Antoni Dąbrowski
Nr. indeksu 317214
Mail: 317214@uwr.edu.pl

Statystyka
Laboratorium - lista IV

2022-01-11

1 Zadanie pierwsze

Podaj przedział ufności dla różnicy dwóch średnich w modelu normalnym o znanych wariancjach na poziomie ufności $1-\alpha$. Uzasadnij jego postać.

1.1 Rozwiązanie

Niech $X_1, ..., X_{n_1} \sim N(\mu_1, \sigma_1)$ oraz $Y_1, ..., Y_{n_2} \sim N(\mu_1, \sigma_1)$. Znamy σ_1, σ_2. Interesuje nas przedział ufności dla parametru $\mu_1 - \mu_2$. Jego estymatorem punktowym jest $\bar{X} - \bar{Y}$, gdzie $\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ oraz $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ Przedziałem ufności na poziomie $1-\alpha$ szukanego parametru jest:

$$[(\bar{X} - \bar{Y}) - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, (\bar{X} - \bar{Y}) + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}],$$

gdzie $\sigma^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

2 Zadanie drugie

Wygeneruj $n_1 = 50$ i $n_2 = 50$ obserwacji z rozkładu

a. normalnego z parametrami przesunięcia μ_1,μ_2 i skali σ_2,σ_2

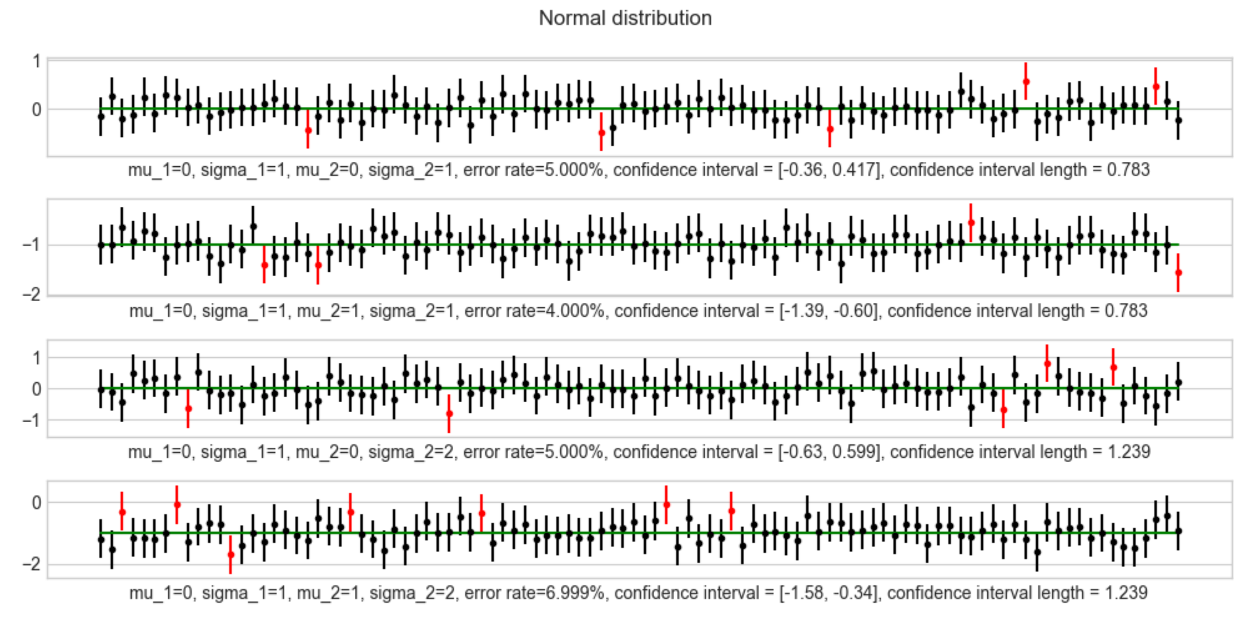
b. logistycznego z parametrami przesunięcia μ_1,μ_2 i skali σ_2,σ_2

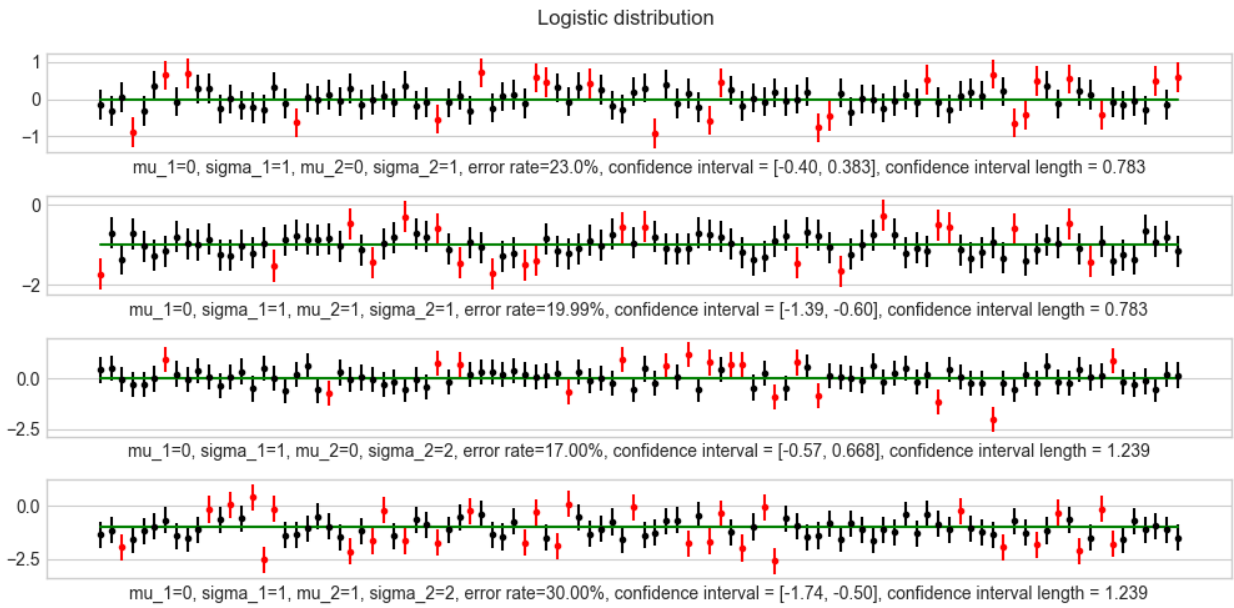
c. Cauchy'ego z parametrami przesunięcia μ_1,μ_2 i skali σ_2,σ_2

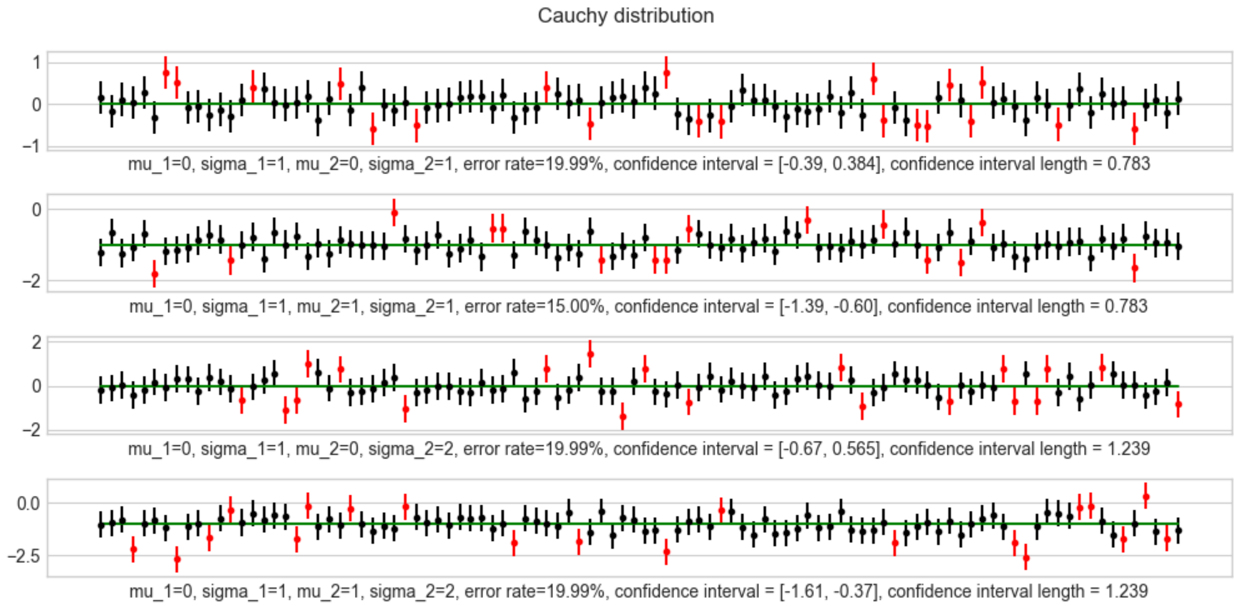
Na tej podstawie wyznacz przedziały ufności z zadania 1 dla parametru $\mu_1 - \mu_2$ na poziomie ufności $1-\alpha = 0.95$ oraz jego długość. Doświadczenie powtórz 10000 razy. Oszacuj prawdopodobieństwo pokrycia nieznanego parametru przez przedział ufności oraz jego długość. Przedyskutuj uzyskane wyniki.

2.1 Rozwiązanie

Poniższe wykresy przedstawiają 1% spośród 10000 losowań 50-elementowych próbek. Zielona linia określa prawdziwą wartość estymowanego parametru. Pionowe słupki to przedziały ufności wokół estymatu punktowego, czerwone słupki oznaczają przedziały nie zawierające prawdziwej wartości. Pod wykresem podaję parametry rozkładu, liczbę błędnie określonych przedziałów ufności (takich, które nie pokrywają wartości rzeczywistej). Dalej podaję przedział ufności i jego długość wyznaczone jako mediana spośród wszystkich prób.







Obserwacje:
Teoria z zadania pierwszego sprawdza się w praktyce. Przedział ufności dla rozkładu normalnego i poziomu ufności 95% faktycznie w takim stopniu pokrywa dokładną wartość. Nie jest tak jednak dla pozostałych rozkładów. Przyczynę tego spodziewam się znaleźć w nie dość dużej próbie, gdyż teoria mówi, że asymptotycznie powinniśmy otrzymać dobry przedział.

3 Zadanie trzecie

Podaj przedziały ufności dla różnicy dwóch średnich w modelu normalnym o nieznanych równych wariancjach na poziomie ufności $1 - \alpha$. Uzasadnij jego postać.

3.1 Rozwiązanie

Tak jak w zadaniu piątym.

4 Zadanie czwarte

Powtórz eksperyment numeryczny z zadania 2 dla wybranych konfiguracji. Na jego podstawie oszacuj prawdopodobieństwo pokrycia nieznanego parametru przez przedział ufności z zadania 3 na poziomie ufności 0.95 oraz jego długość. Przedyskutuj uzyskane rezultaty.

4.1 Rozwiązanie

Tak jak w zadaniu szóstym.

5 Zadanie piąte

Podaj przedziały ufności dla różnicy dwóch średnich w modelu normalnym o nieznanych różnych wariancjach na poziomie ufności $1-\alpha$. Uzasadnij jego postać.

5.1 Rozwiązanie

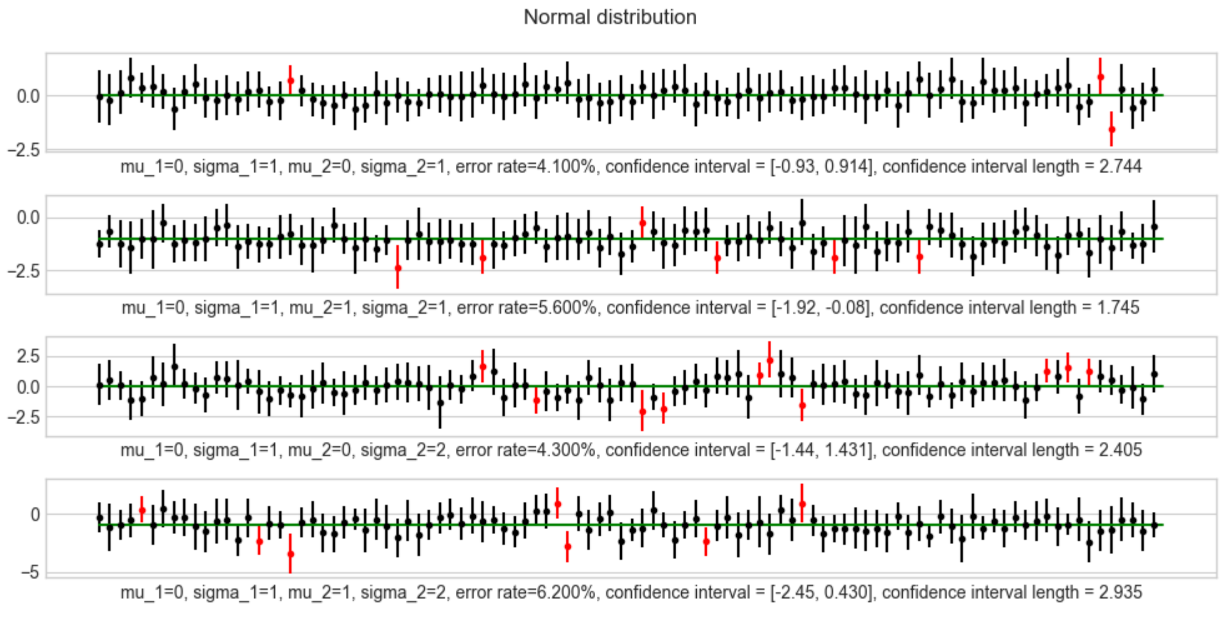
Nie znane σ_1, σ_2.
Estymatory σ_1, σ_2, to $S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\bar{X})^2$ i $S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2$.
Estymator $S_p = \frac{(n_1-1)S_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$
Przedział ufności dla $\mu_1-\mu_2$ ma postać:

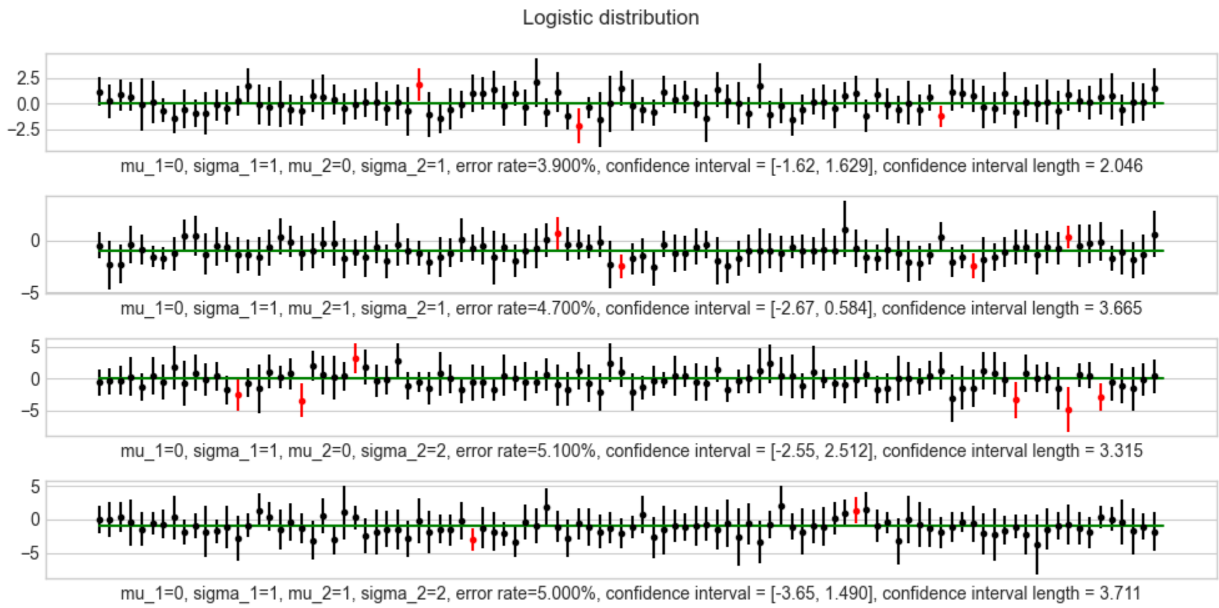
$$[(\bar{X}-\bar{Y})-t_{1-\alpha/2}(n_1+n_2-2)S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, (\bar{X}-\bar{Y})+t_{1-\alpha/2}(n_1+n_2-2)S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}]$$

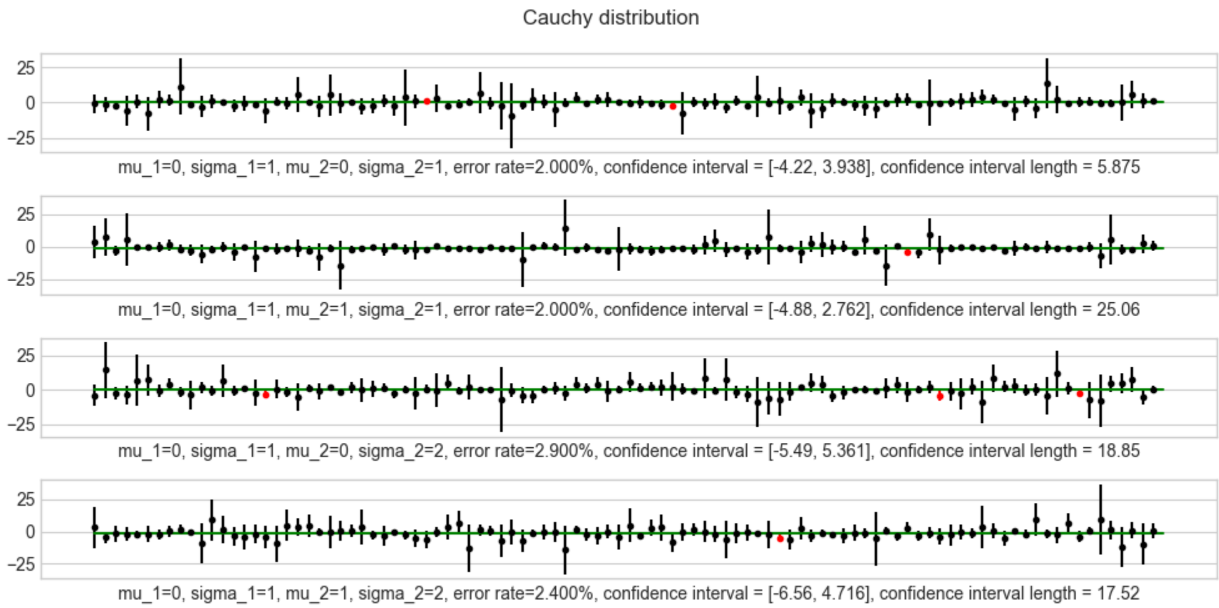
6 Zadanie szóste

Powtórz eksperyment numeryczny z zadania 2 dla wybranych konfiguracji. Na jego podstawie oszacuj prawdopodobieństwo pokrycia nieznanego parametru przez przedział ufności z zadania 5 na poziomie ufności 0.95 oraz jego długość. Przedyskutuj uzyskane rezultaty.

6.1 Rozwiązanie







Znalezione przedziały ufności były znacznie dłuższe od tych wyznaczonych w zadaniu drugim. Mniejsza niepewność w zadaniu drugim prawdopodobnie wynika z dodatkowej wiedzy jaką jest wartość dokładna wariancji.

7 Zadanie siódme

Podaj przedział ufności dla ilorazu dwóch wariancji w modelu normalnym o znanych średnich na poziomie ufności $1 - \alpha$. Uzasadnij jego postać.

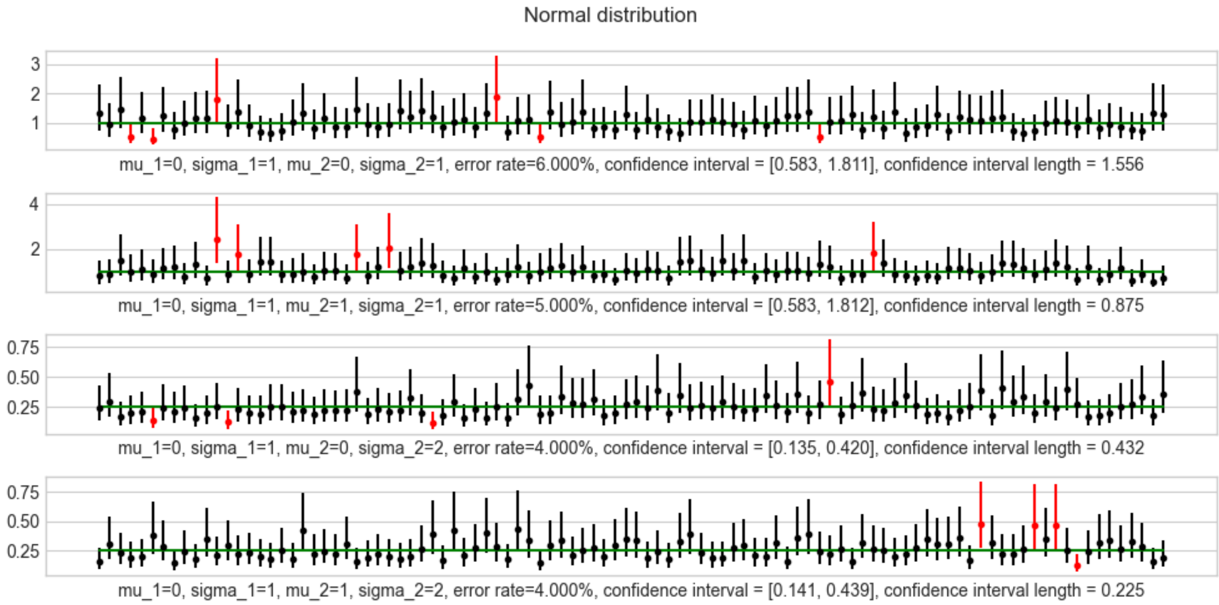
7.1 Rozwiązanie

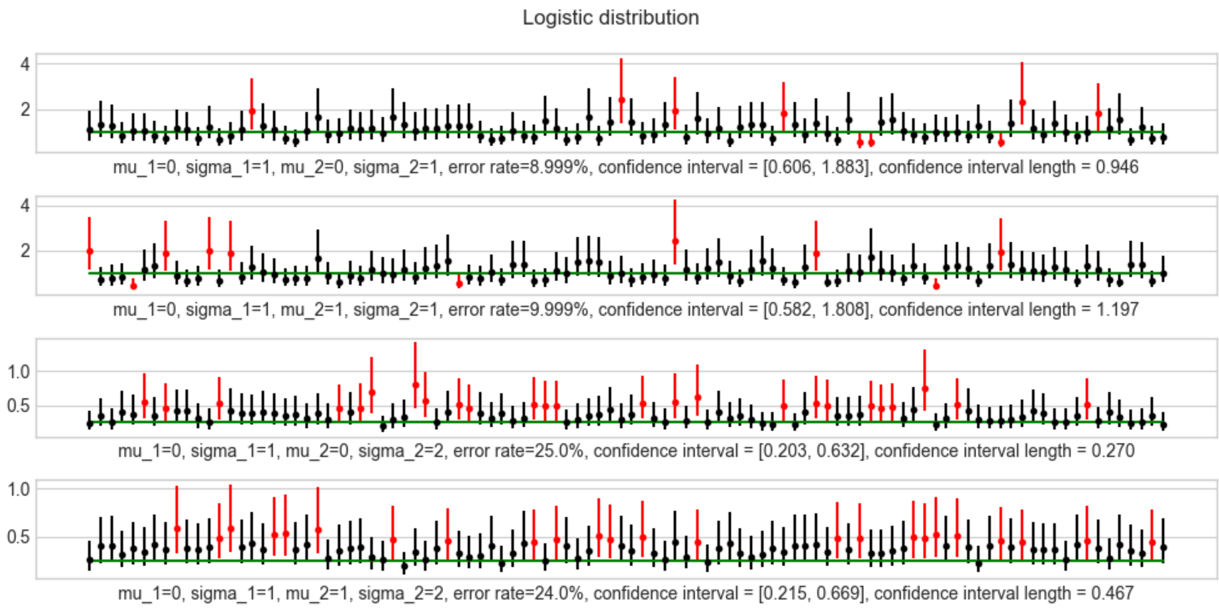
Tak samo jak w zadaniu dziewiątym, jednak korzystam ze znanych średnich do wyznaczenia estymatora wariancji.

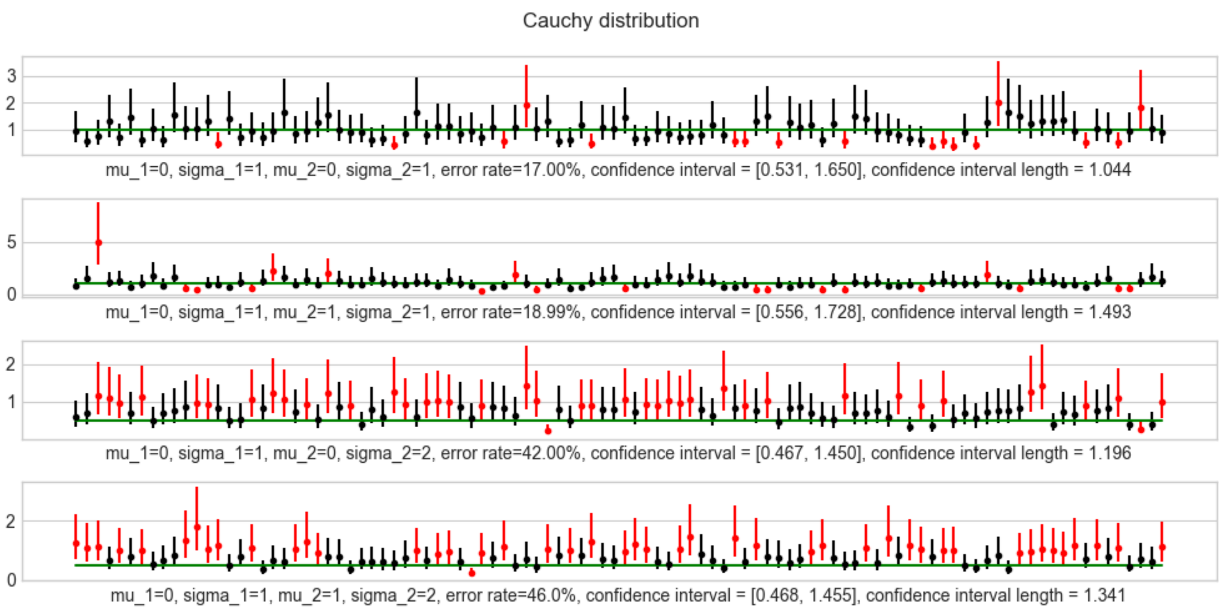
8 Zadanie ósme

Powtórz eksperyment numeryczny z zadania 2 dla wybranych konfiguracji. Na jego podstawie oszacuj prawdopodobieństwo pokrycia nieznanego parametru przez przedział ufności z zadania 7 na poziomie ufności 0.95 oraz jego długość. Przedyskutuj uzyskane rezultaty.

8.1 Rozwiązanie







Rozkład Cauchy'ego stwarza duże problemy w zadaniu estymacji jego parametrów. Jak widać w wielu przypadkach przedziały ufności nie pokrywają faktycznej wartości. Może to wynikać z nie dość licznej próby.

9 Zadanie dziewiąte

Podaj przedział ufności dla ilorazu dwóch wariancji w modelu normalnym o nieznanych średnich na poziomie ufności $1-\alpha$. Uzasadnij jego postać.

9.1 Rozwiązanie

Estymatorem punktowym dla ilorazu $\frac{\sigma_1^2}{\sigma_2^2}$ jest $\frac{s_1^2}{s_2^2}$. Statystyka $\frac{s_1^2}{\sigma_1^2}/\frac{s_2^2}{\sigma_2^2}$ ma rozkład F z n_1-1 i n_2-1 stopniami swobody. Zatem:

$$P(F_{1-\alpha/2} < \frac{s_1^2}{\sigma_1^2}/\frac{s_2^2}{\sigma_2^2} < F_{\alpha/2}) = 1-\alpha$$

$$P(\frac{1}{F_{\alpha/2}}\frac{s_1^2}{s_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{F_{1-\alpha/2}}\frac{s_1^2}{s_2^2}) = 1-\alpha$$

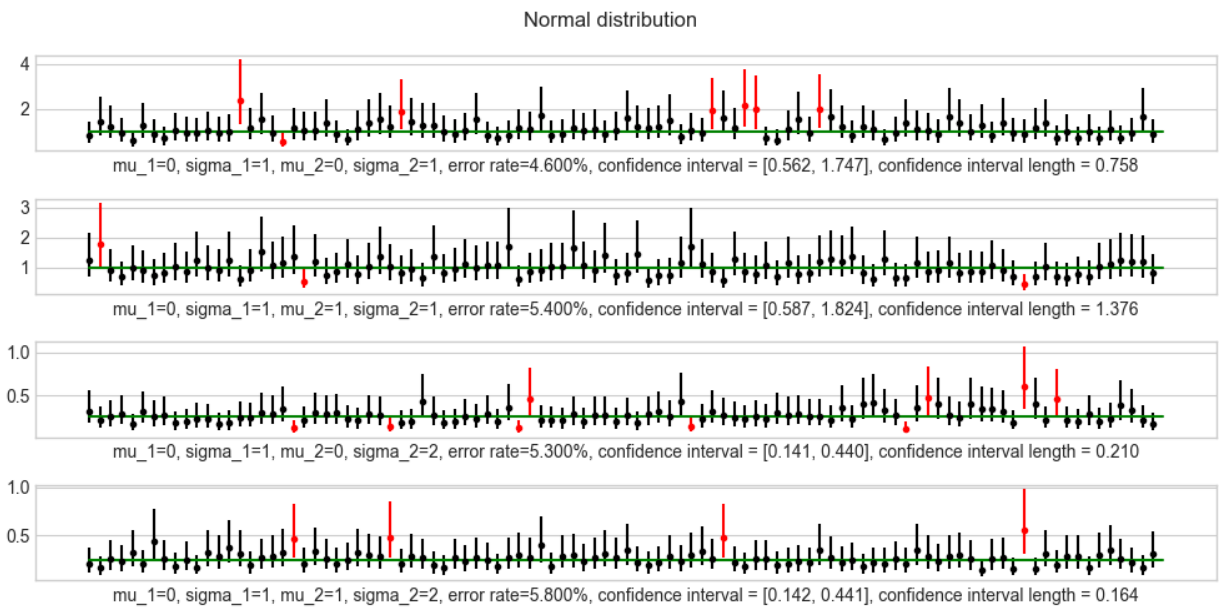
Przedział ufności dla $\frac{\sigma_1^2}{\sigma_2^2}$ ma postać:

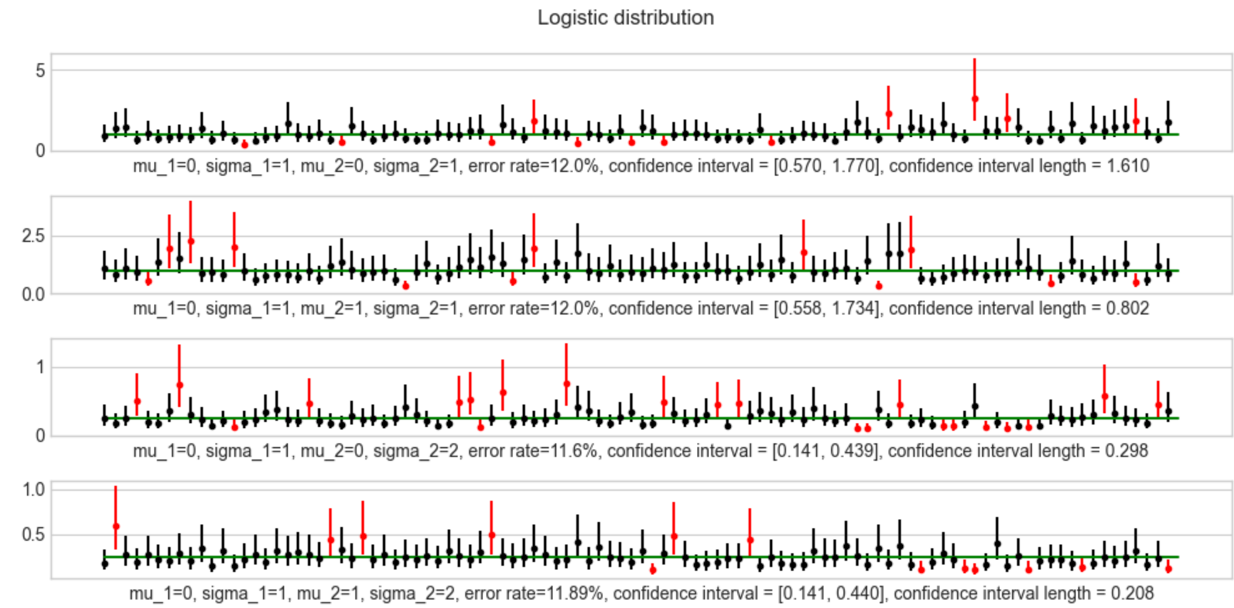
$$\left[\frac{1}{F_{\alpha/2}}\cdot\frac{s_1^2}{s_2^2}, \frac{1}{F_{1-\alpha/2}}\cdot\frac{s_1^2}{s_2^2}\right]$$

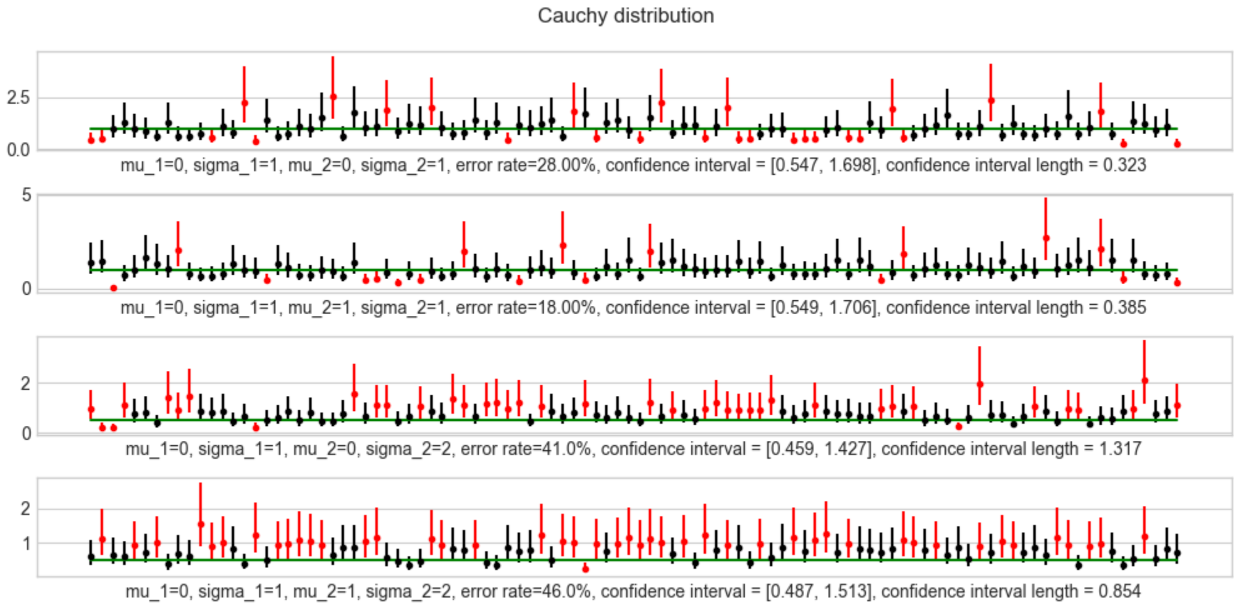
10 Zadanie dziesiąte

Powtórz eksperyment numeryczny z zadania 2 dla wybranych konfiguracji. Na jego podstawie oszacuj prawdopodobieństwo pokrycia nieznanego parametru przez przedział ufności z zadania 9 na poziomie ufności 0.95 oraz jego długość. Przedyskutuj uzyskane rezultaty.

10.1 Rozwiązanie







11 Zadanie jedenaste

Powtórz eksperyment numeryczny z zadań 2, 4, 6, 8, 10, dla $n_1 = n_2 = 20$ i $n_1 = n_2 = 100$. Przedyskutuj uzyskane rezultaty w nawiązaniu do wcześniejszych wyników.

11.1 Rozwiązanie

Dla przejrzystości nie będę zamieszczał wszystkich wykresów, opiszę jedynie główne obserwacje.

- Wraz ze wzrostem liczby obserwacji długość przedziału ufności maleje.
- Eksperymenty dla $n = 20$ cechują się znacznie większą zmiennością wyników.
- Wartości error rate z zadania 2 faktycznie zmniejszają się wraz ze wzrostem liczebności próby.