

Statystyka i rachunek prawdopodobieństwa

Wykład 13: Hipotezy statystyczne—testy nieparametryczne

Mariusz Tarnopolski

Instytut Astronomii UMK

Statystyka ©2026



Służą do weryfikowania hipotezy zerowej, że rozpatrywana zmienna losowa ma rozkład należący do określonej rodziny rozkładów. Np. sprawdzamy czy rozkład obserwowanej/mierzonej wielkości jest normalny.

Służą do weryfikowania hipotezy zerowej, że rozpatrywana zmienna losowa ma rozkład należący do określonej rodziny rozkładów. Np. sprawdzamy czy rozkład obserwowanej/mierzonej wielkości jest normalny.

Takie testy w ogólności można podzielić na dwie główne kategorie:

- 1 gdy wielkości mamy pogrupowane w kategorie—tzw. szereg rozdzielczy (np. test χ^2 Pearsona);
- 2 gdy wielkość jest ciągła i mamy dostęp do *surowych* wyników pomiarów (np. testy Kołmogorowa-Smirnowa lub Andersona-Darlinga).

Służą do weryfikowania hipotezy zerowej, że rozpatrywana zmienna losowa ma rozkład należący do określonej rodziny rozkładów. Np. sprawdzamy czy rozkład obserwowanej/mierzonej wielkości jest normalny.

Takie testy w ogólności można podzielić na dwie główne kategorie:

- 1 gdy wielkości mamy pogrupowane w kategorie—tzw. szereg rozdzielczy (np. test χ^2 Pearsona);
- 2 gdy wielkość jest ciągła i mamy dostęp do *surowych* wyników pomiarów (np. testy Kołmogorowa-Smirnowa lub Andersona-Darlinga).

Badamy hipotezę $H_0 : F(x) = F_0(x)$ —czy dystrybuanta empiryczna F jest zgodna z dystrybuantą rozkładu teoretycznego F_0 .

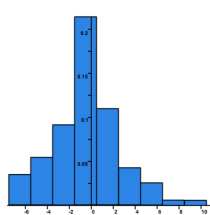
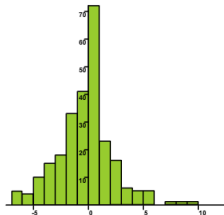
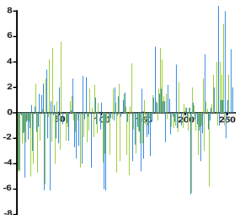
Hipoteza prosta—jeśli parametry są znane z góry; hipoteza złożona—jeśli parametry estymujemy z próby.

Test χ^2 Pearsona

Założenia

- dane są pogrupowane w k klas
- licznosc każdej klasy to n_i
- całkowita licznosc próby wynosi $n = \sum_i^k n_i$
- prawdopodobieństwa teoretyczne zaobserwowania wielkości w danej klasie to p_i
- dla i -tej klasy (tj. dla przedziału $(x_{i-1}, x_i]$):

$$p_i = P(x_{i-1} < X \leq x_i) = \int_{x_{i-1}}^{x_i} dF_0(x) = F_0(x_i) - F_0(x_{i-1})$$



Statystyka testowa to w ogólności empiryczne χ^2 :

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

gdzie O —obserwowane (*observed*), E —spodziewane (*expected*).

Statystyka testowa to w ogólności empiryczne χ^2 :

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

gdzie O —obserwowane (*observed*), E —spodziewane (*expected*).
Dla szeregu rozdzielczego:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$$

Jest to miara rozbieżności danych i modelu. Zachodzi $\chi^2 \geq 0$.

Test χ^2 Pearsona

Rozkład i obszar krytyczny

Asymptotycznie, gdy $n \rightarrow \infty$, TS ma rozkład χ^2 o $(k - p - 1)$ stopniach swobody:

$$\chi^2 \sim \chi^2(k - p - 1)$$

gdzie p jest liczbą wolnych parametrów rozkładu (np. dla rozkładu normalnego $p = 2$), które należy oszacować z próby aby obliczyć prawdopodobieństwa p_i .

Test χ^2 Pearsona

Rozkład i obszar krytyczny

Asymptotycznie, gdy $n \rightarrow \infty$, TS ma rozkład χ^2 o $(k - p - 1)$ stopniach swobody:

$$\chi^2 \sim \chi^2(k - p - 1)$$

gdzie p jest liczbą wolnych parametrów rozkładu (np. dla rozkładu normalnego $p = 2$), które należy oszacować z próby aby obliczyć prawdopodobieństwa p_i .

Prawostronny obszar krytyczny K_α określony jest przez warunek

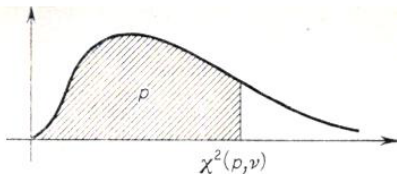
$$\chi^2 > \chi_{1-\alpha, k-p-1}^2$$

ponieważ chcemy zminimalizować błąd II rodzaju odrzucenia hipotezy alternatywnej H_1 , za którą przemawiają duże wartości statystyki χ^2 .

Test χ^2 Pearsona

Rozkład i obszar krytyczny

Tablica 8. Kwantyle $\chi^2(p, \nu)$ rzędu p rozkładu χ^2 o ν stopniach swobody



ν	p							
	0,005	0,01	0,025	0,05	0,95	0,975	0,99	0,995
1	—	—	0,001	0,004	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	11,071	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	18,307	20,483	23,209	25,188

Test χ^2 Pearsona

Uwagi

- **wymaganie praktyczne:** $n_i \gtrsim 5$
- **wymaganie praktyczne:** co najmniej $k = 20$ klas
- **najlepiej działa gdy** estymujemy nie więcej niż $p = 2$ parametry
- często konieczne łączenie klas
- wynik zależy od podziału na klasy

- **wymaganie praktyczne:** $n_i \gtrsim 5$
- **wymaganie praktyczne:** co najmniej $k = 20$ klas
- **najlepiej działa gdy** estymujemy nie więcej niż $p = 2$ parametry
- często konieczne łączenie klas
- wynik zależy od podziału na klasy

Są to zatem dość istotne ograniczenia. Ponadto wadą tej metody jest „marnowanie danych” w przypadku gdy mamy dostęp do surowych (nie podzielonych na klasy) danych; zwłaszcza jeśli mają one naturę ciągłą.

Test χ^2 Pearsona

Uwagi

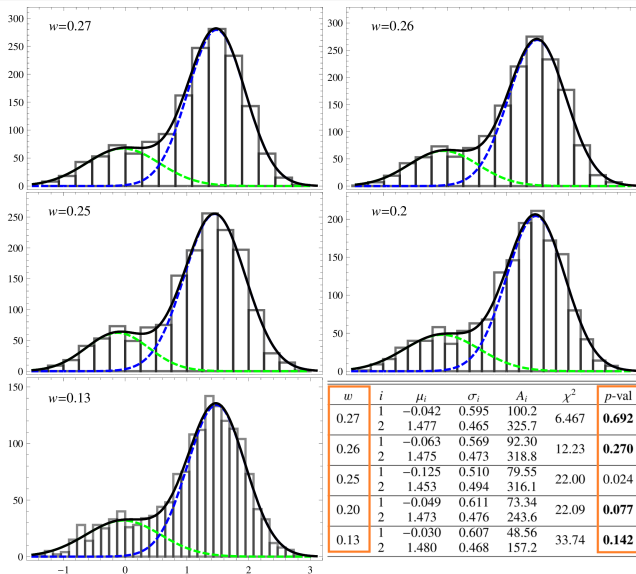
- **wymaganie praktyczne:** $n_i \gtrsim 5$
- **wymaganie praktyczne:** co najmniej $k = 20$ klas
- **najlepiej działa gdy** estymujemy nie więcej niż $p = 2$ parametry
- często konieczne łączenie klas
- wynik zależy od podziału na klasy

Są to zatem dość istotne ograniczenia. Ponadto wadą tej metody jest „marnowanie danych” w przypadku gdy mamy dostęp do surowych (nie podzielonych na klasy) danych; zwłaszcza jeśli mają one naturę ciągłą.

Oczywiście, jeśli zmienna losowa jest dyskretna (o rozkładzie dwumianowym, Poissona, itd.), to test χ^2 jest adekwatny—przy spełnieniu powyższych wymagań.

Test χ^2 Pearsona

Przykład—biny o szerokości w



Test Kołmogorowa-Smirnowa

Założenia i statystyka testowa

- zmienna losowa ciągła
- brak grupowania danych

Hipoteza zerowa:

$$H_0 : F_n(x) = F_0(x)$$

gdzie F_n to dystrybuanta empiryczna z próby o licznosci n . Wprowadźmy uporządkowanie:

$$x_1 \leq x_2 \leq \dots \leq x_n$$

Wtedy

$$F_n(x) = \begin{cases} 0 & x < x_1 \\ \frac{i}{n} & x_i \leq x < x_{i+1} \\ 1 & x \geq x_n \end{cases}$$

Test Kołmogorowa-Smirnowa

Założenia i statystyka testowa

- zmienna losowa ciągła
- brak grupowania danych

Hipoteza zerowa:

$$H_0 : F_n(x) = F_0(x)$$

gdzie F_n to dystrybuanta empiryczna z próby o licznosci n . Wprowadźmy uporządkowanie:

$$x_1 \leq x_2 \leq \dots \leq x_n$$

Wtedy

$$F_n(x) = \begin{cases} 0 & x < x_1 \\ \frac{i}{n} & x_i \leq x < x_{i+1} \\ 1 & x \geq x_n \end{cases}$$

Niech będzie hipoteza prosta H_0 , że zmienna losowa ma dystrybuantę $F_0(x)$ całkowicie określoną. Wtedy statystyka testowa KS to

$$D_n = \sup_x |F_0(x) - F_n(x)|$$

Test Kołmogorowa-Smirnowa

Interpretacja statystyki KS

Rozkład D_n jest stabilizowany.

Duże wartości D_n świadczą o dużej rozbieżności dystrybuanty F_n od F_0 .

Konstruujemy prawostronny obszar krytyczny w oparciu o kwantyl d_n rzędu $1 - \alpha$ (na poziomie istotności α):

$$P(D_n \geq d_{n,1-\alpha}) = \alpha$$

Test Kołmogorowa-Smirnowa

Interpretacja statystyki KS

Rozkład D_n jest stabilizowany.

Duże wartości D_n świadczą o dużej rozbieżności dystrybuanty F_n od F_0 .

Konstruujemy prawostronny obszar krytyczny w oparciu o kwantyl d_n rzędu $1 - \alpha$ (na poziomie istotności α):

$$P(D_n \geq d_{n,1-\alpha}) = \alpha$$

W praktyce:

- obliczamy różnice $\frac{i}{n} - F_0(x_i)$ dla $i = 1, \dots, n$
- oznaczamy $d_n^+ = \max_{1 \leq i \leq n} \left(\frac{i}{n} - F_0(x_i) \right)$
- podobnie $d_n^- = \max_{1 \leq i \leq n} \left(\frac{i-1}{n} - F_0(x_i) \right)$
- $d_n = \max(|d_n^+|, |d_n^-|)$

Jeśli d_n wpada w obszar krytyczny to H_0 odrzucamy na poziomie α .

Test Kołmogorowa-Smirnowa

Duże próby

- Dla stosunkowo dużych prób ($n \gtrsim 40$) można stosować przybliżenie na wartość krytyczną statystyki KS:

$$D_{\text{crit}} = \sqrt{\frac{-\ln\left(\frac{\alpha}{2}\right)}{2n}}$$

Jeśli wartość statystyki z próby D_n spełnia

$$D_n < D_{\text{crit}}$$

to nie ma podstaw do odrzucenia hipotezy H_0 na poziomie istotności α .

Test Kołmogorowa-Smirnowa

Duże próby

- Jeśli próba jest bardzo duża ($n \gg 100$) to:
 - 1 dopuszczalna jest również hipoteza złożona¹—parametry rozkładu można estymować z próby
 - 2 d_n wyznacza się j.w. Graniczny rozkład Kołmogorowa λ (stabilizowany) zmiennej losowej $\sqrt{n}D_n$ nie zależy od n . Obszar krytyczny spełnia

$$P(\sqrt{n}D_n \geq \lambda(1 - \alpha)) = \alpha$$

Zatem jeśli wartość $\sqrt{n}d_n$ z próby jest większa od kwantyla $\lambda(1 - \alpha)$, to hipotezę H_0 odrzucamy na poziomie istotności α .

¹W zasadzie stosuje się test Lilieforsa, który jest modyfikacją testu KS dla tej sytuacji. Tablice statystyczne liczone wyłącznie metodami Monte Carlo.

Test Kołmogorowa-Smirnowa

Duże próby

- Jeśli próba jest bardzo duża ($n \gg 100$) to:
 - 1 dopuszczalna jest również hipoteza złożona¹—parametry rozkładu można estymować z próby
 - 2 d_n wyznacza się j.w. Graniczny rozkład Kołmogorowa λ (stabilizowany) zmiennej losowej $\sqrt{n}D_n$ nie zależy od n . Obszar krytyczny spełnia

$$P(\sqrt{n}D_n \geq \lambda(1 - \alpha)) = \alpha$$

Zatem jeśli wartość $\sqrt{n}d_n$ z próby jest większa od kwantyla $\lambda(1 - \alpha)$, to hipotezę H_0 odrzucamy na poziomie istotności α .

W tym teście dane można pogrupować w klasy i różnice dystrybuant wyznaczać w poszczególnych klasach. Test KS jest czulszy niż test χ^2 Pearsona.

¹W zasadzie stosuje się test Lilieforsa, który jest modyfikacją testu KS dla tej sytuacji. Tablice statystyczne liczone wyłącznie metodami Monte Carlo.

Test Kołmogorowa-Smirnowa

Tabela 12. Kwantyle $d_n(1-\alpha)$ statystyki D_n Kołmogorowa

n	α			n	α		
	0,10	0,05	0,01		0,10	0,05	0,01
1	0,950	0,975	0,995	51	0,168	0,187	0,224
2	,776	,842	,929	52	,166	,185	,222
3	,636	,708	,829	53	,165	,183	,220
4	,565	,624	,734	54	,163	,181	,218
5	,509	,563	,669	55	,162	,180	,216
6	,468	,519	,617	56	,160	,178	,214
7	,436	,483	,576	57	,159	,177	,212
8	,410	,454	,542	58	,158	,175	,210
9	,387	,430	,513	59	,156	,174	,208
10	,369	,409	,489	60	,155	,172	,207

Tabela 13. Wartości $K(y)$ dystrybuanty K statystyki $\sqrt{n} D_n$ Kołmogorowa przy $n \rightarrow \infty$

y	$K(y)$	y	$K(y)$	y	$K(y)$	y	$K(y)$
0,36	0,001	0,72	,322	1,08	0,806	1,44	0,968
0,37	,001	0,73	,339	1,09	,814	1,45	,970
0,38	,001	0,74	,356	1,10	,822	1,46	,972
0,39	,002	0,75	,373	1,11	,830	1,47	,973
0,40	,003	0,76	,390	1,12	,837	1,48	,975
0,41	,004	0,77	,406	1,13	,845	1,49	,976
0,42	,005	0,78	,423	1,14	,851	1,50	,978
0,43	,007	0,79	,440	1,15	,858	1,51	,979
0,44	,010	0,80	,456	1,16	,864	1,52	,980
0,45	,013	0,81	,472	1,17	,871	1,53	,981

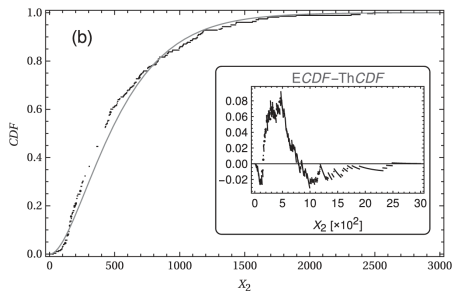
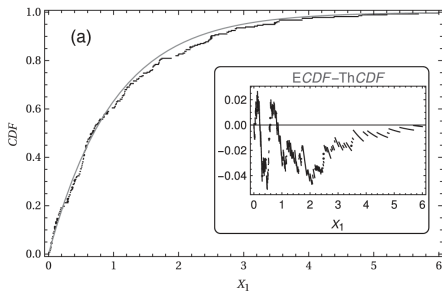
Caveat emptor!

Test KS jest najczulszy na różnice w okolicach mody rozkładu. Może się zdarzyć, że dana próba „zda” ten test mimo nieakceptowalnie dużych różnic w ogonach rozkładu.

Dlatego należy wykonać co najmniej dwa komplementarne testy zgodności (zob. dalej).

Test Kołmogorowa-Smirnowa

Przykład



Sample	n	D_{crit}	D
(a)	278	0.0815	0.0520
(b)		0.0815	0.0915

Test różnicowy Kołmogorowa-Smirnowa

Test KS można stosować do rozstrzygnięcia, czy dwie próby (o licznosciach n_1 i n_2) pochodzą z tego samego rozkładu— $H_0 : F_{n_1}(x) = F_{n_2}(x)$.

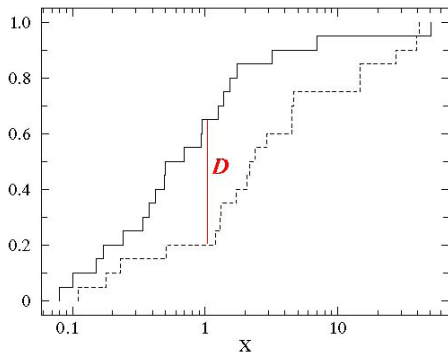
Statystyka testowa to

$$D_{n_1, n_2} = \sup_x |F_{n_1}(x) - F_{n_2}(x)|$$

Zbyt duże wartości świadczą na niekorzyść H_0 . Dla prób $n_i \gtrsim 20$ korzysta się z tego, że statystyka

$$\lambda = \sqrt{n} D_{n_1, n_2}$$

gdzie $n = \frac{n_1 n_2}{n_1 + n_2}$, ma rozkład graniczny Kołmogorowa.



Test Andersona-Darlinga

Założenia i statystyka testowa

Podobnie jak w teście KS, porównujemy dystrybuantę empiryczną F_n z teoretyczną F_0 .

Kluczowa różnica: test AD **silniej uwzględnia różnice w ogonach rozkładu**.

Test Andersona-Darlinga

Założenia i statystyka testowa

Podobnie jak w teście KS, porównujemy dystrybuantę empiryczną F_n z teoretyczną F_0 .

Kluczowa różnica: test AD **silniej uwzględnia różnice w ogonach rozkładu**.

Statystyka testowa:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) [\ln F_0(x_i) + \ln (1 - F_0(x_{n+1-i}))]$$

również mierzy rozbieżność między dystrybuantami F_n i F_0 .

Duże wartości A^2 świadczą o niezgodności z H_0 .

Test Andersona-Darlinga

Założenia i statystyka testowa

Podobnie jak w teście KS, porównujemy dystrybuantę empiryczną F_n z teoretyczną F_0 .

Kluczowa różnica: test AD **silniej uwzględnia różnice w ogonach rozkładu**.

Statystyka testowa:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) [\ln F_0(x_i) + \ln (1 - F_0(x_{n+1-i}))]$$

również mierzy rozbieżność między dystrybuantami F_n i F_0 .

Duże wartości A^2 świadczą o niezgodności z H_0 .

Porównanie KS i AD:

- KS—maksymalna różnica (globalna)
- AD—różnica ważona (większa waga ogonów)

Test Andersona-Darlinga

Obszar krytyczny

Obszar krytyczny (prawostronny):

$$A^2 > A_{1-\alpha}^2$$

Test Andersona-Darlinga

Obszar krytyczny

Obszar krytyczny (prawostronny):

$$A^2 > A_{1-\alpha}^2$$

gdzie $A_{1-\alpha}^2$ to kwantyl rzędu $1 - \alpha$ rozkładu A^2 .

Jeśli wartość statystyki z próby wpada w obszar krytyczny, to odrzucamy H_0 na poziomie istotności α .

Test Andersona-Darlinga

Obszar krytyczny

Obszar krytyczny (prawostronny):

$$A^2 > A_{1-\alpha}^2$$

gdzie $A_{1-\alpha}^2$ to kwantyl rzędu $1 - \alpha$ rozkładu A^2 .

Jeśli wartość statystyki z próby wpada w obszar krytyczny, to odrzucamy H_0 na poziomie istotności α .

- Rozkład statystyki A^2 jest stablicowany, ale
- nie jest uniwersalny: zależy od testowanego rozkładu (hipotezy zerowej) oraz tego, czy parametry są znane czy estymowane z próby.
- PYTHON: `scipy.stats.anderson`
- Dla dużych prób możliwa jest hipoteza złożona.
- Parametry rozkładu można wtedy estymować z próby.

Test Andersona-Darlinga

Obszar krytyczny

Obszar krytyczny (prawostronny):

$$A^2 > A_{1-\alpha}^2$$

gdzie $A_{1-\alpha}^2$ to kwantyl rzędu $1 - \alpha$ rozkładu A^2 .

Jeśli wartość statystyki z próby wpada w obszar krytyczny, to odrzucamy H_0 na poziomie istotności α .

- Rozkład statystyki A^2 jest stablicowany, ale
- nie jest uniwersalny: zależy od testowanego rozkładu (hipotezy zerowej) oraz tego, czy parametry są znane czy estymowane z próby.
- PYTHON: `scipy.stats.anderson`
- Dla dużych prób możliwa jest hipoteza złożona.
- Parametry rozkładu można wtedy estymować z próby.

Test Andersona-Darlinga

Obszar krytyczny

Obszar krytyczny (prawostronny):

$$A^2 > A_{1-\alpha}^2$$

gdzie $A_{1-\alpha}^2$ to kwantyl rzędu $1 - \alpha$ rozkładu A^2 .

Jeśli wartość statystyki z próby wpada w obszar krytyczny, to odrzucamy H_0 na poziomie istotności α .

- Rozkład statystyki A^2 jest stablicowany, ale
- nie jest uniwersalny: zależy od testowanego rozkładu (hipotezy zerowej) oraz tego, czy parametry są znane czy estymowane z próby.
- PYTHON: `scipy.stats.anderson`
- Dla dużych prób możliwa jest hipoteza złożona.
- Parametry rozkładu można wtedy estymować z próby.

Test Andersona-Darlinga

Obszar krytyczny

Obszar krytyczny (prawostronny):

$$A^2 > A_{1-\alpha}^2$$

gdzie $A_{1-\alpha}^2$ to kwantyl rzędu $1 - \alpha$ rozkładu A^2 .

Jeśli wartość statystyki z próby wpada w obszar krytyczny, to odrzucamy H_0 na poziomie istotności α .

- Rozkład statystyki A^2 jest stablicowany, ale
- nie jest uniwersalny: zależy od testowanego rozkładu (hipotezy zerowej) oraz tego, czy parametry są znane czy estymowane z próby.
- PYTHON: `scipy.stats.anderson`
- Dla dużych prób możliwa jest hipoteza złożona.
- Parametry rozkładu można wtedy estymować z próby.

Test Andersona-Darlinga

Obszar krytyczny

Obszar krytyczny (prawostronny):

$$A^2 > A_{1-\alpha}^2$$

gdzie $A_{1-\alpha}^2$ to kwantyl rzędu $1 - \alpha$ rozkładu A^2 .

Jeśli wartość statystyki z próby wpada w obszar krytyczny, to odrzucamy H_0 na poziomie istotności α .

- Rozkład statystyki A^2 jest stablicowany, ale
- nie jest uniwersalny: zależy od testowanego rozkładu (hipotezy zerowej) oraz tego, czy parametry są znane czy estymowane z próby.
- PYTHON: `scipy.stats.anderson`
- Dla dużych prób możliwa jest hipoteza złożona.
- Parametry rozkładu można wtedy estymować z próby.

Caveat emptor!

Test AD jest najczulszy na różnice w okolicach ogonów rozkładu. Może się zdarzyć, że dana próba „zda” ten test mimo nieakceptowalnie dużych różnic w okolicach mody rozkładu.

Dlatego należy wykonać co najmniej dwa komplementarne testy zgodności (AD i KS).