ZADANIE 1.

WARTOŚĆ KRYTYCZNA – wartość statystyki testowej, przy której zmienia się decyzja w teście.

Na liście 5 rozważamy testy prawostronne.

$$\varphi(\mathbb{X}) = \begin{cases} 1, & \text{gdy } S(\mathbb{X}) \geq c, \\ 0, & \text{gdy } S(\mathbb{X}) < c. \end{cases}$$

odrzucenie hipotezy zerowej

brak podstaw do odrzucenia hipotezy zerowej

gdzie $\varphi(\mathbb{X})$ – test, funkcja próby,

$S(\mathbb{X})$ – statystyka testowa,

c – wartość krytyczna

Jak wyznaczyć wartości krytyczne?

Chcemy przeprowadzić test na poziomie istotności α, czyli taki dla którego prawdopodobieństwo popełnienia błędu I-go rodzaju (odrzucenie prawdziwej H_0) wynosi α.

$$\alpha = P_{H_0}(\varphi(X)=1) = P_{H_0}(S(X) \geq c) = 1 - P_{H_0}(S(X) \leq c)$$

↑ p-stwo, gdy H_0 prawdziwe

$$\underbrace{P_{H_0}(S(X) \leq c)}_{\text{dystrybuanta rozkładu } S(X) \text{ przy } H_0} = 1-\alpha \quad \Big/ \text{ nakładam funkcję odwrotną do dystrybuanty (kwantyl)}$$

$$\boxed{c = F^{-1}(1-\alpha)}$$

kwantyl rzędu $1-\alpha$ rozkładu $S(X)$ przy H_0

Kwantyl w pakiecie to linia kodu, ale UWAGA!
– nie znamy rozkładu statystyki testowej przy H_0
(w artykułach możemy znaleźć rozkład asymptotyczny)

Wyznaczamy symulacyjnie wartość krytyczną:

1. Wygeneruj 10000 realizacji statystyki testowej w taki sposób, aby H_0 było prawdziwe. (U nas rozkład X i Y jest taki sam)
2. Policz kwantyl z próby z punktu 1

Czy taki sposób generowania wartości krytycznych jest poprawny?

W zadaniu generujemy z rozkładu normalnego $\mathcal{N}(0,1)$, a H_0 z wstępu do listy (wzór 1) mówi, że mamy generować z dowolnego rozkładu ciągłego.
Czy możemy zawęzić problem? Czy wartość krytyczna wyliczona przy $\mathcal{N}(0,1)$ będzie prawidłową wartością krytyczną dla dowolnego rozkładu ciągłego? Jak wyznaczyć rozkład statystyki $S(X)$ przy dowolnym $F=G$?

Wiemy, że F - dystrybuanta zmiennej losowej X_i będąca jednocześnie dystrybuantą zmiennej Y_i, bo szukamy rozkładu statystyki testowej przy prawdziwej H_0, czyli gdy X i Y mają ten sam rozkład (wzór 1)

Jaki rozkład ma $F(X)$? $F(x) = P(X \leq x) \in \langle 0, 1 \rangle$ więc $F(X) \in \langle 0, 1 \rangle$

dystrybuanta rozkładu $F(X)$:

$$P(F(X) \leq x) = P(X \leq F^{-1}(x)) = F(F^{-1}(x)) = x$$

funkcja F jest rosnąca, bo rozkład zmiennej X jest ciągły

Gęstość rozkładu $F(X)$

$$(P(F(X) \leq x))' = x' = 1$$ - gęstość (pochodna dystrybuanty) rozkładu jest stała, a jedynym rozkładem o stałej gęstości jest rozkład jednostajny => stąd $F(X) \sim U(0,1)$

Wiemy, że F - ciągła (lista) $F(X) \sim U(0,1)$ (poprzedni slajd)

Skoro $F(X) \sim U(0,1)$ oraz $F=G$ to $F(Y) \sim U(0,1)$

Niech Φ - dystrybuanta $N(0,1)$. Wtedy

$\Phi^{-1}(F(X)) \sim N(0,1)$ $\quad \left(P(\Phi^{-1}(F(X)) \le x) = P(F(X) \le \Phi(x)) = \Phi(x) \right)$

$\Phi^{-1}(F(Y)) \sim N(0,1)$

$F(X) \sim U(0,1)$

$P(F(X) \le x) = x$

$\Phi^{-1}(F(x))$ jest rosnąca

Rosnące przekształcenie zmiennych nie zmienia ich rang ani nie wpływa na supremum (wzory 2 i 3), więc rozkład S gdy losujemy z dwóch $N(0,1)$ i gdy losujemy z dwóch dowolnych F jest taki sam.

Nie zmienia się wartość statystyki testowej $\Rightarrow$ nie zmienia się rozkład.

Przykład 1.

$\vec{X} = (1, 7, 3, 10)$

rangi $= (1, 3, 2, 4)$

$2\vec{X} + 7 = (9, 21, 13, 27)$

rangi $= (1, 3, 2, 4)$

Nic się nie zmieniło dla przekształcenia rosnącego.

$-\vec{X} = (-1, -7, -3, -10)$

rangi $= (4, 2, 3, 1)$

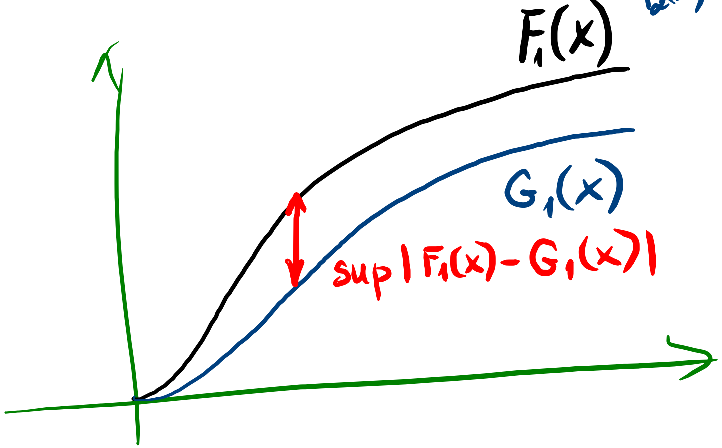
Dla malejącego już tak, dlatego ważne było aby przekształcenie było rosnące

Przykład 2

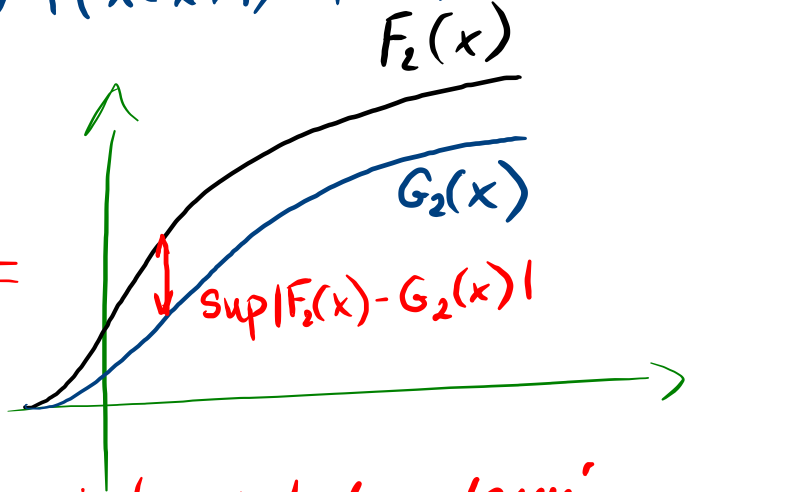
X ma dystrybuantę F_1 $\qquad$ $X-1$ ma dystrybuantę F_2

Y ma dystrybuantę G_1 $\qquad$ $Y-1$ ma dystrybuantę G_2

$$F_2(x)=P(X-1\le x)=P(X\le x+1)=F_1(x+1)$$



$=$



X-Y są różne, ale odległość pomiędzy dystrybuantami pozostaje taka sama.

(Prawda dla każdego rosnącego przekształcenia)

Podsumowując – możemy posłużyć się rozkładem $N(0,1)$, ale moglibyśmy też wybrać dowolny inny ciągły rozkład z dowolnymi parametrami.