dr hab., prof. SGH Barbara Kowalczyk
Zakład Statystyki Matematycznej
Instytut Ekonometrii, KAE SGH

Przypomnienie własności wartości oczekiwanej i wariancji

Twierdzenie 1. Własności wartości oczekiwanej

$X, Y -$ zmienne losowe

$$E(X+Y) = EX + EY$$

$$E(aX) = aEX$$

$$E(X+b) = EX + b$$

Jeżeli zmienne X, Y są niezależne, wtedy $E(X \cdot Y) = EX \cdot EY$

Twierdzenie 2. Własności wariancji $D^2X = E(X - EX)^2 = EX^2 - (EX)^2$

$X, Y -$ zmienne losowe

$$D^2(aX) = a^2 D^2 X$$

$$D^2(X+b) = D^2 X$$

$$D^2(X+Y) = D^2X + D^2Y + 2Cov(X,Y)$$

Jeżeli zmienne X, Y są niezależne, wtedy $D^2(X+Y) = D^2X + D^2Y$

Uwaga Jeżeli zmienne X, Y są niezależne, wtedy $D^2(X-Y) = D^2X + D^2Y$

Proste ćwiczenie 1: korzystając z twierdzenia 2 udowodnić uwagę 1.

Twierdzenie 3.

Niech $X_1, \dots, X_n$ będą niezależnymi zmiennymi losowymi takie, że $EX_i = m_i$, $D^2X_i = \sigma_i^2$. Wtedy:

$$E(a_1X_1 + \dots + a_nX_n) = a_1m_1 + \cdots + a_nm_n,$$

$$D^2(a_1X_1 + \dots + a_nX_n) = a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2.$$

Twierdzenie 4.

Niech $X_1, \dots, X_n$ będą niezależnymi zmiennymi losowymi pochodzącymi z rozkładów normalnych, $X_i \sim N(m_i, \sigma_i^2)$ dla $i = 1, \dots, n$. Wtedy:

$$a_1 X_1 + \ldots + a_n X_n \sim N(a_1 m_1 + \cdots + a_n m_n, a_1^2 \sigma_1^2 + \cdots + a_n^2 \sigma_n^2).$$

Proste ćwiczenie 2: podać przykład rozkładu takiego, że mając niezależne zmienne losowe $X_1, \dots, X_n$ o tym rozkładzie ich suma $X_1 + \cdots + X_n$ nie jest już zmienną losową o tym rozkładzie.

Próba prosta i statystyka

Definicja 1.

Prosta próba losowa (iid – independent identically distributed) jest to ciąg niezależnych zmiennych losowych $X_1, \dots, X_n$ o tym samym rozkładzie.

Definicja 2.

Statystyka $T = T(X_1, \dots, X_n)$ jest to dowolna funkcja próby (zależy tylko od próby, nie może zależeć od nieznanych parametrów).

Rozkład średniej z próby

Twierdzenie 5.

$X_1, \dots, X_n$ – prosta próba losowa z dowolnego rozkładu takiego, że $EX_i = m$ i $D^2 X_i = \sigma^2$. Wtedy dla średniej z próby $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ zachodzi:

$$E(\bar{X}) = m,$$

$$D^2(\bar{X}) = \frac{\sigma^2}{n}.$$

Proste ćwiczenie 3: udowodnić to twierdzenie

Twierdzenie 6.

$X_1, \dots, X_n$ – prosta próba losowa z rozkładu normalnego $N(m, \sigma^2)$, wtedy

$$\bar{X} \sim N\left(m, \frac{\sigma^2}{n}\right)$$

oraz

$$\frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X}-m}{\sigma}\sqrt{n} \sim N(0,1).$$

Proste ćwiczenie 4: udowodnić, że $E\left(\frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}}\right) = 0$ oraz $D^2\left(\frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}}\right) = 1$

Twierdzenie 7. (wniosek z centralnego twierdzenia granicznego)

Niech $X_1, \dots, X_n$ będzie prostą próbą losową z rozkładu takiego, że dla każdego $i = 1, \dots, n$ zachodzi $EX_i = m$ i $D^2X_i = \sigma^2$. Wtedy średnia z próby $\bar{X}$ ma asymptotyczny rozkład normalny $N\left(m, \frac{\sigma^2}{n}\right)$.

Wniosek dla statystyki: dla dużej liczebności próby rozkład średniej z próby można przybliżać rozkładem normalnym.

Definicja 3

$\boldsymbol{Z}$ – zmienna losowa o standardowym rozkładzie normalnym $N(0,1)$. Niech $P(Z \leq z_\alpha) = \Phi(z_\alpha) = \alpha$. Wtedy z_α nazywa się kwantylem standardowego rozkładu normalnego rzędu α.

Definicja 4

$X -$ zmienna losowa o rozkładzie ciągłym. Niech $P(X \leq k_\alpha) = \mathrm{F}(k_\alpha) = \alpha$, gdzie F jest dystrybuantą danego rozkładu. Wtedy k_α nazywa się kwantylem danego rozkładu rzędu α.

Wniosek: Jeżeli X ma rozkład ciągły, wtedy zachodzi $k_\alpha = F^{-1}(\alpha)$.

Proste ćwiczenie 5:

Podać kwantyl rzędu $\alpha = 0{,}5$ w rozkładzie standardowym normalnym

Proste ćwiczenie 6:

Podać kwantyl rzędu $\alpha = 0{,}1$ w rozkładzie a) t-studenta o 5 stopniach swobody b) chi-kwadrat o 6 stopniach swobody

Przykład 1.

$X_1, \ldots, X_{20}$ ~ niezależne zmienne losowe o rozkładzie normalnym $X_i \sim N\left(1, \sigma = \sqrt{40}\right)$, $i = 1, \ldots, 20$. Obliczyć prawdopodobieństwa: $P(\bar{X} < 1)$ oraz $P(\bar{X} > 0)$.

Rozwiązanie:

Najpierw obliczamy, $E\bar{X} = 1$, $D^2\bar{X} = \frac{40}{20} = 2$, czyli $\bar{X} \sim N(1, \sigma^2 = 2)$, czyli $\bar{X} \sim N(1, \sigma = \sqrt{2})$ oraz $\frac{\bar{X}-1}{\sqrt{2}} \sim N(0,1)$. Następnie:

$$P(\bar{X} < 1) = P\left(\frac{\bar{X}-1}{\sqrt{2}} < \frac{1-1}{\sqrt{2}}\right) = P(Z < 0) = \Phi(0) = 0{,}5,$$

$$P(\bar{X} > 0) = P\left(\frac{\bar{X}-1}{\sqrt{2}} > \frac{0-1}{\sqrt{2}}\right) = P(Z > -0{,}707) = P(Z < 0{,}707) = \Phi(0{,}707) \approx 0{,}7611,$$

<u>Rozkład wariancji z próby</u>

Rozkład chi-kwadrat (podstawowa interpretacja)

Jeżeli $X_1, \ldots, X_n$ są niezależnymi zmiennymi losowymi z tego samego rozkładu $N(0,1)$, wtedy

$$X_1^2 + \cdots + X_n^2 \sim \chi^2(n)$$

Twierdzenie 8.

Jeżeli $X \sim \chi^2(n)$, to $EX = n$, $D^2X = 2n$.

Twierdzenie 9.

Jeżeli $Y_1, \ldots, Y_k$ są niezależnymi zmiennymi losowymi z rozkładu chi-kwadrat, $Y_i \sim \chi^2(n_i)$. Wtedy:

$$Y_1 + \cdots + Y_k \sim \chi^2(n_1 + \cdots + n_k).$$

Definicja 6 (wariancja z próby):

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2, \text{ gdzie } m = EX \text{ jest znane}$$

Twierdzenie 10.

Jeżeli $X_1, \dots, X_n$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie normalnym $X_i \sim N(m, \sigma)$, wtedy $\bar{X}$ i $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ są niezależne oraz $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

Wniosek 1:

Jeżeli $X_1, \dots, X_n$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie normalnym $X_i \sim N(m, \sigma)$, wtedy $\bar{X}$ i $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ są niezależne oraz $\frac{n\tilde{S}^2}{\sigma^2} \sim \chi^2(n-1)$.

Twierdzenie 11.

Jeżeli $X_1, \dots, X_n$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie normalnym $X_i \sim N(m, \sigma)$, w którym m jest znane, wtedy są niezależne oraz $\frac{n\hat{S}^2}{\sigma^2} \sim \chi^2(n)$, gdzie $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$

Przykład 3 Niezależne zmienne losowe X, Y, Z, W mają odpowiednio rozkłady $X \sim N(2,1)$, $Y \sim N(0,1)$, $Z \sim \chi^2(4)$, $W \sim \chi^2(2)$. Obliczyć $P((X-2)^2 + Y^2 + Z + W < 20{,}0902)$. **Odp. 0,99**

<u>Związek pomiędzy średnią z próby i odchyleniem standardowym z próby</u>

Twierdzenie 12. (podstawowa interpretacja rozkładu t-studenta)

Niech X, Y będą niezależnymi zmiennymi losowymi takimi, że $X \sim N(0,1)$, $Y \sim \chi^2(n)$. Wtedy

$$\frac{X}{\sqrt{Y/n}} \sim t(n),$$

Twierdzenie 13.

Niech $X_1, \dots, X_n$ będzie prostą próbą losową z populacji $N(m, \sigma)$. Wtedy

$$\frac{\bar{X}-m}{S}\cdot\sqrt{n}\sim t(n-1)$$

gdzie

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

Twierdzenie 14.

Niech $X_1,\dots,X_n$ będzie prostą próbą losową z populacji $N(m,\sigma)$. Wtedy

$$\frac{\bar{X}-m}{\tilde{S}}\cdot\sqrt{n-1}\sim t(n-1)$$

gdzie

$$\tilde{S}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

Twierdzenie 15.

Niech $X\sim t(n)$. Wtedy:

- wartość oczekiwana zmiennej X istnieje dla $n\geq 2$ oraz $EX=0$,
- wariancja zmiennej losowej X istnieje dla $n\geq 3$ oraz $D^2X=\frac{n}{n-2}$.

Przykład 4 Próba prosta $X_1,\dots,X_{16}$ pochodzi z rozkładu $N(1,\sigma^2)$. Obliczyć prawdopodobieństwo $P(\bar{X}>1+0{,}335\cdot S)$. **Odp. 0,1**

Rozkład F-Snedecora – rozkład ilorazu wariancji z próby

Twierdzenie 16. (podstawowa interpretacja rozkładu F-Snedecora):

Założenie: X,Y – niezależne zmienne losowe takie, że $X\sim\chi^2(n_1)$, $Y\sim\chi^2(n_2)$. Wtedy

$$\frac{X/n_1}{Y/n_2}\sim F(n_1,n_2).$$

Twierdzenie 17.

Niech $X\sim F(n_1,n_2)$. Wtedy

- wartość oczekiwana zmiennej X istnieje dla $n_2 \geq 3$ oraz $EX = \frac{n_2}{n_2-2}$
- wariancja zmiennej losowej X istnieje dla $n_2 \geq 5$ oraz $D^2X = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-4)(n_2-2)^2}$.

Przykład 5 Niech $X_1, \dots, X_n$ będzie p.p.l. z rozkładu $N(0, \sigma^2)$. Obliczyć $P\left(\frac{X_1^2+X_2^2}{X_3^2} > 399\right)$

Odp. $P(F(2,1) > 199,5) \approx 0,05$

Statystyki pozycyjne

Definicja 1: Niech $X_1, X_2, \dots, X_n$będzie ciągiem zmiennych losowych, natomiast $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$ ciągiem ich realizacji. Ciąg realizacji (ciąg liczb) porządkujemy niemalejąco:

$$X_{1:n}(\omega) \leq X_{2:n}(\omega) \leq \cdots \leq X_{n:n}(\omega).$$

Zmienne losowe $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ nazywamy statystykami pozycyjnymi lub statystykami porządkowymi.

Twierdzenie 1:

Jeśli $X_1, X_2, \dots, X_n$ są niezależnymi zmiennymi losowymi z rozkładu **ciągłego** o funkcji gęstości $f(x) = \frac{dF(x)}{dx}$, to k-ta $X_{k:n}$ statystyka pozycyjna ma gęstość

$$f_{k:n}(x) = n\binom{n-1}{k-1} f(x)F(x)^{k-1}\left(1-F(x)\right)^{n-k} \qquad (1)$$

Szkic dowodu:

$$F_{k:n}(x) = P(X_{k:n} \leq x) = \sum_{i=k}^{n} \binom{n}{i} (F(x))^i(1-F(x))^{n-i}$$

Następnie należy policzyć pochodną tej dystrybuanty. Pamiętamy, że

$$\left(F^i(x)\right)' = i\left(F(x)\right)^{i-1} \cdot f(x)$$

Twierdzenie 2:

Jeżeli $X_1, X_2, \dots, X_n$ jest prostą próbą losową z rozkładu jednostajnego na przedziale (0,1), czyli $X_i \sim U(0,1)$, to $X_{k:n} \sim Beta(k, n-k+1)$, w szczególności

$$EX_{k:n} = \frac{k}{n+1}$$

$$D^2X_{k:n} = \frac{k(n-k+1)}{(n+1)^2(n+2)}$$

Ćwiczenie 1. Udowodnić twierdzenie 2.

$$f_{k:n}(x) = n\binom{n-1}{k-1} f(x)F(x)^{k-1}\left(1-F(x)\right)^{n-k} \qquad (1)$$

$$f(x) = \begin{cases} 1 & dla \;\; x \in (0,1) \\ 0 & w\, p.p. \end{cases}$$

$$F(x) = \begin{cases} 0 & dla\; x \le 0 \\ x & dla\; x \in (0,1) \\ 1 & dla\; x \ge 1 \end{cases}$$

Podstawiając do wzoru (1) mamy:

Dla $x \notin (0,1)$ $\quad f_{k:n}(x) = 0$

Dla $x \in (0,1)$ $\quad f_{k:n}(x) = n\binom{n-1}{k-1} \cdot x^{k-1} \cdot (1-x)^{n-k}$

Wiemy, że funkcja gęstości rozkładu beta z parametrami (a, b) to

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \qquad gdzie\; x \in (0,1)$$

Mamy pokazać, że

$$f_{k:n}(x) = n\binom{n-1}{k-1} \cdot x^{k-1} \cdot (1-x)^{n-k}$$

oraz

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

to te same funkcje.

$$a = k, \qquad b = n - k + 1$$

Przypomnienie funkcji gamma:

$$\Gamma(a+1) = \Gamma(a) \cdot a$$

$$\Gamma(n) = (n-1)!$$

$$\Gamma(1) = 1 \qquad \Gamma(1/2) = \sqrt{\pi}$$

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} = \frac{n!}{(k-1)!\,(n-k)!} = n\binom{n-1}{k-1}$$

$$EX = \frac{a}{a+b} = \frac{k}{k+n-k+1} = \frac{k}{n+1}$$

$$D^2X = \frac{ab}{(a+b)^2(a+b+1)} = \frac{k(n-k+1)}{(n+1)^2(n+2)}$$