Statystyka matematyczna

Mariusz Przybycień

Wydział Fizyki i Informatyki Stosowanej
Akademia Górniczo-Hutnicza

Wykład 15

Nieparametryczny przedział ufności

W przypadku gdy nie wiadomo z jakiego rozkładu pochodzi próbka losowa (np. mała próbka - nie można skorzystać z CTG) znajdujemy przedziały ufności dla wielkości, które nie są parametrami żadnych konkretnych rozkładów p-twa.

Przedział ufności dla mediany $\boldsymbol{m_{0.5}}$ rozkładu ciągłego zmiennej $\mathbf{X}$:

Dla próbki losowej prostej, $x_1, x_2, ..., x_n$, z populacji o medianie $m_{0.5}$ mamy:

$$P(\mathrm{X} \geqslant m_{0.5}) = P(\mathrm{X} \leqslant m_{0.5}) = \frac{1}{2}$$

Niech n_- oznacza liczbę elementów próby losowej mniejszych od $m_{0.5}$.
Ponieważ, n_- podlega rozkładowi $\mathcal{B}_{n_-}\left(n, p = \frac{1}{2}\right)$, więc dla ustalonego p-twa α można znaleźć takie a i b, że:

$$P(n_- \leqslant a) = \sum_{i=0}^{a} \binom{n}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{n-i} = \sum_{i=0}^{a} \binom{n}{i} \left(\frac{1}{2}\right)^n = \frac{\alpha}{2}$$
$$P(n_- \geqslant b) = \sum_{i=b}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{n-i} = \sum_{i=b}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n = \frac{\alpha}{2}$$

Wartości a i b wybieramy tak, aby obliczone p-twa były jak najbliższe $\alpha/2$. Korzystając z uporządkowanej próbki losowej, $x_{(1)}, ..., x_{(a)}, ..., x_{(b)}, ..., x_{(n)}$, znajdujemy przedział ufności dla mediany: $\boldsymbol{x_{(a)} < m_{0.5} < x_{(b)}}$, gdzie $b = n + 1 - a$.

Przedział ufności dla mediany

Przykład: W dużej firmie uzyskano następujące wyniki w rezultacie losowego sprawdzenia wieku 20 pracowników: 24, 31, 28, 43, 28, 56, 48, 39, 52, 32, 38, 49, 51, 49, 62, 33, 41, 58, 63, 56. Znajdź na poziomie ufności 95% przedział ufności dla mediany wieku pracowników tej firmy.

Porządkujemy rosnąco listę wieku 20 losowo wybranych pracowników:

24 28 28 31 32 33 38 39 41 43 48 49 49 51 52 56 56 58 62 63

Z tablic rozkładu dwumianowego dla $n = 20$ i $p = \frac{1}{2}$ odczytujemy

$$P(\mathrm{X} \leqslant 5) = 0.0207$$

A więc dla $a = 5$ osiągamy wartość prawdopodobieństwa najbliższą $\alpha/2 = 0.025$.

Jako dolną granicę przedziału ufności należy wybrać piąty element w uporzadkowanej rosnąco liście wieku pracowników, $\mathrm{X}_{(5)} = 32$.

Podobnie znajdujemy górną granicę przedziału ufności $\mathrm{X}_{(n+1-a)} = \mathrm{X}_{(16)} = 56$.

Przedział ufności na poziomie ufności 95% dla mediany wieku pracowników to:

$$32 < m_{0.5} < 56$$

Nieparametryczne testy hipotez dla pojedynczej próby

Rozważmy na poziomie istotności α test hipotezy dotyczącej mediany $\mathrm{m}_{0.5}$.
Niech $H_0: \; m_{0.5} = m_0$ przy hipotezie alternatywnej $H_1: \; m_{0.5} > m_0$.

Zakładamy, że próba losowa, $x_1, x_2, ..., x_n$, pochodzi z populacji o rozkładzie ciągłym, tak aby $P(\mathrm{X} < m_{0.5}) = 0.5$.

Jako statystykę testową wybieramy, N^+, oznaczającą liczbę elementów próby losowej większych od m_0. Hipotezę H_0 odrzucamy na rzecz hipotezy H_1, jeśli wartość n^+ statystyki testowej N^+ spełnia relację $n^+ > k$, gdzie k wyznaczamy z warunku:

$$P(N^+ \geqslant k \,|\, H_0) = \sum_{i=k}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n = \alpha$$

Zbiory krytyczne dla innych hipotez alternatywnych mają odpowiednio postać:

Dla $H_1: m_{0.5} < m_0$ mamy $N^+ \leqslant k$, gdzie $\sum_{i=0}^{k} \binom{n}{i} \left(\frac{1}{2}\right)^n = \alpha$

Dla $H_1: m_{0.5} \neq m_0$ mamy $N^+ \geqslant k_1$, gdzie $\sum_{i=k_1}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n = \frac{\alpha}{2}$

lub

$N^+ \leqslant k$, gdzie $\sum_{i=0}^{k} \binom{n}{i} \left(\frac{1}{2}\right)^n = \frac{\alpha}{2}$

W przypadku kiedy nie można osiąnąć dokładnie p-stw α lub $\alpha/2$ należy tak wybrać k (lub k i k_1) aby p-twa były możliwie bliskie α (lub $\alpha/2$).

Test znaków dla mediany

Test znaków: na przykładzie $H_0 : m_{0.5} = m_0$ oraz $H_1 : m_{0.5} > m_0$:

- zastępujemy wszystkie wartości w próbie losowej większe od m_0 przez znak $+$, a mniejsze od m_0 przez znak $-$; jeśli w próbie któraś z wartości jest równa m_0, to ją usuwamy, zmniejszając jednocześnie liczebność próby n,
- niech n^+ oznacza liczbę znaków $+$; obliczamy p-two z rozkładu dwumianowego o parametrach n oraz $p = \frac{1}{2}$: $\gamma = P(N^+ \geqslant n^+)$,
- jeśli $\gamma < \alpha$ to odrzucamy hipotezę H_0 na rzecz H_1.

Przykład (test znaków): Dla następującego zbioru danych eksperymentalnych:

$$1.51,\ 1.35,\ 1.69,\ 1.48,\ 1.29,\ 1.27,\ 1.54,\ 1.39,\ 1.45$$

przeprowadź na poziomie istotności $\alpha = 0.05$ test hipotezy $H_0 : m_{0.5} = 1.4$ przy hipotezie alternatywnej $H_1 : m_{0.5} > 1.4$.

Zastępujemy wartości w próbie losowej większe od 1.4 przez znak plus, a mniejsze przez znak minus: $+ - + + - - + - +$

Dla $n^+ = 5$, $n = 9$ oraz $p = \frac{1}{2}$ obliczamy $\gamma = P(N^+ \geqslant 5) = 0.5$

Ponieważ $\gamma > 0.05$, więc nie ma powodu do odrzucenia hipotezy zerowej, tzn. mediana nie jest większa od 1.4.

Test znaków dla mediany - duża próba losowa

Jeśli liczebność próby losowej jest duża, $n > 10$, można przybliżyć rozkład dwumianowy rozkładem normalnym. Statystyka testowa, N^+, przy założeniu prawdziwości hipotezy zerowej H_0 ma więc rozkład normalny o wartości oczekiwnej $\mu = np = n/2$ oraz wariancji $\sigma^2 = np(1-p) = n/4$.

Zmienna losowa $z = \dfrac{N^+ - n/2}{\sqrt{n/4}} = \dfrac{2N^+ - n}{\sqrt{n}}$ ma więc rozkład $\mathcal{N}(0,1)$.

W zależności od hipotezy alternatywnej, zbiory krytyczne przyjmują postać:

$$H_0 : m_{0.5} = m_0, \quad H_1 : m_{0.5} > m_0, \quad W = \langle z(\alpha), +\infty)$$
$$H_0 : m_{0.5} = m_0, \quad H_1 : m_{0.5} < m_0, \quad W = (-\infty, -z(\alpha)\rangle$$
$$H_0 : m_{0.5} = m_0, \quad H_1 : m_{0.5} \neq m_0, \quad W = (-\infty, -z(\alpha/2)\rangle \cup \langle z(\alpha/2), +\infty)$$

Przykład: 15 pracownikom zlecono przycięcie krzewów winogron. Efektywność wykonanej przez nich pracy mierzona liczbą osobo-godzin/akr wynosiła: $5.2, 5.0, 4.8, 3.9, 6.1, 4.2, 4.4, 5.5, 5.8, 4.5, 4.2, 5.3, 4.9, 4.7, 4.9$. Przetestuj na poziomie istotności $\alpha = 0.05$ hipotezę, że mediana czasu potrzebnego na przycięcia 1 akra winogron wynosi 4.5 h, wobec hipotezy alternatywnej, że ta mediana jest większa.

- $H_0 : m_{0.5} = 4.5, \quad H_1 : m_{0.5} > 4.5, \quad \alpha = 0.05;$
- zapisujemy próbę losową w postaci: $+ + + - + - - + + - + + + + +$
- statystyka testowa: $z = \frac{2N^+ - n}{\sqrt{n}} = 1.6 \notin W = \langle z(\alpha), \infty) = \langle 1.645, \infty)$

Ponieważ $z \notin W$, więc na poziomie istotności $\alpha = 0.05$ nie ma powodu do odrzucenia hipotezy H_0 na rzecz hipotezy H_1.

Test Wilcoxona

Test Wilcoxona służy do testowania hipotezy dotyczącej mediany i jest ulepszeniem testu znaków.

- W teście rang Wilcoxona zastępujemy poszczególne elementy próbki losowej przez rangi uporządkowane zgodnie z wartościami $z_i = |x_i - m_0|$. Najmniejszej wartości z_i przypisujemy rangę 1, itd. Jeśli dwa różne z_i mają tą samą wartość to obu przypisujemy rangi średnie.
- Zastępujemy wszystkie z_i przez ich rangi, a wartościom w próbie losowej większym od m_0 przypisujemy znak "$+$", a mniejszym od m_0 znak "$-$"; jeśli w próbie któraś z wartości jest równa m_0, to ją usuwamy, zmniejszając jednocześnie liczebność próby n.
- Niech W_X^+ będzie sumą rang elementów próbki losowej którym przypisano znak "$+$", a W_X^- sumą rang elementów próbki, którym przypisano znak "$-$".
- Uwaga: Test Wilcoxona może być stosowany tylko do rozkładów ciągłych i symetrycznych.
- W zależności od hipotezy alternatywnej, zbiory krytyczne przyjmują postać:

$$H_1 : m_{0.5} > m_0, \quad W = \langle c, \infty), \quad \text{gdzie} \quad P(W_X^+ \geqslant c) = \alpha$$

$$H_1 : m_{0.5} < m_0, \quad W = (-\infty, c\rangle, \quad \text{gdzie} \quad P(W_X^+ \leqslant c) = \alpha$$

$$H_1 : m_{0.5} \neq m_0, \quad W = (-\infty, c_1\rangle \cup \langle c_2, \infty), \quad \text{gdzie} \quad P(W_X^+ \leqslant c_1) = P(W_X^+ \geqslant c_2) = \frac{\alpha}{2}$$

- Dla dużej próbki losowej ($n > 20$), rozkład W_X^+ przy założeniu prawdziwości hipotezy H_0 ma rozkład normalny o wartości oczekiwanej i wariancji:

$$\mathcal{E}[W_X^+] = \frac{1}{4}n(n+1) \quad \text{oraz} \quad \mathcal{V}[W_X^+] = \frac{1}{24}n(n+1)(2n+1)$$

Test Wilcoxona

- Zmienna losowa $z = \dfrac{W_X^+ - \mathcal{E}[W_X^+]}{\sqrt{\mathcal{V}[W_X^+]}}$ ma więc rozkład $\mathcal{N}(0,1)$.

  W zależności od hipotezy alternatywnej, zbiory krytyczne przyjmują postać:

$$H_0 : m_1 = m_2, \quad H_1 : m_1 > m_2, \quad W = \langle z(\alpha), +\infty)$$
$$H_0 : m_1 = m_2, \quad H_1 : m_1 < m_2, \quad W = (-\infty, -z(\alpha)\rangle$$
$$H_0 : m_1 = m_2, \quad H_1 : m_1 \neq m_2, \quad W = (-\infty, -z(\alpha/2)\rangle \cup \langle z(\alpha/2), +\infty)$$

- Przykład: W eksperymencie uzyskano dane: $1.51, 1.35, 1.69, 1.48, 1.29, 1.27, 1.54, 1.39, 1.45$. Przetestuj na poziomie istotności $\alpha = 0.05$ hipotezę, że mediana wynosi $m_0 = 1.4$ wobec hipotezy alternatywnej, że jest różna od tej wartości.

  - $H_0 : m_{0.5} = 1.4, \quad H_1 : m_{0.5} \neq 1.4, \quad \alpha = 0.05$
  -

Wartość	1.51	1.35	1.69	1.48	1.29	1.27	1.54	1.39	1.45
$z_i = \|x_i - 1.4\|$	0.11	0.05	0.29	0.08	0.11	0.13	0.14	0.01	0.01
Populacja	$+$	$-$	$+$	$+$	$-$	$-$	$+$	$-$	$+$
Ranga	5.5	3	9	4	5.5	7	8	1.5	1.5

  - $W_X^+ = 29, \ n = 9, \ W = (-\infty, 6\rangle \cup \langle 38, \infty), \ W_X^+ \notin W$ - nie ma powodu do odrzucenia hipotezy H_0.

Nieparametryczne testy hipotez dla dwóch niezależnych prób

Chcemy porównać mediany, m_1 i m_2, dwóch niezależnych populacji o rozkładach ciągłych, na podstawie, odpowiednio, n_1 i n_2 elementowych próbek losowych.

Niech N_{1b} będzie zmienną losową oznaczającą liczbę elementów z próbki 1.

Przy założeniu, że mediany obu próbek są takie same ($H_0 : m_1 = m_2$), zmienna N_{1b} ma rozkład hipergeometryczny ($n_1 + n_2 \equiv 2k$):

$$P(N_{1b} = n_{1b}) = \binom{n_1}{n_{1b}} \binom{n_2}{k - n_{1b}} \Bigg/ \binom{n_1 + n_2}{k}, \qquad n_{1b} = 0, 1, 2, ..., n_1$$

W zależności od hipotezy alternatywnej, zbiory krytyczne przyjmują postać:

$$H_1 : m_1 > m_2, \;\; W = \langle 0, c \rangle, \;\; \text{gdzie} \;\; P(N_{1b} \leqslant c) = \alpha$$

$$H_1 : m_1 < m_2, \;\; W = \langle c, n_1 \rangle, \;\; \text{gdzie} \;\; P(N_{1b} \geqslant c) = \alpha$$

$$H_1 : m_1 \neq m_2, \;\; W = \langle 0, c_1 \rangle \cup \langle c_2, n_1 \rangle, \;\; \text{gdzie} \;\; P(N_{1b} \leqslant c_1) = P(N_{1b} \geqslant c_2) = \frac{\alpha}{2}$$

Dla dużych prób losowych, $n_1 > 5$ i $n_2 > 5$, można skorzystać z przybliżenia rozkładem Gaussa.

Niech teraz N_{ib} i N_{ia} będą zmiennymi losowymi oznaczającymi, odpowiednio, liczby elementów poniżej (b) i powyżej (a) mediany w i-tej próbce ($i = 1, 2$).

Standaryzujemy zmienną losową N_{1a} za pomocą wartości oczekiwanej i wariancji:

$$\mathcal{E}[N_{1a}] = \frac{N_a n_1}{n} \qquad \mathcal{V}[N_{1a}] = \frac{N_a n_1 n_2 N_b}{n^2 (n-1)}$$

gdzie $N_a = N_{1a} + N_{2a}, \; N_b = N_{1b} + N_{2b},$ oraz $n = n_1 + n_2$.

Nieparametryczne testy hipotez dla dwóch niezależnych prób

Zmienna losowa $z = \dfrac{N_{1a} - \mathcal{E}[N_{1a}]}{\sqrt{\mathcal{V}[N_{1a}]}}$ ma więc rozkład $\mathcal{N}(0,1)$.

W zależności od hipotezy alternatywnej, zbiory krytyczne przyjmują postać:

$$H_0: m_1 = m_2, \quad H_1: m_1 > m_2, \quad W = \langle z(\alpha), +\infty)$$
$$H_0: m_1 = m_2, \quad H_1: m_1 < m_2, \quad W = (-\infty, -z(\alpha)\rangle$$
$$H_0: m_1 = m_2, \quad H_1: m_1 \neq m_2, \quad W = (-\infty, -z(\alpha/2)\rangle \cup \langle z(\alpha/2), +\infty)$$

Przykład: Dane na temat zużycia opon (w tys. km) dwóch producentów I i II:

$$\mathrm{I}: \quad 34, 32, 37, 35, 42, 43, 47, 58, \quad 59, \quad 62, \quad 69, \quad 71, 78, 84$$
$$\mathrm{II}: \quad 39, 48, 54, 65, 70, 76, 87, 90, 111, 118, 126, 127$$

Przetestuj na poziomie istotności $\alpha = 0.05$ hipotezę, że mediana zużycia opon od producenta II jest większa niż opon od producenta I.

- $H_0: m_1 = m_2, \quad H_1: m_1 < m_2, \quad \alpha = 0.05$
- Ponieważ $n_1 = 14 > 5$ oraz $n_2 = 12 > 5$, więc stosujemy przybliżenie normalne.
- Znajdujemy: $N_{1b} = 10,\ N_{1a} = 4,\ N_{2b} = 3,\ N_{2a} = 9, \quad \Rightarrow \quad N_b = 13,\ N_a = 13$
- Wyznaczamy: $\mathcal{E}[N_{1a}] = 7,\ \mathcal{V}[N_{1a}] = 1.68,\ z_{\mathrm{test}} = -0.023,\ W = (-\infty, -1.645\rangle$
- Ponieważ $z_{\mathrm{test}} \notin W$, więc nie ma powodu do odrzucenia hipotezy o równości median w obu próbach.

Test sumy rang Wilcoxona

Załóżmy, że mamy n_1 elementową próbkę z populacji I oraz n_2 elementową próbkę z populcji II. Testujemy hipotezę o równości median ($H_0 : m_1 = m_2$).

1. Łączymy próbki w jedną o liczebności $n_1 + n_2$ i porządkujemy rosnąco, ale pamiętając, który element pochodzi z której próbki.
2. Sumujemy rangi elementów z próbki II - niech ta suma będzie równa R.
3. Obliczamy wartość statystyki testowej Wilcoxona $W = R - \frac{1}{2}n_2(n_2 + 1)$
4. W zależności od hipotezy alternatywnej, zbiory krytyczne przyjmują postać:

$$H_1 : m_1 > m_2, \quad \mathcal{W} = \langle c, \infty), \quad \text{gdzie} \quad P(W \geqslant c) = \alpha$$
$$H_1 : m_1 < m_2, \quad \mathcal{W} = (0, c\rangle, \quad \text{gdzie} \quad P(W \leqslant c) = \alpha$$
$$H_1 : m_1 \neq m_2, \quad \mathcal{W} = \langle 0, c_1\rangle \cup \langle c_2, n_1 n_2), \quad \text{gdzie} \quad P(W \leqslant c_1) = P(W \geqslant c_2) = \frac{\alpha}{2}$$

Przykład: Ceny opon (w USD) podobnego typu dwóch producentów wynoszą odpowienio:

$$\text{I} : 85,\ 99,\ 100,\ 110,\ 105,\ 87 \quad \text{oraz} \quad \text{II} : 67,\ 69,\ 70,\ 93,\ 105,\ 90,\ 110,\ 115$$

Wykonaj test sumy rang Wilcoxona z $\alpha = 0.05$ do zweryfikowania hipotezy o równości median, przy hipotezie alternatywnej, że mediany są różne.

- $H_0 : m_1 = m_2, \quad H_1 : m_1 \neq m_2, \quad \alpha = 0.05, \quad n_1 = 6, \quad n_2 = 8$
-

Wartość	67	69	70	85	87	90	93	99	100	105	105	110	110	115
Populacja	II	II	II	I	I	II	II	I	I	I	II	I	II	II
Ranga	1	2	3	4	5	6	7	8	9	10.5	10.5	12.5	12.5	14

- $R = 56 \;\Rightarrow\; W = 20 \;\Rightarrow\; W \notin (0, 9\rangle \cup (38, 48)$

Statystyka porządku

Niech $X_1, X_2, \dots$ będą zmiennymi losowymi o rozkładzie zadanym dystrybuantą $F(x)$.

Definicja: Dla $k = 1, 2, \dots, n$ niech $X_{(k)}$ oznacza k-tą najmniejszą wartość spośród $X_1, X_2, \dots, X_n$. Wówczas niemalejący ciąg $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ nazywamy statystyką porządku, a $X_{(k)}$ k-tą zmienną w porządku.

Wniosek: Statystykę porządku otrzymuje się z nieuporządkowanej próbki poprzez permutację taką w wyniku której mamy: $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$

Uwaga: Statystyka porządku zależy także od n: $X_{(k)}$ jest k-tą najmniejszą wartością spośród n zmiennnych $X_1, X_2, \dots, X_n$. Czasem stosuje się bardziej precyzyjną notację $X_{1:n}, X_{2:n}, \dots, X_{n:n}$.

Wartości ekstremalne i ich rozkłady:

$$X_{(1)} = \min\{X_1, X_2, \dots, X_n\} \quad \text{oraz} \quad X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$$

$$\begin{aligned} F_{X_{(1)}}(x) &= 1 - P\left(X_{(1)} > x\right) = 1 - P\left(X_1 > x, X_2 > x, \dots X_n > x\right) = \\ &= 1 - \prod_{k=1}^{n} P\left(X_k > x\right) = 1 - \left(1 - F(x)\right)^n \\ F_{X_{(n)}}(x) &= P\left(X_1 \leqslant x, X_2 \leqslant x, \dots X_n \leqslant x\right) = \prod_{k=1}^{n} P\left(X_k \leqslant x\right) = \left(F(x)\right)^n \end{aligned}$$

Dla rozkładów ciągłych znajdujemy funkcje gęstości prawdopodobieństwa:

$$f_{X_{(1)}}(x) = n\left(1 - F(x)\right)^{n-1} f(x) \quad \text{oraz} \quad f_{X_{(n)}}(x) = n\left(F(x)\right)^{n-1} f(x)$$

Łączny rozkład ekstremów

Przykład: W biegu na 100 metrów czasy uzyskiwane przez zawodników na poziomie olimpijskim mają rozkład płaski na przedziale $(9.6, 10)$ sekund. W finale uczetniczy 8 zawodników. Ile wynosi prawdopodobieństwo, że zwycięzca pobije dotychczasowy rekord 9.69 s?

$$F(x) = 2.5x - 24 \quad \Rightarrow \quad F_{X_{(1)}}(x) = 1 - (25 - 2.5x)^8 \quad \Rightarrow \quad p = F_{X_{(1)}}(9.69) \approx 0.8699$$

Twierdzenie: Łączna gęstość prawdopodobieństwazmiennych $X_{(1)}$ i $X_{(n)}$ dana jest przez:

$$f_{X_{(1)},X_{(n)}}(x,y) = \begin{cases} n(n-1)\left(F(y) - F(x)\right)^{n-2} f(y)f(x), & \text{dla } x < y \\ 0 & \text{dla pozostałych } (x,y) \end{cases}$$

Dowód:

$$P\left(X_{(1)} > x,\ X_{(n)} \leqslant y\right) = P\left(x < X_k \leqslant y,\ k = 1,2,...,n\right) =$$

$$= \prod_{k=1}^{n} P\left(x < X_k \leqslant y\right) = \left(F(y) - F(x)\right)^n, \quad \text{dla } \ x < y$$

$$P\left(X_{(1)} \leqslant x,\ X_{(n)} \leqslant y\right) + P\left(X_{(1)} > x,\ X_{(n)} \leqslant y\right) = P\left(X_{(n)} \leqslant y\right)$$

$$\begin{aligned} F_{X_{(1)},X_{(n)}}(x,y) &= F_{X_{(n)}}(y) - P\left(X_{(1)} > x,\ X_{(n)} \leqslant y\right) = \\ &= \begin{cases} \left(F(y)\right)^n - \left(F(y) - F(x)\right)^n, & \text{dla } \ x < y \\ \left(F(y)\right)^n, & \text{dla } \ x \geqslant y \end{cases} \end{aligned}$$

Aby otrzymać funkcję gęstości należy obliczyc pochodną po x i y.

Łączny rozkład ekstremów, zakres

Gęstość prawdopodobieństwazmiennej losowej $\mathrm{R}_n = \mathrm{X}_{(n)} - \mathrm{X}_{(1)}$ znajdujemy jako rozkład brzegowy dwuwymiarowej zmiennej $(\mathrm{R}_n, \mathrm{U} = \mathrm{X}_{(1)})$:

$$f_{\mathrm{R}_n}(r) = n(n-1)\int_{-\infty}^{\infty} (F(u+r) - F(u))^{n-2} f(u+r) f(u) \mathrm{d}u, \quad \text{dla } \; r > 0$$

Przykład: Dla zmiennej losowej $\mathrm{X} \in u(0,1)$ mamy:

$$f_{\mathrm{R}_n}(r) = n(n-1)\int_0^{1-r} (u+r-u)^{n-2} \cdot 1 \cdot 1 \cdot \mathrm{d}u = n(n-1) r^{n-2}(1-r), \quad 0 < r < 1$$

Wartość oczekiwana: $\mathcal{E}\left[\mathrm{R}_n\right] = \dfrac{n-1}{n+1}$.

Przykład: Niech $\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n$ będą niezależnymi zmiennymi losowymi z rozkładu wykładniczego $\mathcal{E}(x;\lambda) = \lambda e^{-\lambda x}$. Znajdź (a) $f_{\mathrm{X}_{(1)},\mathrm{X}_{(n)}}(x,y)$, (b) $f_{\mathrm{R}_n}(r)$.

$$\begin{aligned} f_{\mathrm{X}_{(1)},\mathrm{X}_{(n)}}(x,y) &= n(n-1)\left(1 - \mathrm{e}^{-\lambda y} - (1 - \mathrm{e}^{-\lambda x})\right)^{n-2} \cdot \lambda \mathrm{e}^{-\lambda y} \cdot \lambda \mathrm{e}^{-\lambda x} = \\ &= n(n-1)\lambda^2 \left(\mathrm{e}^{-\lambda x} - \mathrm{e}^{-\lambda y}\right)^{n-2} \mathrm{e}^{-\lambda(x+y)}, \quad \text{dla } \; 0 < x < y \end{aligned}$$

$$\begin{aligned} f_{\mathrm{R}_n}(r) &= n(n-1)\lambda^2 \int_0^{\infty} \left(\mathrm{e}^{-\lambda u} - \mathrm{e}^{-\lambda(u+r)}\right)^{n-2} \mathrm{e}^{-\lambda(2u+r)} \mathrm{d}u = \\ &= (n-1)\lambda^2 \left(1 - \mathrm{e}^{-\lambda r}\right)^{n-2} \mathrm{e}^{-\lambda r}, \quad \text{dla } \; r > 0 \end{aligned}$$