frac{(n-1)s_x^2}{\sigma^2} \leq u_+\right) = \mathbf{P}\left(z_- \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq z_+\right) \mathbf{P}\left(u_- \leq \frac{(n-1)s_x^2}{\sigma^2} \leq u_+\right) = 1 - \alpha$$

$$u_- = u\left(\frac{\alpha}{4}, n-1\right) \qquad u_+ = u\left(1 - \frac{\alpha}{4}, n-1\right)$$

$$z \equiv z_+ = z\left(1 - \frac{\alpha}{4}\right) = -z\left(\frac{\alpha}{4}\right) = -z_-$$

$$\mathbf{P}\left(z_- \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq z_+\right) = \sqrt{1 - \alpha} \cong 1 - \frac{\alpha}{2}$$

$$\mathbf{P}\left(u_- \leq \frac{(n-1)s_x^2}{\sigma^2} \leq u_+\right) = \sqrt{1 - \alpha} \cong 1 - \frac{\alpha}{2}$$



$$\sigma_{\pm}^2 = \frac{(n-1)s_x^2}{u_{\mp}} \qquad \text{oraz} \qquad \left| \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right| \leq z \qquad \Rightarrow \qquad \sigma^2 \geq \frac{n(\bar{x} - \mu)^2}{z^2}$$

# Obszar ufności dla parametrów $\mu$ i $\sigma$

**Przykład:** Zmierzono średnice 51 drzew wybranych losowo z lasu sosnowego i otrzymano średnią średnicę 37.3 cm oraz wariancję 13.77 cm<sup>2</sup>. Zakładając, że średnice drzew mają rozkład normalny, wyznaczyć 90%-ową realizację zbioru ufności dla wartości przeciętnej i wariancji średnicy drzewa z tego lasu.

Wielkości dane:  $n = 51$     $\alpha = 0.10$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 37.3 \text{ cm} \qquad s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 13.77 \text{ cm}^2$$

Wyznaczamy wartości odpowiednich kwantyli:

$$u_- = u\left(\frac{\alpha}{4}, n-1\right) = u(0.025, 50) = 32.357 \qquad z = z\left(1 - \frac{\alpha}{4}\right) = z(0.975) = 1.96$$

$$u_+ = u\left(1 - \frac{\alpha}{4}, n-1\right) = u(0.975, 50) = 71.420$$

90%-owa realizacja zbioru ufności dla wartości przeciętnej i wariancji średnicy drzewa dana jest przez:

$$\sigma_-^2 = \frac{(n-1)s_x^2}{u_+} = 9.64 \text{ cm}^2 \qquad \sigma_+^2 = \frac{(n-1)s_x^2}{u_-} = 21.28 \text{ cm}^2$$
$$\sigma^2 \geq \frac{n(\bar{x} - \mu)^2}{z^2} = 13.28 \times (37.3 - \mu)^2$$

# Estymacja parametru $p$ rozkładu $\mathcal{B}_k(N,p)$

**Przykład:** Spośród  $n = 100$  wirów w wannie zaobserwowano  $k = 43$  wiry zgodne z kierunkiem ruchu wskazówek zegara. Znajdź przedział ufności na poziomie ufności 95% dla parametru  $p$  rozkładu dwumianowego z tych danych.

$$P(k \leq 43; p_+, n = 100) = \sum_{k=0}^{43} \binom{100}{k} p_+^k (1 - p_+)^{100-k} = 0.025 \quad \Rightarrow \quad p_+ \approx 0.533$$

$$P(k \geq 43; p_-, n = 100) = \sum_{k=43}^{100} \binom{100}{k} p_-^k (1 - p_-)^{100-k} = 0.025 \quad \Rightarrow \quad p_- \approx 0.332$$

A więc przedział ufności dla parametru  $p$  na poziomie ufności  $1 - \alpha = 95\%$  dany jest przez  $(0.332, 0.553)$  i obejmuje  $p=0.5$ , co oznacza, że żaden kierunek wiru nie jest uprzywilejowany.

# Estymacja parametru $p$ rozkładu $\mathcal{B}_k(N, p)$

Chcemy znaleźć przedział ufności na poziomie  $1 - \alpha$  dla parametru  $p$  rozkładu dwumianowego  $\mathcal{B}_k(N, p)$ , przy znanym parametrze  $N$ , jeśli próbka danych licząca  $n$  liczb jest na tyle duża ( $n \geq 100$ ) że możemy odwołać się do centralnego twierdzenia granicznego ( $\mu = Np$     $\sigma = \sqrt{Np(1-p)}$ ).

$$\mathcal{B}_{k_i}(N, p) = \binom{N}{k_i} p^{k_i} (1-p)^{N-k_i} \quad k = \sum_{i=1}^n k_i \quad \bar{k} = \frac{1}{n} \sum_{i=1}^n k_i$$

Stosujemy CTG przybliżając rozkład dwumianowy rozkładem Gaussa:

$$\mathbf{P} \left( \left| \frac{\bar{k} - Np}{\sqrt{Np(1-p)} / \sqrt{n}} \right| \leq z \left( 1 - \frac{\alpha}{2} \right) \right) = 1 - \alpha \quad \Rightarrow \quad \left( \frac{\bar{k} - Np_{\pm}}{\sqrt{Np_{\pm}(1-p_{\pm})} / \sqrt{n}} \right)^2 = z^2$$

Rozwiązujemy równanie kwadratowe względem  $p$ :

$$(nN + z^2)p^2 - (2n\bar{k} + z^2)p + \frac{n}{N}\bar{k}^2 = 0$$

$$\Delta = 4n^2\bar{k}^2 + 4nz^2\bar{k} + z^4 - 4n^2\bar{k}^2 - 4z^2\frac{n}{N}\bar{k}^2 = 4z^2 \left[ n\bar{k} \left( 1 - \frac{\bar{k}}{N} \right) + \frac{z^2}{4} \right]$$

# Estymacja parametru p rozkładu $\mathcal{B}_k(N,p)$

Poszukiwany przedział ufności na poziomie  $1-\alpha$  dany jest przez:

$$\begin{aligned} p_{\pm} &= \frac{2n\bar{k} + z^2 \pm 2z\sqrt{n\bar{k}\left(1 - \bar{k}/N\right) + z^2/4}}{2(nN + z^2)} = \\ &= \left(1 + \frac{z^2}{nN}\right)^{-1} \left( \frac{\bar{k}}{N} + \frac{z^2}{2nN} \pm z\sqrt{\frac{1}{nN} \frac{\bar{k}}{N} \left(1 - \frac{\bar{k}}{N}\right) + \frac{z^2}{4(nN)^2}} \right) \cong \\ &\cong \left\| z^2 \ll nN \right\| \cong \frac{\bar{k}}{N} \pm z\sqrt{\frac{1}{nN} \frac{\bar{k}}{N} \left(1 - \frac{\bar{k}}{N}\right)} \equiv \hat{p} \pm z\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\hat{p}(1 - \hat{p})}{nN}} \end{aligned}$$

**Przykład:** Wśród 350 wybranych losowo wyrobów znaleziono 31 wyrobów wadliwych.

Wykorzystując wynik badania kontrolnego podać 95%-ową realizację przedziału ufności dla frakcji wyrobów dobrych w całej partii wyprodukowanych wyrobów.

$$n = 1 \quad N = 350 \quad k = 350 - 31 = 319$$

$$\alpha = 0.05 \quad z\left(1 - \frac{\alpha}{2}\right) = z(0.975) = 1.96$$

$$p_- \cong 0.88$$

$$p_+ \cong 0.94$$