

Hipotezy statystyczne

- Testowanie hipotez statystycznych pozwala na sprawdzenie na podstawie wyników próby, przy zadanym poziomie ufności, czy jakieś twierdzenie (hipotezę) dotyczące populacji generalnej jest prawdziwe. Taką procedurę nazywamy **testem statystycznym**.
- Przykłady:
 - Czy próbka pochodzi z rozkładu normalnego?
 - Czy na podstawie próby można powiedzieć, że wartość oczekiwana jest większa niż pewna wybrana liczba?
 - Czy wariancje dwóch rozkładów, z których mamy dwie różne próbki, są sobie równe czy nie?
- Jak widać mamy zawsze do czynienia z pewną **hipotezą zerową** H_0 , którą sprawdzamy i **hipotezą alternatywną** H_1 , którą przyjmujemy gdy w wyniku testu odrzucamy H_0 .

RPIS 2024/2025 1

1

Testy statystyczne

- Schemat postępowania:
 - Sformułowanie hipotezy zerowej H_0 .
 - Określenie rozmiaru próbki n .
 - Ustalamy **poziom istotności** α , zwany również **poziomem błędem testu**. Jest to prawdopodobieństwo popełnienia **błędu I-go rodzaju**, a więc odrzucenia, w wyniku testu, hipotezy H_0 , podczas gdy jest ona w rzeczywistości prawdziwa. Zazwyczaj $\alpha < 0.1$.
 - Sformułowanie hipotezy alternatywnej H_1 .
 - Zakładamy chwilowo, że H_0 jest prawdziwa, i wybieramy statystykę testową δ , zależną od próby, o znanym rozkładzie prawdopodobieństwa.
 - Wyznamy **obszar krytyczny W** dla wartości δ , zależny od α , H_1 , i n , w taki sposób, że prawdopodobieństwo, że $\delta \in W$ wynosi α .
 - Wyliczamy, otrzymaną na podstawie pobranej próby, wartość statystyki δ . **Gdy należy ona do obszaru krytycznego W to odrzucamy H_0 , w przeciwnym razie twierdzimy, że nie ma podstaw do odrzucenia H_0 .**

RPIS 2024/2025 2

2

Testy statystyczne - uwagi

- Określenie obszaru krytycznego na podstawie α jest niejednoznaczne. Można to poprawić wprowadzając prawdopodobieństwo **błędu II-go rodzaju**: przyjęciu fałszywej H_0 , gdy w rzeczywistości prawdziwa jest hipoteza alternatywna H_1 . Prawdopodobieństwo to oznaczamy β , a $1-\beta$ nazywamy **mocą testu**.
- Jeżeli znamy $f(\delta|H_0)$ (α) i $f(\delta|H_1)$ (β) to obszar krytyczny wyznaczamy jako zakres wartości δ , który minimalizuje β przy zadanym α . Zazwyczaj znamy tylko $f(\delta|H_0)$ i nie znamy mocy testu co oznacza, że nie możemy twierdzić, że H_0 jest prawdziwa, a tylko, że nie odrzucamy H_0 . Test statystyczny, przy nieznaności β , nazywamy **testem istotności**.
- W praktyce opracowane są zestawy statystyk testowych i zakresy obszaru krytycznego dla różnych rodzajów hipotez H_0 i H_1 . W większości testów zakładamy, że próbka pochodzi z rozkładu normalnego.

RPIS 2024/2025 3

3

Testy statystyczne dotyczące E(X)

- Zał. Próbka o liczebności n pochodzi z rozkładu $N(E(X), \sigma^2)$.
- Badamy $H_0: E(X) = X_0$.
- Statystyką testową jest:

$$U = \frac{(\bar{X} - E(X))\sqrt{n}}{\sigma(X)}$$
 Gdy znamy odchylenie standardowe σ :

$$t = \frac{(\bar{X} - E(X))\sqrt{n}}{S(X)}$$
 Ma ona rozkład t-Studenta o $n-1$ stopniach swobody.

H_1	W (znane σ)	W (nieznane σ)
$E(X) \neq X_0$	$ u > u_{1-\alpha/2}$	$ t > t_{1-\alpha/2}$
$E(X) > X_0$	$u > u_{1-\alpha}$	$t > t_{1-\alpha}$
$E(X) < X_0$	$u < -u_{1-\alpha}$	$t < -t_{1-\alpha}$

RPIS 2024/2025 4

4

Testy statystyczne dotyczące równości dwóch wartości oczekiwanych

- Zał. Dwie populacje i dwie próbki: pierwsza o liczebności n_x pochodzi z rozkładu $N(E(X), \sigma_x^2)$, druga o liczebności n_y pochodzi z rozkładu $N(E(Y), \sigma_y^2)$.
- Badamy $H_0: E(X) = E(Y)$.
- Gdy σ_x^2 i σ_y^2 są znane statystyką testową jest $U = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$. Ma ona rozkład $N(0,1)$.

$$E(U) = \left(\frac{\sigma_x^2(X) + \sigma_y^2(Y)}{n_x + n_y} \right)^{-1/2} (E(X) - E(Y)) = 0$$

$$var(U) = \left(\frac{\sigma_x^2(X) + \sigma_y^2(Y)}{n_x + n_y} \right)^{-1} var(\bar{X} - \bar{Y}) = \left(\frac{\sigma_x^2(X) + \sigma_y^2(Y)}{n_x + n_y} \right)^{-1} (var(\bar{X}) + var(\bar{Y})) = \left(\frac{\sigma_x^2(X) + \sigma_y^2(Y)}{n_x + n_y} \right)^{-1} \left(\frac{\sigma_x^2(X)}{n_x} + \frac{\sigma_y^2(Y)}{n_y} \right) = 1$$

H_1	W
$E(X) \neq E(Y)$	$ u > u_{1-\alpha/2}$
$E(X) > E(Y)$	$u > u_{1-\alpha}$

RPIS 2024/2025 5

5

Testy statystyczne dotyczące równości dwóch wartości oczekiwanych

- Gdy σ_x^2 i σ_y^2 są nieznane lecz równe, statystyką testową jest

$$t = \frac{\bar{X} - \bar{Y}}{S(X, Y) \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

$$S(X, Y) = \sqrt{\frac{(n_x - 1)S^2(X) + (n_y - 1)S^2(Y)}{n_x + n_y - 2}}$$
 Ma ona rozkład t-Studenta o $(n_x + n_y - 2)$ stopniach swobody.

H_1	W
$E(X) \neq E(Y)$	$ t > t_{1-\alpha/2}$
$E(X) > E(Y)$	$t > t_{1-\alpha}$

RPIS 2024/2025 6

6

Testy statystyczne dotyczące równości dwóch wartości oczekiwanych

- Gdy σ_x^2 i σ_y^2 są nieznanne i są różne, statystyką testową jest

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S^2(X)}{n_x} + \frac{S^2(Y)}{n_y}}}$$

Ma ona rozkład przybliżony przez rozkład t-Studenta o efektywnej liczbie stopniach swobody $\text{int}(n_{\text{eff}})$.

$$n_{\text{eff}} = \frac{\left(\frac{S^2(X)}{n_x} + \frac{S^2(Y)}{n_y}\right)^2}{\frac{1}{n_x+1} \left(\frac{S^2(X)}{n_x}\right)^2 + \frac{1}{n_y+1} \left(\frac{S^2(Y)}{n_y}\right)^2} - 2$$

H_1	W
$E(X) \neq E(Y)$	$ t > t_{1-\alpha/2}$
$E(X) > E(Y)$	$t > t_{1-\alpha}$

RPIS 2024/2025 7

7

Wartość P (P-value)

Zaś: na podstawie próby uzyskaliśmy δ^{exp}

→ przyjmujemy H_0

RPIS 2024/2025 8

8

Wartość P (P-value)

Zwiększając poziom istotności α : δ^{exp} może znaleźć się w obszarze krytycznym W

→ odrzucamy H_0
Wartość P to graniczny poziom istotności α pomiędzy odrzuceniem a przyjęciem H_0

RPIS 2024/2025 9

9

Wartość P (P-value)

Konkretna próbka wyznacza jednoznacznie wartość statystyki δ na której opieramy test. Porównanie jej z granicami obszaru krytycznego prowadzi, przy zadanym poziomie istotności testu, do przyjęcia lub odrzucenia hipotezy H_0 . Zwiększenie poziomu istotności α zwiększa obszar krytyczny W. Przy odpowiednio dużym α wartość statystyki δ z próbki trafi do obszaru krytycznego.

Można zatem zapytać:
Jakiej wartości istotności testu α odpowiada otrzymana z próbki wartość statystyki δ ? Istotność testu α obliczoną na podstawie statystyki δ nazywamy **wartością P (P-value)**.

Czyli P-value jest najniższym poziomem istotności wymuszającym odrzucenie H_0 . Równocześnie jest to najwyższy poziom istotności wymuszający przyjęcie H_0 .

Używanie pojęcia P-value pozwala zatem na równoczesne testowanie hipotezy zerowej na wszystkich poziomach istotności:
dla $\alpha < P$ przyjmujemy H_0
dla $\alpha > P$ odrzucamy H_0

RPIS 2024/2025 10

10

Wartość P (P-value)

W praktyce:
dla $P < 0.01$ odrzucamy H_0 (0.01 jest najmniejszym zwyczajowym α)
dla $P > 0.1$ przyjmujemy H_0 (0.1 jest największym zwyczajowym α)
dla $0.01 < P < 0.1$ widać, że ustalenie α może zmienić konkluzję testu (zatem np. pożądane byłoby zebranie większej ilości danych).

Rola P-value w analizie danych jest szczególnie istotna, gdy δ leży (przy założonym wcześniej poziomie istotności α) blisko obszaru krytycznego.

P-value jest prawdopodobieństwem otrzymania (z rozkładu zgodnego z H_0) wartości co najmniej tak ekstremalnej jak δ .

Związek powyższej definicji z metodą największej wiarygodności.

Związek obszaru krytycznego z przedziałem ufności.

RPIS 2024/2025 11

11

Testy statystyczne rodzaju rozkładu

Test zerowania się współczynnika asymetrii i kurtozy (warunek konieczny rozkładu normalnego).

$$\beta_1 = \frac{E((x-E(X))^3)}{(\sigma(X))^3} \quad \beta_2 = \frac{E((x-E(X))^4)}{(\sigma(X))^4} - 3$$

- $H_0 : \beta_1 = \beta_2 = 0$
- Korzystamy z estymatorów

$$T_n(\beta_1) = \frac{M_3}{(M_2)^{3/2}} \quad T_n(\beta_2) = \frac{M_4}{(M_2)^2} - 3 \quad M_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^k$$

- Dla dużych n statystyki $Q_1 = \frac{\sqrt{n}T_n(\beta_1)}{\sqrt{6}}$ i $Q_2 = \frac{\sqrt{n}T_n(\beta_2)}{\sqrt{24}}$ mają rozkłady $N(0,1)$.
- $H_1 : \beta_1 \neq 0$ i $\beta_2 \neq 0$
- Obszarem krytycznym na poziomie istotności α jest przedział:
 $|Q_1| > u_{1-\alpha/2} \cup |Q_2| > u_{1-\alpha/2}$

RPIS 2024/2025 12

12

Wykres normalny (normal probability plot)

- Służy do zbadania, czy próbka pochodzi z rozkładu normalnego.
- Syt: dysponujemy n-elementową próbą x_1, \dots, x_n ; porządkujemy ją dostając $x_{(1)}, \dots, x_{(n)}$.
- Na jednej osi (najczęściej OY) odkładamy $x_{(i)}$.
- Na drugiej osi (najczęściej OX) odkładamy kwantyle rozkładu normalnego $N(0,1)$ na poziomach:
 - a) k/n dla $k=1, \dots, n$
 - lub b) $k/(n+1)$ dla $k=1, \dots, n$
 - lub c) $(0.5)^{(1/n)}$ dla $k=n$
 $(k-0.3175)/(n+0.365)$ dla $k=2, \dots, n-1$
 $1-(0.5)^{(1/n)}$ dla $k=1$
 - lub d) $(k-a)/(n+1-2*a)$ dla $k=1, \dots, n, a=3/8$ w języku R
 (najczęściej b) lub c) – różnice praktycznie nieistotne dla dużych n

RPiS 2024/2025 13

13

Wykres normalny (normal probability plot)

Oś Y: próbka pochodząca z rozkładu:

N(0,1)

Weibulla (a=1.1, b=4)

jednorodnego (-0.5, 0.5)

Kwantyle rozkładu normalnego N(0,1) na poziomie wg. podpunktu c)

- Można wyznaczyć współczynnik korelacji i porównać z wartościami krytycznymi odpowiedniego testu statystycznego – na tej podstawie przyjmujemy lub odrzucamy hipotezę mówiącą, że próbka pochodzi z rozkładu normalnego.
- Analogicznie można badać pochodzenie próbki z innych rozkładów.

RPiS 2024/2025 14

14

Testy statystyczne rodzaju rozkładu

Test Shapiro-Wilka (n≤50)

- H_0 – próbka pochodzi z rozkładu normalnego.
- Statystyką testową (dla uporządkowanej próbki) jest

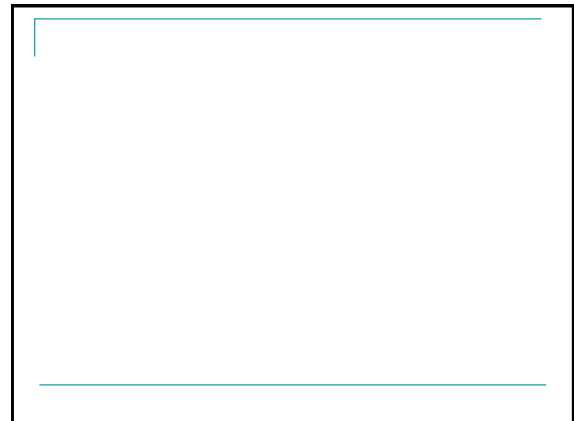
$$W = \frac{\left(\sum_{i=1}^{\text{int}(n/2)} a_i(n)(x_{n-i+1} - x_i) \right)^2}{\sum_{j=1}^n (x_j - \bar{X})^2}$$

Różnice $x_{n-i+1} - x_i$ nazywamy quasi-rozstępami rzędu i.

- Gdzie $a_i(n)$ to znane (stabilcowane) stałe współczynniki zależne od i i n.
- Obszar krytyczny:
 $W \in (-\infty, W_{\alpha/2}(n)) \cup (W_{1-\alpha/2}(n), +\infty)$
 gdzie $W_\beta(n)$ to kwantyle rzędu β rozkładu W (dostępne w tablicach).
- Przykład aplet: test Shapiro-Wilka dla próbek z różnych rozkładów.

RPiS 2024/2025 15

15



16

Generatory rozkładu jednorodnego

- Rozkład jednorodny** na przedziale (a,b)
 $E(X)=(b+a)/2$ $var(X)=(b-a)^2/12$
- Realizacja numeryczna:
- Najbezpieczniej używać sprawdzonych procedur (np. biblioteki CERN)
- 1) Metoda kongruencyjna – generator liniowy

$$x_{n+1} = (a_0 x_n + \underbrace{a_1 x_{n-1} + \dots + a_k x_{n-k}}_{k+1 \text{ poprzednich}} + a_{k+1}) \pmod{M} \quad a_i, x_i \in [0, M)$$

K=0 – generatory multiplikatywne $x_{n+1} = (a_0 x_n) \pmod{M}$
 K=1 – generatory Fibonacciego $x_{n+1} = (a_0 x_n + a_1 x_{n-1}) \pmod{M}$

Generatory mieszane $x_{n+1} = (a_0 x_n + a_1) \pmod{M}$
 (W języku C funkcja srand daje $a_1 =$ liczba sekund od 1.1.1970)
 Np. $a_0=69069$ $a_1=1$ $M=2^{32}$ (Marsaglia 1972)

RPiS 2024/2025 17

17

Generatory rozkładu jednorodnego

- Przykład - generator mieszany:
 $x_n = (5x_{n-1} + 1) \pmod{16}$

Okres = 16
 $x_i = x_{i+16}$

- Dane wejściowe: $x_0=5$:
- Kolejne otrzymywane liczby:
 10, 3, 0, 1, 6, 15, 12, 13, 2, 11, 8, 9, 14, 7, 4, 5, 10, 3, 0, 1, 6, 15, 12, 13, 2, 11, 8, 9, 14, 7, 4, 5.
- Po podzieleniu przez 16:
- 0.6250, 0.1875, 0.0000, 0.0625, 0.3750, 0.9375, 0.7500, 0.8125, 0.1250, 0.6875, 0.5000, 0.5625, 0.8750, 0.4375, 0.2500, 0.3125, 0.6250, 0.1875, 0.0000, 0.0625, 0.3750, 0.9375, 0.7500, 0.8125, 0.1250, 0.6875, 0.5000, 0.5625, 0.8750, 0.4375, 0.2500, 0.3125.

RPiS 2024/2025 18

18

Generatory rozkładu jednorodnego

- 2) Uogólnione generatory Fibonacciego (przypominają ciąg Fibonacciego $f_0=f_1=1$ $f_n=f_{n-2}+f_{n-1}$ dla $n>1$)

$$x_n = (a_0x_{n-2} + a_1x_{n-1}) \pmod M \rightarrow$$

$$x_n = (a_0x_{n-r} + a_1x_{n-k}) \pmod M \rightarrow$$

$$x_n = (a_0x_{n-r} \oplus a_1x_{n-k}) \pmod M$$
 gdzie $\oplus \equiv +, -, \cdot, xor, \dots$

Inny zapis: $F(r, s, \oplus)$

Np.
 $F(17, 5, +)$ dla $M=2^{32}$ ma okres $(2^{17}-1)2^{31}$
 $F(17, 5, \cdot)$ dla $M=2^{32}$ ma okres $(2^{17}-1)2^{29}$

RPIS 2024/2025 19

19

Generatory rozkładu jednorodnego

- 3) Generatory oparte na mnożeniu z przeniesieniem
 Inicjujemy podając wartości $x_1, x_2, \dots, x_k \in (0, M)$ i $c=0$ c-parametr przeniesienia.

$$x_n = (c + a_1x_{n-1} + \dots + a_kx_{n-k}) \pmod M \quad a_i, x_i \in [0, M)$$
 gdzie w kolejnych iteracjach
 $c \rightarrow \text{Int}[(c + a_1x_{n-1} + a_2x_{n-2} + \dots + a_kx_{n-k}) / M]$
 Np.

$$x_n = (c + 1941x_{n-1} + 1860x_{n-2} + 1812x_{n-3} + 1776x_{n-4} + 1492x_{n-5} + 1215x_{n-6} + 1066x_{n-7} + 12013x_{n-8}) \pmod{2^{16}}$$
- 4) Generatory nieliniowe np. $x_n = (a_i(x_{n-1})^{-1} + c) \pmod M$

$$x_n = (a(n+b+c)^{-1}) \pmod M$$

$$x_n = (x_{n-1}^{-1})^2 \pmod M$$

x_n nie zależy od poprzednich x_k → dobry do programowania równoległego

RPIS 2024/2025 20

20

Generatory rozkładu jednorodnego

- 5) Generatory oparte na rejestrach przesuwnych

$$x_n = (a_1x_{n-1} + \dots + a_lx_{n-l}) \pmod 2 \quad a_i, x_i \in \{0,1\}$$

Ponieważ $(a+b) \pmod 2 = a \text{ xor } b$ to

$$x_n = x_{n-j_1} \text{ xor } x_{n-j_2} \text{ xor } \dots \text{ xor } x_{n-j_l}$$

a	b	a xor b
0	0	0
1	0	1
0	1	1
1	1	0

$\begin{cases} a_{jm} = 1 \text{ dla } m \leq l \\ a_{jm} = 0 \text{ dla } m > l \end{cases}$

Tworzymy ciąg bitów x_n i z niego liczby pseudolosowe
 $u_i = 0.x_{i+1}x_{i+2} \dots x_{i+L}$ dla $s \leq L$ oraz $i = 0, 1, \dots$
 gdzie s i L to wybrane parametry.
 Np. $L=8, s=3$

$$u_0 = 0.x_1x_2 \dots x_8$$

$$u_1 = 0.x_{3+1}x_{3+2} \dots x_{3+8} = 0.x_4x_5 \dots x_{11}$$

$$u_2 = 0.x_{6+1}x_{6+2} \dots x_{6+8} = 0.x_7x_8 \dots x_{14}$$

Konstrukcja ta zwana jest również schematem Tauswortha.

RPIS 2024/2025 21

21

Generatory rozkładu jednorodnego

- Przykład: generator MZT (Marsaglia, Zaman, Tsang 1990), okres 2^{144} .
 Składa się z dwóch generatorów:
 1) $F(97, 33, \oplus)$ o okresie 2^{120} generuje liczby $V_n \in [0,1)$

$$V_n = V_{n-97} \oplus V_{n-33} \quad x \oplus y \equiv \begin{cases} x - y & \text{dla } x \geq y \\ x - y + 1 & \text{dla } x < y \end{cases}$$

Inicjalizacja polega na nadaniu wartości liczbom V_1, \dots, V_{97} . Robimy to generując ich bity b_n generatorami liczb całkowitych

$$y_n = (y_{n-3} \cdot y_{n-2} \cdot y_{n-1}) \pmod{179}$$

$$z_n = (52z_{n-1} + 1) \pmod{169}$$

$$b_n = \begin{cases} 0 & \text{dla } (y_n \cdot z_n) \pmod{64} < 32 \\ 1 & \text{dla } (y_n \cdot z_n) \pmod{64} \geq 32 \end{cases}$$

Czyli użytkownik podaje $y_1, y_2, y_3 \in \{1, 2, \dots, 178\}$, $z_1 \in \{1, 2, \dots, 168\}$

RPIS 2024/2025 22

22

Generatory rozkładu jednorodnego

- 2) Generator liniowy o okresie $2^{24}-3$ generuje liczby $c_n \in (0,1)$

$$C_n = C_{n-1} \oplus a, \quad n \geq 2$$
 gdzie $a = 7654321/16777216$, $C_1 = 362436/16777216$,

$$x \oplus y \equiv \begin{cases} x - y & \text{dla } x \geq y \\ x - y + 16777213/16777216 & \text{dla } x < y \end{cases}$$

Ostatecznie generator MZT zwraca liczbę

$$U_n = V_n \oplus C_n$$

gdzie ponownie

$$x \oplus y \equiv \begin{cases} x - y & \text{dla } x \geq y \\ x - y + 1 & \text{dla } x < y \end{cases}$$

Powyższe przykłady i więcej informacji: R.Wieczorkowski, R.Zieliński „Komputerowe generatory liczb losowych”, WN-T 1997

RPIS 2024/2025 23

23

Generatory – inne metody

- Metoda superpozycji**
 Stosujemy tam, gdzie trudno odwrócić dystrybuantę, ale można ją podzielić na części, które łatwo odwracać.

$$F_x(x) = \sum_k c_k F_k(x) \quad 0 < c_k < 1 \quad \sum_k c_k = 1$$
 Przykład:

$$f_x(x) = \frac{5}{12}(1 + (x-1)^4) \quad x \in (0,2)$$
- Metoda eliminacji**: stosujemy gdy $f(x)$ jest niezerowa tylko w przedziale (a,b) i ograniczone (przez stałą $<c$). Metoda postępowania: losujemy punkt (x,y) w prostokącie o bokach (a,b) i $(0,c)$, jeżeli leży on pod $f(x)$ to akceptujemy x , w przeciwnym wypadku losujemy kolejny punkt.
- Metody z przekształceniem** – szukamy innej zmiennej losowej o tej samej dystrybuancie (funkcji gęstości prawdopodobieństwa).
 Przykład: $f_x(x) = nx^{n-1} \quad x \in (0,1)$
 generujemy jako $x = \max(\gamma_1, \gamma_2, \dots, \gamma_n) \quad \gamma_i \in (0,1)$

RPIS 2024/2025 24

24

Metoda superpozycji - przykład

Niech $f_x(x) = \frac{5}{12}(1+(x-1)^4)$ $x \in (0,2)$
 wtedy $F_x(x') = \frac{5}{12} \int_0^{x'} (1+(x-1)^4) dx = \frac{5}{12} (x + \frac{(x-1)^5}{5}) \Big|_0^{x'} = \frac{1}{12} [5x' + (x'-1)^5 + 1]$
 (odwracanie uciążliwe bo wymaga wyliczenia y z $\frac{1}{12} [5y + (y-1)^5 + 1] = x$)

Ale:
 $f_x(x) = \frac{5}{12}(1+(x-1)^4) = \frac{5}{6} \cdot \frac{1}{2} + \frac{1}{6} \cdot \frac{5}{2}(x-1)^4$
 $F_x(x) = \sum_k c_k F_k^{(k)}(x)$
 $F_x(x) = \frac{5}{6} \cdot \frac{1}{2} x + \frac{1}{6} \cdot \frac{5}{2} [(x-1)^5 + 1]$
 $0 < c_k < 1 \quad \sum_k c_k = 1$

Na tym opieramy algorytm:
 1. Losujemy γ_1 z rozkładu jednorodnego $[0,1]$ – będzie decydować o wyborze c_i
 2. Losujemy γ_2 z rozkładu jednorodnego $[0,1]$ – wartość dystrybuantry $F_x(x')$
 $\gamma_1 < \frac{5}{6} \rightarrow x = \frac{1}{2} y \rightarrow y = 2x \rightarrow y = 2\gamma_2$
 $\gamma_1 \geq \frac{5}{6} \rightarrow x = \frac{1}{2} [(y-1)^5 + 1] \rightarrow y = 1 + \sqrt[5]{2x-1} \rightarrow y = 1 + \sqrt[5]{2\gamma_2-1}$
3. Ostatecznie zwracamy to lub to

RPIS 2024/2025 25

25

Metoda z przekształceniem - przykład

Niech $f_x(x) = nx^{n-1}$ $x \in (0,1)$
 Wtedy $F_x(x) = \int_0^x n(x')^{n-1} dx' = x^n$
 $\theta(z) = \begin{cases} 1 & \text{dla } z \geq 0 \\ 0 & \text{dla } z < 0 \end{cases}$
 $\rightarrow \theta(x-x') = 1$
 $\rightarrow x-x' \geq 0$
 $\rightarrow x \geq x'$

Pokażemy, że $x = g(\gamma_1, \gamma_2, \dots, \gamma_n)$ gdzie γ_i mają niezależne rozkłady jednorodne na $(0,1)$ dla pewnej funkcji g mają ten sam rozkład prawdopodobieństwa co wyjściowy X .

$$F_x(x) = \int_{\Omega: x' \leq x} f_x(x') dx' = \int_{x=g(\gamma_1, \dots, \gamma_n) \leq x} f_1(\gamma_1, \dots, \gamma_n) d\gamma_1 d\gamma_2 \dots d\gamma_n = \dots$$

Z niezależności $f_{\Gamma}(\gamma_1, \dots, \gamma_n) = f_1(\gamma_1) \cdot f_2(\gamma_2) \dots f_n(\gamma_n) = 1^n = 1$
 wtedy $\dots = \int_0^1 \dots \int_0^1 1 \cdot \theta(x - g(\gamma_1, \gamma_2, \dots, \gamma_n)) d\gamma_1 d\gamma_2 \dots d\gamma_n = \dots$

Wybieramy $g(\gamma_1, \gamma_2, \dots, \gamma_n) \equiv \max(\gamma_1, \gamma_2, \dots, \gamma_n)$
 co prowadzi do $\dots = \int_0^1 \dots \int_0^1 \theta(x - \max(\gamma_1, \gamma_2, \dots, \gamma_n)) d\gamma_1 d\gamma_2 \dots d\gamma_n = \dots$

RPIS 2024/2025 26

26

Metoda z przekształceniem - przykład

Wyrażenie pod całką
 $\int_0^1 \dots \int_0^1 \theta(x - \max(\gamma_1, \gamma_2, \dots, \gamma_n)) d\gamma_1 d\gamma_2 \dots d\gamma_n = \dots$
 jest niezerowe gdy $\theta(x - \max(\gamma_1, \gamma_2, \dots, \gamma_n)) \neq 0$
 $\rightarrow x \geq \max(\gamma_1, \gamma_2, \dots, \gamma_n)$
 $\rightarrow x \geq \gamma_1 \quad i \quad x \geq \gamma_2 \quad i \dots i \quad x \geq \gamma_n$
 $\rightarrow \gamma_1 \leq x \quad i \quad \gamma_2 \leq x \quad i \dots i \quad \gamma_n \leq x$

Zatem
 $\dots = \int_0^x \dots \int_0^x d\gamma_1 d\gamma_2 \dots d\gamma_n = x^n$

Czyli zmienna $x = \max(\gamma_1, \gamma_2, \dots, \gamma_n)$ ma rozkład tak jak zmienna dana interesującą nas funkcją gęstości prawdopodobieństwa $f_x(x) = nx^{n-1}$
 \rightarrow zmienną X możemy generować poprzez szukanie maksimum z $(\gamma_1, \gamma_2, \dots, \gamma_n)$.

RPIS 2024/2025 27

27